organized by

Faculty of Communications and Communication Technologies
Technical University of Sofia, Bulgaria

Faculty of Technical Sciences
University "St. Kl. Ohridski", Bitola, Macedonia

Faculty of Electronic Engineering
University of Niš, Serbia and Montenegro

with support of

- *Union of Scientists in Bulgaria*
- *Ministry of Transport and Communications*
- *Bulgarian Telecommunication Company*
- *Ericsson Telecommunication Bulgaria*

in co-operation with

- *IEEE Bulgaria Section*
- *IEEE Macedonia Section*
- *IEEE Serbia and Montenegro Section*

ICEST HISTORY

The ICEST Conference appears to succeed a series of conferences started from 1963 at the Technical University of Sofia under the name "Day of the Radio". In 1977 the name of the Conference was changed into "Communication, electronic and computer systems".

Since 2000 it has become an international conference under the name EIST (Energy and Information Systems and Technologies). The first two EIST Conferences were organized by the Faculty of Communications and Communication Technologies, Sofia and the Faculty of Technical Sciences, Bitola.

In 2002 the Faculty of Electronic Engineering, Niš joined successfully the Conference organizers. Again, the Conference changed its name becoming ICEST (International Scientific Conference on Information, Communication and Energy Systems and Technologies).

This year the host of the ICEST Conference is the Faculty of Communications and Communication Technologies, Sofia.

HONORARY COMMITTEE

O. Kouzov
Ministry of Transports and Communications

Iv. Spasov
Bulgarian Telecommunication Company

R. Engman
Ericsson Telecommunications, Bulgaria

V. Botusharov
Ericsson Telecommunications, Bulgaria

P. Antonov
IEEE Communication Chapter Chair

S. Bogdanova
IEEE SP Joint Chapter Chair

J. Kolev
IEEE SSC/ED Chapter Chair

G. Mollova
IEEE SP Joint Chapter Member

Ph. Philipov
IEEE MTT/ED/AP/CPMT Chapter Chair

V. Sgurev
IEEE SC/IM/SMC Chapter Chair

S. Stojtchev
IEEE Computer Chapter Chair

PROGRAM COMMITTEE

General Chairman:
R. Arnaudov
Technical University of Sofia, Bulgaria

Vice Chairmen:

B. Milovanović
University of Niš, Serbia and Montenegro

C. Mitrovski
University "St. Kliment Ohridski" – Bitola, Macedonia

Members:

E. Altimirski
Technical University of Sofia, Bulgaria

A. Bekiarski
Technical University of Sofia, Bulgaria

G. Dimirovski
University of Skopje, Macedonia

B. Dimitrijević
University of Niš, Serbia and Montenegro

D. Dimitrov
Technical University of Sofia, Bulgaria

R. Dinov
Technical University of Sofia, Bulgaria

D. Dobrev
Technical University of Sofia, Bulgaria

N. Dodov
Technical University of Sofia, Bulgaria

E. Ferdinandov
Technical University of Sofia, Bulgaria

H. Hristov
Technical University of Sofia, Bulgaria

M. Hristov
Technical University of Sofia, Bulgaria,
IEEE CAS Chapter Chair

L. Jordanova
Technical University of Sofia, Bulgaria

M. Kaneva
Technical University of Sofia, Bulgaria

A. Kirij
Technical University of Sofia, Bulgaria

R. Kountchev
Technical University of Sofia, Bulgaria

S. Lishkov
Technical University of Sofia, Bulgaria

P. Merdjanov
Technical University of Sofia, Bulgaria

B. Milovanović
University of Niš, Serbia and Montenegro

S. Mirtchev
Technical University of Sofia, Bulgaria

M. Momtchejikov
Technical University of Sofia, Bulgaria

I. Nedelkovski
University "St. Kl. Ohridski" – Bitola, Macedonia

L. Nikoloski
IEEE Macedonia Section Chair

H. Oskar
Technical University of Sofia, Bulgaria

B. Pankov
Technical University of Sofia, Bulgaria

E. Pentcheva
Technical University of Sofia, Bulgaria

A. Popova
Technical University of Sofia, Bulgaria

V. Poulkov
Technical University of Sofia, Bulgaria

D. Sotirov
Technical University of Sofia, Bulgaria

B. Spasenovski
University of Skopje, Macedonia

M. Stefanović
University of Niš, Serbia and Montenegro

N. Stojadinović
IEEE Serbia and Montenegro Section Chair

D. Stojanović
University of Niš, Serbia and Montenegro

M. Stojcev
University of Niš, Serbia and Montenegro

I. Stojanov
University Polytechnica, Bucharest, Romania

G. Stoyanov
IEEE Bulgaria Section Chair

St. Tabakov
Technical University of Sofia, Bulgaria

Lj. Trpezanovski
University "St. Kl. Ohridski" – Bitola, Macedonia

K. Zaharinov
Technical University of Sofia, Bulgaria

L. Zieleznik
Brookes – Oxford University, UK

ORGANIZING COMMITTEE

Chairman:
D. Dimitrov
Technical University of Sofia, Bulgaria

Members:

S. Bekiarska
Technical University of Sofia, Bulgaria

D. Dimitrov
Technical University of Sofia, Bulgaria

I. Dochev
Technical University of Sofia, Bulgaria

V. Georgieva
Technical University of Sofia, Bulgaria

R. Goleva
Technical University of Sofia, Bulgaria

G. Iliev
Technical University of Sofia, Bulgaria

I. Iliev
Technical University of Sofia, Bulgaria

S. Kolev
Technical University of Sofia, Bulgaria

P. Koleva
Technical University of Sofia, Bulgaria

L. Lubih
Technical University of Sofia, Bulgaria

G. Marinova
Technical University of Sofia, Bulgaria

R. Mironov
Technical University of Sofia, Bulgaria

T. Mitsev
Technical University of Sofia, Bulgaria

E. Pentcheva
Technical University of Sofia, Bulgaria

P. Petkov
Technical University of Sofia, Bulgaria

A. Popova
Technical University of Sofia, Bulgaria

P. Rizov
Technical University of Sofia, Bulgaria

A. Tsenov
Technical University of Sofia, Bulgaria

CONFERENCE SECRETARIAT

V. Georgieva Technical University of Sofia, Bulgaria
L. Lubih Technical University of Sofia, Bulgaria

Faculty of Communications and Communication Technologies
Kl. Ohridski Blvd. 8, 1000, Sofia, Bulgaria
Phone: (+359 2) 965 3998; Fax: (+359 2) 965 3095; E-mail: icest@tu-sofia.bg

CONFERENCE INTERNET SITE

For further information, please visit the Conference Internet Site: **http://radio.tu-sofia.bg/icest**

LIST OF ICEST 2003 REVIEWERS

Prof. Dr. **Altimirski, Emil**
Technical University of Sofia, Bulgaria

Prof. Dr. **Andonova, Anna**
Technical University of Sofia, Bulgaria

Prof. Dr. **Angelov, Angel**
Technical University of Sofia, Bulgaria

Prof. Dr. **Arnaudov, Rumen**
Technical University of Sofia, Bulgaria

Prof. Dr. **Bekiarski, Alexander**
Technical University of Sofia, Bulgaria

Prof. Dr. **Bichev, Georgi**
Technical University of Sofia, Bulgaria

Prof. Dr. **Bumbarov, Ognyan**
Technical University of Sofia, Bulgaria

Prof. Dr. **Darjanov, Petar**
Technical University of Sofia, Bulgaria

Prof. Dr. **Demirev, Vesselin**
Technical University of Sofia, Bulgaria

Prof. Dr. **Dimitrov, Dimiter**
Technical University of Sofia, Bulgaria

Dr. **Dimitrov, Marin**
Bulgarian Academy of Sciences

Prof. Dr. **Dineff, Peter**
Technical University of Sofia, Bulgaria

Prof. Dr. **Dinov, Rangel**
Technical University of Sofia, Bulgaria

Prof. Dr. **Djamiykov, Todor**
Technical University of Sofia, Bulgaria

Prof. Dr. **Dobrev, Dobri**
Technical University of Sofia, Bulgaria

Prof. Dr. **Dodov, Nikola**
Technical University of Sofia, Bulgaria

Prof. Dr. Sc. **Ferdinandov, Ervin**
Technical University of Sofia, Bulgaria

Prof. Dr. **Gadjeva, Elissaveta**
Technical University of Sofia, Bulgaria

Prof. Dr. **Goranov, Petar**
Technical University of Sofia, Bulgaria

Prof. Dr. Sc. **Hristov, Hristo**
Technical University of Sofia, Bulgaria

Dr. **Hristov, Hristo R.**
Bulgarian Academy of Sciences

Dr. **Iliev, Georgi**
Technical University of Sofia, Bulgaria

Dr. **Iliev, Ilia**
Technical University of Sofia, Bulgaria

Prof. Dr. **Ivanov, Emil**
Todor Kableshkov Higher School of Transport – Sofia, Bulgaria

Prof. Dr. **Ivanov, Todor**
University "Assen Zlatarov", Burgas, Bulgaria

Prof. Dr. Sc. **Kountchev, Rumen**
Technical University of Sofia, Bulgaria

Prof. Dr. **Lazarov, Vladimir**
Technical University of Sofia, Bulgaria

Prof. Dr. **Lishkov, Slavcho**
Technical University of Sofia, Bulgaria

Prof. Dr. **Maltchev, K.**
Technical University of Sofia, Bulgaria

Prof. Dr. **Marković, Vera**
University of Niš, Serbia and Montenegro

Prof. Dr. **Merdjianov, Pavel**
Technical University of Sofia, Bulgaria

Prof. Dr. **Mihov, Georgy**
Technical University of Sofia, Bulgaria

Prof. Dr. **Milovanović, Bratislav**
University of Niš, Serbia and Montenegro

Prof. Dr. **Mirtchev, Seferin**
Technical University of Sofia, Bulgaria

Prof. Dr. **Mitrovski, Cvetko**
Faculty of Technical Sciences, Bitola, Macedonia

Prof. Dr. **Momchedjikov, Michael**
Technical University of Sofia, Bulgaria

Prof. Dr. **Nikolić, Zorica**
University of Niš, Serbia and Montenegro

Prof. Dr. **Nikolov, Dimitar**
Technical University of Sofia, Bulgaria

Dr. **Nikolov, Tashko**
Technical University of Sofia, Bulgaria

Prof. Dr. **Pankov, Borislav**
Technical University of Sofia, Bulgaria

Prof. Dr. **Pentcheva, Evelina**
Technical University of Sofia, Bulgaria

Prof. Dr. **Perić, Zoran**
University of Niš, Serbia and Montenegro

Prof. Dr. **Popova, Antoaneta**
Technical University of Sofia, Bulgaria

Prof. Dr. **Poulkov, Vladimir**
Technical University of Sofia, Bulgaria

Algorithm for Optimal Receiving Signals of Jump-Like Change of Carrying Frequency with Rapid Fluctuations of Time Lag

Antonio V. Andonov[1]

Abstract – **The paper presents an algorithm for optimal receiving signals with jump-like change of frequency with random lag. The direct examination on the random lag allows obtaining an algorithm involving a wide range of tasks. For example, such tasks are receiving signals for which the random lag occurs not only in moving away the limits of the time tact intervals but also in receiving under the conditions of changing lag.**

Keywords – **Signals of jump-like change of carrying frequency, random time lag**

I. Problem Setting

As it is known, with radio signals of a jump-like change of carrying frequency [1-3] the value of the carrier in the tact intervals is given in a code sequence. Without considering the information modulation, such signals are described analytically, using a complex presentation as follows:

$$s(t) = \mathrm{Re}\{U_0 c(t) \exp[j(\omega_0 t + \varphi)]\} \quad (1)$$

where:

$$c(t) = \sum U_0[t - (i-1)T] \exp[j(\omega_i t - \varphi_i)] \quad (2)$$

is the modulating function where ω_i is the value of the frequency jump in the i-th interval and φ_i is the value of the phase in the same interval.

Let one of the n-number element signals, which are due for those identically equal to zero out of the interval, is marked with $s_k\ [0, T]$. Then the signal transmitted can be described with the expression:

$$s(t) = \sum_{k=0}^{n} s_k(t - iT) \quad (3)$$

With the presence of random time lag, the useful signal received is:

$$s(t, \tau(t)) = s(t - \tau(t)) = \sum_{i=0}^{n} s_i(t - iT - \tau(t)) \quad (4)$$

Let present it in the kind of a known function of information discrete parameter $\theta(t)$ and random time lag $\tau(t)$, i.e.:

$$s(t, \theta, \tau) = s[t - \tau(t), \theta(t - \tau(t))] \quad (5)$$

The discrete parameter takes constant values on tact intervals $\theta(t) = \theta_i$, $t \in [t_i, t_{i+1}]$. The values of the information parameter on the various tact intervals form a simple Markov's chain θ_i, $i = 0, 1, \ldots n$ with n states and a known matrix of the transitions from the i-th into the j-th state $\Pi = \pi_{i,j}$, and vector of the initial states $p = p_1$. The limits of the tact intervals are determined by random time lag $\tau(t)$, i.e. $t_i = t_i(\tau)$. With the time lag realization specified, the limits of tact interval are:

$$t_i = iT + \tau(t_i)$$

On tact interval t_i, t_{i+1}, signal $s(t, \theta, \tau)$ coincides with elementary signal $s_i(t - iT - \tau(t))$ if $\theta(t) = \theta i$. Random time lag $\tau(t)$ corresponds to the signal lag caused by the relative motion of the receiver and the transmitter and in the common case it can be examined as a component of the diffusion Markov's process:

$$\lambda(t), \tau(t) = \lambda_1(t)$$

Process $\lambda(t)$ satisfies the system of stochastic differential equations:

$$\frac{\partial \lambda_i(t)}{\partial t} = f_i(t, \lambda) + n_i(t) \quad (6)$$

Here $f_i(t, \lambda)$ are functions satisfying the condition of Lipshits [4] and $n_i(t)$ are Gauss noises with intensity $b_{ij}(t, \lambda)$. The a priori probable features of process $\lambda(t)$ are determined by the equation of Kolmogorov–Foker–Plank [4]:

$$\frac{\partial W}{\partial t} = -\sum_{\alpha=1}^{n} \frac{\partial}{\partial \lambda_\alpha}[a_\alpha(t, \lambda) W] + \\ + \frac{1}{2} \sum_{\alpha=1}^{n} \sum_{\gamma=1}^{n} \frac{\partial^2 b_{\alpha\gamma}(t, \lambda) W}{\partial \lambda_\alpha \Lambda_\gamma} \equiv L[W] \quad (7)$$

where $W = W(t, \lambda)$ is the a priori probable density of process $\lambda(t)$; $a_\alpha(t, \lambda)$ are the diffusion coefficients of Markov's process $\lambda(t)$.

The observation on signal $s(t, \theta, \tau)$ is realized on the background of noise, i.e. it has the kind of

$$\xi(t) = s(t, \theta, \tau) + n(t) \quad (8)$$

where $n(t)$ is non-correlated with $\theta(t)$ and $\tau(t)$ is white noise of feature $M\{n(t)\} = 0$.

II. Optimal Assessment of the Information Parameter

With a certain realization of lag $\tau(t)$, the optimal (according to the criterion of the error probability minimum) assess-

[1]Antonio V. Andonov, Department of Communication and Safety Equipment and Systems, Todor Kableshkov Higher School of Transport, Geo Milev 158, 1574 Sofia, Bulgaria, E-mail: a.andonov@infotel.bg

ment of the information discrete parameter constant in interval $[t_i, t_{i+1}]$ is determined with expression [5]:

$$\theta_i^* = \max^{-1}\{P(\theta_i = i)\} = \\ = \max_i^{-1}\{P[(i+1)T + \tau, i]\} \quad (9)$$

With random values of tact interval t_{i+1}, when posteriori probabilities $P(\theta_i = i)$ have to be compared, in the moments of time the conditional probability has to be examined with fixed τ, i.e.

$$P(\theta_i = i) = P[t_{i+1}(\tau), i/\tau] \quad (10)$$

The optimal assessment of the discrete parameter can be examined as a non-conditional discrete parameter posteriori probability at the end of the interval under observation. The probability can be determined by the expression:

$$P[t_{i+1}(\tau), i/\tau] = P[i/\tau, \xi^{t_{i+1}(\tau)}] \quad (11)$$

made average with weight corresponding to posteriori probable density of random lag $P(t, \tau)$. Then in the i-th tact interval the algorithm of assessment of the information discrete parameters takes the kind of:

$$\theta_i^* = \max_i^{-1}\{P(\theta_i = i)\} = \\ = \max_i^{-1}\Big\{ \int P[t_{i+1}(\tau), i/\tau] P[t_{i+1}(\tau), \tau] d\tau \Big\} \quad (12)$$

III. Conclusions

The paper presents an algorithm for optimal receiving signals with jump-like change of frequency with random lag. The direct examination on the random lag allows obtaining an algorithm involving a wide range of tasks. For example, such tasks are receiving signals for which the random lag occurs not only in moving away the limits of the time tact intervals but also in receiving under the conditions of changing lag.

The algorithm obtained (12) presents a summary of an algorithm for assessment of a constant parameter in a certain interval with the presence of indeterminacy at the moment of signal appearing.

References

[1] Cooper G.R., C.D. McGillem, Modern Communications and Spread Spectrum. New York, NY, 1986.

[2] Holmes, J.K. Coherent Spread Spectrum Systems. New York, NY; Wiley, 1982.

[3] Kumar, P.R., Varaiya. Stochastic Sustem Estimation, Identification and Adaptive Con-trol. Englewood Cliffs, NJ Prentice-Hall, 1984

[4] Moser, R.J. Spread Spectrum Techniques Microwave Journal, 8/1982.

[5] Spread Spectrum Communication Systems. Course notes. Integrated Computer Systems.Publishing Company, Brussels, 1992

Bit Error Probability of the QPSK System in the Presence of Ricean Fading and Pulse Interference

Milan S. Milošević, Mihajlo Č. Stefanović[1]

Abstract – **The purpose of this paper is to provide the theoretical approach for determining the bit error probability in detecting a coherent quaternary phase shift keyed signal in the presence of the Ricean fading, Gaussian noise and pulse interference, as well as the noisy carrier reference signal. Phase locked loop, as the constituent part of the receiver, is used in providing the synchronization reference signal extraction, which is assumed to be imperfect in this paper. The determined results are based on the non-linear phase locked loop model with primary emphasis on the degradations in the system performance produced by the imperfect carrier signal extraction.**

Keywords – **QPSK, PLL, Pulse interference, Ricean fading**

I. Introduction

The performance evaluation of binary and M-ary ($M > 2$) phase-shift-keying communication systems have been analyzed in a great variety of papers, which have appeared in the literature [1-7]. In the majority of this papers, the system performance has been determined under the condition of the perfect signal extraction. Quaternary phase-shift-keying (QPSK or 4-PSK) systems have the greatest practical interest of all nonbinary (multiposition) systems of digital transmission of messages by phase modulated signals. Currently, QPSK is one of the prevalent modulations in use for digital communication systems (since bandwidth efficiency). The only significant penalty factor is an increased sensitivity to carrier phase synchronization error.

Any successful transmission of information through a digital phase-coherent communication system requires a receiver capable of determining or estimating the phase and frequency of the received signal with as few errors as possible. In practice, quite often the phase locked loop (PLL) is used in providing the desired reference signal [1-4]. Frequently, a PLL system must operate in such conditions where the external fluctuations due to the additive noise are so intense that classical linear PLL theory neither characterizes adequately the loop performance, nor explain the loop behavior [5]. Numerical results for QPSK system is presented so that this results combined with the characteristic of the phase recovery circuit will enable the best practical design of a QPSK system.

It is well known that the certain components that appear in telecommunication channels are very often with a pulse characteristics, i.e. noise can be described as a sum of peaks of large amplitudes in comparison with the common noise level. This channels are often narrowband, so it follows narrowband systems are considered. Poisson pulse noise model is used for modelling. Samples consist of a random delta functions. This model gives a very good approximation of the most important natural pulse noise features [6,7].

The error probability, as a measure of systems quality, is an important issue. Noise influence and pulse interference are often fundamental limiting factors in digital transmission systems. An expression for the bit error probability was calculated when the signal and Gaussian noise are applied at the input of the QPSK system [4]. QPSK system performance when the signal, Gaussian noise, pulse interference, Ricean fading and imperfect carrier phase recovery are considered as source of degradation, are determined in this paper.

II. System Features

The model for the communication system to be considered in this paper is shown in Fig. 1. Let the input signal at QPSK receiver consists of the signal, pulse interference and Gaussian noise:

$$r(t) = A\cos(\omega_0 t + \phi_0) + A_1 n_i \cos(\omega_0 t + \theta) + m(t), \quad (1)$$

where A is a signal amplitude, ϕ_0 can be $\pi/4$, $3\pi/4$, $-\pi/4$ or $-3\pi/4$ depending upon which symbol is transmitted, ω_0 is a constant carrier frequency, $m(t)$ is a Gaussian noise, A_1 is a pulse interference amplitude, θ is the uniformly distributed phase with the probability density function $p(\theta) = 1/2\pi$, $\{-\pi \le \theta \le \pi\}$. Cos wave is modulated by the pulses with the form:

$$n_i = \sum_{i=-\infty}^{\infty} a_i \sqrt{\frac{2}{T}} \cos\omega_1 (T - t_i), \quad (2)$$

where a_i represents a random area where i pulses appears at the random time t_i, $\omega_1 = 2\pi n/T$, n is an integer and T is a pulse duration. Moments of pulses appearing is presented as a Poisson process.

It is assumed that the signal is corrupted by the pulse interference with the following probability density function:

$$p(n_i) = (1-\gamma)\delta(t) + \left(\frac{\gamma}{\pi}\right)^{3/2} \int_0^{\pi/2} \frac{e^{-\frac{\gamma n_i^2}{4\sigma_i^2 \cos^2(y)}}}{\sigma_i^2 \cos(y)} dy \quad (3)$$

where γ represents average number of pulses which appears in the signal time duration T, σ_i is the spectral density of the pulse modulated interference.

[1]The authors are with the Faculty of Electronic Engineering, Dept. of Telecommunication, Beogradska 14, 18 000 Niš, Yugoslavia, E-mail: milan@elfak.ni.ac.yu

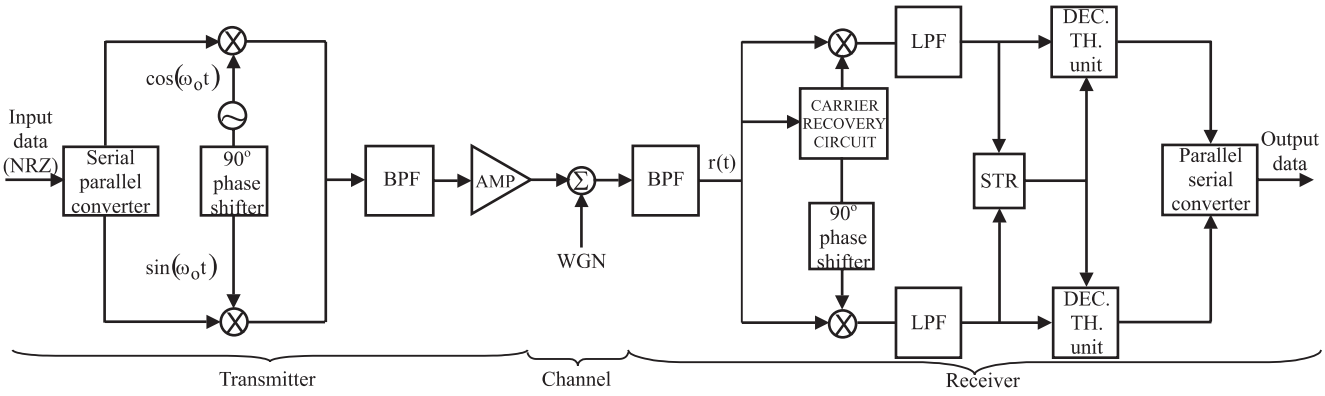


Fig. 1. QPSK communications modem

It is assumed that the signal is affected by the Ricean fading:

$$p(A) = \frac{A}{\sigma_f^2} e^{\frac{A^2+B^2}{2\sigma_f^2}} I_0\Big(\frac{AB}{\sigma_f^2}\Big), \quad A \geq 0,\ B = 2\sigma_f. \tag{4}$$

Now, input signal can be also written with in the form:

$$r(t) = AR\cos(\omega t + \psi) + m(t), \quad \eta = \frac{A_1}{A},$$
$$R = \sqrt{1+\eta^2+2\eta\cos\theta}, \quad \psi = \text{arctg}\frac{\eta\sin\theta}{1+\eta\cos\theta} \tag{5}$$

From now on, pulse interference, additive Gaussian noise Ricean fading and imperfect phase carrier recovery, are taken into account in our detection analysis. All other functions are considered ideal. The block diagram of a QPSK receiver would be adopted is shown in Figure 1. The recovered carrier signal is assumed to be in the form of the sin wave. Also, it would be adopted that a original message is in binary form.

Under the assumption of a constant phase in the symbol interval, the conditional error probability for the given phase error ϕ (the phase error ϕ is the difference between the receiver incoming signal phase and the voltage controlled oscillator output signal phase) can be written as [9],

$$P_{e/\phi}(\phi) = \frac{1}{4}\text{erfc}\Big\{\Big[\sqrt{2R_b}(\cos\phi - \sin\phi)\Big] + \text{erfc}\Big[\sqrt{2R_b}(\cos\phi + \sin\phi)\Big]\Big\} \tag{6}$$

where the function $\text{erfc}(x)$ is the well known complementary error function defined as:

$$\text{erfc}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-z^2/2} dz. \tag{7}$$

The received signal to noise spectral density ratio in the data channel (demodulator) denoted by R_b, is given by $R_b = E/N_0$, where E is a signal energy per bit duration T. N_0 represents the normalized noise power spectral density in W/Hz, referenced to the input stage of the demodulator, since the signal to noise ratio is established at that point. The signal detection in receiver is accomplished by cross-correlation-and-sampling operation. The effect of filtering due to $H(f)$ in Figure 1 is not considered here.

The conditional steady state probability density function, for the non-linear PLL model with a known signal and noise at the PLL input, of modulo 2π reduced phase error is given by the following approximation [10]:

$$p(\phi) = \frac{e^{\beta\phi+\alpha\cos\phi}}{4\pi^2 e^{-\pi\beta}|I_{j\beta}(\alpha)|^2} \int_{\phi}^{\phi+2\pi} e^{-\beta x+\alpha\cos x} dx, \tag{8}$$

$I_{j\beta}(\alpha)$ is the modified Bessel function of complex order $j\beta$ and real argument α. The range of definition for ϕ in the previous equation is any interval of width 2π centered about any lock point $2n\pi$, with n an arbitrary integer. The parameters α and β, that characterize Eq.(8), for the first order non-linear PLL model in this case are:

$$\alpha = \alpha_0 R, \qquad \beta = \beta_0 \Omega, \tag{9}$$

where α_0 and β_0 are constants [10,11]. The parameter α is a measure of the loop signal to noise ratio in the sense that the larger the value of α, the smaller are the deleterious effects due to noise reference signal. The parameter β is a measure of the loop stress. Ω is the loop detuning, i.e. the frequency offset defined by:

$$\Omega = \frac{d}{dt}(\omega_0 t + \psi) - \omega_0 = \frac{\eta(\eta+\cos\theta)}{R^2}\frac{d\theta}{dt}. \tag{10}$$

Since $(d\theta/dt) = 0$, it follows $\Omega = 0$, i.e. $\beta = 0$ and Eq.

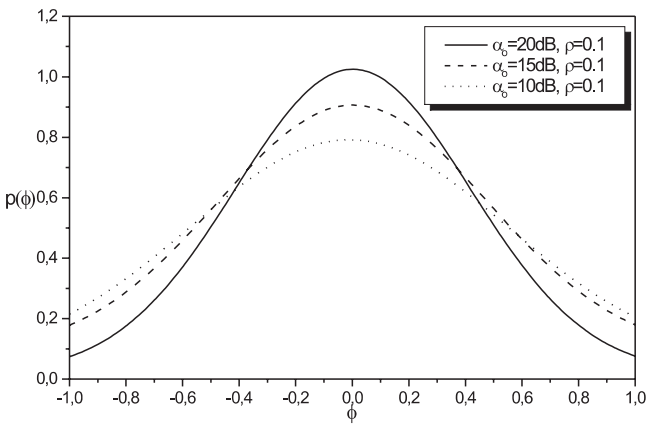


Fig. 2. Probability density function of phase error for the nonlinear first order PLL model

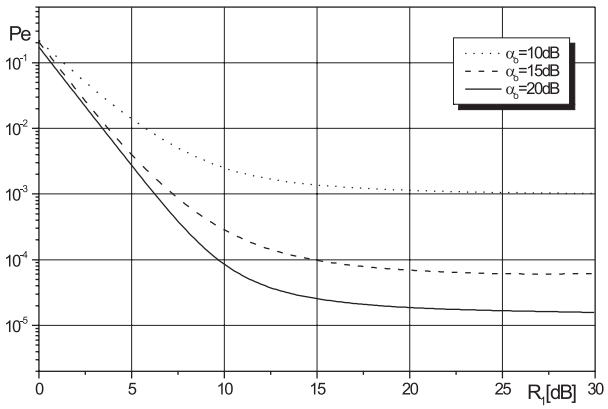


Fig. 3. Average bit error probability performance of a QPSK coherent system with noisy carrier synchronization reference signal in the presence of both, pulse interference and Ricean fading

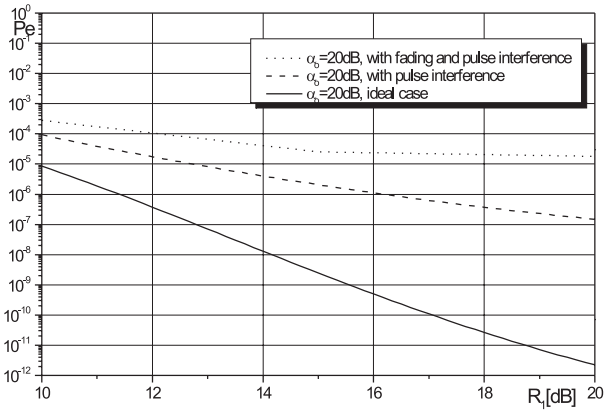


Fig. 4. Comparation of the QPSK system performances

(8) takes the form

$$p(\phi) = \int_A \int_\theta \int_{n_i} p(\phi/A, n_i, \theta) p(A) p(n_i) p(\theta) dA d\theta dn_i. \quad (11)$$

Calculating the previous equations yields the probability density function of the phase error shown in Fig. 2. In order to obtain numerical results the following is supposed $\gamma = 1$, $A_1 = 1$, $\sigma_i = 1$, $\sigma_f = 1$.

III. System Performance

Substituting $R_b = R_1 R^2$ in Eq. (6), where R_1 corresponds to the case when there is no interference, the conditional bit error probability is determined. The total error probability is determined by averaging the conditional error probability over random variable ϕ:

$$P_e = \int_A \int_\theta \int_{n_i} \int_\phi P_{e/\phi} p(\phi/A, \theta, n_i) p(A) p(\theta) p(n_i) dA d\theta dn_i d\phi \quad (12)$$

The total error probability is computed on the basis of the Eq. (12) and is plotted versus signal to noise ratio (R_1 [dB]) in Fig. 3. The values are given in figures.

The total bit error probability, when the signal, Gaussian noise and pulse interference are applied at the input of the receiver, as a function of the signal to noise ratio for the Ricean fading, is shown in Fig. 3 and Fig. 4. From figures follows that the system error probability decreases with the increase of the signal to noise ratio.

Figure 3 shows the influence of the PLL parameter α_0 on the bit error probability of the observed QPSK system. On the basis of the numerical results and shown figures it can be seen that the bit error probability increases with the decrease of the PLL parameter α_0. Also, in the ideal case, when the fading and pulse interference are absent, system has better performance rather than in the presence of both, pulse interference and Ricean fading.

The influence of the pulse interference and Ricean fading is evident from Fig. 4. The following observation is significant. One can see that the bit error probability has changed 809.7, for SNR = 15 dB, if at the receiver input is present pulse interference. But, if the fading appeared together with interference the bit error probability has changed $11.298\,10^3$ times. On the basis of the above discussion can be concluded that the pulse interference and Ricean fading have a great influence on the QPSK system performances. Therefore it is justified to increase the parameter α_0 to a certain limit when the error probability decreases considerably with its increase, but there is no justification for the parameter α_0 increase then that value because certain great variations can cause the insignificant error probability reductions.

IV. Conclusion

The QPSK system is analyzed by means of the system error probability, in this paper. Noise influence, interference, fading and imperfect carrier phase recovery are the limiting factors in the observed system performance. The interference is represented by cosinusoidal signal with the uniform distributed phase. The influence of the imperfect reference signal extraction is expressed by the probability density function of the PLL phase error.

The detailed analysis of the obtained numerical results is performed in this paper. Case when the signal, Gaussian noise and pulse interference are applied at the input of the receiver, as a function of the signal to noise ratio for the Ricean fading, is considered in this paper. On the basis of the shown analysis can be concluded that the system has better performance if both, the PLL parameter α_0 and signal to noise ratio has a greater value. The influence of the pulse interference as well as the influence of fading on the system error probability are especially considered. On the basis of the shown analysis one can conclude that the system has better performances if both, pulse interference as well as fading have a smaller values.

However, from Fig. 3, for the large signal to noise ratio system error tends to a constant value (BER floor). In the BER floor area, the signal to noise ratio is relatively large with respect to parameter α_0 and has therefore a small influence on the system error probability. It is seen from figures that this BER floor can be reduced by increasing the parameter α_0 which depends on the applied PLL loop. On the basis

of the shown analyze it is possible to determine the QPSK system parameter α_0 and useful signal power necessary to compensate the imperfect carrier extraction. This means that the presented conclusions can be useful in QPSK system design.

References

[1] Stefanović M., Djordjević G., Djordjević I., Stojanović N., "Influence of Imperfect Carrier Recovery on Satellite QPSK Communication System Performance", *International Journal of Satellite Communications*, Vol. 17, pp. 37-49, 1999.

[2] Stefanović M., Drača D., Vidović A., Milović D., "Coherent Detection of FSK signal in the presence of Cochannel interference and noisy carrier reference signal", *International Journal of Electronics and Communications*, No. 2, pp. 77-82, 1999.

[3] Stefanović M., Djordjević G., Djordjević I., "Performance of Binary CPSK Satellite Communication System in the Presence of Noises and Noisy Carrier Reference Signal", *International Journal of Electronics and Communications*, No. 2, pp. 70-76, December 1999.

[4] Prahbu. K. V., "PSK Performance with Imperfect Carrier Phase Recovery, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-12, No.2, pp. 275-285, March 1976.

[5] Okunev Y., *Phase and Phase-Difference Modulation in Digital Communications*, Artech House, Inc., London, 1997.

[6] R.E. Ziemer, "Character error probabilities for M-ary Signaling in impulsive noise environments", *IEEE Trans. Commun. Technol.*, Vol. COM-15,1967.

[7] R.E. Ziemer, "Error probabilities due to additive combinations of Gaussian and impulsive noise", *IEEE Trans. Commun. Technol.*, Vol. COM-15,1967.

[8] Drača, D., Stefanović M., "Bit error probability of phase coherent communication systems in presence of noise and interference", *Electronics Letters* 26, No. 16, pp. 1234-1235, 1990.

[9] Lindsey, C., "Nonlinear analysis and synthesis of generalized tracking system", *Proc. IEEE*, 57, pp. 1705-1722, 1969.

[10] Rosenbaum, S., Glave, E., "An error probability upper bound for coherent phase shift keying with peak-limited interference", *IEEE Trans.*, COM-22, (1), pp. 6-16, 1974.

[11] Lindsey, C., *Synchronization systems in communication and control*, Prentice-Hall, Englewood Cliffs, 1972.

[12] Viterby, J., *Principles of coherent communication*, McGraw-Hill, New York, 1966.

[13] Abramowitz, M., Stegun, A., *Handbook of mathematical functions with formulas, graphs and mathematical tables*, New York, Dover Publications, Inc., 1970.

[14] M. Stefanović, M. Milošević, "QPSK performance with imperfect carrier phase recovery in the presence of atmospheric noise and interference", *Facta Universitatis – Series Electronics and Energetics*, Vol. 14, No. 1, pp. 55-66, April 2001.

[15] M. Stefanović, M. Milošević, "QPSK performance in the presence of Rayleigh fading and atmospheric noise", *International Conference of Communications*, pp. 252-255, Bucharest, December 2000, Romania.

Diversity Systems Performance in the Presence of Shadowing

Milan Živković[1], Nenad Milošević[2], Zorica Nikolić[3]

Abstract – **In this paper we compare several diversity reception techniques in an additive white Gaussian noise channel in the presence of Rayleigh fading and log-normal shadowing employing coherent (BPSK) and noncoherent (DPSK) digital signaling. Dependence to the required SNR over MR combining of the number of branches to produce the same bit error rate (BER) is used as the measure of the performances. It is shown that the effects of log-normal shadowing and Reyleigh fading affect almost identically both of the signaling cases, coherent (BPSK) and noncoherent (DPSK).**

Keywords – **BPSK, DPSK signaling, Rayleigh fading, log-normal shadowing, diversity combining**

I. Introduction

Binary digital signaling is often followed by presence of fading and shadowing. Fading is the term used to describe the rapid fluctuations in the amplitude of the received radio signal over a short period of time caused due to the interference between two or more versions of the transmitted signals which arrive at the receiver at slightly different times. The resultant received signal can vary widely in amplitude and phase, depending on various factors such as the intensity, relative propagation time of the waves, bandwidth of the transmitted signal etc... In mobile environments transmitted signal can be also affected by effect of shadowing which results in the long-term attenuation of received signal due to specific propagation environment (vegetation, buildings). Therefore, the mobile communication channel can be modelled as additive white Gaussian noise channel subject to Rayleigh fading (received amplitudes has Rayleigh distribution) and log-normal shadowing (the mean of signal-to-noise ration has log-normal distribution). A powerful communication receiver technique that provides wireless channel improvement at relatively low cost is a well-known as diversity reception. Diversity techniques are based on the notion that errors occur in reception when the channel attenuation is large (when channel is in a deep fade). Supplying to the receiver several replicas of the same information signal transmitted over independently fading channels, the probability that all the signal components will fade simultaneously is reduced considerably [1] and therefore, instant and mean SNR can be increased. The diversity reception can be categorized as microscopic and macroscopic.

Microscopic diversity is a method for reducing the effect of instantaneous fading in which several uncorrelated faded signals are received at a radio port. There are several techniques for evaluating transmitted signal at the receiver. For the coherent digital signaling (CFSK, BPSK) with independent branch fading, achieved by separating receiver antennas at least 10 wavelenghts, the optimum diversity technique is known as Maximal Ratio Combining (MRC). In Maximal Ratio Combining (MRC), the signals from all the branches are co-phased and individually weighed by fading factor to provide the optimal SNR at the output. But it is seldom implementable in a multipath fading channel because the receiver complexity for MRC is directly proportional to the number of branch signals L available at the receiver. Since L may vary with location as well as time, it is undesirable to have receiver complexity dependent on a characteristic of the physical channel from a production and implementation point of view. Similarly, for the noncoherent digital signaling (NCFSK, DPSK) the commonly used technique is Equal Gain Combining (EGC), where all available branches are equally weighted and then added incoherently. It is clear that this technique is analogous to MRC in the sense that all available branches are used, therefore it has the same undesirable feature of having receiver complexity dependent on L. So it is very desirable to implement some other suboptimal diversity techniques in order to evaluate transmitted signal. The simplest suboptimal technique is the Traditional Selection Diversity Model (SC) that selects, among the L diversity branches, the branch providing the largest signal-to-noise ratio (or largest fading amplitude). Clearly, SC and MRC represent the two extremes in diversity combining strategy with respect to the number of signals used for demodulation. Consequently, other techniques representing compromise between this two were developed. One of them is S + N Selection Model, where S + N denotes a signal-plus-noise sample (i.e., not a power measurement), and noise is treated as random variable. The selective technique, which selects the branch providing the largest LLR (log-likelihood ratio) developed for BPSK signaling, has the closest performance to MRC. Also, combining diversity techniques that use two (SC2) or three (SC3) branches with largest amplitudes (or signal-to-noise ratio) for getting transmitted signal were developed. In this paper we compare this several diversity combining techniques for a Rayleigh faded channel in the presence of white Gaussian noise (AWGN) employing one coherent (BPSK) and one noncoherent (DPSK) digital signaling.

Macroscopic diversity, again, is a method for reducing the

[1]Milan Živković is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: zivkovic8@ptt.yu

[2]Nenad Milošević is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: nemilose@ni.ac.yu

[3]Zorica Nikolić is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: zora@ni.ac.yu

effect of shadowing, in which several signals are received at different radio-ports, with differently experienced long-term shadowing. The most common technique of macroscopic diversity is macrodiversity selection where selected signal originates from the port where the smallest long-term shadowing (the largest mean SNR) is present. In this paper we consider different systems with implemented macroscopic and microscopic techniques and compare different microscopic selection methods where dependence to the required SNR over MR combining of the number of branches to produce the same bit error rate (BER) is used as the measure of the performances.

II. System Model

Consider the system with K different radio-ports forming the microscopic diversity group. In order to mitigate the effect of shadowing (long-term attenuation) one can select signal originated from those port where the largest mean SNR is present. Consider, again, L independent microdiversity branches at every radio-port employ one of considered microdiversity techniques. If the transmitted signal is $x(t)$, the low-pass equivalent received signal l-th branch of the k-th port [2]

$$\omega_{kl} = \alpha_{kl} e^{j\phi_{kl}} x(t) + \eta_{kl} \quad k = 1, ..., K \quad l = 1, ..., L \quad (1)$$

where α_{kl} – fading amplitude (factor) in the l-th branch of the k-th port (nonnegative number); ϕ_{kl} – fading phase in the l-th branch of the k-th port; η_{kl} – additive complex Gaussian noise in the l-th branch of the k-th port; E_b – bit energy.

Corresponding signal in the l-th branch of the k-th port after cophasing is

$$r_{kl}(t) = \mathrm{Re}\{\omega_{kl} e^{j\,\phi_{kl}}\} = \alpha_{kl} x(t) + n_{kl} \quad (2)$$
$$k = 1, ..., K \quad l = 1, ..., L$$

where $n_{kl} = \mathrm{Re}\{\eta_{kl} e^{j\,\phi_{kl}}\}$. We assume that $E\{n_{kl}^2\} = N_0/2$ for every k and l.

Let's $\alpha_{k1}, \alpha_{k2}, ..., \alpha_{kL}$ denote fading amplitudes correspondent to microdiversity branches of the k-th port. We assume that they are statistically independent with Rayleigh probability density function (pdf) of the instant SNR, $\gamma_{kl} = \alpha_{kl}^2 \frac{E_b}{N_0}$, which has the form

$$p_{\gamma_{kl}}(\gamma_{kl}/\gamma_k) = \frac{1}{\gamma_k} e^{\frac{\gamma_{kl}}{\gamma_k}} \quad (3)$$

which is conditioned on the local mean SNR at the k-th port $\gamma_k = E\{\alpha_k^2\}\frac{E_b}{N_0} = z_k \frac{E_b}{N_0}$, where $z_k = E\{\alpha_k^2\}$ denotes mean-square amplitude values at the k-th port. We assume, also, that, the local mean-square amplitude values at given radio-ports are statistically independent and have log-normal pdf [2]

$$p_{z_{k\ \max}}(z_k) = \frac{10/\ln 10}{\sqrt{2\pi}\sigma_s z_k} \exp\Big(-\frac{(10\log_{10} z_k - \mu_k)^2}{2\sigma_s^2}\Big) \quad (4)$$

where μ_k (dB) denotes mean, and σ_s (dB) denotes standard deviation of the quantity $10\log_{10} z_k$. Let $z_{k\ \max}$ be the largest local mean-square value selected from the K radio-ports, that is $z_{k\ \max} = \max\{z_1, z_2, ..., z_k\}$. One can show that the pdf of $z_{k\ \max}$ has the form

$$p_{z_{k\ \max}}(z_{k\ \max}; K) = \sum_{j=1}^{K} \frac{10/\ln 10}{\sqrt{2\pi}\sigma_s z_{k\ \max}} \times \exp\Big(-\frac{(10\log_{10} z_{k\ \max} - \mu_j)^2}{2\sigma_s^2}\Big) \times \prod_{k\neq j, k=1}^{K} \Big(1 - \frac{1}{2}\mathrm{erfc}\Big(\frac{10\log_{10} z_{k\ \max} - \mu_k}{\sqrt{2}\sigma_s}\Big)\Big) \quad (5)$$

where the mean μ_k (dB) is generally dependent on the distance between the port and the user's location. For any given arrangement of the macrodiversity ports, an average probability density function can be calculated by averaging (5) over all possible locations within the serving cell. At the point which is equidistant from the serving ports which make up a macrodiversity group, area mean of each port can be assumed identical, that is $\mu_k = \mu$ for $k = 1, 2, ..., K$. Therefore (5) becomes [2]

$$p_{z_{k\ \max}}(z_{k\ \max}; K) = \frac{K.10/\ln 10}{\sqrt{2\pi}\sigma_s z_{k\ \max}} \times \exp\Big(-\frac{(10\log_{10} z_{k\ \max} - \mu)^2}{2\sigma_s^2}\Big) \times \Big(1 - \frac{1}{2}\mathrm{erfc}\Big(\frac{10\log_{10} z_{k\ \max} - \mu}{\sqrt{2}\sigma_s}\Big)\Big)^{K-1}. \quad (6)$$

III. The System Performance

The bit error probability (BER) is derived by determining $P_b(\gamma_b; L/z_{k\ \max})$, the bit error probability conditioned on $z_{k\ \max}$, and averaging it over the pdf given in (6)

$$P_b = \int_0^\infty P_b(\gamma_b; L/z_{k\ \max}) p_{z_{k\ \max}}(z_{k\ \max}; K) dz_{k\ \max}. \quad (7)$$

For BPSK, MR Combining is the optimal microdiversity combining technique for fading overcoming. The matched filter outputs at every branch of the k-th port are multiplied with corresponding factor $\alpha_{kl} e^{-j\phi_{kl}}$ (cophasing and weighting branch signals) and then summed at the combiner. The BER conditioned on $\gamma_{k\ \max}$ has the form [2]

$$P_{bmrc}(\gamma_b; L/z_{k\ \max}) = \Big(\frac{1-m}{2}\Big)^L \sum_{i=0}^{L-1} \binom{L-1+i}{i} \Big(\frac{1+m}{2}\Big)^i \quad (8)$$

where $m = \sqrt{\frac{\gamma_{k\ \max}}{1+\gamma_{k\ \max}}}$ and $\gamma_{k\ \max} = z_{k\ \max}\frac{E_b}{N_0}$. This microdiversity technique is optimum if it is assumed that channel parameters $\alpha_{kl} e^{-j\phi_{kl}}$ is estimated perfectly. Otherwise, if fading fluctuations is sufficiently fast to preclude the implementation of coherent detection, the implementation of noncoherent detection or selection diversity techniques may be more adequate.

In the case of the traditional selection diversity technique (SC), the one branch with the largest SNR is selected on each radio-port. BER conditioned on $\gamma_{k\ \max}$ has the form [2]

$$P_{bSC}(\gamma_b; L/z_{k\ \max}) = \\ = \frac{L}{2}\sum_{i=0}^{L}\binom{L}{i}(-1)^i\sqrt{\frac{\gamma_{k\ \max}}{i+\gamma_{k\ \max}}} \quad (9)$$

for BPSK signaling, and [3]

$$P_{bSC}(\gamma_b; L/z_{k\ \max}) = \\ = \frac{L}{2}\sum_{k=0}^{L-1}\binom{L-1}{i}(-1)^k\frac{1}{1+i+\gamma_{k\ \max}} \quad (10)$$

for DPSK signaling.

For practical implementations, however, measurement of SNR may be difficult or expensive, especially for high signaling rates. For this reason, the branch with the largest signal-*plus*-noise is often chosen. We use S + N to denote a signal-plus-noise sample (i.e., not a power measurement). When physically realizing this technique, by sampling the output of a matched filter, the noise is a random variable (SC assumes that noise is constant in all branches). Consequently, this model perform better than traditional SC model because there is opportunity for at least one sample to be better (less noisy) than the average of the samples.The exact expressions for BER conditioned on $\gamma_{k\ \max}$ can be found in [4].

One of the latest modification of the selective microscopic combining technique employing with BPSK signaling is "ar" selection where the signal from the microdiversity branch with the largest value of log-likelihood ratio (LLR) is selected at the each port. This value is equal to the product of fading amplitude α_{kl} and output of the matched filter r_{kl}. The exact expression for BER conditioned on $\gamma_{k\ \max}$ is provided in earlier author's work [2].

Combining microscopic diversity techniques that use two (SC2) or three (SC3) branches with largest amplitudes (or signal-to-noise ratio) for getting transmitted signal were developed. This techniques, denoted as second or third order selection combining is a compromise between MR or EG Combining and traditional SC model and requires a less complex receiver than MR or EG EG Combining, therefore may be implemented regardless of the number of resolvable branch signals available and, consequently, offer better performance (BER) than traditional SC model. The exact expressions for BER conditioned on $\gamma_{k\ \max}$ for SC2 and SC3 microscopic diversity techniques can be found in [3].

The average BER for proposed systems consisted of K radio ports with L microdiversity branches can be obtained by substituting derived conditioned BERs and (5) in (7).

IV. Numerical Results

We compare different systems with macrodiversity selection technique and different microscopic diversity techniques considering the two cases, employing coherent BPSK signaling and noncoherent DPSK signaling. The impact of different values of long-term attenuation (effect of shadowing), as well as different microdiversity techniques, is considered in both cases. We use the value of required SNR over MR com-

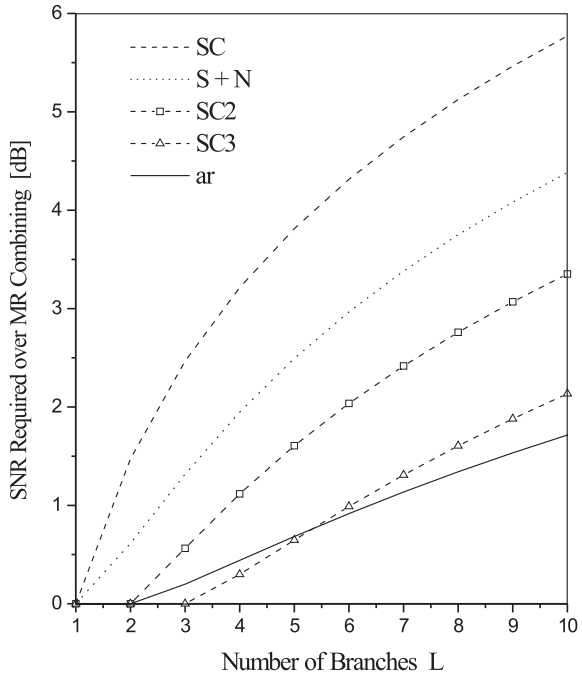


Fig. 1. Required SNR per bit over MR combining for various microdiversity techniques and specified number of diversity branches for average BER of $P_b = 10^{-3}$ employing BPSK signaling

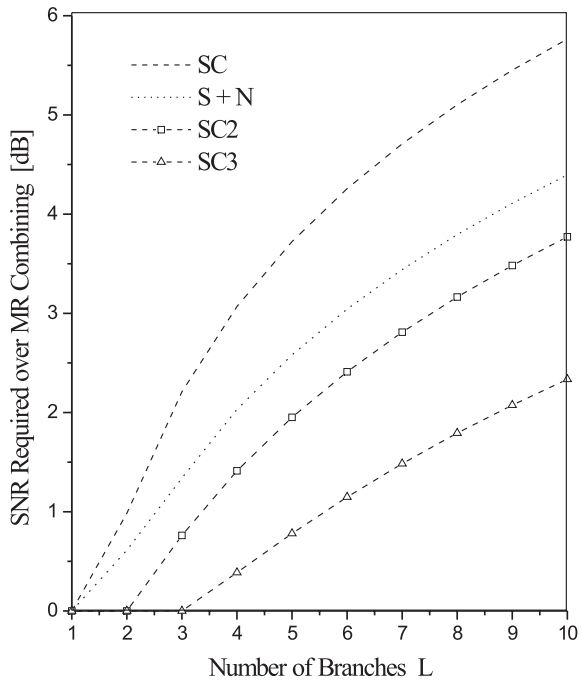


Fig. 2. Required SNR per bit over EG combining for various microdiversity techniques and specified number of diversity branches for average BER of $P_b = 10^{-3}$ employing DPSK signaling

bining, which correspond to the different number of branches that produce the same bit error rate (BER), as the measure of the performances. Fig. 1 and Fig. 2 compare different microscopic selection combining techniques to the optimum combining technique, MR combining in the case of BPSK and DPSK signaling, respectively, where an average BER of $P_b = 10^{-3}$ is chosen. The MRC curve is effectively the horizontal axis. As the number of microdiversity branches increases, the selection combining curves are shown to deviate. The SC technique gives the worst performance while the modified selection (SC, S + N, SC2, SC3, "ar") curves fall intermediately between the SC and MRC. In the case of BPSK signaling it is shown that SC3 has the best performance if the number of branches is 5 or less. As the number of microdiversity branches increases "ar" technique outperforms other selection techniques. The same performance trend is present at different magnitudes of shadowing, as it is shown in [3]. It is also shown that the same performance trend is present in the case of DPSK signaling (barring "ar" selection, which is purely coherent technique). Therefore, the effects of log-normal shadowing and Reyleigh fading affect almost identically both of the signaling cases, coherent (BPSK) and noncoherent (DPSK).

V. Conclusion

In this paper we compare several diversity reception techniques in an additive white Gaussian noise channel in the presence of Rayleigh fading and log-normal shadowing employing coherent (BPSK) and noncoherent (DPSK) digital signaling. Dependence to the required SNR over MR combining of the number of branches to produce the same bit error rate (BER), $P_b = 10^{-3}$, is used as the measure of the performances. The SC technique gives the worst performance while the modified selection (SC, S + N, SC2, SC3, "ar") curves fall intermediately between the SC and MRC. In the case of BPSK signaling it is shown that SC3 has the best performance if the number of branches is 5 or less. As the number of microdiversity branches increases "ar" technique outperforms other selection techniques. The same performance trend is present at different magnitudes of shadowing, as it is shown in [3]. It is also shown that the same performance trend is present in the case of DPSK signaling (barring "ar" selection, which is purely coherent technique). Therefore, the effects of log-normal shadowing and Reyleigh fading affect almost identically both of the signaling cases, coherent (BPSK) and noncoherent (DPSK).

References

[1] J.G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[2] M. Živković, N. Milošević, Z. Nikolić, "Performanse BPSK diverziti sistema u prisustvu fedinga i efekta senke", *Zbornik radova, TELFOR 2002*, str. 303-306, Beograd, 2002.

[3] M. Živković, N. Milošević, Z. Nikolić, "Performanse BPSK diverziti sistema u prisustvu fedinga i efekta senke", *Zbornik radova na CD-u, YUINFO 2003*, Kopaonik, 2003.

[4] E.A. Naesmith, N. C. Beaulieu "New Results on Selection Diversity", *IEEE Trans. Commun.*, vol. 46, pp. 695-704, May 1998.

QPSK System Performance Using Fading Identification Circuit

Zorica Nikolić[1], Nenad Milošević[2], Bojan Dimitrijević[3]

Abstract – **In this paper we consider performance evaluation of QPSK transmission system operating in a frequency nonselective fading channel in the persence of QPSK interference. A algorithm that combines decision feedback and adaptive linear prediction (DFALP) [1] by using tentative coherent detection is used for tracking the phase and amplitude of the fading channel. The channel gain is predicted by adaptive linear predictor employing LMS algorithm, by Kalman filter and by smoother. It will be shown that system employing Kalman filter has the best performance, as expected, over wide range of system, channel and interference parameters.**

Keywords – **fading channels, adaptive filtering, channel identification**

I. Introduction

To detect an information sequence transmitted coherently and reliably over a fading channel, it is necessary to estimate the channel phase and amplitude. This is motivated by the fact that coherent detection of signals over fading channels is superior to non-coherent detection if accurate channel state information (CSI) is available.

One approach was proposed by Moher and Lodge [2] to track frequency nonselective fading channels, where one training symbol is sent for every $K_t - 1$ data symbols, and linear interpolation is used to estimate channel gains. This idea was extended by Irvine and McLane [3] using decision-feedback and noise smoothing filters. However, such filtering results in large decision delay.

It is well-known that fading channels are correlated. Therefore, past channel gain estimates may be used to predict the channel gain using linear prediction theory. This paper investigates fading channel amplitude and phase prediction in the presence of QPSK interference. Three different predictors are used: adaptive linear predictor employing LMS algorithm, Kalman filter, and smoother.

It should be noted that a non-adaptive linear predictor was used by Young and Lodge [4]. However, the algorithm reported in [4] may not outperform a conventional differential detector when the signal to noise ratio (SNR) is less than 20 dB.

In the next Section, we review the fading channel model, and describe the DFALP algorithm using both predictors. Simulation results are presented in Section III.

[1]Zorica Nikolić is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: zora@ni.ac.yu

[2]Nenad Milošević is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: nemilose@ni.ac.yu

[3]Bojan Dimitrijević is with the faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: dbojans@ptt.yu

II. Fading Channel Model

Let I_k denote a binary information sequence, and x_k a low-pass equivalent discrete-time output of the encoder/modulator. The complex signal x_k is transmitted over a frequency-nonselective Rayleigh or Rician fading channel. The received low-pass equivalent discrete-time signal is then

$$y_k = x_k c_k + i_k c_{I_k} + n_k \tag{1}$$

where i_k is complex QPSK interference, c_k, and c_{I_k} are channel gains, complex Gaussian processes with memory. The mean of c_k is $a = Ec_k$. When $a = 0$, the fading channel is Rayleigh. Otherwise it is Ricean. The covariance function of c_k, is

$$r_{k,k-n} = r_n = E\{(c_k - a)(c_{k-n} - a)^*\} \tag{2}$$

in general case. A special case of the above model is the Jakes-Reudink fading channel with r_n given by

$$r_n = r_0 J_0(2\pi f_m n T) \tag{3}$$

where $J_0()$ is the zeroth order Bessel function, T is the symbol period and f_m is the maximum Doppler frequency given by $f_m = v/\lambda$, with v and λ defined as mobile vehicle speed and transmission wavelength, respectively.

The the channel gain c_k, can be divided into two parts: the line-of-sight (LOS) part with average power a^2 and the random scattering part with average power r_0. The K factor is defined as the ratio $K = a^2/r_0$. If r_0 is normalized to 1, then $K = a^2$. The K factor is equal to zero for Rayleigh fading channels and is greater than zero for Rician fading channels. The average signal-to-noise (SNR) ratio per symbol is then

$$\gamma_s = \frac{a^2 + r_0}{\sigma_n^2} \tag{4}$$

where σ_n^2 is the variance of the additive white Gaussian noise (AWGN) n_k.

We now describe the algorithm using decision feedback and adaptive linear prediction (DFALP) [1] to track frequency nonselective (flat) fading channels. If x_k is a known training symbol and if the signal-to-noise ratio (SNR) and signal-to-interference ratio (SIR) are high, a good estimate of c_k can be easily computed as

$$c_k = \frac{y_k}{x_k} = \tilde{c}_k \tag{5}$$

according to Eq.(l), where y_k is the received signal. However, most of the received symbols are not training symbols.

In these cases the available information for estimating c_k can be based upon prediction from the past detected data-bearing symbols $\bar{x}_i$ $(i < k)$. Since a fading channel is usually correlated, it is possible to use an adaptive linear filter to estimate the current complex channel gain c_k using the past detected symbols $\bar{x}_i$ $(i < k)$ and the current observed signal y_k.

The block diagram of the receiver is shown in Fig. 1.

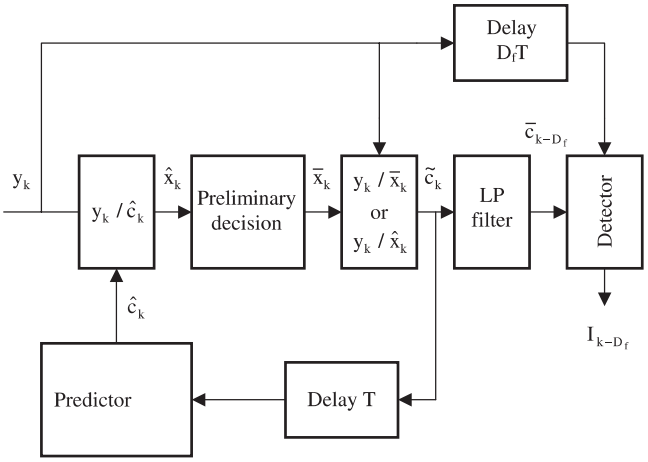


Fig. 1. Receiver block diagram

First, we estimate the data symbol using the predicted channel gain

$$\hat{x}_k = \frac{y_k}{\hat{c}_k} \tag{6}$$

where y_k, is the current received signal plus noise, and $\hat{c}_k$ is a channel estimate given by the linear predictor. Second, we use the minimum distance decision rule

$$\min_{x_k \in D} |\hat{x}_k - x_k| \tag{7}$$

where D is the signal constellation of the modulated complex low-pass equivalent signal x_k. For QPSK, $D = \{e^{jn\pi/4}, n = 1, 3, 5, 7\}$. Let $\bar{x}_i$ denote the detected data symbol, i.e.,

$$|\hat{x}_k - \bar{x}_k| = \min_{x_k \in D} |\hat{x}_k - x_k| \tag{8}$$

Using the detected data symbol $\bar{x}_k$, we formulate a new estimate of the channel gain

$$\frac{y_k}{\bar{x}_k} \tag{9}$$

There exist two possibilities for the decision rule (8). One possibility is that the decision is correct, i.e., $\bar{x}_k = x_k$. Then the estimate $y_k/\bar{x}_k$ would be reliable. On the other hand, if the decision is wrong, i.e., $\bar{x}_k \neq x_k$, the estimate $y_k/\bar{x}_k$ will certainly be very poor. To solve this problem, we use a thresholding idea. In most cases, if the decision is correct, the distance between the predicted channel gain $\hat{c}_k$ and the decision feedback estimate $y_k/\bar{x}_k$ would not be large, i.e., the probability that $|\hat{c}_k - y_k/\bar{x}_k| < \beta$ would be high, where β is a chosen threshold. On the other hand, if the decision is wrong, the distance between the predicted channel gain $\hat{c}_k$, and the decision-feedback estimate $y_k/\bar{x}_k$ would be large, i.e., the probability that $|\hat{c}_k - y_k/\bar{x}_k| > \beta$ would be high. Therefore the corrected channel estimate may be expressed as

$$\tilde{c}_k = \begin{cases} y_k/\bar{x}_k & |\hat{c}_k - y_k/\bar{x}_k| < \beta \\ \hat{c}_k & |\hat{c}_k - y_k/\bar{x}_k| \geq \beta \end{cases} \tag{10}$$

There exists no analytical approach to choosing the threshold β. In our experiments we determined optimal value for the threshold β to be 0.7 for all predictors.

The predicted fading channel gain, for adaptive linear predictor with LMS algorithm, at time k is

$$\hat{c}_k = \sum_{i=1}^{N_{LMS}} b_i^* \tilde{c}_{k-i} \tag{11}$$

where

$$(\tilde{c}_{k-1}, \tilde{c}_{k-2}, ..., \tilde{c}_{k-N})^T = \tilde{\boldsymbol{c}}(k) \tag{12}$$

is a vector of past corrected channel gain estimates and

$$(b_1, b_2, ..., b_N)^T = \boldsymbol{b}(k) \tag{13}$$

are the filter (linear predictor) coefficients at time k. The superscript T stands for transpose. The constant N_{LMS} is the order of the linear predictor. The LMS algorithm computes the filter coefficients $\boldsymbol{b}(k+1)$ of the next time-step using the current filter coefficients $\boldsymbol{b}(k)$ and the estimation error $\tilde{c}_k - \hat{c}_k$. Formally, the algorithm is

$$\boldsymbol{b}(k+1) = \boldsymbol{b}(k) + \mu(\tilde{c}_k - \hat{c}_k)^* \tilde{\boldsymbol{c}}(k) \tag{14}$$

where μ is the adaptation parameter controlling the convergence rate and steady-state error of the algorithm.

In case of Kalman filter, prediction is done in the following manner

$$\hat{c}_{tmp} = \rho \hat{c}_k \tag{15}$$

$$M_{tmp} = \rho^2 M_k + (1 - \rho^2) \tag{16}$$

$$K = \frac{M_{tmp}}{\sigma^2 + M_{tmp}} \tag{17}$$

$$\hat{c}_{k+1} = \hat{c}_{tmp} + K(\tilde{c}_k - \hat{c}_{tmp}) \tag{18}$$

$$M_{k+1} = (1 - K) M_{tmp} \tag{19}$$

The smoother predicts next channel gain as the average value of the pervious channel gains:

$$\hat{c}_k = \frac{1}{N_s} \sum_{i=1}^{N_s} \tilde{c}_{k-i} \tag{20}$$

where N_s is the smoother length.

The corrected channel estimate $\tilde{c}_k$ is then low-pass filtered using a linear phase low-pass filter (LPF) with $2D_f + 1$ taps to reduce the noise. That is, the final channel gain estimate is

$$\bar{c}_{k-D_f} = \sum_{i=0}^{2D_f} h_i \tilde{c}_{k-i} \tag{21}$$

where h_i is the impulse response of a LPF with $2D_f + 1$ taps. The filter cutoff frequency is equal to Doppler frequency.

III. Numerical Results

In following figures we simulated a typical digital cellular telephone channel, where the carrier frequency is 800 MHz, symbol rate is 24000 symbols/sec, and the fading channel is Rayleigh. The low-pass filter length is set to $D_f = 10^3[-7.526(f_mT)^3 + 3.6729(f_mT)^2 - 0.3981(f_mT) + 0.0153]$, which is determined in [1] to be optimal. The interference rate is the same as the rate of useful signal.

Figure 2. shows the error probability of the system using adaptive linear predictor with LMS algorithm as a function of LMS algorithm length. Signal to interference ratio is set to SIR = 20 dB. It can be seen that for high receiver speed (curve b) a optimal value for filter length is $N_{LMS} = 2$. For low receiver speed (curve a) optimal value is $N_{LMS} = 4$. However, minimal error probability is only slightly lower than the error probability for $N_{LMS} = 2$. Because of this fact we chose $N_{LMS} = 2$ for all the following simulations due to simplicity of the receiver.

Error probability as a function of smoother length is shown in Fig. 3. The error probability depends on the smoother length in similar way as in Fig. 2. So, similar conclusions can be made, and we chose $N_s = 2$ for all the following simulations.

Figure 4. shows the error probability as a function of the receiver speed for all the three predictors. On can see that Kalman filter has the best performance regardless of signal to interference ratio and receiver speed. Since it is a optimal structure, the error probability increases with the receiver speed. On the other hand, smoother and LMS algorithm filter are not optimal structures and they have the best performance for particular receiver speed and SIR, i.e. their parameters are best suited for one combination for receiver speed and SIR. Therefore, in order to have the best performance, smoother and LMS algorithm filter should change its parameters, such as length and adaptation factor, adaptively, which can be very difficult.

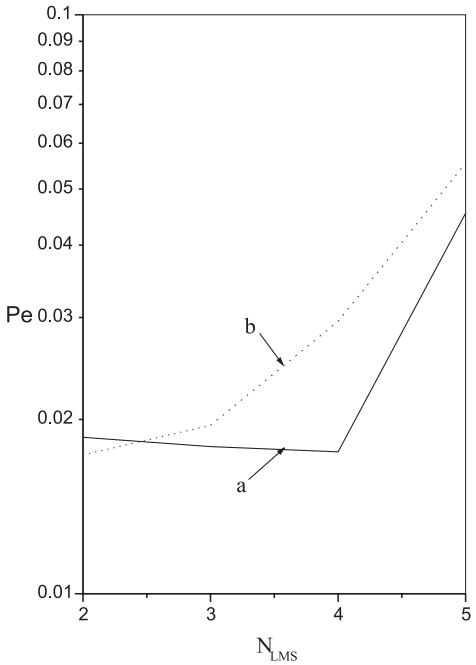


Fig. 2. Error probability as a function of LMS algorithm length: a – $v = 5$ km/h, b – $v = 50$ km/h

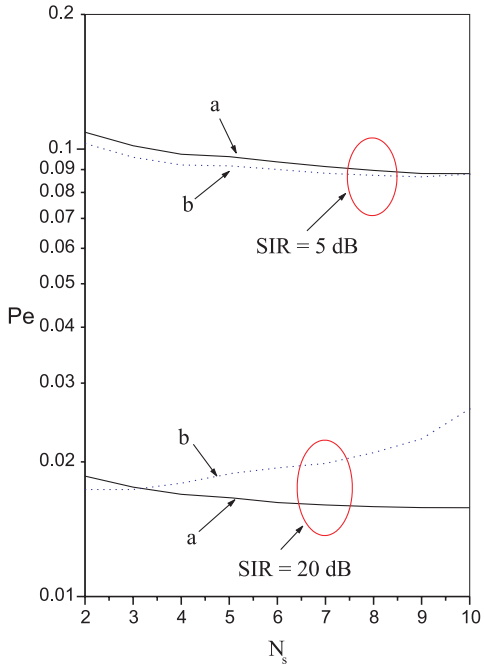


Fig. 3. Error probability as a function of smoother length: a – $v = 5$ km/h, b – $v = 50$ km/h

Error probability as a function of signal to interference ratio is shown in Fig. 5. Kalman filter again has better performance than the other two predictors. For high SIR, smoother and LMS algorithm filter have the same error probability. If

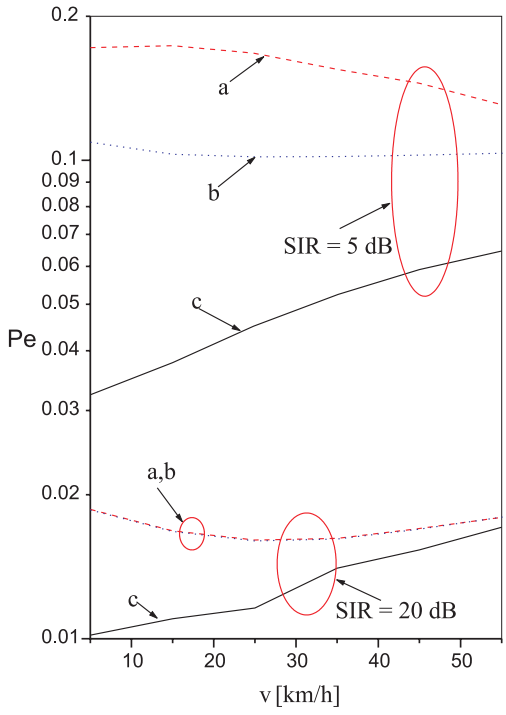


Fig. 4. Error probability as a function of receiver speed: a – LMS algorithm, b – smoother, c – Kalman filter

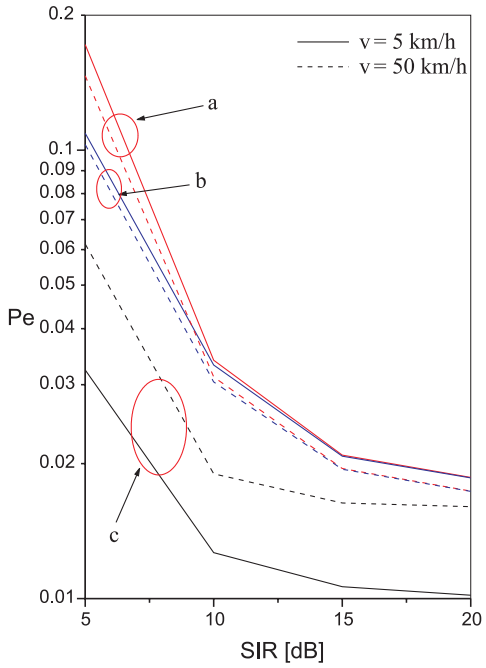


Fig. 5. Error probability as a function of signal to interference ratio: a – LMS algorithm, b – smoother, c – Kalman filter

we consider the influence of receiver velocity, the same conclusions can be made as in Fig. 4.

IV. Conclusion

In this paper we considered performances of QPSK transmission system operating in a frequency nonselective fading channel and in the presence of QPSK interference. The channel gain is predicted by adaptive linear predictor employing LMS algorithm, Kalman filter, and smoother. It was shown that both LMS algorithm and smoother length may be set to 2 since the error probability for this length is not much higher than the minimal error probability. Kalman filter is a optimal structure and has the best performance, as expected. On the other hand, smoother and LMS algorithm filter are not optimal structures and, in order to have the best possible performance, they should change its parameters, such as length and adaptation factor, adaptively, which can be very difficult.

References

[1] Y. Liu and S. D. Blostein, “Identification of Frequency Nonselective Fading Channels Using Decision Feedback and Adaptive Linear Prediction”, *IEEE Transactions on Communications*, Vol.43, No.2/3/4, 1995, pp.1484-1492.

[2] M. L. Moher and J. H. Lodge, “TCMP-A modulation and coding strategy for Rician fading channels”, *IEEE Journal on Selected Areas in Communications*, Vol.7, No.9, 1989, pp.1347-1355.

[3] G. T. Irvine and P. J. McLane, “Symbol-aided plus decision-directed reception for PSK/TCM modulation on shadowed mobile satellite fading channels”, *IEEE Journal on Selected Areas in Communications*, Vol.10, No.8, 1992, pp.1289-1299.

[4] R. J. Young and J. H. Lodge, “Linear-prediction-aided differential detection of CPM signals transmitted over Rayleigh flat fading channels”, in *IEEE Vehicular Tech. Conf.*, 1990, pp.437-442.

[5] R. Haeb and H. Meyr, “A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels”, *IEEE Transactions on Communications*, Vol.37, No.7, 1989, pp.748-754

Precision and Noise Immunity of the Communication Systems

Ilia G. Iliev[1], Slavtcho D. Lishkov[2]

Abstract – **The article proposed an estimation of the functional precision and noise immunity of a communication system, taking in to account the instability factors. Graphic dependencies for error probability, as a function of the reference signal phase dispersion in coherent demodulation of 2PSK, are derived. The results can be used in designing, exploitation and adjusting of communication systems.**

Keywords – **Noise immunity, Communication system**

I. Introduction

The main factor impacting on the precision and the noise immunity of the communication system are: schematics, production tolerance and the operating conditions [1,4]. The system precision is defined as the ability of the system to execute the given functions at a definite proximity to its ideal parameters. The functional dependence of the system may be show as a function of:

$$y = \varphi(S, X) \tag{1}$$

where S are the input signals, X – the system parameters. Through introducing the concept of ideal system, an estimation of the reproduction accuracy of (1) may be done. Formula (1) is modelled when absolute accuracy is set:

$$\Delta y = y_n - y = \varphi(S_n, X_n) - \varphi(S, X) \tag{2}$$

where $y_n = \varphi(S_n, X_n)$ is the nominal value of y. The functional precision is determined in the field of limit $[a, b]$ of y:

$$a \leq y \leq b. \tag{3}$$

If the function $\varphi(x_1, x_2, ..., x_n)$ is derivate, the absolute error Δy of the function φ is:

$$\Delta y = \sum_{i=1}^{n} \Big|\frac{\partial y}{\partial x_i}\Big| \Delta x_i \tag{4}$$

where Δx_i usually in the design process is given the boundary value of Δy, that may be done by determining the boundary values of Δx_i, satisfying the inequality:

$$\Delta y \leq \sum_{i=1}^{n} \Delta x_i. \tag{5}$$

If it is accepted that the influence of all x_i is equal and for a given Δy:

$$\Delta x_i = \Delta y \Big/ \sum_{i=1}^{n} \Big|\frac{\partial y}{\partial x_i}\Big|, \qquad \Delta x_i = \Delta y \Big/ n \Big|\frac{\partial y}{\partial x_i}\Big|. \tag{6}$$

[1]Ilia Georgiev Iliev, Dept. of Radiotechnique in Faculty of Communications and Communication Technologies in TU - Sofia E-mail igiliev@vmei.acad.bg

[2]Slavtcho Dimitrov Lishkov, Dept. of Radiotechnique in Faculty of Communications and Communication Technologies in TU - Sofia

II. Calculation of the System Parameters Limits

The calculation of the limits of the system parameters may be done by using experimental-statistic modelling. The following problem is solved:

$$y = y_n + \Delta y. \tag{7}$$

For a given Δy and known nominal characteristics, from equation (7) are derived the necessary changes of the factors Δx_i $(i = 1, ..., n)$.

There could be the following cases:

A.) The output index (signal magnitude, energy) depends on only one factor by a linear law:

$$y = \varphi(x_1) \rightarrow y = b_0 + b_1 x_1 \rightarrow \Delta y = b_1 \Delta x_1. \tag{8}$$

B.) Linear dependence of two factors:

$$y = \varphi)x_1, x_2) \rightarrow y = b_0 + b_1 x_1 + b_2 x_2 \rightarrow \\ \rightarrow \Delta y = b_1 \Delta x_1 + b_2 \Delta x_2. \tag{9}$$

Equation (10) is transformed in:

$$\frac{\Delta y}{b_1 b_2} = \frac{1}{b_2}\Delta x_1 + \frac{1}{b_1}\Delta x_2. \tag{10}$$

The parameter Δx_2 is defined with:

$$\Delta x_2 = \frac{b_2 \Delta y - b_1 \Delta x_1}{b_2}. \tag{11}$$

By analogy, Δx_1 is determined. It is necessary to remark that the computed values of the parameters must satisfy the technology limits:

$$-c_1 \leq \Delta x_1 \leq c_1, \quad -c_2 \leq \Delta x_2 \leq c_2.$$

For the general case for n factors, the equation is:

$$\Delta y = \sum_{i=1}^{n} b_i \Delta x_i. \tag{12}$$

Solution of (12) is possible, when it is used the method of limit increasing or the method of succession approximation where are computed repeatedly the values of b_i. These are the values of partial derivatives and define the order of impact every variable Δx_i.

III. The Noise Immunity as a Criterion for Functional Precision

The system noise immunity may be used as a criterion for defining its functional precision. In this case the main parameters is the relation between the average power of the signal

and noise power – signal-to-noise ratio (SNR) at the which is obtained a given error probability $Pe = \varphi(q)$. At the practical realization and the exploitation of the communication system is observed worsening of the noise immunity in comparison with the "ideal" (theoretical) system. The reason for this, are divergence of the real characteristics from the ideal under the influence of instability factors. These differences between the ideal and the real system lead to an energy and reliability loss:

$$\gamma_E = \frac{E_R}{E_N} > 1 \qquad \gamma_P = \frac{P_R}{P_N} > 1 \quad \text{for}$$
$$h_R^2 = \frac{E_R}{N_0} \qquad h_n^2 = \frac{E_n}{N_0}. \tag{13}$$

E_R and E_n are the signal energies of the real and ideal system, and N_0 is the power spectral density of the noise. If K_R and K_n are transmission coefficients of the real and ideal channel with additive white Gaussian noise (AWGN), the error probability is given by:

$$Pe_R = 0.5\text{erfc}\Big[K_R\sqrt{0.5h_R^2(1-\rho)}\Big] \tag{14}$$

for a real communication system and

$$Pe_N = 0.5\text{erfc}\Big[K_N\sqrt{0.5h_N^2(1-\rho)}\Big] \tag{15}$$

for an ideal communication system and

$$\text{erfc}(u) = \frac{2}{\sqrt{\pi}}\int_u^\infty \exp(-x^2)dx.$$

If $Pe_R = Pe_N$, the necessary SNR is:

$$q = \frac{K_R}{K_N}\Big[\sqrt{0.5h_N^2(1-\rho)}\Big].$$

ρ is a correlation coefficient of the modulated signals and error probability is:

$$Pe_N = 0.5\text{erfc}\Big[\frac{K_R}{K_N}\sqrt{0.5h_N^2(1-\rho)}\Big]. \tag{16}$$

Figure 1 shows relation $Pe = \varphi\Big(\frac{K_R}{K_N}\Big)$ for $\rho = -1$ 2PSK.

The error in the recovering the phase of the reference signal at the coherent demodulation impacts on the noise immunity. If $\Delta\varphi$ is the phase error of the reference signal; the phase of received symbol $\varphi_i \in [0.\pi]$ and the phase of reference signal φ_0, consequently the signal after the correlator of the ideal receiver is:

$$Z_s(\varphi) = \frac{2}{N_0}\int_0^T S(t,\varphi_i)S_0(t,\varphi_0+\Delta\varphi)dt.$$

The error probability as a function of the phase error is equal to:

$$Pe(\Delta\varphi) = 0.5\text{erfc}\Big(\sqrt{h^2}\cos(\Delta\varphi)\Big). \tag{17}$$

The average error probability for all values of $\Delta\varphi$ in $-\pi \leq$

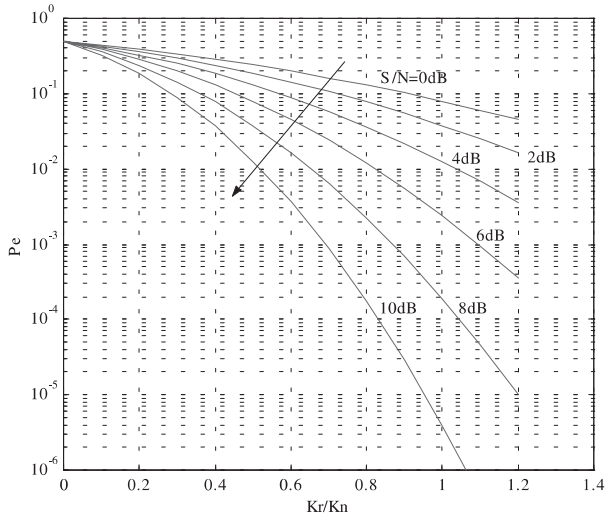


Fig. 1. Another example is related to the error correction coding. The code parameters (n,k) and the SNR is reduced in $h^2 = \frac{k}{n}h_N^2$. In that way the equivalent error probability is computed

$\Delta\varphi \leq \pi$ is:

$$Pe = \int_{-\infty}^{+\infty} W(\Delta\varphi)Pe(\Delta\varphi)d\Delta\varphi =$$
$$= \int_{-\pi}^{\pi} W(\Delta\varphi)Pe(\Delta\varphi)d\Delta\varphi. \tag{18}$$

The distribution $W(\Delta\varphi)$ of $\Delta\varphi$ may by uniform or normal. Because of the filtering used in optimal receiver, the distribution density approximates with the Gaussian low with dispersion σ_φ^2 and mean value $m_\varphi = 0$:

$$W(\Delta\varphi) = \frac{1}{\sqrt{2\pi}\sigma_\varphi}\exp\Big(-\frac{\Delta\varphi}{2\sigma_\varphi^2}\Big). \tag{19}$$

The mean value of $\cos(\Delta\varphi)$ is:

$$\overline{\cos(\Delta\varphi)} = \int_{-\infty}^{\infty}\cos(\Delta\varphi)W(\Delta\varphi)d\Delta\varphi = \exp\Big(\frac{\sigma_\varphi^2}{2}\Big). \tag{20}$$

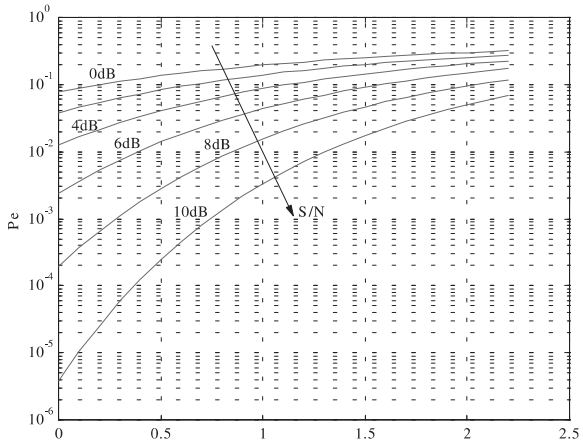


Fig. 2.

Figure 2 show the error probability by demodulation of 2PSK modulated signal in relation to the phase error dispersion by a parameter SNR.

If it is assumed that the signal magnitude after the receiver correlator and $\Delta\varphi$ has a Gaussian distribution, consequently the probability for falling into the interval $E_1 < E < E_2$ and $\varphi_1 < \Delta\varphi < \varphi_2$ is:

$$P(E, \Delta\varphi \in R) = \frac{1}{4}\Big[\text{erf}\Big(\frac{E_2 - \bar{E}}{\sqrt{2}\sigma_E}\Big) - \text{erf}\Big(\frac{E_1 - \bar{E}}{\sqrt{2}\sigma_E}\Big)\Big] \times \Big[\text{erf}\Big(\frac{\varphi_2 - \varphi}{\sqrt{2}\sigma_\varphi}\Big) - \text{erf}\Big(\frac{\varphi_1 - \varphi}{\sqrt{2}\sigma_\varphi}\Big)\Big].$$

IV. Conclusion

The article proposed an estimation of the functional precision and noise immunity of a communication system, taking in to account the instability factors. Graphic dependencies for error probability, as a function of the reference signal phase dispersion in coherent demodulation of 2PSK, are derived. The results can be used in designing, exploitation and adjusting of communication systems.

References

[1] Венцель Е. С, *Теория вероятности*, Висшая школа, Москва, 2000

[2] Ненов Г., Лишков Сл., Бърз алгоритъм за определяне разпределението на параметрите на комуникационни устройства и системи, *Юбилейна научна сесия, 50 години ТУ - София*, 1998

[3] Ненов Г., Лишков Сл., Статистически екстремални задачи при разработване на радиоелектронни устойства, КЕКС, ТУ - София, 1999

[4] Шапиро С., Хаи Г., *Статистически модели в инженерных задачах*, Мир, Москва, 1969

[5] Жовинскии, А. Н. *Инженерны експрес-анализ случайных процесов*, Энергия, 1988

Functional Precision and Work Ability of the Communication System

Ilia G. Iliev[1], Slavtcho D. Lishkov[2]

Abstract – **The article discusses the communication system precision and work ability problems in their designing, manufacturing and exploitation. It is connected with the defining the fields of functioning, when techno-economical factors are given. In the article, it is proposed the using of Shapiro-Wilk test for express estimation of the normal density of distribution. The method is illustrated by an analysis of the precision and work ability of UHF radio transmitters.**

Keywords – **Work ability, Shapiro-Wilk test**

I. Introduction

Every task for communication system design, exploitation, and manufactory is related to the defining of the fields of functioning, when techno-economical factors are given. For example they could be: data errorless transmission, noise performance, reliability, electromagnetic compatibility etc.

In the general case the system S is assumed, that functions if the condition is true:

$$a_i \le y_i \le b_i, \qquad i = 1, 2, 3, ...$$

where y_i is the nominal value of the system indicator (power magnitude, frequency etc.). The interval $[a, b]$ defines the field limits – the deviation from the nominal value of y_i. It is expressed by the reliability probability:

$$P_D = p(a < \bar{y} < b) = \int_a^b W(y)dy. \tag{1}$$

Let the following probabilities are given:

$$\begin{aligned} Pa_i &= p(y < a_i) = \int_{-\infty}^{a_i} W(y)dy, \\ Pb_i &= p(y < b_i) = \int_{-\infty}^{b_i} W(y)dy. \end{aligned} \tag{2}$$

The precision and work ability estimation of the system S, is made by checking two alternative hypothesis:

H_1 – system indicators are in the field of limits;

H_2 – system indicators are out of the $[a, b]$ interval.

[1]Ilia Georgiev Iliev, Dept. of Radiotechnique in Faculty of Communications and Communication Technologies in TU - Sofia E-mail igiliev@vmei.acad.bg

[2]Slavtcho Dimitrov Lishkov, Dept. of Radiotechnique in Faculty of Communications and Communication Technologies in TU - Sofia

II. Precision and Work Ability

The problem for defining the work ability is in obtaining the one dimensional distribution law $W(y_i)$, as a function of the parameters x_1, $x_2, ..., x_n$. Because of the great number of factors x_i $i = 1, 2, ..., n$, no one among them cannot be dominating. Consequently, the random parameters are distributed in a Gaussian law $W(y_i)$ according to the Central Limit Theorem

$$W(y_i) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\Big(-\frac{(y_i - \bar{y})^2}{2\sigma_i^2}\Big), \tag{3}$$

where $\bar{y}$ is the mean value of y_i. If the parameters $\{x_1, x_2, ..., x_n\}$ are measured the check if the probability density has a normal distribution may be done through the Shapiro-Wilk algorithm [3,7]:

1. From a sample of N elements a variation sequence is build: $\{x_1, x_2, ..., x_n\}$;
2. The weight sum b is defined by the coefficients a_{n-i+1}:

$$\begin{aligned} b &= a_n(x_n - x_1) + a_{n-1}(x_{n-1} - x_2) + \cdots + a_{n-k+1}(x_{n-k+1} - x_k) \\ &= \sum_{i=1}^{k} a_{n-i+1}(x_{n-i+1} - x_i), \end{aligned} \tag{4}$$

   where $k = N/2$ for even N and $k = (N-1)/2$ for odd N.
3. The statistic criterion is:

$$V = \frac{b^2}{S_x}, \quad S_x = \sum_{i=1}^{N} x_i^2 - \frac{1}{N}\Big(\sum_{i=1}^{N} x_i\Big)^2; \tag{5}$$

4. The reliability probability P_D is given and it is compared to the critical value $P_V(N, P_D)$. If $V > P_V(N, P_D)$ it is assumed the normal distribution; for $V < P_V(N, P_D)$ – the hypothesis is denied.

For the given sample, it is computed the statistic estimations – mean value μ_y and dispersion σ_y.

The allowable boundaries of the parameters with the reliability P_D are determined. For a normal distribution law the probability for falling of the y_i in the field of limits $[a, b]$ is:

$$\begin{aligned} P_D &= P(a \le \bar{y} \le b) \\ &= \frac{1}{2}\Big|\Big[\mathrm{erf}\Big(\frac{b - \mu_y}{\sqrt{2}\sigma_y}\Big) - \mathrm{erf}\Big(\frac{a - \mu_y}{\sqrt{2}\sigma_y}\Big)\Big]\Big| \end{aligned} \tag{6}$$

for $\mathrm{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u \exp(-x^2)dx$.

If the changing of the output parameters is symmetrical $b - \mu_y = \mu_y - a = \Delta$ for probability is derived:

$$P(a \le \bar{y} \le b) = \mathrm{erf}\Big(\frac{\Delta}{\sqrt{2}\sigma_y}\Big). \tag{7}$$

For a given σ_y and P_D, the unknown quantities Δ are computed:

$$\begin{aligned} \Delta_1 &= b_1 - \mu_y = \sigma_y \quad a_1 = \mu_y - \sigma_y, \quad \text{for } P_D = 0.68 \\ \Delta_2 &= b_2 - \mu_y = 2\sigma_y \quad a_2 = \mu_y - 2\sigma_y, \text{ for } P_D = 0.95 \end{aligned} \tag{8}$$

III. Results

The task consists of defining the precision and work ability of UHF radio transmitters. They are controlled in two basic parameters – output power P and carrier frequency f. If the two dimensional distribution of P and f is normal, so it is equal to:

$$W(P,f) = \frac{1}{2\pi\sigma_P\sigma_P\sqrt{1-r^2}} \times \exp\left[-\frac{\frac{(P-P_0)^2}{\sigma_P^2} - \frac{2r(P-P_0)(f-f_0)}{\sigma_P^2\sigma_f^2} + \frac{(f-f_0)^2}{\sigma_f^2}}{2(1-r^2)}\right],$$

where r is a correlation coefficient. If the frequency changes are in small interval in which the output power stays unchanged, consequently these two parameters are independent and $r = 0$. Then the probability of the power and frequency for being in the field R (Fig. 1) is:

$$P(E, f \in R) = \frac{1}{4}\Big|\Big[\mathrm{erf}\Big(\frac{P_h - P_0}{\sqrt{2}\sigma_P}\Big) - \mathrm{erf}\Big(\frac{P_l - P_0}{\sqrt{2}\sigma_P}\Big)\Big] \times \Big[\mathrm{erf}\Big(\frac{f_h - f_0}{\sqrt{2}\sigma_f}\Big) - \mathrm{erf}\Big(\frac{f_l - f_0}{\sqrt{2}\sigma_f}\Big)\Big]\Big|.$$

The results from the power and frequency measurements of the sets of $N = 10$ are given in Table 1.

The distribution of $y(P,f)$ is checked through Shapiro-Wilk algorithm. For $N = 10$ and $a_{10} = 0.574$, $a_9 = 0.33$, $a_8 = 0.21$, $a_7 = 0.12$, $a_6 = 0.04$, the relation of $P_V(N, P_D)$ is shown in Fig. 2 [3].

Making the computation in the algorithm, given in the previous section, for the parameter V is derived $V_P = 0.9345$.

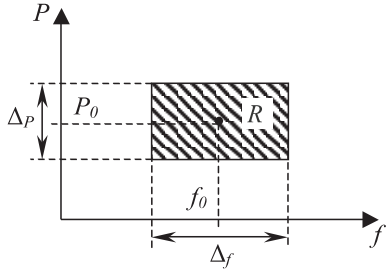


Fig. 1.

Table 1.

No	P [W]	f [MHz]
1	2.16	53.3257
2	1.65	53.3262
3	2.31	53.3259
4	2.7	53.3258
5	1.76	53.3261
6	2.39	53.3267
7	2.55	53.3248
8	1.38	53.3264
9	1.5	53.3256
10	2.2	53.3265
	$\mu_P = 2.06$ $\sigma_P = 0.4576$	$\mu_f = 53.3260$ $\sigma_f = 5.4507\mathrm{e}-04$

In the conditions $V_P > P_V(N, P_D)$ and power values distributed in a Gaussian law, from Fig. 2 is found that the reliability probability must be greater than $P_D \ge 0.605$.

The power deviation from the nominal value $P_n = 2$ W is equal to ± 1 dB. Then $P_l = 1.59$ and $P_h = 2.41$ W. The reliability interval $\Delta_P = 0.41$. For the view of the mean value of the power is approaching to the nominal value so that the formula (7) is used. The reliability probability of the power is equal to $P_D = 0.6297$, when the dispersion $\sigma_P = 0.4576$ (Table 1).

In the same way is computed $V_f = 0.9473$. If the frequency accuracy is ± 100 Hz and the nominal value is $f_n = 53.325$ MHz, then the reliability interval is $\Delta_f = 1\mathrm{e}\text{-}4$ MHz. For a dispersion $\sigma_f = \Delta_f$ (formula 8), the reliability probability is $P_D = 0.68$ and $P_V(N, P_D) = 0.9234$(Fig. 2). The condition $V_f > P_V(N, P_D)$ is true, so that it is assumed that the frequency values have normal distribution.

For this parameter, the mean value and nominal value of the frequency are different. The upper and bottom limit of the frequency precision are equal to $f_h = 53.3251$ MHz and $f_l = 53.3249$ MHz. The reliability probability is the defined by (6) for $f_0 = \mu_f = 53.326$ MHz $P_D = 0.0308$.

The results show the following:

1. The probability of the output power of transmitters sample for falling in the limits interval is equal to $P_D =$

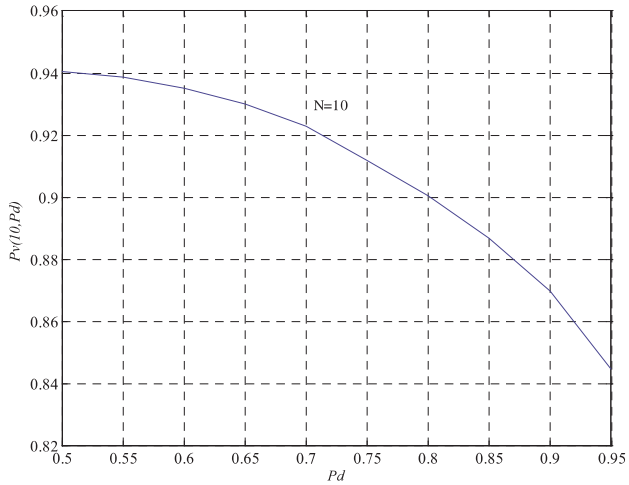


Fig. 2.

0.6297 – 62% from manufacturing devices will have the necessary work ability with pre-adjustment of the amplification;

2. The carrier frequency must be adjusted additionally because of the deviation from the nominal value is more than 1 KHz. The reliability probability is less than the necessary $P_D = 0.68 - P_D = 0.0308$.

IV. Conclusion

The proposed algorithm allows express statistic estimation of the precision and the work ability of communication systems and devices. It is used Shapiro-Wilk's method for defining the type of distribution of the controlled parameters. The algorithm is illustrated thought a study of an UHF radio transmitters' sample. The given algorithm can be used in design, exploitation and adjustment of communication devices and systems

References

[1] Венцель Е. С, *Теория вероятности*, Висшая школа, Москва, 2000

[2] Ненов Г., Лишков Сл., Бърз алгоритъм за определяне разпределението на параметрите на комуникационни устройства и системи, *Юбилейна научна сесия, 50 години ТУ - София*, 1998

[3] Шапиро С., Хаи Г., *Статистически модели в инженерных задачах*, Мир, Москва, 1969

[4] Жовинскии, А. Н. *Инженерны експрес-анализ случайных процесов*, Энергия, 1988

[5] Кривошейкин А.В., *Точность параметров и настройка радиоэлектронных цепей*, Москава, Радио и Звязь, 1988

[6] Иванов А.З., Круг К.Г., *Статистические методы в инженерных исследованях*, Москава, Радио и Звязь, 1989

[7] Shapiro S. S. and Wilk M. B., An analysis of variance test for normality (complete samples), *Biometrika*, 1965, 52, 3 and 4, pages 591-611.

Synchronization in TETRA Networks

Plamen Hristov Vanchev[1], Marin Simeonov Marinov[2]

Abstract – **Digital FIR-Filter with integration on synchronization word is synthesized for communication between MSs-BSs in TETRA Networks. For solvability of the comparative level of the filter output signals and the digital integrator, the Neiman -Pierson criterion is used. In the latter according to the standards the misrouting probability (probability of false alarm) is given and the probability of successful service completion (probability of true detection) is maximized. The filter coefficients are calculated with invariant transformation of filter pulse characteristic. The sampled frequency is obtained by virtue of the criterion for receiving of maximum information from the input signal. The values of digital and analogue thresholds of the filter and the accumulator are presented for reliable synchronization making, according to the ETSI standards. Conclusions and recommendations are made for the use of the presented synchronization method in TETRA Networks.**

Keywords – **TETRA, FIR-Filter, p/4-DQPSK, Synchronization, Air Interface**

TETRA (Terrestrial Trunked Radio) is an all-digital spectrum efficient trunked LMR (Land Mobile Radio) radio system that uses a 4-slot TDMA (Time Division Multiple Access) technology. This technology provides in a 25 kHz physical radio channel 4 simultaneous logical voice and/or data paths. Alternatively, slots can be concatenated for high-speed data. The on-air data rate in the 25 kHz channels is 36 kbit/s. TETRA uses 7.2 kbit/s speech CODEC, providing clear speech quality [5-7].

I. Introducion

TETRA is defined by ETSI (the European Telecommunication Standards Institute) and is designed for PMR (Private Mobile Radio) and PAMR (Public Access Mobile Radio) utilization.

ETSI provides a suite of standards that includes Design Guides, Conformance Specifications, Air Interface Specifications and Interoperability Specifications, which describe both trunked and direct mode operations. TETRA provides voice and data communication with short data services, circuit mode and packet mode data services.

As well as ETSI a number of major TETRA users have generated specific-to-system specifications that further define the standard. TETRA is aimed at markets that range from very large regional and national public, secure and emergency service networks to on-site systems. Customers include network providers, police, fire and ambulance services, security services, gas, water and electricity utilities, mass transit authorities and operators, airports, ports and the general professional radio market.

The TETRA technology provides one-to-one or one-to-many voice and/or data communication with 'traditional' simplex PMR push-to-talk operation or duplex, cellular type, operation. Trunked mode (TMO) operation is the normal state but various managed and unmanaged direct modes (DMO) for direct communication or, via gateways, subscriber to system communication are possible. Very fast call set-up times (< 0.3 ms) are standard.

Standardization for TETRA commenced in 1990 with first phase standards completed in 1995. Harmonized frequencies across Europe were allocated by CEPT in 1996.

TETRA is now an established European standard, as well as an accepted standard in Russia, China, in many Pacific Rim and South American countries and it is within the standard acceptance procedure in the US. In particular TETRA Network will be built for Bulgarian Ministry of Defense and Bulgarian Ministry of Interior. The project will be implemented in a couple of years. This project will allow organizing 24 hours secure and reliable connection between network subscribers. The interfaces will be suitable for all types of Radio Network [7].

II. TETRA Air Interface

There are many possible variations in TETRA network topology ranging from small, on-site applications to very large, regional and national systems. In general, however, the systems have many common elements [2].

The simplest TETRA network sites consist of one or more base station transmitter/receivers, linked to a local switching center (LSC) via modems. The next most complex network site consists of one or more base station transmitter/receivers linked to the LSC directly, gateways to other systems, databases and a gateway to the rest of the TETRA network. For larger systems, a main switching center (MSC) is used where centralized network and subscriber management sub-systems are connected. Each of these basic elements is further described below.

The linking between MS to BS in TETRA is into $(380 \div 520)$ MHz Frequency Band.

When used in dedicated TETRA frequency bands, TETRA MSs shall transmit in the TETRA uplink frequency band, and TETRA BSs shall transmit in the TETRA downlink frequency band. The uplink and downlink frequency bands are

[1]Plamen Hristov Vanchev, National Military University, Aviation Faculty, Department of Electronics, Communication and Navigation Equipment of the Air Craft, Dolna Mitropolia-5856, Pleven, Bulgaria, E-mails: pvanchev@af-acad.bg, pvanchev@yahoo.com

[2]Marin Simeonov Marinov, National Military University, Aviation Faculty, Department of Electronics, Communication and Navigation Equipment of the Air Craft, Dolna Mitropolia-5856, Pleven, Bulgaria, E-mails: mmarinov@af-acad.bg, mmarinov2000@yahoo.com

of equal width. Their edges shall be as follows [6]:

$$\begin{cases} F_{up,min} - F_{up,max} \text{ (MHz): } \textit{mobile transmit, base receive;} \\ F_{dw,min} - F_{dw,max} \text{ (MHz): } \textit{base transmit, mobile receive.} \end{cases} \quad (1)$$

The TETRA RF carrier separation shall be 25 kHz. The uplink and downlink bands are divided into N RF carriers. In order to ensure compliance with the radio regulations outside the band, a guard band of G kHz is needed at each side of both uplink and downlink bands.

The basic radio resource is a timeslot lasting 14,167 ms (85/6 ms) transmitting information at a modulation rate of 36 kbit/s. This means that the time slot duration, including guard and ramping times, is 510 bit (255 symbol) duration.

The following subsections briefly introduce the structures of hyper-frame, multi-frame, frame, timeslot, and burst, as well as the mapping of the logical channels onto the physical channels, shown in Fig. 1 [5].

The center frequencies of uplink RF carriers, $F_{up,c}$ shall be given by [6]:

$$F_{up,c} = F_{up,min} + 0,001G + 0,025(c - 0,5) \text{ (MHz)}, \quad \text{for} \quad c = 1, ..., N \quad (2)$$

and the corresponding center frequency of downlink RF carriers, $F_{dw,c}$, shall be given by [6]:

$$F_{dw,c} = F_{up,c} + D \quad \text{for} \quad c = 1, ..., N \quad (3)$$

When a TETRA system is operated in frequency bands used for analogue Private Mobile Radio (PMR), the uplink and downlink transmit and receive center frequencies and the duplex spacing (D) will be allocated by the National Regulatory Administration (NRA).

In all frequency bands, the TETRA stations use a fixed duplex spacing D.

The access scheme is TDMA with 4 physical channels per carrier. The carrier separation is 25 kHz.

For synchronization procedure we must know the radio interface between MS and BS (AI – Air Interface between MS and BS) for a user. The TDMA frame is separated into four

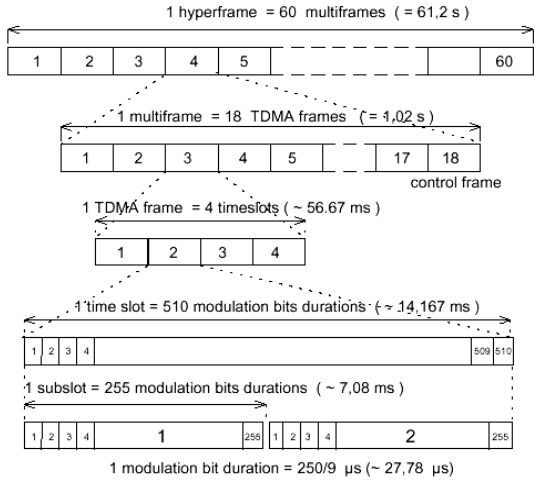


Fig. 1. TETRA TDMA structure

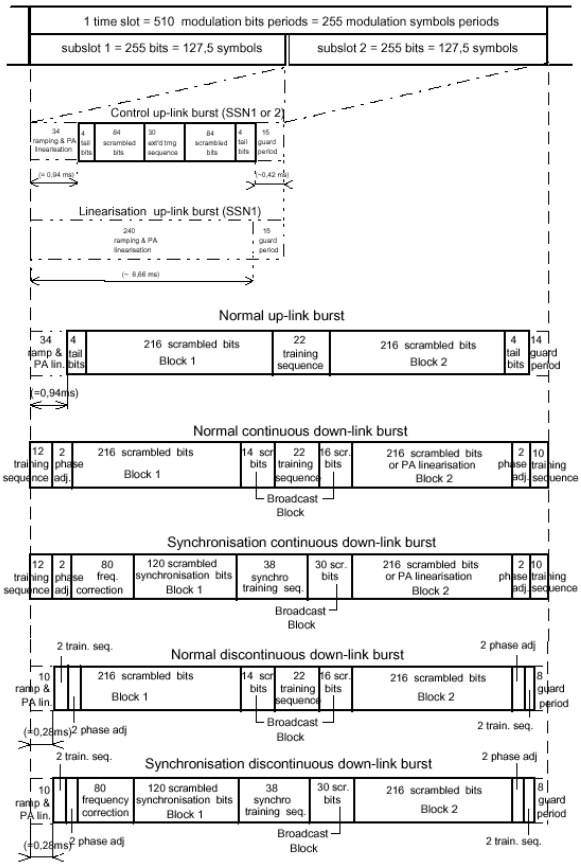


Fig. 2. Types of bursts for each TDMA Frame

time intervals, each corresponding to the client, connected with the cell (Fig. 1 and Fig. 2). The duration (type of bursts) for single TDMA frame is presented in Fig. 2 [6].

In all frequency bands, the TETRA stations use a fixed duplex spacing D.

III. TETRA Frame Synchronization Algorithm

The modulation scheme is $\pi/4$-shifted Differential Quaternary Phase Shift Keying. ($\pi/4$-DQPSK) with root-raised cosine modulation filter and a roll-off factor of 0,35 and the modulation rate is 36 kbit/s. In a $\pi/4$-QPSK transmitter, the input bit stream is partitioned by a serial-to-parallel converter into two parallel data streams $m_{I,k}$ and $m_{Q,k}$, each with a symbol rate equal to half of the incoming bit rate. The k-th in-phase (I_k) and quadrature pulses (Q_k) are produced at the output of the signal mapping circuit over time. They represent rectangular pulses over one symbol duration having amplitudes given by [8]

$$\dot{S}(t) = \begin{cases} I_k = I_{k-1}\cos\phi_k - Q_{k-1}\sin\phi_k, \\ Q_k = I_{k-1}\sin\phi_k + Q_{k-1}\cos\phi_k, \end{cases} \quad (4)$$

where the phase shift ϕ_k is related to the input symbols $m_{I,k}$ and $m_{Q,k}$, according to table 1.

The signal after demodulator can be presented with (4). This analytical expression allows using non-recursive algorithm for synchroword processing. According to the ETSI

Table 1.

Information bits $m_{I,k}$ and $m_{Q,k}$	11	01	00	10
Phase shift ϕ_k	$\pi/4$	$3\pi/4$	-$3\pi/4$	-$\pi/4$

references, multipath delays are not greater than 5 μs. It requires foreseeing the same duration into digital filter synthesis. The time interval can be taken from synchroword length, which is different, according to the information type as TSI (TETRA Subscriber Identities) and TMI (TETRA Management Identity, presented in Fig. 3 [6].

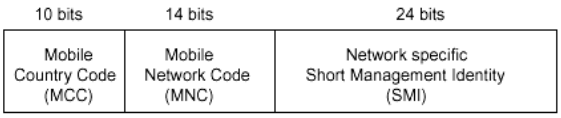


Fig. 3. Contents of TSI and TMI

To make reliable synchronization between MS and BS it is necessary to find synchroword, to process and to get a decision about its type. Synchronization signal comes in the receiver input at fixed time for every user (Fig. 2 and Fig. 3). Signal processing can be performed for whole word or for different code segments. It allows the utilization of digital filtration for synchronization procedure as data rate and modulation scheme reading.

Optimal synchrosignal finding is is possible to realize by Nayman-Pirson criterion as determination of the probability of false alarm (misrouting probability) P_F and maximization of the probability of true detection (probability of successful service completion) P_D [4]. The FIR filter syntheses can be made by invariant impulse response characteristic transformation. The filter is presented as linear time invariant (LTI) circuit and its impulse response characteristic $h(k)$ is obtained from the expected input synchronization signal as the approach for its defining is that first input sample is the final LTI circuit reaction [3]:

$$\dot{h}(k) = \dot{S}(N-1-k). \tag{5}$$

The quantized sample of input signal is made by sampled frequency $f_d = (3 \div 5)f_N$ where f_N is the Nyquist frequency. It is an implementation of the requirement about maximum amount information gain from received signal [2]. The quantized samples of synchroword can be presented as follows:

$$M = m.N. \tag{6}$$

The optimal filtration solves the convolution integral between input quantized samples and filter impulse response characteristic. When non-recursive algorithm is used output signal is obtained from the mathematical expression [1,3]:

$$\dot{y}(t) = \dot{S}(t) * \dot{h}(t). \tag{7}$$

Output complex filter signal can be calculated as follows [4]:

$$\dot{y}(t) = \sum_{n=0}^{N-1} \dot{S}(nT_i)\dot{h}(N-1-nT_i). \tag{8}$$

The decision about signal presence and signal missing is determined with the value of the output filter signal, compared with threshold h. It can be calculated with basic equations for P_F and P_D, the probabilities of false alarm and of true detection [4].

The problem of reliable synchrosignal processing can't be solved without digital integration of output FIR filter signal. After L integrations posterior misrouting probability is calculated by the expression [1,4]:

$$P_{MPO} = \sum_{k=K_0}^{L} \binom{L}{k} P_F^k (1-P_F)^{L-k}, \tag{9}$$

and posterior probability of successful service completion is defined with the equation [1,4]:

$$P_{PSSCO} = \sum_{k=K_0}^{L} \binom{L}{k} P_D^k (1-P_D)^{L-k}, \tag{10}$$

where K_0 is the minimum integration numbers after which the input probability of false alarm is obtained necessary value, according to the threshold h.

IV. Numerical Results and Simulations

Computer simulations for different SNR and different length of the synchronization sequence have been done. The results have been obtained, using 10000 realizations for each value of SNR and length of synchronization sequence. Studies have been done in the presence of noise and other $\pi/4$ QPSK modulated signals simultaneously.

Some results for estimated misrouting probabilities (MP), depending on threshold, are shown on Fig. 4 and Fig. 5. The results for typical values of SNR are represented. The solid curves illustrate the estimated MP when 8 frames are integrated. The dash curves show the estimated MP when 16 frames are integrated.

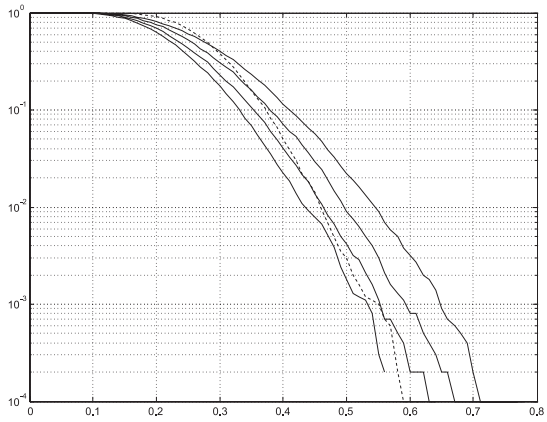


Fig. 4.

Fig. 4 shows the estimated MP for 8 bits synchroword. It's obvious that the threshold have to be determined for the lowest SNR. The determinate threshold for normalized output signal is 0.8.

Fig. 5 shows the estimated MP for 16 bits synchroword. It's obvious that the threshold have to be determined for the

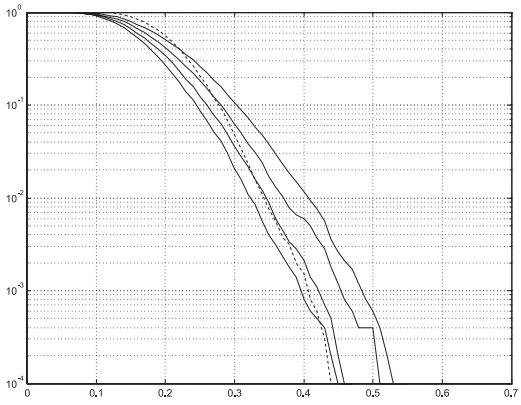


Fig. 5.

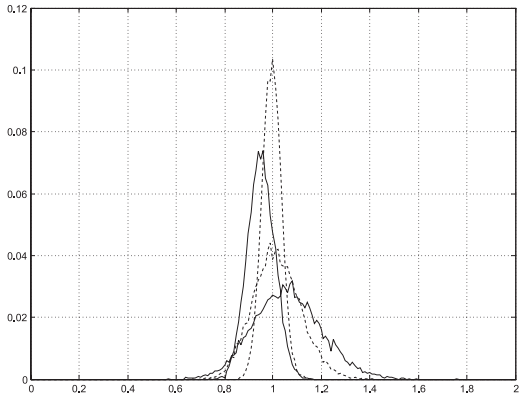


Fig. 6.

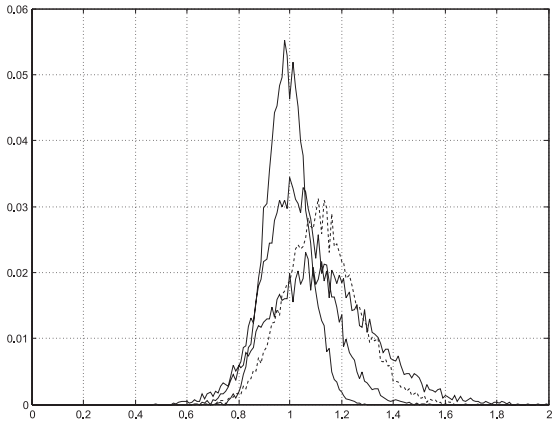


Fig. 7.

lowest SNR. The determinate threshold for normalized output signal is 0.6.

Some results of estimated probabilities of maximum output signal level (PMOSL) are shown on Fig. 6 and Fig. 7. The values of SNR are the same as on Fig. 4 and Fig. 5.

Fig. 6 shows the estimated PMOSL for 8 bits synchroword. The solid curves illustrate the estimated PMOSL when 8 frames are integrated. The dash curves illustrate the estimated PMOSL when 16 frames are integrated. Fig. 7 shows that the probability of lower maximum output signal level then threshold is negligible. The estimated probability of detection is 1 (one) for SNR > 8 dB. The cases of false alarm are 0(zero) for each set of 10000 realizations.

Fig. 8 shows that the probability of lower maximum output signal level then threshold is negligible. The estimated probability of detection is 1(one) for all SNR. The cases of false alarm are 0(zero) for each set of 10000 realizations.

Fig. 7 shows estimated PMOSL for 16 bits synchroword. These results substantiate that for secure and reliable synchronization the threshold value must exceed 0.6 or 0.8, according to the sychroword length and integration number.

V. Conclusions and Discussions

The proposed algorithm for synchronization in TETRA networks has a high reliability in the presence of noise and other $\pi/4$ QPSK modulated signals. The results of simulations shows that the performances of algorithm are very good and they meet the requirements of TETRA networks.

The core of algorithm is a digital matched filter, which can be implemented as a digital finite impulse response filter. Because the filter doesn't have to be of high order there is no need of complex software or hardware using.

References

[1] Vanchev P.H., "Frame synchronization method in VHF Frequency Hopping Spread Spectrum Radios", *Annual Reports of XXXIV Scientific Conference Communication, Electronic and Computer Systems '99*, Sofia, Bulgaria, 1999. (in Bulgarian);

[2] Marinov M.S., S.N. Ivanova, "Research of the relation to Signal to Noise Ratio on quantized frequency by transmission of PSK signals in Constricted Pass Band", *Annual Reports of XXXIII Scientific Conference Communication, Electronic and Computer Systems '98*, Sofia, Bulgaria, 1999. (in Bulgarian);

[3] Rabiner L.R., B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1975;

[4] Tihonov V.I., V.N., Harisov, *Statistical Analysis and Synthesis of Radio engineering Equipment and Systems, Radio and Communication*, Moscow, 1991, (in Russian);

[5] ETS 300 392-1 Reference: DE/RES-06001-1 Radio Equipment and Systems (RES); Trans-European Trunked Radio (TETRA); Voice plus Data (V+D); Part 1: General network design, February 1996

[6] ETS 300 392-2 Reference: DE/RES-06001-2 Radio Equipment and Systems (RES); Trans-European Trunked Radio (TETRA); Voice plus Data (V+D); Part 2: Air Interface (AI), March 1996

[7] TETRA Overview, Securicor Wireless technology, 2001.

[8] ITU, Document 1/BL/7-E, Spectra and Bandwidth of Emissions, Annex 6, Digital Phase Modulation, 10 September 1999

The Ultra-Wideband Technology in Europe: Regulatory Issues and Research Activities

Ludwig Lubich[1]

Abstract – **We have reviewed the European regulatory status and research activities in the sphere of UWB technology. Evaluations have been made of the achievable performance and the perspectives before UWB technology depending on the possible decisions of the European regulatory authorities.**

I. Introducion

Usually a system is termed ultra-wideband (UWB), if it has a fractional bandwidth $B/f_c > 0.25$ or if it occupies bandwidth wider than 500 MHz. In the past UWB signals have been used in radar applications and for military purposes. Recently, we witness an increased interest in them. The use of the so-called impulse radio (IR) – a form of UWB spread spectrum signaling, is most often proposed. The signal used in IR is a train of base-band pulses with duration in the order of tenth of the nanoseconds, which leads to spreading of its energy in a range of near d. c. to several GHz. The information is conveyed through the use of different modulation schemes, such as for example Pulse position (PPM) or pulse-amplitude modulation (PAM), combined with pseudo-random time hopping, in order to allow multiple access and provide spectral smoothing. This signal is transmitted without the conventional up-conversion, so it is sometimes called "carrier-free". The latter leads to possibility for low-cost low-complexity transceiver design. [1,2]

The UWB signals have extremely large bandwidth, therefore have low power spectral density and are usually noise-like. They contain also spectral components with relatively low frequencies. These peculiarities lead to some of their distinguishing features and capacities: They enable communication with very high data rates. They have low susceptibility to multipath fading and immunity to interference from conventional radio communication systems. They allow covert communication and reusing of spectrum already allocated to the established services, without causing significant interference. Thus the UWB signals are very appropriate for short-range high data rate (100 s of Mb/s) communications. They allow very fine time resolution; therefore they are suitable for precision positioning systems and radar applications. In particular, the presence of low-frequency spectral components allows penetrating in materials – this leads to their use and application in ground penetrating radars (GPR), wall- and through-wall and medical imaging systems. Another significant application is also the short-range automotive radars.

[1]L. Lubich is with the Faculty of Communication at Technical University - Sofia, E-mail:lvl@tu-sofia.bg

II. Main Objectives

This article aims at reviewing the state of UWB technology in Europe and assessing its further prospects. For the purpose, a rough assessment of achievable performance is going to be made, according to possible regulatory constraints.

III. Coexistence Issues And Regulatory Status

The fact that the UWB systems reuse the spectrum, given to other users, leads to the problem for interference protection of the existing conventional radio communication systems, especially of systems, related to the security and safeness of the flights, Search and Rescue Satellite (SARSAT), enhanced 911 etc. Several years ago in the USA, one of the first alarming signs was that, that the UWB emissions could lead to lowering in performance of Global Positioning System (GPS) receivers and selected Federal Government radar systems operating below 3.1 GHz, even when the transmitted power is lower than the limits set by the then operating FCC Part15, referring to the non licensed emissions. These issues were widely discussed in USA. The National Telecommunication and Information Administration (NTIA) made researches, which confirm that the concern shown is quite reasonable [3-5]. It is good mentioning that the Earth Exploration Satellite Service (EESS) and the passive radio astronomy are threatened, especially by the automotive short range radars (SRR), planned to work roughly about 24 GHz. It is completely clear that rules are needed, specially elaborated for the UWB. On 14.02.2002 FCC launched the well known "First Report and Order" (FRO), which determines the rules and regulations for work of the UWB devices. [6] Quoting one of the FCC-members FRO is "ultra-conservative". On 13.02.2003 FCC confirmed, the announced one year earlier with a few slight amendments and relieving corrections [7]. FRO is a first cautious step. It is expected that after experimental and experience data be gathered from the work of the UWB devices, which will very soon emerge on the market, additional alleviations in the limitations to be undertaken.

FCC divides the UWB devices into imaging systems, vehicular radar systems, indoor systems and hand held systems. Imaging systems are divided in low frequency ($f < 960$ MHz), mid-frequency (1.99-10.6 GHz) and high frequency (3.1-10.6 GHz) imaging systems. The spectral masks for indoor and handheld devices are given on figure 1.

IEEE works at the standard IEEE 802.15 for wireless personal area network (PAN). The IEEE 802.15 High Rate Alternative PHY Task Group (SG3a) deals with UWB and in

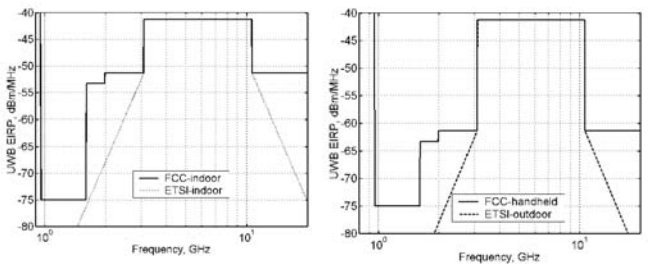


Fig. 1. UWB spectral masks

the end of 2003 will issue a draft standard [8].

In Europe, still[2] the regulatory issues are not solved. They are responsibility of the European Conference of Postal and Telecommunications Administrations (CEPT) and the European Telecommunications Standard Institute (ETSI). The data, which could be found, are very few. On behalf of CEPT with UWB deal the Work Groups Spectrum Engineering SE21 and SE24. Their tasks, connected to the UWB are choice of appropriate interference scenarios and path loss models and to assess required distance and interference margins, which to be imposed to the UWB devices. On behalf of ETSI a task group was established, specially formed for UWB: TC ERM (Technical Committee Electromagnetic compatibility and Radio spectrum Matters) TG31A.

For the moment in Europe the UWB devices are divided into indoor and portable [9] (in some sources one can also find "outdoor" instead of "portable" [10]. For imaging applications such as GPR nothing could be heard for the moment. By ETSI (in some sources the authorship is accredited to CEPT [10]) sloped masks are proposed, which very well overlap those of the FCC, as is demonstrated on figure 1. The slope is $\pm 87 \log(f)$. The difference will not lead to lowering in performance of the UWB devices, at the same time it will result in better interference protection. The fact that, according to SE24, some conventional radio communication systems need from 10 dB to 30 dB more protection over the FCC in-band limit causes uneasiness in the UWB proponents. According to [11,9] this is referred to separate frequencies, but according to [12] there is risk the emission limits to be lowered for all frequencies. This could happen if when evaluating the potential interference theoretical worst-case models are used instead of models based on "real world data", and also due to the shown over CEPT SE24 strong pressure from participating administrations (regulatory agencies) to protect primary services such as fixed services, mobile services, etc.

Another issue at question is the car SRR, working on 24 GHz, which could disturb the Earth Explore Satellite Service (EESS) and the passive radio astronomy. Suitable frequencies are being searched. Candidate frequencies are 26.5 and 35 GHz where the occupied bandwidth will be 4.2 GHz [13]. At least till 20.02.2003 decision has not been taken.

More clarity over the UWB-regulatory issues in Europe is expected in the end of 2003. In particular, in November when the publishing of the EN 302 065 and EN 302 066 standards of ETSI, is expected.

[2]The article is written in April 2003

IV. Performance and Interference Assessment

If at the end the proposed by ETSI mask is approved, in Europe data rates and ranges will be accessible as these in USA. Whereas in the sphere of applications such as GPR and Wall-imaging there will remain lost opportunities.

More fears arouse the possibility to impose in-band emission limits, with 10 to 30 dB lower than those allowed by FCC. If this occurs only for specific frequencies, bit rates and ranges are still possible, comparable to those achieved in USA. These limitations, especially if they tear the frequency band apart, they will probably lead to significant changes in the conception of the UWB transmission. One possible way of development is to go on the track of multi band IR, [14] or to leave the carrier free signals and to pass trough the "Spectrally filtered UWB", which allows a precise control of the radiated spectrum, but leads to the conventional up/down-conversion architectures [15]. Suitable candidate will be the OFDM [16], which also allows precise PSD tailoring. The last two solutions increase the complexity, therefore the price of the transceivers, and thus lower they competitiveness.

If the limitations with 10-30 dB below those permitted by FCC (-41 dBm/MHz) encompass the whole range from 3.1 to 10.6 GHz, in that case the future of the UWB technologies in Europe will become problematic.

A rough estimation will be made of the upper bound of the accessible bit rates/ranges in that case. For that purpose, the well-known theorem of Shannon [17] will be applied. We assume the most optimistic scenario: There are no interferers, but only AWGN, we dispose of ideal receiver with noise figure $NF = 0$ dB, which captures the whole energy that comes to it. It is necessary to find SNR. For those purpose the following path loss model will be used [18]. Similar is also the dependence given in [10].

$$\begin{aligned} PL(d,f)_{[\mathrm{dB}]} &= PL(d_0,f)_{[\mathrm{dB}]} + 10n\log(d/d_0), \\ PL(d_0,f)_{[\mathrm{dB}]} &= 20\log(4\pi f d_0/c). \end{aligned} \quad (1)$$

$PL(d_0)$ is the mean path loss at the reference distance d_0 (in our case $d_0 = 1$ m), at which the propagation can be considered to be close enough to the transmitter such that multipath and diffraction are negligible and the link is approximately that of free space. In indoor environment we can assume that the path loss exponent $n = 2$ in conditions of line of sight (LOS) and $n = 4$ in non-line of sight of sight of sight of sight (NLOS) between the transmitter and receiver.

In order to find the received power P_r, when using the whole band allowed from 3.1 to 10.6 GHz, integration will be needed. Then assuming isotropic transmit- and receive antennas we have:

$$P_r = \int_{f_l}^{f_h} \frac{PSD_{\max}(f)}{PL(d,f)} df, \quad (2)$$

where $PSD_{\max}$ is the maximum admissible emitted power in unity bandwidth, and f_l and f_h are the limits of the used frequency range, respectively 3.1 GHz and 10.6 GHz.

After simple transformations, assuming $N_0 = -174$ dBm/Hz and $B = 7.5$ GHz and replacing in the Shannon's

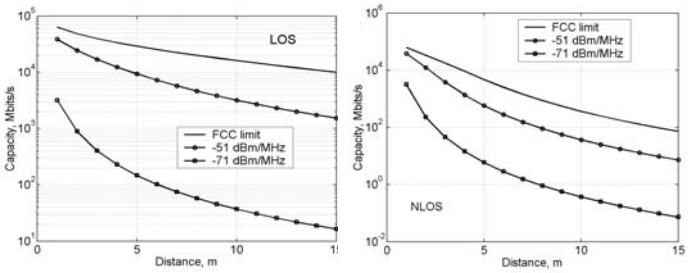


Fig. 2. Channel capacity

formula, we obtain the channel capacity. On figure 2 its dependency is shown from the distance and in different limitations of the transmitted power for LOS and NLOS condition, respectively.

The issue is to what extent the proponents of strict restrictions are correct. The rough influence of the UWB devices upon the conventional victim receiver will be estimated. When B $\ll$ PRF, where B is the IF Bandwidth of the influenced receiver, and PRF is the pulse repetition frequency of the UWB transmitter, the interference, caused by the UWB is similar to white gaussian noise [19]. This is often carried out. One figure of merit commonly used to assess interference issues is the distance between the UWB transmitter and the victim receiver, at which the interference will cause an effective rise in the noise floor at 1 dB [20]. It is easy to prove that in order to happen this, the interference level must be 6dB below the current noise floor.

Or for interference power P_i we can write:

$$\begin{aligned} P_{i\,[\mathrm{dBm}]} &= PSD_{t\,[\mathrm{dBm/Hz}]}(f) + 10\log(B) - PL(f, d_{+1dB})_{[\mathrm{dB}]} \\ &= -174[\mathrm{dBm/Hz}] + 10\log(B) + NF - 6\mathrm{dB}, \end{aligned} \quad (3)$$

where $PSD_t(f)$ is the transmitted power in a unity bandwidth, NF is the noise figure (in dB) of the victim receiver, and d_{+1dB} is the distance between the UWB- and the victim device. After simple transformations, assuming isotropic antennas, for worst case (LOS) we have:

$$\log(d_{+1dB}) = 13.35 + 0.05(PSD_{t\,[\mathrm{dBm/MHz}]} - NF) - \log(f) \quad (5)$$

According to data of the FCC spectral masks and the ETSI spectral masks, using (5), the shown in figure 3 graphs are obtained. They represent dependency of the distance d_{+1dB}, which leads to an effective rise in the noise floor of 1 dB, from the working frequency of the victim receiver. For the latter it is assumed to be $NF = 8$ dB.

One can find out that in some circumstances, the interference, caused by the UWB transmitters can be felt at very large distances and is difficult to underestimate this risk. Even more serious are the issues for the aggregated effect from the operation of several such devices. The statement that we some times come across, that the effect could be taken into account only from the closest emitter, in the common case is not true. In certain conditions aggregate effect from multiple, high PRF emitters could be significantly more deleterious than the effects of the closest, single emitter [5]. It is expected a quick proliferation of the UWB devices and it is

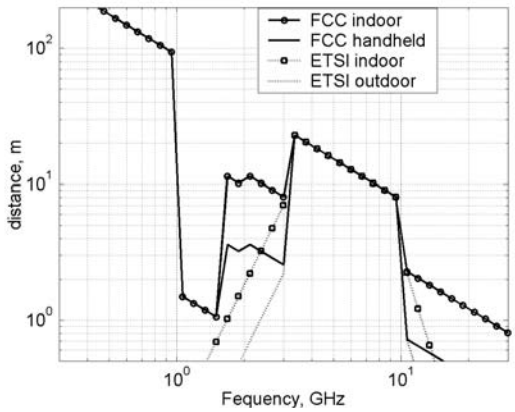


Fig. 3. The distance d_{+1dB}

possible, especially in office buildings to have several UWB devices operating at a place of several square meters. One can see that the, spectrum masks, proposed by ETSI, present significantly better protection of the systems, working on frequencies under 2 GHz, in comparison to FCC mask. In that range many radio navigation and safety-of-life systems are operating as well the mass widespread systems as DVB-T, T-DAB, GSM, PCN.

V. European UWB Research Activities

European research activities are structured around consecutive five-year programs, or so-called Framework Programs (FP). They have multi-theme structure and contain in themselves Thematic Programs One of them is the program Information Society Technologies (IST). This program at the time of FP5 (1998-2002) initiated work on 4 projects, connected to the UWB technologies. Some leading companies and participants from various universities take part in them:

Whyless.com – the open mobile access network. "Whyless.com will research scalable radio technology and network resource trading principles in order to avoid gigantic infrastructure paradigm shifts caused by current network development principles, while enabling steady evolutionary growth and 'swift' adaptation to future user and business requirements." [15]

Ultra-wideband Concepts for Ad-hoc Networks (U.C.A.N.). Overall objective is to develop and demonstrate a complete ultra-wideband (UWB) system demonstrator. They treat the air channel characterization, the coexistence issues, and the communication system: physical layer (RF and baseband), medium access control (MAC) and network layer. Ad-hoc networking and positioning aspects will be demonstrated, as a potential component of the future 4G-communication infrastructure [21].

Universal Remote Signal Acquisition For hEalth (U-R-SAFE). The objective of the U-R-SAFE project is to propose a Personal Health Care system, allowing convalescent and elderly a quasi-normal life, providing continuous monitoring. A Wireless Personal Area Network based on the UWB will be used on the patient himself. This Network will be inter-

faced with the Home, Fixed and mobile Networks [22].

ULTRA Wideband Audio Video Entertainment System (ULTRAWAVES). Main Objectives: To provide a high performance and low cost wireless home connectivity solution for multi-streaming of high quality video and broadband multimedia; design and implement a complete UWB based system [23].

A perspective proposal for FP 6 is *PULSERS (Pervasive Ultra-wideband Low Spectral Energy Radio System)*. It is an inter-national consortium, established in 2002 as a joint venture of IBM Research (Zurich) and Philips Research (Redhill, England). The purpose is to create a short-range wireless system, based on UWB, which allows high-rate transmission in multimedia applications (hundreds of Mb/s on a few meters), and also low-rate transactions, together with position location, sensing or identification (hundreds of Kb/s to several tens of meters) [12].

It became also known that, Philips Semiconductors and General Atomics develop in cooperation UWB wireless communication chipsets [24].

VI. Conclusions

The state of the UWB technology in Europe was reviewed and its further prospects were assessed. For this purpose, a rough assessment of the achievable performance was implemented, according to the possible regulatory constraints.

One notices that Europe works very seriously in the sphere of the UWB-technologies. Great importance for its future will have the decisions of the regulatory authorities. There is a risk some far too conservative regulations to delay the development of the UWB technologies, by limiting the achievable performance, questioning their practical usefulness. At the same time, The introduction of strict emission limits would require approaching the theoretical limits of UWB device performance, resulting in UWB devices that are unlikely to be low-cost low-complexity. Probably the applications would be limited to low-rate communications, for instance information transfer from remote sensors, covert communication, or connection to extremely short distances, for example in systems for contact-less identification. Such a development could lead to a backwardness of the European producers from their American rivals. From other hand, by gathering experience from the exploitation of the first UWB devices and accumulating data for the interference, caused by them in real life, it is possible the limitations to be lessened. Of great importance would be the activity of the interested producers, as this happens in USA.

For the time being, there is no talk at all about allowing the UWB systems, working under 1 GHz, or in other words for the moment we cannot think of applications such as GPR. For the automotive SRR, it is obvious that a suitable solution is searched and their implementation will not be hindered. As it was shown above, the cautiousness in the regulatory sphere is not completely groundless. What remains is to hope that a wise compromise will be found between the protection of the nowadays-existing non-UWB systems and the creation of conditions for development of the promising and prospective UWB technology.

References

[1] Win, M.Z., Scholtz, R., “Impulse Radio: How it Works”, *IEEE Communications Letters*, Feb. 1998.

[2] Mitchell T., “Broad is the way”, IEE REVIEW, January 2001

[3] NTIA Report 01-383 “The Temporal and Spectral Characteristics of Ultrawideband Signals”, January 2001;

[4] NTIA Report 01-384 “Measurements to Determine Potential Interference to GPS Receivers from Ultrawideband Transmission Systems”, February 2001

[5] NTIA Special Publication 01-43, “Assessment Of Compatibility Between Ultrawideband Devices And Selected Federal Systems”, January 2001

[6] “First Report And Order”, Federal Communications Commission, Washington, D.C., Released April 22, 2002

[7] “Memorandum Opinion and Order”, FCC, Washington, D.C., 13.02.2003

[8] http://www.ieee802.org/15/pub/TG3a.html

[9] Huang B, “UWB Regulatory Overview”, Sony AWT Group, 3 October 2002

[10] ITU, Radiocommunication study groups, Document 1-8/6-E, 10 October 2002

[11] Huang B, “Status Report of European UWB Regulatory Activities”, IEEE 802.15 WG for WPAN, September 17, 2002

[12] Hirt, W., “European UWB regulatory status and research initiatives”, UWB Seminar, Singapore, 2003.

[13] Short range devices Maintenance Group, Notes of the 22nd Meeting, 5-7 June, Bern

[14] Aiello R., J. Ellis, U. Kareev, K Siwiak, L. Taylor, “Understanding UWB – Principles and implications for low-Power communications – A Tutorial”, IEEE Working group 802.15, March 2003.

[15] www.whyless.org

[16] Gerakoulis D., P. Salmi, S. S. Ghassemzadeh, “An Ultra Wide Bandwidth System for In-Home Wireless Networking” European Wireless 2002 February 25-28, 2002- Florence, Italy.

[17] J. Wozencraft, I. Jacobs, “Principles of communication engineering”, John Wiley & sons, New York, 1965.

[18] “Propagation overview for TM-UWB”, Time Domain Corporation.

[19] Fontana, R., “An Insight into UWB Interference from a Shot Noise Perspective ”, *IEEE Conference on Ultra Wideband Systems and Technology*, 2002

[20] Shilvely D, “Ultra-Wideband Radio – The New Part 15”, Microwave Journal, 2003, vol. 47, No.2, pp132-146

[21] www.ucan.biz

[22] http://ursafe.tesa.prd.fr

[23] www.ultrawaves.org

[24] http://www.semiconductors.philips.com/news/publications/content/file_1126.html

Questions of the Coexistence of the Ultra-Wideband Systems with the Conventional Radio Communication Systems

Ultra-Wideband Interference on a Non-UWB System

Ludwig Lubich[1]

Abstract – **The present article sets forth the effects, caused by UWB interference upon the conventional receivers.**

I. Introducion

Recently we witness an increased interest in the Ultra-wideband (UWB) communications and especially in the so-called impulse radio (IR). For carrying information it uses train of extremely short base-band pulses. The key concept is through the use of signal with very large bandwidth, respectively very low power spectral density, to reuse the spectrum, occupied by the already existing users without seriously lowering their performance [1,2]. An important area of research is the exploration of the effects, which will have the UWB-transmissions upon the non-UWB systems and vice versa, as well as the ways to avoid and suppress the mutual interference. Further in this article for the sake of brevity under "UWB system" we shall understand only UWB Impulse Radio.

II. UWB Signal Description

One of the general mathematical descriptions of a UWB signal is the following [3]:

$$s(t)=\sum_{l=-\infty}^{\infty}\frac{1}{\sqrt{R}}\sum_{i=0}^{R-1}A_l g_{pulse}(t-lT_S-iT_F-c_{(lR+i)_P}T_c-B_lT_{PPM}),$$

where $(lR+i)_P \equiv (lR+i) \mod P$ and g_{pulse} is the basic transmitted pulse, e.g., a monocycle. A_l and B_l represent the data modulation and they are constant in the frames of the symbol time T_S. A symbol is transmitted through R in number monocycles, where their average repetition time is T_F, and the exact position of each monocycle in its frame is determined by a pseudo-random time-hopping (dithering) sequence $\boldsymbol{c} = (c_0, c_1, ..., c_{P-1})$. The latter lowers the probability for catastrophic collisions of pulses, when operate more than one UWB transmitters and results in a more smooth power spectral density (PSD) of the UWB signal. The PSD is very important when considering the interference issues. The most important is, PSD to be if possible flat over the occupied bandwidth and what is most important not to have concentration of considerable power in discrete spectral lines.

[1]L. Lubich is with the Faculty of Communication at Technical University - Sofia, E-mail:lvl@tu-sofia.bg

Many researchers have derived for PSD analytical expressions, proved by computer simulations [3-6]. It is characteristic that PSD is a sum of continuous component and a discrete component. In general, the shape of $PSD(f)$ is determined of PSD of the monocycle. If M, N is the smallest pair of integers, for which $N.R = M.P$, the discrete component of PSD is comprised of discrete spectral lines at frequencies $k/(NT_S)$. IF for $\forall l \neq l'$ the expectation $E_{l\neq l'}\{A_l A_{l'}\} = 0$, the discrete part vanishes [3,4]. This is valid in binary pulse-amplitude modulation (PAM), as well as in its combination with the pulse position modulation (PPM) (Fig. 1) .

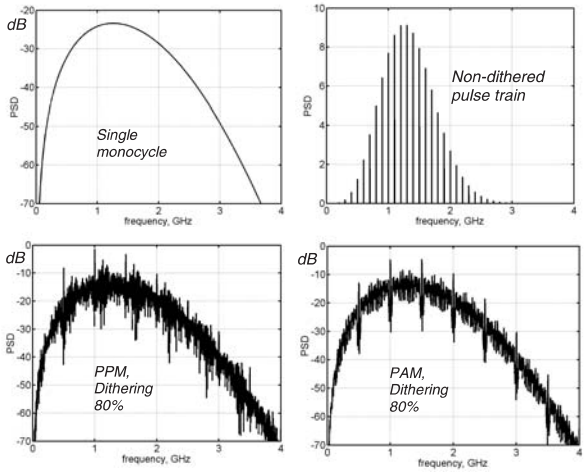


Fig. 1. UWB PSD examples

III. Influence of the UWB Emissions upon the Non-UWB Systems

This topic has been seriously researched, by both the regulatory authorities in respect to the necessity of creating regulations, treating the UWB devices, and the UWB proponents [7-14, etc.]. The greatest interest is the research of the level and the nature of the UWB interference at the IF- and the demodulator's output of a victim receiver, their dependence by the UWB signal parameters and the aggregate effect of multiple UWB emitters, especially in the expected proliferation of UWB devices. Based on these studies one can find appropriate UWB emission limits, UWB signal parameters and rules for operating of UWB devices, where the deleterious action upon the conventional receivers could be acceptable.

According to some early statements of UWB proponents, the influence of UWB signal upon a non-UWB receiver is similar to white gaussian noise. It turns out that this is not always the case. Generally, the level and the nature of the UWB interference at the output of a non-UWB receiver depend both by the parameters of the UWB signal, and by the receiver, especially by its IF bandwidth (IFBW). The most significant parameters of the UWB signal, affecting UWB waveform and power level at the receiver's IF output are the pulse repetition frequency (PRF = $1/T_F$), the pulse width T_m, use of time dithering/and/or gating of the UWB device and the type of the data modulation [8].

The power of a non dithered signal is concentrated in discrete spectral lines at frequencies divisible by PRF, where its potential to disturb the operation of the victim receivers is great, especially if any line of the UWB spectrum align with the non-UWB carrier. The experimental results [10] confirm this. And vice versa dithering results in a spectral smoothing, which is a prerequisite for getting at receiver's IF output more noise like waveform.

A. *Character of the UWB interference at the victim receiver IF output*

The ratio PRF/IFBW is decisive for the character of the waveform at receiver's IF output [8,15,16]. In practice, the response of the IF filter on each impulse, due to the very little Tm is equal to the impulse response of the IF filter. For $PRF < B_{IF}$, the responses, caused by the separate pulses remain distinguishable. The waveform at the receiver IF output will be pulse-like. For $PRF > B_{IF}$ the waveform will be similar to a continuous wave (CW), if the UWB signal is non-dithered, and noise-like in dithered UWB. The degree, to which receiver output response appears noise like, depends on the value of $c_{\max} T_c/T_F$ and the randomness of the dithering sequence. In [16] it is shown, that in PRF, several times bigger than BIF, at the demodulator output of a victim receiver in practice we receive Gaussian distribution of the signal, as it could be expected from the central limit theorem.

For a description of a time-domain characteristics and in particular for the evaluation of the degree, in which the waveform at the IF output is noise-like, is proposed the use of Amplitude probability distribution (APD) [7,10]. It is measurable, from it can be obtained some statistical values and can be used in receiver performance prediction. The APD express the probability that signal amplitude excess a threshold. The APD graph displays amplitude on the y-axis and probability on the x-axis. The probability is so scaled; that in gaussian noise APD becomes a straight line [10, Appendix E].

B. *UWB interference levels at the victim receiver IF output*

The dependency of the UWB interference level at the IF output is reviewed in several places [8,15]. The level depends basically on the IF bandwidth, PRF and on the presence or absence of dithering and/or gating. A simplified dependence is given in [17,18].

In [16] the same topic is also discussed, but some moments there are problematic. Probably the best notion about the problem could be acquired by [8,15], despite the rather more practical, than strictly scientific character of these two sources. Two cases are being discussed

Non-dithered signal: Its spectrum is built of discrete lines at frequencies spaced at $\Delta f = PRF$. When $\boldsymbol{IFBW < PRF}$, in the worst case one single spectral line will enter the IFBW. Then the *average power* P_{IAV} at the IF output would be independent by B_{IF}. In $B_{IF} > PRF$, P_{IAV} will be proportional to the number of UWB spectral lines, entering the IFBW, which is proportional to B_{IF}, therefore, P_{IAV}[dB] $= P_{AV0} + 10\log(B_{IF}/PRF)$, where P_{AV0} is the power in $B_{IF} < PRF$. *Peak power* P_{IP}: When $\boldsymbol{IFBW \ll PRF}$, the waveform at the IF output is CW-like. Then, obviously $P_{IP} \approx P_{IAV}$. In wider IFBW, waveform will be pulse-like. Then the ratio P_{IP}/P_{IAV} increases proportionally to B_{IF}. In the transitional area, when B_{IF} and PRF are from one and the same order, the dependence is more complex. But as a threshold we can accept $B_{IF} = 0.45PRF$, e.g. in $IFBW > 0.45PRF$ P_{IP}[dB] $= P_{AV0} + 20\log[B_{IF}/(0.45PRF)]$.

Dithered signal: In this case $PSD(f)$ is generally flat within the frames of IFBW of a conventional receiver, therefore P_{IAV} is proportional to IFBW, i.e. P_{IAV}[dB] $= P_{IAV}(B_{ref}) + 10\log(B_{IF}/B_{ref})$, B_{ref} is some IFBW, from which we know P_{IAV}. The dependence of P_{IP} by IFBW changes due to the dither percentage $c_{\max} T_c/T_F$. In a reasonable value of 50%, as a boundary value of IFBW one can accept with quite a good exactness $B_{IF} = 0.2PRF$: P_{IP}[dB] $= P(0.2PRF) + 20\log[B_{IF}/(0.2PRF)]$ when $B_{IF} > 0.2PRF$ and $P_{IP} \approx P_{IAV}$ when $B_{IF} < 0.2PRF$. These dependencies are shown on figure 2.

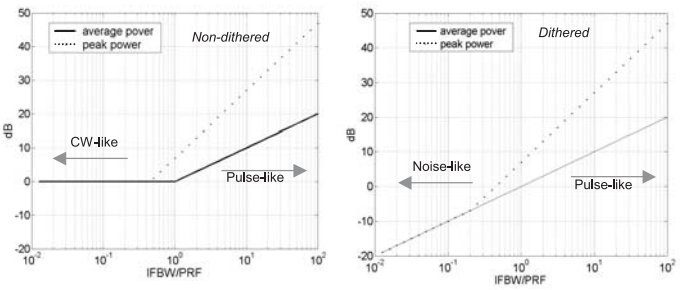


Fig. 2. UWB power at the receiver's IF output

They have been used by NTIA for the deriving of the so-called bandwidth correction factors (BWCF) [8,15], which allow to estimate average and peak power for various IFBW of the victim receiver from average power measurements made in a 1MHz reference bandwidth.

More often the answer to another question is much more: How the interference power at the IF output of a victim receiver depends on the UWB PRF (IFBW is fixed). In the literature I did not come across a thorough answer to this question. Let us review two cases:

Non-dithered UWB signal. The power of a non-dithered UWB signal is concentrated in discrete lines, spaced

at $\Delta f = PRF$ one from another, and their number will be approximately $N_L = B_{UWB}/PRF$. Then the power in each line will be $P_L = P_{UWB}/N_L = PRF.P_{UWB}/B_{UWB} = PRF.PSD_{AV}$, where P_{UWB} is the power of the UWB signal, that falls on the input of the receiver, and $PSD_{AV} = P_{UWB}/B_{UWB}$ is average PSD of the UWB signal. The UWB power at IF output is proportional to the number of the spectral lines, that fall in the IFBW and the power of one spectral line, e.g. $P_{IAV} = a.(B_{IF}/PRF).PRF.PSD_{AV} = a.B_{IF}.PSD_{AV}$ in $PRF \ll B_{IF}$, where a is a given constant. In $PRF > B_{IF}$, in the worst case in IFBW will fall only one spectral line. Then $P_{IAV} = P_L = a.PRF.PSD_{AV}$, i.e. increase in PRF, the UWB signal becomes more deleterious.

Dithered UWB signal. In this case PSD is relatively even. Therefore $P_{IAV} = a.B_{IF}.PSD_{AV}$. For peak power could be expected $P_{IP} \approx P_{IAV}$ when $PRF \gg B_{IF}$ and increase in P_{IP} caused by the lowering of PRF in the transition area around $PRF = B_{IF}$ due to two factors: transition from noise/CW-like waveform to pulse-like and increased energy in one pulse, proportionally to $1/PRF$. When $PRF \ll B_{IF}$ the impulses are completely separated and only the latter factor is valid, therefore P_{IP} will increase proportionally to $1/PRF$.

In [8], Appendix D, the dependence is shown from PRF of a peak UWB power in a 50MHz bandwidth related to the average UWB power in a reference bandwidth of 1 MHz. It has been established to the considering of how to regulate the allowed peak power. A bandwidth of 50 MHz has been chosen, since it is comparable to the widest victim receiver IFBW. It can be determined that the setting of a limit for the maximum allowed P_{IP}/P_{IAV}, on behalf of the regulatory authorities, leads to the necessity to use PRF beyond a set limit. For example, if P_{IP}/P_{IAV} is limited to 20 dB, PRF > 10 MHz for non-dithered UWB signal, and for dithered signal there is no value of PRF, that could satisfy the preset condition.

In [16] something more is done: analytical expression for the power caused by UWB interference has been given at the detector output of a victim receiver, but still in the article there are some dubious moments.

C. *Victim receiver performance degradation*

The victim receiver performance degradation is difficult to be foreseen exactly, because to a great extent it is dependent on the specific signal processing, used in the receiver. Many of the receivers are optimized to give the best performance, assuming that they work in conditions of white gaussian noise. Then in some cases, serious performance degradation can be expected, when UWB interference appear pulse-like at the IF output.

In [17,18] there is derived a simple expression for the signal/noise ratio and its use for determining of BER of the victim receiver is proposed. However, about the latter, it is necessary to know also the statistical properties of the UWB signal after its transition through the receiver's front-end. In [14] analysis is made about the influence of the UWB interference on a NB receiver and expression for BER has been found, which however, is difficult to apply. In [8], Appendix. E, is shown how easily from APD could predict bit error rate in non-coherent binary FSK.

Especially useful could be the gathered multiple results from simulations and measurements of the interference, caused by the UWB signals upon non-UWB receivers. [19,20,8,9,10]. In [19,20] mainly simulation results are presented. More interesting however are the given results from experiments conducted with GPS receivers [10]. In this article I have tried to summarize some of the more important results from [10]. Most informative appear to be the dependence of Reacquisition time (RQT) and pseudo-range (PSR) accuracy by the UWB power and the interference power at the break lock point (BL), measured in various UWB signal parameters. For a comparison measurements have been made with AWGN too. An interesting fact has been found out, that the deleterious influence of UWB interference is growing up with the increase of PRF. In the highest used PRF (20 MHz) the values for BL and RQT are almost the same as in AWGN, when the UWB signal is dithered, and considerably worse in non-dithered signal, as it was expected. It also interesting that in difference to the AWGN case, the PSR error does not change significantly in case of a change in UWB interference level and is about 2-3 times bigger than the error measured without additionally injected noise or UWB signal. Here also the non-dithered UWB signal is more deleterious especially in high PRF.

It is not exactly clear why the influence from the UWB signal gets bigger by the increase of PRF. The above described statements for the interference power at the IF output give partial explanation, since this phenomenon is witnessed in a dithered signal too. An explanation can be found in the presence of the non zero discrete component in the used in experiments UWB signals. In [16] an explanation is given, but with problematic correctness. Interesting is the article of [21], where based on the careful insight in the experimental data, one comes to the idea that significant role could have the participation of the nonlinearities in the victim receiver. Namely, a compressed receiver stage acts as a bandpass limiter, which is well known to help reduce the effects of pulsed interference. A lower PRF results in a higher peak power, which would be compressed in an earlier receiver stage and thus would produce less output interference. In general the question for the participation of the receiver's nonlinearities in the presence of UWB interference is interesting and poorly investigated.

IV. Aggregation Of Multiple UWB Signals

The issue has been discussed in many places [8,10,14,16,22]. Not until its answer the interference potential of the UWB systems will be fully and completely revealed. It is possible in urban areas hundreds, thousands or even more of UWB devices per square kilometer to be employed. The main questions are: how the power is accumulated from multiple UWB emitters and what is the nature of the resultant signal, and at the same time to have in mind the peculiarities of the radiopropagation. There are 2 opposite opinions: that decisive is the effect of the single nearest UWB-emitter and that there is

no aggregate effect and vice versa.

In general for stationary, stochastic processes, average power from multiple sources do add linearly. It could be expected that the RMS UWB power also aids linearly. For the peak power things are more complex. Then amplitude statistics could be useful. In [8] experimental data are given, from combining of two UWB signals under various signal parameters. In practice upon turning on the second transmitter (having the same P_{AV}), P_{AVaggr} approximately doubles in size. Different are the results in [22], where a radiated measurement is carried out with measurement of SNR loss in a GPS receiver. Unfortunately not enough details are given about the conditions of the measurement. One can presume that the differences are caused by the effects in the radio propagation. Generally, the numerous experiments show that the average (RMS) power emitted by UWB devices is linearly additive in a receiver.

For evaluation of the possible interference, caused by the operation of a great number of UWB devices different models have been worked out: analytical and statistical. In [8] a comparison is made between the results, obtained using different models. More or less the results agree closely within 2 dB.

In NTIA is developed the "UWBrings" model, with which a number of experiments have been carried out [8]. The results can prove that:

Under given conditions some UWB emitter density exists, and above which aggregate interference begins significantly to exceed that from a single UWB emitter. Under different conditions, this emitter density is of few active emitters per square kilometer to greater than 1000.

The summation of numerous independent UWB signals must lead to a noise-like signal at the output of a narrow-band receiver. In [16] it is shown under what conditions aggregate UWB interference lead to Gaussian process at receiver's output.

As far as the aggregate UWB signal at receiver's input is concerned, in [14] is established analytically that the aggregate received signal is heavy-tailed.

The peculiarities of the radio propagation have a significant role for the aggregate interference. Factors such as obstructions due to terrain irregularities, foliage, buildings and UWB antenna directivity have also a very powerful influence [8]. It comes out that in urban environment; UWB emitters in a radius of 1 kilometer are decisive for the aggregate effect upon terrestrial victim receiver. In the experiment, described in [22] interesting effects have been witnessed, related probably to the reflections of near to the victim receiver objects. These effects could significantly increase the harmful effect of the UWB interference.

A great number of results from experiments and computations are given in [8,10]. In [10] the effect of various interference scenarios upon GPS receivers is given. In brief, the results presented, show that the influence of the aggregate UWB interference does not lead to results significantly worse than those in AWGN, except the case with the non-dithered signals. By increasing the number of the UWB emitters to more than 2, in fixed aggregate power, in practice the results cease to change with the number of the emitters and converge on these, received in AWGN, which agree with [16].

V. Conclusions

The mechanisms of influence of UWB signals on a non-UWB receiver were above exposed as well as some of the more important experimental results. It is obvious that it is a vast area for a research. For now on it has been studied as far as, it could give the right to accept the existence of the UWB technology, to find out appropriate UWB parameters, modulation schemes and regulations, so that the existing nowadays non-UWB systems could be adequately protected. Probably in future development of non-UWB receivers, when the proliferation of UWB transmitters will be a fact, methods for signal processing will be searched for purposeful suppression/mitigation of the UWB-interference. For that purpose a more serious clarification of the mechanisms of UWB signal's influence upon the non-UWB receivers will be needed.

References

[1] Win, M.Z., Scholtz, R., "Impulse Radio: How it Works", *IEEE Communications Letters*, Feb. 1998.

[2] Mitchell T., "Broad is the way", IEE REVIEW, January 2001

[3] Lehmann N., A. Haimovich, "New approach to control the power spectral density of a timehopping UWB signal", Conf. on Information Sciences and Systems, The Johns Hopkins University, March 12-14, 2003

[4] Romme J, L. Piazzo, "On the power spectral density of time-hopping impulse radio", *IEEE Conference on Ultra Wideband Systems and Technology*, 2002.

[5] Fontana R., "A note on Power Spectral Density Calculations for Jittered pulse trains", Multispectral solutions Inc., 2000.

[6] A. Alvarez, J. Garcia, "Improved Fast Fourier Transform Algorythm for the eastimation of the power spectral density of UWB signals", June 2002.

[7] Kissick, W.A., editor, "The Temporal and Spectral Characteristics of Ultrawideband Signals," NTIA Report 01-383, January 2001.

[8] Brunson, L.K. et al., "Assessment of Compatibility Between Ultrawideband Devices and Selected Federal Systems," NTIA Special Publication 01-43, January 2001.

[9] Anderson, D.S., E.F. Drocella, S.K. Jones and M.A. Settle, "Assessment of Compatibility between Ultrawideband (UWB) Systems and Global Positioning Systems (GPS) Receivers", NTIA Special Publication 01-45, February 2001.

[10] NTIA Report 01-384 Measurements to Determine Potential Interference to GPS receivers from Ultrawideband Transmission Systems, February 2001.

[11] RTCA Paper No. 086-01/PMC-139, "Second Interim Report to the Department of Transportation: Ultra-Wideband Technology Radio Frequency Interference Effects to Global Positioning System Receivers and Interference Encounter Scenario Development", RTCA Special Committee 159, March 27, 2001

[12] The Radiocommunications Agency (www.radio.gov.uk), "UWB Compatibility with TDAB and DVB-T. An RTCG Project Report", Project No 739, Date 22nd April 2002

[13] The Radiocommunications Agency (www.radio.gov.uk), "UWB interference to Bluetooth and GSM DCS 1800. An RTCG Project Report", Project No 639, Date 22nd April 2002

[14] Swami, A., Sadler, B., Turner, J., "On the Coexistence of Ultra-Wideband and Narrowband Radio Systems". *IEEE Military Communications Conference: Communications for Network-Centric Operations: Creating the Information Force*, 2001, vol. 1, pp. 16-19.

[15] Quincy E. A., "Victim Receiver Response To Ultrawideband Signals", *IEEE Military Communications Conference: Communications for Network-Centric Operations: Creating the Information Force*, 2001, vol. 1, pp. 20-24.

[16] Fontana, R., "An Insight into UWB Interference from a Shot Noise Perspective ". *IEEE Conference on Ultra Wideband Systems and Technology*, 2002.

[17] L. Piazzo, "Some basic facts about UWB EM compatibility and UWB spectrum", Internal report for the Whyless.com project, May 2001.

[18] L. Piazzo, "UWB EM compatibility and coexistence issues", IINFOCOM dept., University of Rome "La Sapienca", 2001.

[19] Hamalainen, M., Iinatti, J., Hovinen, V., Latva-aho, M., "In-Band Interference of Three Kind of UWB Signals in GPS L1 Band and GSM900 Uplink Band". *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept. 2001, vol. 1, pp. 76-80.

[20] Hamalainen, M., Hovinen, V., Iinatti, J., Latva-aho, M., "In-Band Interference Power Caused by Different Kinds of UWB signals at UMTS/WCDMA Frequency Bands". *IEEE Radio and Wireless Conference*, 2001, pp. 97-100.

[21] R. D. Wilson, R. D. Weaver, M.-H. Chung And R. A. Scholtz, "Ultra Wideband Interference Effects On An Amateur Radio Receiver", *IEEE Conference on UWB Systems and Technologies*, Baltimore May 2002.

[22] Cummings, D.A., "Aggregate Ultra Wideband Impact on Global Positioning System Receivers". *IEEE Radio and Wireless Conference*, 2001, pp. 101-104.

A Possibility for UWB Spectrum Shaping

Ludwig Lubich[1]

Abstract – **Below is reviewed an opportunity for UWB spectrum shaping though pulse synthesis in discrete time. Assessment of the obtained increase of the UWB systems performance was carried out.**

I. Introducion

Recently the interest in the ultra-wideband (UWB) technology, especially in the so-called impulse radio (IR) has increased. IR is a promising candidate for application in precise positioning systems, radar applications, as well as in short range communications, because it allows achieving of high data rates (hundreds of MBPS), while at the same time it permits reusing of the spectrum, occupied by the already existing users. IR conveys information through train of extremely short baseband pulses (pulse-width in the order of tenth of nanosecond) [1,2].

Below is a very simplified mathematical description of the used IR signal:

$$s(t) = \sum_{k=0}^{\infty} A_k g_t(t - kt_f - t_k), \quad (1)$$

where the frame duration t_f determines the mean pulse repetition frequency, and $g_t(t)$ is the basic transmitted pulse. The time shift t_k determines the position of the k-th pulse in the k-th frame and depends both on the data modulation, and on a pseudo random time-hopping sequence. The amplitude A_k of the k-pulse is determined by the data modulation.

II. Purpose of the Article

An opportunity for UWB spectrum shaping will be studied. Its practical feasibility will be considered and some possible implementations will be suggested. The improvement of the performance of the UWB communication systems will be assessed.

III. UWB Spectrum

The UWB signal may be presented as follows:

$$s(t) = g_t(t) \otimes M(t), \quad (2)$$

where $M(t)$ is a pulse excitation process, which includes both the time hopping and the data modulation. The power spectral density (PSD) of the UWB signal is then given by

$$S_{UWB}(f) = |G_t(f)|^2 . S_M(f), \quad (3)$$

where $G_t(f)$=*$[g_t(t)]$ and S_M is the PSD of $M(t)$. If there is an appropriate signal design $S_M(f)$ is generally flat. Consequently the spectrum shape of the UWB signal is determined by $G_t(f)$ [3,4].

For now pulse generation is mainly implemented through impulse excitation of the transmitting antenna, where there are not great opportunities for effective pulse spectrum control.

Precise spectrum control, however, is highly desirable. Firstly, It would provide the opportunity for PSD shaping, maximally corresponding to the emission masks, imposed by the regulatory authorities. Consequently, there would appear an opportunity to increase the transmitted power. It should be noted that it would be possible, through appropriate pulse spectrum shaping, to compensate the frequency dependence of the antenna gain.

Furthermore, an opportunity to adapt the UWB systems to the changing interference scenarios would be created. In the UWB correlation receivers the generation of template waveforms with a precisely controlled spectrum, would facilitate optimal filtering in the conditions of non-white noise and interference.

IV. Essence of the Proposed Approach

This article proposes pulse generation in discrete time. According to the desired spectrum, the required pulse form can be found through inverse Fourier transform. Then it may be approximated through its discrete samples. In the definition of the desired spectrum shape, the frequency dependence of the antenna gain, and the fact that non-ideal pulses would represent the samples could be taken into account. The generation of the samples may be performed in various ways depending on the technology state and the requirements to spectrum shaping precision. The most radical approach is digital as much as possible. A sample implementation, similar to that of the frequency synthesizers, employing direct digital synthesis, is displayed at Fig. 1a. If it is necessary to achieve higher frequencies, which at the moment cannot be achieved by the DDS technology and in case of a simpler pulse shape, the approach displayed by Fig. 1b may be applied. In this case however, flexibility is much lower.

It should be noted that once being calculated the values of the samples remain unchanged until changes in the interference scenario. If the UWB device does not provide for adaptation to the changing interface scenarios, these values may be hardware preset.

Suppose that the desired spectrum of the transmitted pulse is $G_t(\omega)$, and the transfer function of the transmitting antenna plus the additional filter (if any) is $H_A(\omega)$. Further-

[1]L. Lubich is with the Faculty of Communication at Technical University - Sofia, E-mail:lvl@tu-sofia.bg

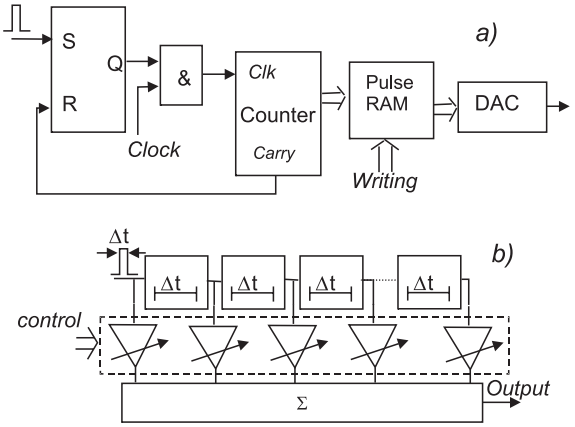


Fig. 1.

more, $G_S(\omega)$ is the spectrum of the elementary pulse, representing the samples of the transmitted pulse. Then, the pulse to be generated may be described as:

$$\begin{aligned} g(t) &= \Phi^{-1}\Big[\frac{G_t(\omega)}{H_A(\omega).G_S(\omega)}\Big] \\ &= \Phi^{-1}\Big[\frac{|G_t(\omega)|}{|H_A(\omega)|.|G_S(\omega)|}e^{j[\varphi_t(\omega)-\varphi_A(\omega)-\varphi_S(\omega)]}\Big] \\ &= \Phi^{-1}[|G(\omega)|e^{\varphi(\omega)}], \end{aligned} \quad (4)$$

where $\varphi_A(\omega)$ and $\varphi_S(\omega)$ are known and $\varphi_t(\omega)$ may be arbitrary. Consequently, there is an opportunity to select $\varphi(\omega)$ where it is possible to optimize the transmitted pulse.

V. Feasibility

Obviously the proposed approach would complicate and make more expensive the UWB devices. Currently permits FCC operation of UWB devices in the frequency ranges 0...0.9 GHz and 3.1...10.6 GHz. Therefore the minimal sample rates are 1.8 GSPS and 21.2 GSPS respectively. The implementation shown in Fig. 1a for devices operating below 0.9 GHz is completely possible with respect to the current state of technology.

It should be noted that the pulse formation requires, as may be seen below, several dozens of samples, so the necessary memory is incomparably smaller than the memory employed in a DDS frequency synthesizer. The needed clock frequency instability is in the order of several percent, which also simplifies the implementation. The high cost is a problem, however it is expected to go down. With respect to the devices operating in the range 3.1 and 10.6 GHz, initially, the approach displayed at Fig. 1b seems more realistic.

In the last several years, the semiconductor technology has progressed rapidly and the achievable clock frequencies have increased dramatically [5]. It has been reported that SiGe bipolar transistors with $f_t = 350$ GHz and a 4.7 ps gate delay had been achieved [6,7].

A method for ultra-fast, highly reproducible signal synthesis and sampling, based on the so-called “Libove Gate architecture”, is presented in [8]. Thus bandwidth up to 20 GHz (40 GS/second) is achievable using a slightly complicated scheme.

Therefore the completely digital pulse synthesis will be widely accessible very soon, and its price will considerably go down.

VI. Results

Computations have been made showing a possible increase of UWB device performance, using the above approach. Pulses with a various number of samples N were generated, in compliance with FCC spectral masks for indoor and handheld UWB devices operating in 3.1...10.6 GHz range (Fig. 2). Fig. 3 shows PSD of the generated pulse for one of the cases ($N = 64$). A comparatively small number of samples are enough to achieve a sufficiently good adjustment of pulse PSD to the emission mask.

In the pulse generation it is admitted that $H_A(\omega)$ is the same like the one used in the generation of second derivative of the gaussian pulse following the classical approach. Sampling frequency $f_s = 25$ GHz was selected so as to make possible a good suppression of the aliases. The sam-

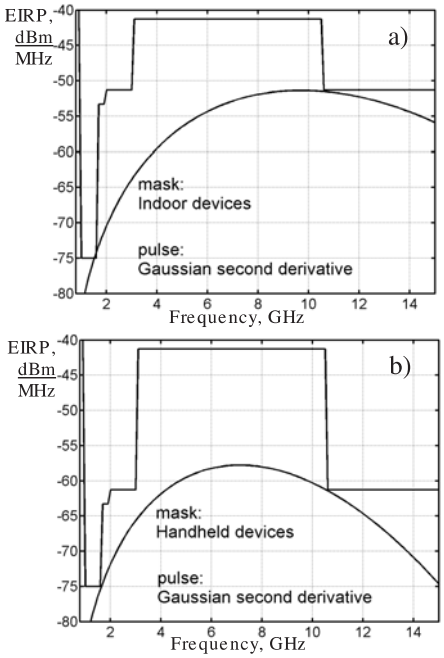


Fig. 2. UWB spectral masks

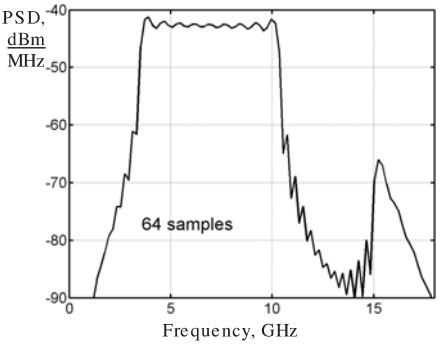


Fig. 3.

ples are presented by idealized rectangular pulses with duration $\Delta t = 1/f_s$.

Assuming that $S_M(f) = \text{const}$, the maximal possible transmitted power is calculated. The latter is compared to those, which is achievable with the use of the most frequently proposed second derivative of the gaussian pulse, suitable for generation, following the traditional approach. The results are provided in Table 1.

Table 1.

Pulse	Feasible power, dBm		Improvement, dB	
	indoor	handheld	In-door	Hand -held
Gaussian 2-nd derivative	-14.6 (-7.5)*	-20 (-17.6)*		
Synthesized, N=32	-5.1	-5.1	9.5 (2.4)	14.9 (12.5)
Synthesized, N=48	-4.53	-4.53	10.07 (2.97)	15.47 (13.7)
Synthesized, N=64	-4.27	-4.27	10.33 (3.23)	15.73 (13.3)
Synthesized, N=128	-3.95	-3.95	10.65 (3.55)	16.05 (13.7)

Ideal case: -2.55 dBm in 3.1–10.6 GHz

*Assuming an additional rejection in the GPS band is implemented

The calculation of the power, emitted in the employed frequency range, is made through numerical integration, according to:

$$P_t = \int_{f_B}^{f_H} S_{UWB}(f)df, \tag{5}$$

where f_H and f_B are the boundaries of the frequency range, object of our interest.

The results show that through the reviewed method there is a considerable increase of the transmitted power as compared to the case when the most frequently proposed second derivative of the gaussian pulse is used.

In order to obtain a desired bit error rate (BER), it is necessary to provide a certain value of the signal to noise ratio per bit in the receiver.

$$q_0 = E_b/N_0 = P_r T_b/N_0 = P_r/(R_b N_0) = P_t/(R_b.N_0.PL),$$

where P_r is the received power and PL is path loss. Then $R_b = P_t/(PL.q_0.N_0)$.

Consequently by obtaining an increase of the transmitted power by approximately 10 dB for indoor devices and 15 dB for handheld devices (Table 1), the increase of the achievable bit rate would be 10 and 32 times, respectively. In case of a fixed bit rate it is possible to increase the value of PL, and as a result the UWB system range in line of sight conditions will widen 3.2 and 5.6 times, respectively.

Some recent publications propose the use of higher derivatives of the Gaussian pulse. For example PSDs of the fourth and seventh derivatives fit best to the masks for indoor and handheld devices, respectively (Fig. 4). They rely on the traditional method for pulse generation.

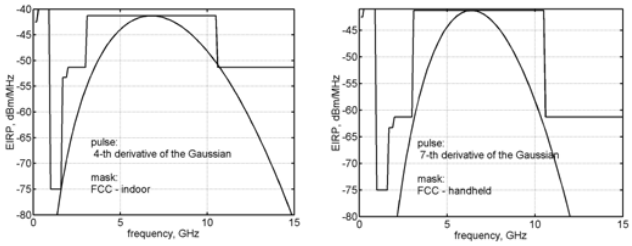


Fig. 4. PSD of the 4-th and 7-th derivative of the Gaussian pulse

The achievable transmitted power is -5.1 for indoor and -6.5 dBm for handheld devices, which is quite close to the possibility of the approach proposed in this article. The latter, however, as mentioned before, provides some additional advantages.

VII. Conclusions

An approach for UWB spectrum shaping was reviewed and some possible implementations were suggested. The performance improvement of the UWB communication systems, obtained through this approach, was assessed. The improvements are considerable, and the technical implementation is completely feasible, although at a higher price. The latter, however, will go down in near future. Further research will focus on the sensitivity of the obtained PSD to the inaccuracies of the technical implementation and on optimization of the generated pulses.

References

[1] Win, M.Z., Scholtz, R., "Impulse Radio: How it Works", *IEEE Communications Letters*, Feb. 1998.

[2] Mitchell T., "Broad is the way", IEE REVIEW, January 2001

[3] Lehmann N., A. Haimovich, "New approach to control the power spectral density of a timehopping UWB signal", Conf. on Information Sciences and Systems, The Johns Hopkins University, March 12-14, 2003

[4] L. Piazzo, "Some basic facts about UWB EM compatibility and UWB spectrum", Internal report for the Whyless.com project, May 2001.

[5] Asbeck P., Ian Galton, K. Wang, J. Jensen, K. Oki, C. Chang, "Digital signal Processing – Up to Microwave Frequencies", *IEEE Transactions on Microwave Theory and Techniques*, vol. 3, March 2002

[6] http://www.compoundsemiconductor.net/magazine/article/9/1/2/1, "IEDM highlights include SiGe HBTs operating at 350 GHz"

[7] http://www-3.ibm.com/chips/news/2002/0225_fast.html, "IBM creates world's fastest semiconductor circuits"

[8] Furaxa, Inc., "Next-Generation Ultra-Fast Signal Synthesis and Sampling", www.furaxa.com

Overview of Ultra Wideband Multiple Access Systems

Dobri M. Dobrev[1], Nikola K. Stanchev[2]

Abstract – **In this paper in brief is presented multiple access system realized by a new technology known as Ultra Wide-band (UWB) radio. The signals, especially the pulses, used in UWB communications are described. Two very important conceptions for realization of UWB multiple-access systems are presented - the classical with Time - Hopping Pulse Position Modulation (TH PPM) and the Delay-hopped transmission - reference (DHTR) systems. The advantage of using chaotic signals in these systems is proposed.**

Keywords – **Ultra wide band, pulse position modulation, time hopping, delay hopping**

I. Introducion

The UWB radio or so called Impulse Radio (IR)[1,2] uses a train of very short pulses with pulse duration in order of nanosecond. In frequency domain the pulses occupy very large bandwidth, greater than 1 GHz. Further, the pulse train has a very low duty cycle, less than 1%. This is the very important property of UWB signals that results in a very low average power. The main advantages of UWB follow:

- The pulse train is directly fed to transmitted antenna and because of that an UWB is known as a " carrier- free" communication. Hence, no need of RF up-converters and down-converters, the both the UWB receivers and transmitters are with lower complexity and are cheaper than conventional systems.
- UWB is very suitable for dense multipath (inherent in indoor as office buildings). For example, if the pulse width is 1 ns, then no fading when differences in the paths are greater than 1 foot (30 cm). Combining of delayed signals with a RAKE receiver is a very simple task and it could obtain an excellent system performance. System designers no need of great margin in link budget.

II. Common Characteristics of UWB Multiple Access Systems

Data modulation in UWB is performed by changing the pulse position, shape or polarity (phase). Multiple access capability of the UWB systems due to nature of UWB signals is realized by spread spectrum Time Hopping (TH) technique. In the common case are used pseudorandom sequences for TH. When the low probability of detection and interception are required then a chaotic signal are well suited.

In essence, by PN sequences or chaotic signals are performed a randomly time shifts of the pulses and this results in two consequences. The first is a smooth spectrum shape without a great spike. Second, the probability of collision between the signals of different users is decreased.

The receivers in the IR are optimal (correlation) receivers. After despreading in the receiver is performed correlation between the received signal and a template signal. The template signal must be aligned with the received signal in the time and must be with identical waveform. In other words, a synchronization process is realized in the receiver and system designers must perform channel estimation. The synchronization process take a finite time and increases software complexity.

Other approach to realize an UWB multiple access system is Delay-hopped Transmission Reference (DHTR) technique. In this system is transmitted a reference signal that undergoes the same distortion and delay as the information signal. Hence, no need of synchronization of the individual pulses and any channel estimations. Multiple-access capability is obtained trough the principle of code division multiple access (CDMA) and usage of Delay-Hopping technique.

III. Basic UWB Pulse Shapes

A. Gaussian pulse shape – monocycle

The basic pulse shape proposed for use in UWB communication system is the Gaussian pulse. The pulse generator produces a smooth impulse, which is approximately Gaussian pulse. The effect of antennas over the pulse is that of differentiation. Hence, the transmitted pulse is first derivative of Gaussian function. This is so called monocycle [3]

$$\nu(t) = A\sqrt{2e}\frac{t}{\tau}\exp\Big(-\frac{t^2}{\tau^2}\Big) \qquad (1)$$

where τ is time decay constant that defines the pulse width; A is the peak amplitude of the pulse. On the other hand, the pulse width specifies the bandwidth and center frequency. The monocycle is shown on Fig. 1.

In the frequency domain, the Gaussian monocycle envelope is:

$$\nu(w) = Aw\tau^2\sqrt{2\pi e}\exp\Big(-\frac{w^2\tau^2}{2}\Big). \qquad (2)$$

It can be seen on Fig. 2.

The center frequency is then:

$$f_c = \frac{1}{2\pi\tau},\ \text{Hz}. \qquad (3)$$

[1]Dobri M. Dobrev is with Faculty of Communications and Communication Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: dobrev@tu-sofia.bg

[2]Nikola K. Stanchev is with Faculty of Communications and Communication Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: stanchev_n@abv.bg

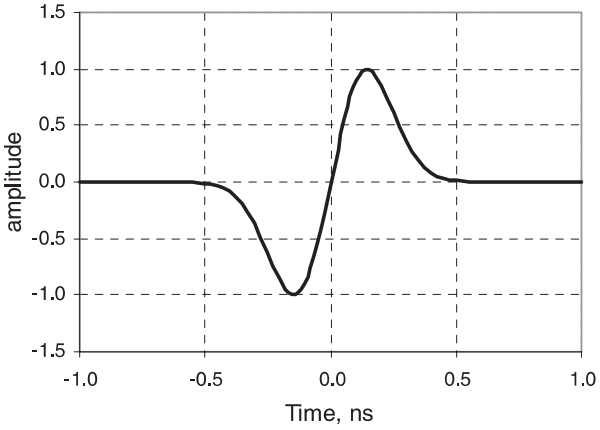


Fig. 1. The monocycle

B. Hermite-Based pulse shapes

Recently, in [4,5] was proposed a novel pulse shapes based on a modified Hermite polynomials (MHP) that can be presented by

$$h_n(t) = e^{-\frac{t^2}{4}} h_{e_n}(t) = (-1)^n e^{\frac{t^2}{4}} \frac{d^n}{dt^n}(e^{-\frac{t^2}{2}}) \qquad (4)$$

where $n = 0, 1, 2...$ and $-\infty < t < \infty$. The MHP pulses are shown on Fig. 3.

The variation of pulse width is small when the order is changed. Hence, the bandwidth is not varied significantly. Due to the MHP are orthogonal functions; the pulses could be used for a M-ary modulation or for MA system. For example, 00,01,10,11 can be represented by MHP pulses of order $n = 1, 2, 3, 4$. In this way higher data rates can be achieved simply by sending different pulse shapes. This can be extended to a multi-user system, assigning for example MHP pulses of order $n = 1, 2$ to user 1 for 0,1 and $n = 3, 4$ to user 2 for 0,1.

It can be seen in [6] that the MHP pulses of order equally to or smaller than 8 are applicable. The higher order pulses have very short autocorrelation pick. Hence, they are more sensitive to pulse jitter and it would result in degradation in the system performance.

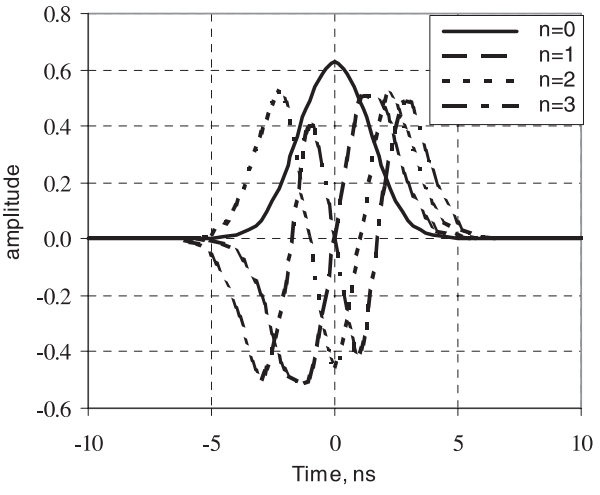


Fig. 3. The MHP pulses of order $n = 0, 1, 2, 3$

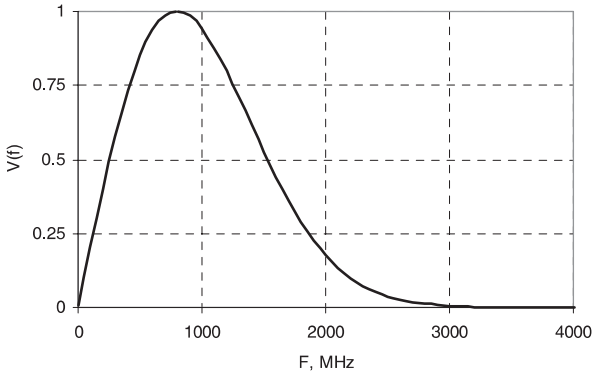


Fig. 2. The spectrum of monocycle

IV. Basic UWB Multiple Access Systems

A. UWB multiple access system with time-hopping impulse modulation

This is the first proposed multiple-access (MA) system for UWB communications [7]. In this system multiple access capability is realized by spread spectrum time-hopping (TH) technique and a pulse position modulation (PPM) is used for data modulation. A typical hopping format with data modulation is given by

$$s^{(k)}(t) = \sum_{j=0}^{\infty} w(t - jT_f - c_j^{(k)} Tc - \delta d_{[j/N_s]}^{(k)}. \qquad (5)$$

Here $w(t)$ represents the transmitted monocycle waveform and superscript (k) indicates transmitter-dependent quantities. The j-th monocycle of user (k) nominally beginning at time $jT_f + c_j^{(k)} Tc + \delta d_{[j/N_s]}^{(k)}$. The time shift T_f is the frame (pulse repetition) time and equals the average time between pulse transmissions. The frame time T_f may be a hundred to a thousand times the monocycle width.

Pseudorandom Time-Hopping: To eliminate catastrophic collisions in multiple accessing, each link (indexed by k) is assigned a distinct time-hopping code $\{c_j^{(k)}\}$. These hopping codes $\{c_j^{(k)}\}$ are periodic pseudorandom codes with period N_p. Each code element is an integer in the range $0 < c_j^{(k)} < N_h$. It assumed that $N_h T_c < T_f$. On the other hand, using pseudorandom time hopping the shape of spectrum is smoothed.

Data modulation: The data sequence $\{d_i^{(k)}\}$ of transmitter (k) is a binary (0 or 1) symbol stream. In essence, a pulse position modulation used in this system means that, when data symbol is 0, no additional time shift, but time shift of δ is added to a monocycle when the symbol is 1. In this system is used oversampled modulation with N_s monocycles transmitted per symbol. In the receiver's correlator is performed an integration over N_s hops and it suggests that the signal to noise ratio (S/N) is increased. Usually, the number of pulse per symbol N_s is a few hundreds and the system performance is improved significantly.

This system uses simple approach to realize basic advantages of UWB for Multiple Access (MA) system in dense multipath environment.

B. *UWB multiple access system with time-hopping and M-ary PPM modulation*

Further investigation of MA system described above results in a Block Waveform or M-ary PPM modulation [8-11]. Let the TH PPM signal is

$$x^{(\nu)}(t) = \sum_{k=0}^{\infty} w(t - kT_f - c_k^{(\nu)} Tc - \delta^k_{d^{(\nu)}_{[j/N_s]}}). \quad (6)$$

Where the superscript (ν) indicates user-dependent quantities $1 < \nu < N_u$ and the number of simultaneous active users is N_u. The index (k) is the number of pulses that has been transmitted. The time shift corresponding to the data modulation is $\delta^k_{d^{(\nu)}_{[j/N_s]}} \in \{\tau_1 = 0 < \tau_2 < \cdots < \tau_\eta\}$ with $\eta \geq 2$ an integer. The data sequence $\{d_m^{(\nu)}\}$ of user ν is an M-ary symbol stream.

If we define a hopping function

$$H_m^{(\nu)}(t) = \sum_{k=mN_s}^{(m+1)N_s-1} T_c c_k^{(\nu)} p(t - kT_f)$$
$$p(t) = \begin{cases} 1, & \text{if } 0 \leq t \leq T_f \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

and a signal set

$$S_i(t) = \sum_{k=0}^{N_s-1} w(t - kT_f - \delta_i^{(k)}) \quad (8)$$

for $i = 1, 2, ..., M$, then (6) can be rewritten

$$x^{(\nu)}(t) = \sum_{m=0}^{\infty} S^{(\nu)}_{d_m}(t - mN_sT_f - H_m^{(\nu)}(t)) \quad (9)$$

where m indexes the transmitted symbols. The PPM signal $S_i(t)$ represents the i-th signal in an ensemble of M information signals, each signal completely identified by the pulse shape w(t) and the sequence of time shifts $\{\delta_I^k\}$, $k = 0, 1, ..., N_s - 1$. In [8] are represented M-ary PPM sets, namely orthogonal (OR), equally correlated (EC) and N-orthogonal (NO) signal sets. For example, EC M-ary PPM signal set is defined by time shift pattern $\delta_i^k = a_i^k . \tau_2$, with $\tau_2 \in (0, T_w]$, where the pulse duration is T_w and $a_i^k \in \{0, 1\}$, $k = 0, 1, 2, ..., N_s - 1$ the i-th cyclic shift of the m-sequence, for $i = 1, 2, ..., M$. It can be shown that the EC M-ary PPM signals have normalized correlation values $\alpha_{ii} = 1$ and $\alpha_{ij} = \lambda$, $i \neq j$, where $\lambda < 1$.

The simulation results in [8,11] depicts that using higher values of M other than 2, it is possible either to improve the probability of detection for a fixed number of users N_u, or to increase the number of users for a fixed probability of error, without increasing each user's signal power.

This very attractive property of M-ary PPM modulation could be proved by following considerations. If the pulse frame time T_f and the bit rate are fixed, then the number of pulses N_s used in the M-ary modulation $N_s = \log_2(M).N_s^{(2)}$ are increased with respect to the number of pulses used in a binary modulation. In this way $E_s = N_sE_w = \log_2(M)N_s^{(2)}E_w = \log_2(M)E_b$, where E_s and E_b are the energy of symbol after receiver's correlator in M-ary and binary communications, respectively. Hence, after integration in an M-ary PPM receiver correlator we have energy greater than those in the binary receiver.

Using the M-ary PPM sets with large values of M, it is possible to increase the number of users supported by the system for a given multiple-access performance and bit transmission rate, making efficient use of the signal to noise ratio available.

C. *Delay-hopped transmitted-reference UWB multiple access system*

The Delay-Hopped Transmitted-Reference (DHTR) is the last proposed technique for UWB communications [12,13]. As shown on Fig. 4, a DHTR UWB signaling scheme is implemented by transmitting pairs of identical pulses (called doublets) separated by a time interval D, known to both the receiver and the transmitter. The transmitted data is encoded by the relative phase of the two pulses; in this scheme is used binary phase modulation, e.g. a changing of pulse polarity is performed. More than one doublet is associated with each information bit, as long as all the associated doublets have the same time interval D between pulses and the same relative polarity of pulses.

The interval between doublets, called the pulse repetition time (PRT), may be varied in order to shape the spectrum of the transmission.

The sequence of early pulses in the sequence of doublets on Fig. 4, since it contains no information, it should be considered as a carrier (reference) signal. The sequence of second pulses would then be a modulated (information) signal.

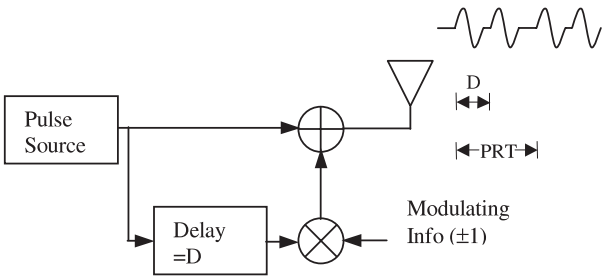


Fig. 4. Generation of DHTR signal

The Delay-Hopping refers to the method of varying the delay used in DHTR transmission according to a fixed pattern known to the transmitter and to the receiver. This pattern constitutes a code word, and multiple-access capability is obtained trough the principle of code division multiple access (CDMA).

On the signal level, a DHTR code word consists of N_c chips, transmitted sequentially. Each chip is composed of N_p doublets, with interpulse delay D_i, for $i = 1, 2, ..., N_c$ and with the same polarity.

The DHTR receiver consists of a bank of pulse-pair correlators connected trough a LNA to an antenna. Each correlator in the bank of pulse-pair correlators responds to a different delay. A CDMA code word correlator follows the bank of pulse-pair correlators. The code word correlator adds

the individual chip waveforms in phase, producing an output, which is a high SNR version of the individual chip waveform.

In contrast to conventional pulsed UWB receivers, the use of the transmitted-reference technique makes synchronization with the individual pulses unnecessary.

D. UWB MA system with chaotic pulse position modulation (CPPM)

In the classical UWB MA system is used a pseudorandom sequences for a time hopping technique. In the CPPM, which is proposed in [14], is used chaotic signal for this purpose.

In the more chaos-based communication schemes synchronization is very sensitive (susceptible) to distortions and noise. In the CPPM the information about the state of the chaotic system is contained entirely in the timing between pulses, the distortion that affect the pulse shape will not significantly influence the ability of the chaotic-pulse generators to synchronize. Furthermore, using chaotic-pulse generators the system reacquires synchronization automatically, without any specific "hand-shaking" protocol. The decoder only need to detect two correct consecutive pulses in order to reestablish synchronization.

In the previous described UWB systems the use of pseudorandom sequences results in the presence of characteristics frequency in transmitted signal spectrum. Chaotically varying spacing between the pulses enhances the spectral characteristics of the system by removing any periodicity from the transmitted signal. Due to the absence of characteristic frequencies, chaotically positioned pulses are difficult to observe and detect for an unauthorized user.

V. Conclusions

In this paper was presented the main UWB multiple access communication systems. The M-ary PPM modulation has a system performance better than this in binary PPM system for a given signal to noise ratio (SNR). The Delay-Hopped Transmission-Reference (DHTR) technique is very suited for burst mode communications where synchronization process must be take up very short time. It was represented that the usage of chaotic signal in UWB systems is viable and has many advantages over the usage of classical PN sequences.

References

[1] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works?", *IEEE Commun. Letters*, vol. 2, no. 2, pp. 36-38, February 1998

[2] M. Z. Win and R. A. Scholtz, "Impulse Radio", invited paper, in *Proceeding of IEEE PIMRC'97*, Sep. 1997

[3] U.S. patent 6430208: Ultra-wide Band Communication System and Method

[4] M. Ghavami, L. B. Michael, R. Kohno, "A Novel UWB Pulse Shape Modulation System", *Wireless Personal Communications Journal*, vol.23, issue 1, pp. 105-120, October 2002

[5] M. Ghavami, L. B. Michael, R. Kohno, "Hermite Function Based Orthogonal Pulses for Ultra Wideband Communication", in *Proceeding of WPMC'01*, 2001

[6] M. Ghavami, L. B. Michael, R. Kohno, "Effect of Timing Jitter on Hermite Function Based Orthogonal Pulses for Ultra Wideband Communication", in *Proceeding of WPMC'01*, 2001

[7] R. A. Scholtz, "Multiple Access with Time Hopping Impulse Modulation", invited paper, *Proceeding of IEEE MILCOM'93*, Oct. 1993

[8] F. Ramirez-Mireles, "Performance of Ultra Wideband SSMA Using Time Hopping and M-ary PPM", *IEEE JSAC*, Vol. 19, No. 6, June 2001

[9] F. Ramirez-Mireles and R. A. Scholtz, "Multiple Access with Time Hopping and Block Waveform PPM", in *Proceeding of IEEE ICC Conference*, pp. 775-779, June 1998

[10] F. Ramirez-Mireles and R. A. Scholtz, "Multiple Access Using SS Time Hopping and Block Waveform Pulse position Modulation, Part 1: Signal design", in *Proceeding of ISITA Symposium*, October 1998

[11] F. Ramirez-Mireles and R. A. Scholtz, "Multiple Access Using SS Time Hopping and Block Waveform Pulse position Modulation, Part 2: System performance", in *Proceeding of ISITA Symposium*, October 1998

[12] Hoctor, R., Tomlinson, H., "Delay-Hopped Transmitted- Reference RF Communications", *Proc. IEEE UWBST 2002 Conf.*, May 2002

[13] Hoctor, R., Tomlinson, H., "Delay-Hopped Transmitted- Reference RF Communications Experimental Results", *Proc. IEEE UWBST 2002 Conf.*, May 2002

[14] N. Rulkov, M. Sushchik, L. Tsimring and Al. Volkovskii, "Digital Communication Using Chaotic-Pulse-Position Modulation", *IEEE Trans. on Circuits and Systems*, vol. 48, no. 12, pp. 1436-1444, Dec. 2001

Maximum Operating Range for RF Communication of Tactical VHF/UHF Networks

Plamen Hristov Vanchev [1]

Abstract – **For analysis of Maximum Operating Range of VHF/UHF Communication, the Hata-Okumura propagation model is used. In the latter, according to the military standards, different modulation techniques are researched, which are used for Tactical Communication. In this paper the mainly used modes, such as voice and data communication and synchronization, are presented. The values of receiver sensitivities of the radios are taken from MIL-STD Standards and the characteristics of the most popular tactical radios, which are used around the world. Conclusions and recommendations are made for the building of the Radio Cannels in Security Services Control and Management.**

Keywords – **Ultra wide band, pulse position modulation, time hopping, delay hopping**

I. Introducion

Nowadays the communication between different people (or groups of people) is a integral part of human life. For instance are used improved and newer systems especially for security service management after the world changes after the attacks against World Trade Center in New York in 2001.

These changes in the world have increased the security service requirements too. A support of reliable and secure performance for each operation is usually determined by stable communication with maximum operation range. This would be realized by means of interaction in very good quality between different participants and high coordination on operation stage. It sets a part radiocommunication as one of fast growing branches of high technology. This approach provides secure communication between mobile objects on huge distances with radio networks. VHF/UHF is used Band for security services managements, according to the ITU standards. The main radio links of Police and Army are established in these bands.

II. VHF/UHF Point-to-Point Communication

Single channel VHF/UHF radios are used for establishment of direct connection (point-to-point) between mobile users and command. They work in $100 \div 512$ MHz Frequency Band. In the most cases these radios are unique communication equipment of Special Forces, according to the requirements that they are communicated into the area of special operation. This demands high quality of tactical & technical parameters as the most important are [5]:

- Maximum protection against radio intelligence of the enemy (ECCM – electronic contra-contra measure);
- Communication service without searching and adjustment;
- Work in high jamming conditions (ECM – electronic contra measure);
- Minimum communication duration without stops and repeats;
- Information security transmission;
- Selected user calling or selected user groups.

All of these hard technical requirements can be met using of different operation modes. Incapability of old radios to satisfy high tactical & technical parameters, connected mainly with ECM independent of the enemy or crime and current technology development, has brought a qualitative leap to contemporary radio communication equipment development. DSP is implemented everywhere and every module uses computer management. Standards with high requirements against RF jamming (MIL-STDs) were created for improvement of ECCM and increase of battle possibility of RF communication in ECM conditions in NATO.

One of the fastest and secure ways for quick transport of special groups for a mission is used by Air Craft. Air Traffic Control and the linking between command point and special group is usually organized with VHF/UHF radios. For example many VHF/UHF Air Craft Radios are utilized in NATO Air Force and national polices. Harris radios are built in AH-64D Apache, Bell 206(406), Sikorsy and etc. Rohde & Schwarz is the basic contactor for Eurofighter-2000. Besides other world leaders of RF production present VHF/UHF radios on the market. The most popular radios, their modes, the used modulations and the basic technical parameters are presented in Table 1 [4, 6-11].

For hidden management of any mission it's quite important to know the maximum operation range where reliable and secure RF communication can be provided. It depends on the following:

- Carrier frequency;
- Conditions of radio wave propagation;
- Output power;
- Receiver's sensitivity;
- Operation modes;
- Modulation and etc.

The most important circumstance for establishment of any RF connection is to appear information signal with necessary power, according to the chosen mode, on the receiver's

[1]Plamen Hristov Vanchev, National Military University, Aviation Faculty, Department of Electronics, Communication and Navigation Equipment of the Air Craft, Dolna Mitropolia-5856, Pleven, Bulgaria, E-mails for contact: pvanchev@af-acad.bg, pvanchev@yahoo.com

Table 1.

Radio Manufacturer (country)	BASIC SPECIFICATIONS					
	Frequency Range (MHz) Power Output (W)	Channel Spacing (kHz)	Emission Modes (Functions) / Modulations	Data Interface	Sensitivity (dBm)	Data Rate
RF-5800U-MP Harris (USA)	90 ÷ 420 MHz 20 W FM; 10 W AM	5, 6.25, 8.33, 12.5, 25 kHz	Analog Voice, 16 kbps CVSD Voice, 16 kbps Data / AM, FM, ASK/FSK	Synchronous or asynchronous 75 ÷ 115 kbps RS232/422	FM– 116 dBm; AM –110 dBm for 70% modul.; minimum for 10 dB SINAD	16 kbps. Options for 48 and 64 kbps
TRC 6020 Thales (UK/France)	30 ÷ 400 MHz 15 W FM; 10 W AM	8.33 and 25 kHz	Voice: clear / secure, Data: point-to-point, option (TDMA, L22) / AM, FM, MSK, FSK	ARINC 429 or RS 422	FM– 106 dBm; AM –103 dBm for 30% modul.; FFH –100 dBm	16 kbps
MRR Ericsson (Sweden)	30 ÷ 400 MHz 5 W, 50 W	45 kHz	CVSD Voice, 16 kbps Data / GMSK	RS-232 by EURO-COM D/1	- 116 dBm	16 kbps
M3TR Rohde&Schwarz (Germany)	1,5 ÷ 400 MHz 10 W; 15 W	5, 8.33, 12.5, 25 kHz	Analog Voice, 16 kbps CVSD Voice, 16 kbps Data	Internet/ Intranet access via IP-interface (UDP/TCP)	FM– 117 dBm; AM –110 dBm	Up to 64 kbps
Spectre Datron (USA)	30 ÷ 88 MHz, Option up to 512 MHz 10 W	12.5, 25 kHz	Voice: FF FM, simplex or half-duplex, Data: 16 kbps output, internal FEC / AM, FM, ASK/FSK	Interface: RS232, 100 bps to 56 kbps, asynchronous	- 106 dBm;	Up to 64 kbps
CNR - 9000 Tadiran (Israel)	30 ÷ 88 MHz, Option up to 512 MHz 10 W	25 kHz	Analog Voice, 16 kbps CVSD Voice, 19.2 kbps Data / AM, FM, FSK	Automatic data rate adaptation	-108 dBm	Up to 32 kbps

input. Therefore it's necessary to know the receiver's input signal intensity. Radio wave propagation is related to propagation loss in environment. This means that the receiver's power would be rather different from the transmitted power.

III. VHF/UHF Propagation Model

The increasing demand for mobile communication has led to the need of more efficient propagation prediction models as one of the essential parts of the radio network planning tools and operation range determination. The basic equation, which is used for an obtaining of input receiver's power P_r is the following [3]:

$$P_r = P_t\Big(\frac{G_r G_t \lambda^2}{16\pi^3 r^2}\Big), \qquad (1)$$

where P_t is transmitted power in dB, G_r (G_t) indicates the rate of Antenna Directivity Gain of receiver (transmitter), r is the path and λ is wavelength.

More propagation models, which are suitable for some different radio waves and propagation conditions, are used into practise. The large-scale propagation models give results as path loss versus range. The characteristics of the basic point-to-point path loss prediction models widely used in generating signal coverage map, co-channel interference area map and handoff occurrence map follow below.

Log-distance path loss model, theoretical and experimental propagation models indicate that the average received signal power decreases logarithmically with distance and [1]:

$$PL(d) = \overline{PL}(d_0) + 10n \lg\Big(\frac{d}{d_0}\Big) + X_\sigma, \qquad (2)$$

where path loss, PL, is in dB, n indicates the rate at which the path loss increases with distance d, d_0 is the reference distance, which is determined by measurements close to the transmitter, X_σ is a zero-mean Gaussian distributed random variable (in dB) with standard deviation σ. X_σ accounts for the variation in average received power due to the shadowing. The values of n and σ? are derived from measured data. A smaller value of σ means more accurate path loss prediction. Typical values for n are: $n = 2$ (free space), $n = 2.7$ to 3.5 (urban area), $n = 3$ to 5 (shadowed urban area). However this model requires a huge number of empirical data and it's used when it is not necessary to make any frequency planning.

When it is looked at the RF links usually the antennas are isotropic. The propagation loss is the difference between the radiated power and the receiver power. It can be obtained with the equation [1]:

$$L_p(\text{dB}) = P_t - P_r = P_t(\text{dBW}) - \\ - E(\text{dB}\mu\text{V/m}) - 10\log_{10}\left(\frac{\lambda^2}{4\pi}\right) + 145.8, \quad (4)$$

where E(dBμV/m) is received field strength of an isotropic antenna. The most popular signal prediction modelling is Hata-Okumura model [2]. It is based on extensive measurements in urban area over a quasi-smooth terrain using vertical omni-directional antennas at the base and mobile stations assuming $h_t = 1000$ m and h_r(mobile)= 1 m. This model is a proper model for an analysis of propagation loss when it is researched cellular networks and communication quality between ground-to-air. The median attenuation relative to free space $A(f,d)$ was presented by Okumura as a family of curves plotted as function of frequency (100 MHz $< f <$ 1920 MHz) and as function of distance from the base station (1 km $< d <$ 100 km). The model previews correction factor accounting for the terrain type, G_{area}, which is given in another set of curves, as well as different expressions for antennas height gain factors, $G(h_t)$, $G(h_r)$, and for values of h_t, h_r, others than the assumed in the curves (the antenna pattern is not taken into account). The median value of the path loss following the model can be expressed as [2]:

$$L = L_0 + A(f,d) - G(h_t) - G(h_r) - G_{area}, \quad (5)$$

where L_0 is the free space propagation loss. The formula for median path loss is given by (6), where f is in MHz, h_t is the effective base station height in meters ($30 < h_t < 200$ m), h_r is the effective mobile antenna height ($1 < h_r < 10$), the distance d is in km, $a(h_r)$ is a correction factor related to the mobile antenna height. The values of $a(h_r)$ for small to median sized cities can be found in [6]. The model is applicable for frequencies from 100 MHz to 1500 MHz and $1 < d < 100$ km.

$$L = \begin{cases} 69.55 + 26.16\log_{10} f - 13.821\log_{10} h_t - \\ \quad -a(h_r) + \log_{10} d(44.9 - 6.55\log_{10} h_t), \\ \qquad \text{for urban area} \\ L(\text{urban}) - 2\left[\log_{10}\left(\frac{f}{28}\right)\right]^2, \\ \qquad \text{for suburban area;} \\ L(\text{urban}) - 4.78(\log_{10} f)^2 - \\ \quad -18.33\log_{10} f - 40.98, \\ \qquad \text{for open rural area.} \end{cases} \quad (6)$$

The above-described model is among the best (and simplest) models, it has standard deviation between predicted and measured path loss value about 10 to 14 dB, [1], and is used in practically all cellular and land mobile radio systems planning tools. This model can be used for calculation of propagation loss of air-to-ground VHF/UHF communication. It's important when it is necessary to obtain the minimum distance for reliable, but hidden, operation management. This model allows determining where the staff must dispose the Air C4I point.

IV. Simulations and Conclussions

When the security and hidden management is organized, Air Command Point usually situated into helicopters is used. The manager staff officer must know the maximum operation range for reliable communication. It can be calculated by (6). The simulation results for propagation losses are presented on figure 1, figure 3 and figure 2. Onto all of the figures propagation losses in Open Area are presented with red curves, propagation losses in Sub-urban Area are presented with green curves and propagation losses in Urban Area are presented with blue curves. There is a research about three typical Helicopters Output Powers such as 5 W, 10 W and 15 W.

The propagation loss for carrier frequency 220 MHz is presented on fig. 1. This frequency is in the beginning of Security Air Traffic Control Frequency Band. A simulation of propagation loss for carrier frequency 300 MHz can be shown on fig. 2. This frequency is in the middle of the Band. The flight altitude for these cases is 100 m. A research of propagation loss for carrier frequency 400 MHz is presented on fig.3. This frequency is in the end of the Band and the

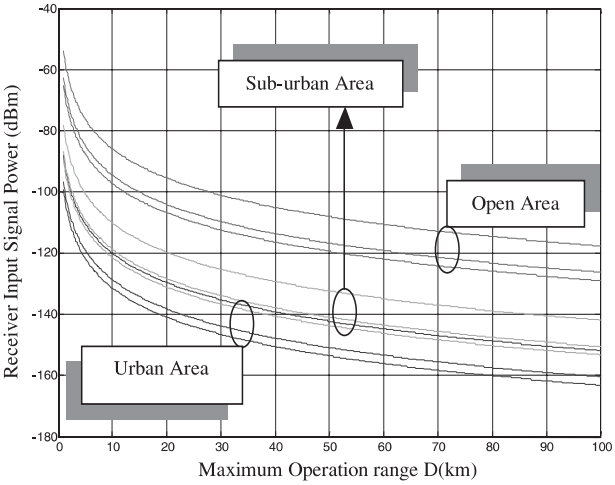


Fig. 1. Propagation Losses for Carrier Frequency 220 MHz

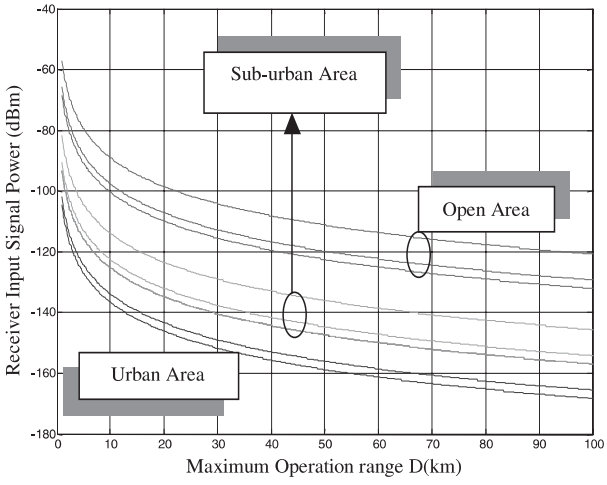


Fig. 2. Propagation Losses for Carrier Frequency 300 MHz

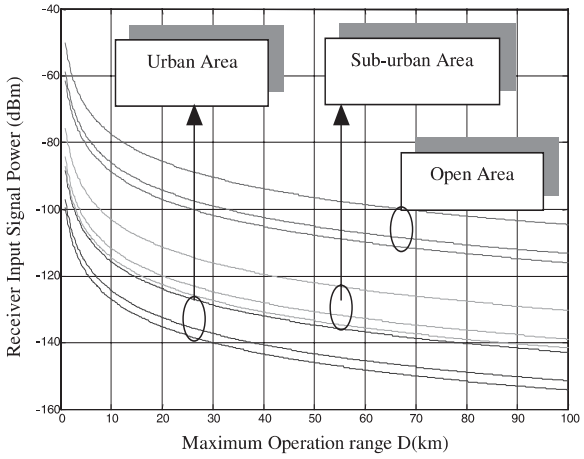


Fig. 3. Propagation Losses for Carrier Frequency 400 MHz

flight altitude is 500 m. The higher altitudes are not usually used for Helicopter Flight.

The researches lead to the following conclusions:

1. It's possible to organize hidden operation management by Air Command Point;

2. In open area the management with operation range for over 30 km can be organized independent of operation mode and carrier frequency;

3. For Urban and Sub-urban areas the maximum operation range is about 10 km. This is sufficient range for hidden management;

4. It is too important to make quality frequency panning due to escape used channels and to establish communication between command point and the mission with maximum covert to maximum operation range;

5. Reliable radio links for over 50 km can be established for low band frequencies. This can be used for military command;

6. The combination between hidden management and hidden Air Command Point is the best approach special service.

Present research shows the possibility for hidden C4I application. It is very important nowadays when the crimes can get special equipment via illegal market. The application of hidden command is a problem, which can be solved by tactical specialists and operation management staffs and officers.

References

[1] W. C. Y. Lee, *Mobile Cellular Telecommunications*, 2nd Edition, New York, McGraw-Hills, 1995

[2] Masaharu Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Service", *IEEE Trans. on Vehicular Technology*, vol. VT-26, no.3, pp. 317-325, August 1980

[3] H. Hristov, E. Altimirsky, "Radio engineering electrodynamics and Radio wave propagation", Techniques, Sofia, 1990;

[4] JANE'S Military Communication, 1999-2000, fourteenth edition

[5] Radio Communications in the Digital Age, Volume Two: VHF/UHF Technology, By Harris Corporation, RF Communications Division, Library of Congress Catalog Card Number: 00 132465, First Printing, June 2000

[6] www.titan.com

[7] www.thalesgroup.com

[8] www.rsd.de

[9] www.harris.com

[10] www.ericsson.com

[11] www.tadiran-com.co.il

Optimisation of Parameters of Output Filters for Direct Digital Synthesizers

Ognian I. Petkov[1]

Abstract – **Investigation of the possibility for optimum choice of phase and amplitude characteristics of the output filters of direct digital synthesizers for better noise and spurious performance.**

Keywords – **Direct Digital Synthesizers, noise, spurious and alias frequencies, output filters, amplitude and group delay time characteristics**

I. Introducion

Direct Digital Frequency Synthesis (DDFS) becomes a very popular technique for generating frequencies whenever very precise frequency resolution and fast frequency switching is needed. The most common DDFS architecture includes a periodically overflowed phase accumulator, for generating and storing phase information and uses a ROM based look up table to compute the sine function. The digitally generated sine from the ROM is next passed to a digital to analog converter (DAC) where it is converted to a pulsed analog form and finally this form is filtered by a low pass filter. The frequency of the generated sine wave is controlled by the Frequency Control Word (FCW). The output frequency could be defined from Eq. (1):

$$f_{\text{out}} = \text{FCW}\,\frac{f_{\text{clk}}}{2^j} \tag{1}$$

where f_{out} is the synthesized frequency, f_{clk} is the clock frequency and "j" is the word length of the phase accumulator. The spectral density of the generated signal could be expressed with the help of the formulae, given in Eq. (2):

$$S(f) = e^{-j\pi f/f_{\text{clk}}}\frac{\sin(\pi f/f_{\text{clk}})}{\pi f/f_{\text{clk}}} \times \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} c_m\,\sigma(f - nf_{\text{clk}} - mf_{\text{out}}) \tag{2}$$

where n and m are the numbers of harmonics of the clock and output frequencies.

As can be seen from Fig. 1 the DDFS causes all frequency components residing in the generated signal to be aliased about every harmonic of the clock frequency f_{clk}. If the output is a perfect sine wave, then the spectrum contains the frequencies $nf_{\text{clk}} \pm f_{\text{out}}$. The component, corresponding to n=0 is the desired sine wave and the others are the images or alias frequencies.

[1]Ognian I. Petkov is with the Faculty of Electronics, Technical University – Varna, Studentska 1, 9010 Varna, Bulgaria, E-mail : oipbg@yahoo.com

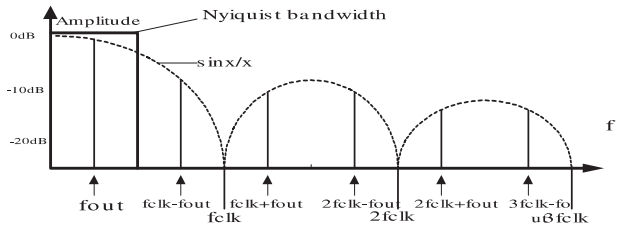


Fig. 1.

For many applications, the DDS solutions have distinct advantages over the equivalent frequency synthesizers, employing PLL circuitry. This advantages could be used more completely if we know the real nature of the synthesized output sine wave in DDS. One important subject is the generation of spurious spectrum lines and the noise at the output of the DDS and the appropriate measures for diminishing their influence on the purity of the output spectrum.

In this paper the sources of the noise and spurious frequencies will be analyzed and some solution for the output filter design will be proposed.

II. Noise and Spurious Sources in DDS

As we have seen before the most dangerous and difficult to handle are the spurious frequencies, arising from the digital nature of the generated signal. But the output spectrum contains additional spurious frequencies and wideband noise, which had to be taken in consideration. The most important of them are:

1. Spurious lines and noise due to the truncation of the phase accumulator bits addressing the sine ROM.

2. Noise due to the distortion from the compression of the sine ROM.

3. Errors, due to the limited precision of the samples, stored in the ROM.

4. Noise from the nonlinearities in the digital to analog conversion.

We need to know the carrier to noise (C/N) ratio for every one of these noise and spurious sources. From [2] it is calculated that the worst case C/N caused from the truncation of the least significant bits in the phase accumulator is from Eq. (3):

$$C_t/N \approx 6.02 - 3.92\ \text{dBc} \tag{3}$$

where k is the number of most significant bits of the phase accumulator, used for programming the ROM table.

The finite character of the quantization in the sine ROM values also leads to the output DDFS spectrum impairments.

The carrier to noise ratio for this case could be calculated from Eq. (4):

$$C_q/N \approx 6.02m + 1.76 \text{ dBc} \quad (4)$$

In high speed and high resolution (> 10 bits, > 50 MHz) DDFS most of the spurs are generated from the analog errors in the DAC. In the advanced high speed DDFS the DAC is the most critical component. The anomalies in the output spectrum, caused by the DAC do not follow the $\sin(x)/x$ roll off response.

The carrier to noise ratio on the output of the DAC could be defined from the well known formulae in Eq. (5):

$$C_q/N \approx 6.02n + 1.8 \text{ dBc} \quad (5)$$

where n is the number of bits in the DAC. In almost every state of the art DDS the number of bits after truncation in the phase accumulator is big enough in comparison with the bits in the DAC, so here one can assume that the main source of wideband noise at the output of the DDS is the quantizing noise of the DAC, which can be well predicted. For an example if we take the AD9854, which has 48 bits phase accumulator, 17 bits after truncation and 12 bits DAC. The C/N from the truncation can be calculated and is more than 102 dB and the C/N from the DAC in this case is 72 dB.

The nonlinear process in the DAC is generating harmonically related spurs in the output spectrum. The amplitude of the spurs is difficult to predict, but the location of the spurs is harmonically related to the output frequency of the synthesizer. A detailed analysis could be done for every particular DAC . At this stage of the development of the direct digital synthesis it could be seen that only the lack of high frequency and linear DACs is that stops their more intense application in the frequency bands for portable and computer wireless communications. The comparatively high level of wideband noise and the presence of alias and image frequencies prevents their applications in precise measurement equipment.

If we summarize the analysis we can make the conclusion that the main sources of spurious frequencies and noise at the DDS output are the digital nature of the signal at the output and the quantizing and nonlinearities noise from the DAC.

III. Output Filter Consideration

For every particular application of the DDS these spurious frequencies have to be minimized to a levels that are compatible with the spectrums of the other frequency sources. In almost every DDS this is accomplished from the lowpass filter after the DAC. The frequency response of an ideal lowpass filter would be 1 over the Nyquist band ($0 \leq f \leq F_s/2$) and 0 elsewhere. Such a filter is not physically realizable and this results in the sacrifice of some portion of the available output bandwidth in order to match the non-ideal response of the antialias filter.

The parameters of the output filters are very important for the overall performance of the DDS and the requirements they have to met define the purity of the synthesized signal.

The proper choice of the filter is connected with the reaching of some compromise between the contradictionary requirements for the great steepness of the transitional part of the frequency response of the filter and the flatness of the group delay time characteristic, which is needed for digital communication systems.

For a proper choice of a filter we need to define the different applications of the DDS and we can divide the applications in three types:

Direct digital synthesizers used individually or as a part of hybrid synthesizer where the whole possible frequency band has to be utilized.

DDS, used for a reference oscillators for hybrid DDS-PLL combination, where the only a part of the possible band of DDS is used and no need for wide band digital modulation.

DDS for a relatively narrow frequency band in mixer type hybrid synthesizers where wide band digital modulations have to be implemented.

Every one of this types has a need particularly designed low pass filter. The most serious demands are for the first application where the aim for achieving a wider bandwidth leads to reaches to the Nyquist limit of $f_{\text{clk}}/2$ in which case the first image $f_{\text{clk}} - f_{\text{out}}$ is getting closer to f_{out} and it becomes impossible to separate these two frequencies with the known filter approximations.

Eq. (6) could be used for exact calculation of the amplitude difference in dB for f_{out} and ($f_{\text{clk}} - f_{\text{out}}$). The calculations show that there difference when $f_{\text{out}} = 0.4 f_{\text{clk}}$ reaches approximately 4 dB.

$$\Delta A = 10 \lg \left(\frac{\sin\left(\frac{2\pi f_{\text{out}}}{f_{\text{clk}}}\right)}{\frac{2\pi f_{\text{out}}}{f_{\text{clk}}}} \right) - 10 \lg \left(\frac{\sin\left(\frac{2\pi (f_{\text{clk}} - f_{\text{out}})}{f_{\text{clk}}}\right)}{\frac{2\pi (f_{\text{clk}} - f_{\text{out}})}{f_{\text{clk}}}} \right) \quad (6)$$

The choice of a proper filter depends on its behavior in the frequency and time domains. Usually when the parameters of filters are analyzed the stress is put on frequency domain but there are some special cases where the time domain behavior is more important. The knowledge of the relation between the time and frequency characteristics of the filter is very important for the DDS, because depending on the application's demand one of them could be optimized. It is impossible to reach at the same time good selectivity and good pulse response from one filter. Typical filter parameters are the cut off frequency fc, the frequency where the minimum specified attenuation is reached fs, maximum ripple in the passband ($A_{\max}$) and minimal attenuation in the stopband ($A_{\min}$).

An important parameter of the filter is the group delay time or GDT, which is the first derivation of the phase and is a measure for the delay of different frequencies, when passing through the filter. GDT could be comparatively easily measured and used for criteria for the time domain characteristics of the filter.

There are a lot types of filters known from the theory [3] and practice but the most popular are Butterworth, Chebyshev and Kauer (elliptic) approximations.

On Fig. 2 from [3] a comparison can be made for fifth order filters of different types. It can be seen that only the

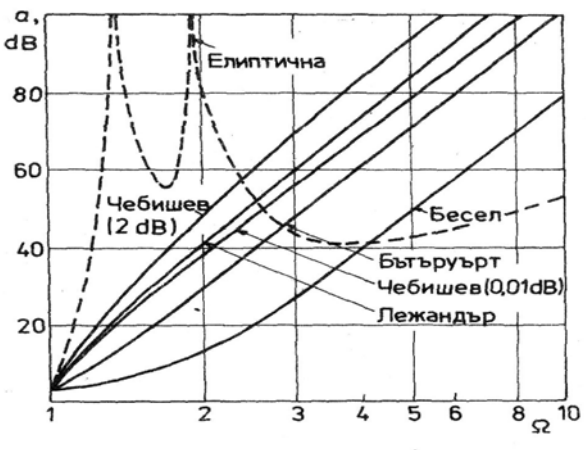


Fig. 2.

elliptic approximation of fifth order is near the needed characteristics for the filter for DDS applications.

IV. Filter Design and Simulations

On the basis of these analysis and consideration a program in Matlab was designed where the input parameters are the $(f_{clk} - f_{out}/f_{out}) - \Omega$ ratio, the ripple in the passband and the minimal attenuation in the stopband. The definition of the limits for all these parameters depends exclusively on the specific application of the synthesizer but for the start a ripple of 1 dB, Ω of 1,2 and attenuation of 60 dB and 80 dB were chosen. Parts of the results are given in Table I (1).

Table 1.

	type	Elliptic	Chebishev 1	Chebishev 2	Butterworth
60dB	order	5	9	9	14
	GDT_{100M}	2÷18	2÷14	2÷14	2÷23
	GDT_{80M}	3	5	5	7
	ripple dB	1	1	0	0
80dB	order	7	13	13	18
	GDT_{100}	2÷40	2÷45	2÷15	2÷30
	GDT_{80}	4	7	6	10
	ripple	1	1	0	0

The results from the simulation confirm the most of the intuitive made assumption and practical knowledge on this matter. As can be seen from the results for the attenuation of 60 dB filters with Chebyshev approximation still can be used with some concern on the complexity of the circuit, but for 80 db attenuation the only reasonable choice is the elliptic filter of seventh and higher order.

From the table could be seen two values for the GDT -one for the cut off frequency of 100MHz and one for 80MHz. One conclusion can be done that even the high order elliptic filter can be used near the cut off frequency with comparatively low level of GDT distortion. Another conclusion is that for equal attenuation and different order, the GDT fluctuations of the elliptic filters are similar those of Chebishev 1 and Butterworth filters.

A seventh order practical filter design was synthesized with the help of program Serenade7.5 for the outputs of the quadrature DDS AD 9854 . The circuit diagram of the filter is shown on Fig. 3.

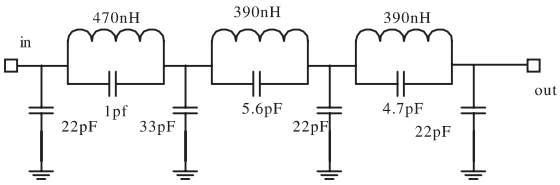


Fig. 3.

When the requirements to the purity of the output spectrum of the signal of DDS are more stringent there is a possibility for further improvement with the use of tuneable low-pass filters.

The analysis of the requirements for the characteristics of the output filters for other two DDS applications shows that good improvement can be achieved with the addition of narrow band bandpass filter. A quartz or SAW filter can filter the parasitics and improve the carrier to noise ratio at the output enough for wireless application.

References

[1] A.V. Oppenhaim, and R.W. Schafer, “Digital Signal Processing”, Prentice Hall Englewood Cliffs, New Jersey, 1975.

[2] Y .C. Jeng, “Digital Spectra of Nonuniformly Sampled Signals – Digital Look-Up Tunable Sinusoidal Oscillators”, *IEEE Trans. On ins. And Meas.*, Vol. 37, No.3, pp 358- 362, Sept. 1988.

[3] G. Stojanov, “Theoretical Foundations of telecommunication technics”, Sofia, Technica, 1993.

The Express Method for Pattern Recognition in Production and Maintenance of Electronic Devices

Georgi D. Nenov[1], Borislav D. Boiadjiev[2]and Borislav P. Avramov[3]

Abstract – **A pattern recognition method with statistical row, which is got with designed experiment, is offered in this paper. The row processing is used for determination of given specimen to one of ranks formed in training.**

Keywords – **Recognition, statistical row designed experiment, training, ranks, electronic devices**

Pattern recognition is used in manufacturing, management, medicine, criminology and other branches of the science. The requirements for recognition reliability are in wide limits and first of all are depends on formulation of the task.

The offered express method might be used in production and maintenance of electronic devices. The training for forming the ranks is made with representative sample from general collection. A statistical row, developed with design on experiment, is used for the recognition. To define proximity of tested specimen to corresponding rank, the members of the row are processed by an algorithm.

The purpose of training is to form the necessary number of specimen ranks. It can make in relation to one or more parameters of the device. In the second case the task is complicated and the decision is made with help of experts, who know the devices of recognized type.

In cases with one parameter, the creation of ranks can make with representative sampling from general collection. Depending on the purpose of the recognition, the number of ranks and extreme values are defined. In the simplest case they are two – with a sharp limit or with a limit strip between ranks.

The sorted by ranks specimens, are tested with full factorial design, fractional factorials or with saturated designs [3]. Parameter's values are averaged for each rank for respective trials. Thus the patterns-samples are forming and they are used in recognition.

The trials are reduced by means of design of experiments, and are definite combinations of measurements of the factors (variables), which are controlled variables. Their influence over the parameter is reported for each trial. For instance controlled variables might be – the values of passive elements of the electronic device, power supplies, external influences and other factors, and parameters are – gain coefficient, actuation threshold, cut-off frequency etc.

The output parameter's values which are got from trials, forms a statistical row which is processed by different ways depending on formulated task.

For recognition of a specimen from given device are necessary analogous trials as these which are made for forming the ranks (training). The statistical row, which is got, and statistical rows of patterns-samples are compared and according to the established proximity, the decision for pertaining of recognized specimen to a respective rank is made.

The proximity might be established by different methods. For the purposes of express analysis the using of rank correlation, created by Spearman is offered. It is not necessary to know distributions of actions and reactions for it application. And when are the small number of statistical rows' members, the estimations are very simple, since is operated with small integers. This convenience is a premise for simplifying of operations and their registration

The assessment of rank correlation is reduced to:

1. The members from both statistical rows are compared for each separate trial.

2. The members from one of the rows (in this case the row of the pattern!) are arranged by ascending or descending order, and the rank for each member is assigned by ascending order.

3. The ranks for the other row are assigned in the same way, but the correspondence of the members from both rows by trials is preserved.

4. The rank correlation coefficient R is defined by:

$$R = 1 - \frac{6 \sum_{u=0}^{n} (x_u - y_u)^2}{n(n^2 - 1)}, \qquad (1)$$

when x and y are the ranks of the two rows for separate trials, n is number of members for each of the rows.

Depending on the R value the following stages of correlation are differed:

a) $R = 0.7 - 0.9$ – high;
b) $R = 0.5 - 0.7$ – noticeable;
c) $R = 0.3 - 0.5$ – moderate;
d) $R = 0.1 - 0.3$ – weak.

For recognition the high stage is preferable. Since the method is statistical, factors' values have accidental kind and

[1]Georgi Dimitrov Nenov is with the Department of Communication and Railway Signalling Engineering in "Todor Kableshkov" Higher School of Transport. 158 Geo Milev Street, Sofia 1574, Bulgaria.

[2]Borislav Draganov Boiadjiev is with the Department of Communication and Railway Signalling Engineering in "Todor Kableshkov" Higher School of Transport. 158 Geo Milev Street, Sofia 1574, Bulgaria. E-mail: bboiadjiev@vtu.bg.

[3]Borislav Petrov Avramov is with the Department of Communication and Railway Signalling Engineering in "Todor Kableshkov" Higher School of Transport. 158 Geo Milev Street, Sofia 1574, Bulgaria. E-mail: b.avramov@vtu.bg.

probability of occurrence in the high stage reduces with decreasing of the number of trials. On the other hand in electronics the measurements are rather precision, these ensure the appropriate linear models [3].

The method and algorithm are illustrated by following example. Amplifier with powerful integrated circuit is the object. Controlled parameters (factors) are the load resistance R_L and the power supply E. The output power P_L is the parameter. In the training stage the full factorial design is used. It has $N = 2^2 = 4$ trials, also the trial in the "central" point is added, and it corresponds to the nominal values of R_L and E. In the other four trials two settings are given for the factors. For the R_L the nominal value is $R_{L0} = 4\ \Omega$ and the two settings are about 10% above and below it, i.e. $\Delta R_L = 0.5\ \Omega$. The standardized value is signed with x_1 and is as follows:

$$x_1 = \frac{R_L - R_{L0}}{\Delta R_L}. \qquad (2)$$

When $R_L = R_{L0} + \Delta R_L - x_1$ is set to 1, and when $R_L = R_{L0} - \Delta R_L - x_1$ is respectively (-1). The power supply E is standardized in the same way and is signed with x_2. The design matrix is given, in table 1.

Table 1.

Trial	Factors		P_L, W	P_L, W	P_L,W	P_L,W
	x_1	x_2	Pattern	Sample 1	Sample 2	Sample 3
0	0	0	6.00	6.10	5.95	4.80
1	1	1	5.40	5.30	5.35	5.20
2	1	-1	5.30	5.40	5.40	5.60
3	-1	1	6.90	6.50	6.85	6.20
4	-1	-1	6.70	6.80	6.75	6.90

The first for P_L column corresponds to the pattern's statistical row. It is created by averaging of the representative sample.

The next three columns contain statistical rows made by measuring of three samples. The recognition is made with two ranks. The first one is with the high correlation between the pattern and verified sample. In the second rank are the samples, for which $R < 0.7$.

Recognizing the first sample (Sample1), the rows of the pattern and Sample1 with the ranks in increasing order are as follows:

P_L, Pattern	6.00	5.40	5.30	6.90	6.70
Rank	3	2	1	5	4

P_L, Sample 1	6.10	5.30	5.40	6.50	6.80
Rank	3	1	2	4	5

The coefficient R is estimated through Eq. (1) with $n = 5$:

$$R = 1 - \frac{6[(1-2)^2 + (2-1)^2 + (3-3)^2 + (4-5)^2 + (5-4)^2]}{5(5^2 - 1)}$$

$$R = 0.8.$$

By the same way R is estimated for the second sample (Sample 2) and is $R = 0.9$. Therefore the both samples belong to the first rank, which consists of a population with the near to the pattern's qualities (characteristics).

But the third sample (Sample 3) is different. The recognition of it shows that:

P_L, Pattern	6.00	5.40	5.30	6.90	6.70
Rank	3	2	1	5	4

P_L, Sample 3	4.80	5.20	5.60	6.20	6.90
Rank	1	2	3	4	5

$$R = 1 - \frac{6[(1-3)^2 + (2-2)^2 + (3-1)^2 + (4-5)^2 + (5-4)^2]}{5(5^2 - 1)}$$

$$R = 0.5.$$

When the respective sample (Sample3) belongs to the second rank, its population consists of disqualified elements or with elements which don't meet the requirements.

Conclusion

The offered method is economical; its procedure is compact and is suitable for automatic monitoring. These and other advantages make it suitable for other applications.

References

[1] Avaliani, G.V. Evristicheskie metody v raspoznavanii obrazov. Tbilisi, Metzniereba, 1988 (in Russian).

[2] Gaskarov, D.V., T.A. Golinkevich, A.V. Mozgalevskiy. Prognozirovanie tehnicheskogo sostoianiya u nadezhnosti radioelectronnoy apparatury. Moskva, Sov. Radio, 1974 (in Russian).

[3] Nenov, G.D. modelirane i optimizirane na radiotehnicheski verigi. Sofia, Tehnika, 1977 (in Bulgarian).

[4] Dean, A., D. Voss. Design and Analysis of Experiments. New York, Springer-Verlag, 1999.

[5] Hahn, G., S. Shapiro. Statistical Models in Engineering. New York, John Wiley, 1994.

Possibilities for Encoding Analaogue TV Signal in Cable TV Networks

Stanimir Sadinov[1], Kiril Koitchev[2]and Stefan Nemtsov[3]

Abstract – **Cable TV causes a lot of interest which brought to searching for methods and ways of limiting the access to individual "hit" programs. This access control can be effected by encoding the analogue TV signal. The problem that demands a solution here is how to satisfy both sides, in other words the system coder-decoder should be effective and with low cost. This, in short, is the aim of the paper: to analyse analogue methods of encoding and compare them with the current conditions of utilization of cable TV networks in Bulgaria. The paper suggests a specific version which takes into account the above requirements and the "hacker" capabilities as well.**

Keywords – **CATV, TV signals, coding**

I. Introduction

At present there are several technologies for encoding analogue TV signals which are distinguished by their genuine technical solutions. In Bulgaria experiments are confined to two major systems ACS-500 and Crypt On which use Syns Supression technologies [1,2]. They feature the following advantages

- low cost of subscriber's decoder (up to $20) allowing for individual(address) encoding of the channels
- There is no deterioration in the quality of encoding
- Good compatibility of the decoder with the modulator of the main station of the cable operators and the mass TV sets.

The disadvantages include unsettled varieties of schemes which are employed by different cable operators and the lack of coordination with the current law concerning licensing and network running.

II. Presentation

A. Basic principles

With analogue methods for encoding the scrambling technology is used by means of which level slip is effected (row synchronizing pulses – RSP) In the decoder the synchro-pulses should be restored or generated.

[1]Stanimir Sadinov is with the Department of Communications Technology and Equipment, Technical University of Gabrovo, str. "Hadji Dimitar" No 4, 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

[2]Kiril Koichev is with the Department of Communications Technology and Equipment, Technical University of Gabrovo, str. "Hadji Dimitar" No 4, 5300 Gabrovo, Bulgaria, E-mail: koichev@tugab.bg

[3]Stefan Nemtsov is with the Department of Communications Technology and Equipment, Technical University of Gabrovo, str. "Hadji Dimitar" No 4, 5300 Gabrovo, Bulgaria, E-mail: stefan@tugab.bg

Fig. 1 shows oscillograms of the initial signal (at the input of the coder), the encoded signal (at the output of the coder), the decoded signal(at the output of the decoder) and oscillograms of the signal with the carrying frequency which is modulated by the corresponding video signals.

Encoding and decoding can be done at two points of the tract: lower frequency (in the video signal) or higher frequency (in radio signal)

The case with lower frequency (LF):

encoding: video signal is imposed with a sequence of right angled pulses which coincide in time with RSP. As a result RSP slips in level and I the pulses go to the "grey" level;

decoding: the executing circuit of the decoder bypasses (shunts) the encoded signal thus inserting a pulse of zero level in the video signal.

With higher frequency (HF):

encoding: for a while the executing circuit of the coder decreases to the "grey" level;

decoding: during synchro-pulses (RSP) the executing circuit id the decoder makes a jump increase of the coefficient of transmission in such a way that the level of the radio signal may correspond to the maximum level.

High frequency encryption ensures the correct operation of the device for automatic gain control of the input signal (AGC), and recovery of the constant component of the signal (RCC) in the modulator. For this reason, the encryption device must be connected to the modulator using the medium frequency (specification D-38.9 MHz), i.e. after the modulator. Unfortunately, many modulators accommodate this type of connection. This is due to the fact that the decoder needs to have its own de-modulator and all related support components – a device for tuning, channel memory, remote control, etc. As a result these decoders are expensive (around $100). It is possible to make use of the TV de-modulator. However, it is then necessary to connect the decoder between its audio-channel and block of coloured. Too few TV sets have this capability, including those that have a video in/out. Typically, if there a signal detected at the video input, the processor turns off the audio signal.

The schematics discussed earlier are organized according to the "LF encryption – HF decryption" principle, what needs to be in the front is shifted to the rear and vice versa. Encryption is performed at the low frequency and decryption at

the audio frequency. The channel decoder is connected between the source of the video signal and the modulator input. It alters the level of the regular synchronization signal (ACS-500), or altogether removes the synchronization signal and inserting the constant “grey” level (Crypt On). In addition, the control signals for subscriber decoders are inserted in the last rows of every image, at frequency of about 3MHz. This signal, along with the video signal is fed into the modulator of the CRT (Cathode Ray Tube) of the TV set at 20-30 V. This enables the subscriber decoder, placed next to the TV set, to pick up the control signal using a special antenna. Therefore, the decoder need not be connected to the internal schematic of the TV set. The decoder is connected between the subscriber’s connection to the cable network and the TV set’s antenna input. The decoder’s antenna is placed in the rear of the TV set along with the board of the CRT. When the TV set is switched to any channel, the decoder accepts only the inputs of the encryption device for that channel. This the decoder can restore the signal if the signal ”permitted” is received (the subscriber has paid for the program), or to pass through the encrypted signal if the signal “forbidden” is received, or in the absence of the signal “permitted”. The design of the decoder modulates the audio signal using the control signals (Fig. 1), which is equivalent to shifting the level of the video signal. Thus, RSP are “inserted” directly in the audio frequency signal.

Fig. 1 shows the audio frequency signal of a given television channel. In fact, signals from several channels are present at the input of the decoder and the decoder has no frequency selection ability, which is why RSP are inserted simultaneously in the signal of all channels. However, the output signals at the base station are not synchronized and therefore, the encryption devices are also not synchronized. As a result the decoder recovers only the signal on the channel to which the TV set is tuned. On the other channels, RSP are inserted not at their appropriate place, but in a random position. The result is that the other encrypted signals at the output of the decoder are “even more encrypted” – in addition to their “own” RSP, they now also contain “foreign” RSP. Furthermore, if the TV set is tuned to an encrypted channel, then all unencrypted channels at its input (the decoder output) will also be “encrypted”. It is therefore not possible to use the same decoder with several TV sets simultaneously.

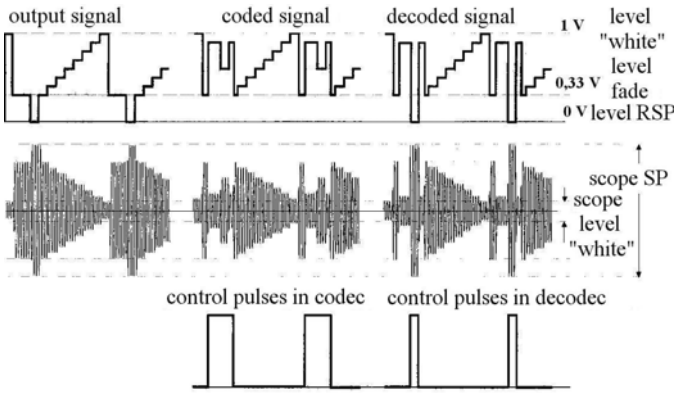


Fig. 1. Oscillogram of codec and decodec input/output TV signals

B. Possible problems in the transmission and reception area

This type of “reverse” system has two large advantages: the encryption device can be connected to every modulator and the decoder to any TV set. However, there are shortcomings. At the transmission end some parts of the modulator may interact with the encryption device signal in unintended ways:

— The automatic gain control of the input signal (AGC). It is designed to maintain a constant level of the video signal at the input of its own modulator, despite any random fluctuations in the input signal. However, the level in the active area of the horizontal rows may change from a low “fade out” level to a high “white” level depending on the contents of the image. This is why AGC regulates the overall level, taking into account the range of RSP. However, in the encrypted signal RSP exist only in the frame fade pulse (FFP) and are not present in the active horizontal rows. If the time constant of AGC is small enough, in the active area of an image AGC will register the RSP level as too low and will amplify the image. Therefore, if the modulator has an AGC device at the input, it must be disabled.

— Recovery of the constant component of the signal (RCC). The signal between the source and the modulator (i.e. modulator as a functional element and not an actual device), passes through a number of distribution capacitors, which, along with the input resistors connected to them constitute discriminating units. If the time constant of these units is small, the constant component of the video signal is lost. As a result, the video signal with a range of 0 to 1 V becomes a signal with a range of -0.5 V to 0.5 V. At the same time the level of the synchronizing signal, which is constant in the initial signal, will change from pulse to pulse depending on the contents of every row. Figure 2 shows a ”black and white horizontal bands” signal that has passed through the discriminating units (extreme distortion is apparent). The carrier frequency of such a signal must not be modulated. Instead, the constant signal component must be restored – the low levels of the synchronizing pulses must be “placed” on a horizontal line at 0V. Because of this, it is necessary to use RCC schematic, during RSP the signal is connected to “ground”. Since in the encrypted signal RSP are not present, the RCC schematic needs to be modified.

These problems are only relevant in rare special cases.

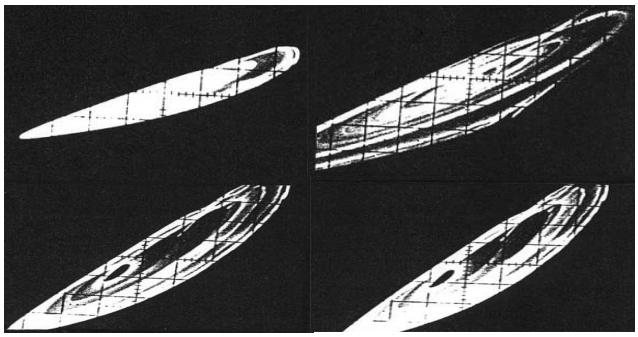


Fig. 2. Video signal at the input and output of the discriminating unit

Typically, the modulation system is implemented with inexpensive modulators, where RCC and AGC are simply not present. The modulator could function without AGC since the input resistance of the steps, when build with modern, components is high. In addition using the appropriate values for the discriminating capacitors can eliminate the need for RCC.

— The decoder receives control signals from the CRT modulator using its antenna. However, during FFP the modulators are turned off – this is necessary, or else all control signals contained in FFP (teletext, signals for test rows, image and color synchronization signals SECAM, etc.) would be visible during the reverse phase of the vertical signal. Therefore, the control data for the decoders are transmitted not through FFP, but in the active rows, the last few rows in every image. As a result, during certain TV setups (vertical sizing) the control data are visible on the screen – an array of apparently random black and white bands at the bottom of the screen.

— In some cases, the above-mentioned method is inappropriate. It is then necessary to hook up a connection at the port for video output of the TV set. If such a port is not available, the hook up is directly with the internal circuits – this method is neither convenient nor safe.

C. Consistent and reliable protection from "unauthorized" access

Despite the inexpensive implementation the solution discussed is pirate proof and can be used in networks where the access tariffs are not too high.

When designing encryption systems one must take into account its cost, and in addition, the demand for this type of technology and its potential for enhancements.

Fig. 3 shows a schematic of a simple pirate decoder suitable for the early types of encryption, where the row synchronizing pulses (RSP) are mixed at level [3]. The graphs in Fig. 1 indicate that in order to recover the signal, it is sufficient to increase the transmission coefficient in the decoder during the reverse phase of the horizontal signal. This is accomplished through the use of a variable attenuator (R4) and resistance of diode VD1. Through the sequence R1, C1, R2 the reverse phase of the signal in L1, is applied at the base of transistor VT1. L1 essentially consists of several coils placed directly over the coils of the output transformer for row scanning.

The polarity of L1 must be such that, during the reverse phase of the signal, the voltage at the base of VT1 must be positive. During the regular phase VT1 does not let the signal through, the collector voltage is the same as the emitter voltage and the current flows through R3, VR1 and VD1. The resistance of VD1 and therefore the rate of transmission between RF-in and RF-out is also small.

The reverse phase of the signal turns VT1 on, and the voltage at the collector is close to zero. The diode VD1 does not let current through and as its resistance rises, so does the transmission coefficient of the decoder. In practice the decoder does not distinguish between modified RSP (active

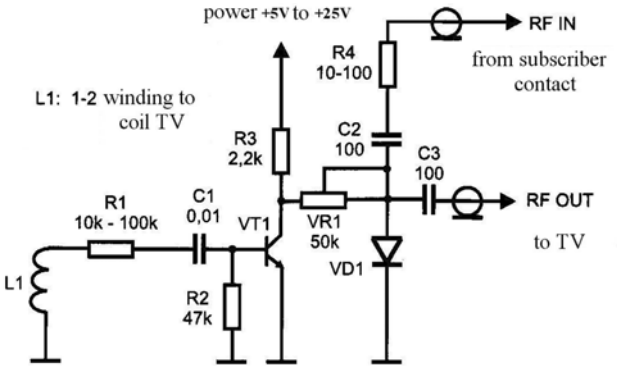


Fig. 3. Schematic of an illegal ("pirate") decoder

field) and RSP with image fade-out interval, which is constant. In addition, it is very difficult and dangerous to connect L1 to the TV set. In conclusion Fig. 3 represents a rather simple illegal decoder.

The "Crypt On" encryption devices separate RSP from the signal and instead replace them with a signal with a constant level. The recovery of such a signal is rather complex as it is necessary to generate RSP all over again. In order to recover RSP it is possible to use a phase locked loop generator (PLL). The generator is synchronized using the RSP during the image fade-out interval, which are not removed, or modified.

Similar technology is used to recover the color carriers in PAL decoders. They take up 3.5% of the length of a row, which is sufficient to synchronize the signal phase and frequency of the PLL and the junction (support) generator. Then, the entire active segment of the support generator row signal is used by two synchronous phase detectors for color discrimination signals. IFS take up 12.5% of the field, and as a result the phase and frequency synchronization using PLL in the active segment of the field is more stable then when using a PAL decoder.

Various websites list detailed information on illegal decoder schematics for this type of encryption. Therefore, this method can be enhanced to provide a higher degree of protection by adding an image synchronizer to each decoder. This will allow small changes in the length of the row. The image

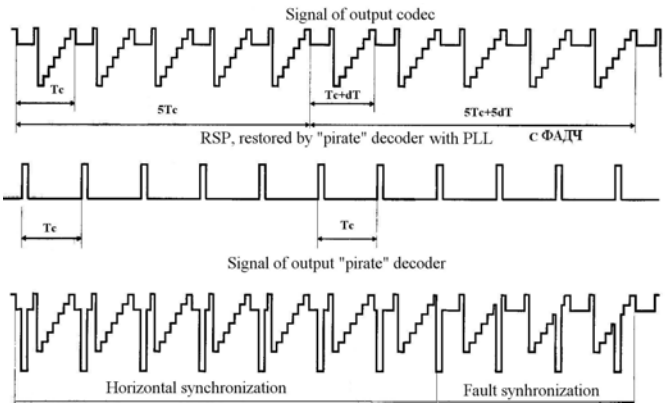


Fig. 4. Enhancement to the encryption system

synchronizer is controlled by the decoder. As a result the encrypted signal is not only separated from RSP, but the row length may be change by small amounts. The change is done by a pseudo-random process generated by the encryption device. For example, beginning with the i-th row, the length N of the row is increased by dT1 and beginning with $i + N$ row the length of the row M is reduced by dT2, etc. The change is small enough so that geometric changes are visible on the screen.

Using the variable row length method of encryption above, it is difficult to decode the signals through RSP recovery and PLL. This is evident in Fig. 4. In section 5T the row length is constant and all decoders whether legal or illegal operate normally. In the area from 5Tc+5dT the length of the row varies by dT. The legal decoder, will be instructed by the encryption device to change the frequency of the recovered RSP. The PLL system dynamically changes the frequency of the RSP so that they are in-synch with the decoder and the image is synchronized. The illegal decoder will continue to generate RSP with phase Tc and thus the RSP will be "mixed" and as a result the horizontal synchronization will be disrupted.

III. Conclusion

The reviewed system for encrypting analog TV signal has several advantages including inexpensive implementation and satisfactory operation, which would make illegal access difficult.

At the moment, however, many cable operators are faced with problems related to the changes in the telecommunications and licensing legislation. In addition, the creditworthiness of many subscribers is poor, which slows down or even obviates the need for encryption at this stage. Hopefully, in the future both the legal and economic environment will improve, along with new and ingenious solutions to the encryption-decryption problem. One such solution, which can be used to implement a stable and inexpensive encryption system, was presented here.

References

[1] Visockii G. Particularity system coding ACS-500 and Cripton, "Telesputnik Media", St. Petersburg, 9-2000.

[2] Visockii G. System Pay per View, "Telesputnik Media", St. Petersburg, 8-2000.

[3] Visockii G. System Restrictive access in CATV, "Telesputnik Media", St. Petersburg, 7-2000.

Investigation of the Impact of the Supply Voltage Form in Cable TV Amplifiers

Stanimir Sadinov[1], Kiril Koitchev[2]and Stefan Nemtsov[3]

Abstract – **The supply block of highway cable TV amplifiers is a source of electromagnetic disturbances which spread throughout the entire cable TV network. It is proved that the most expedient approach is to use converters in power blocks, commutation in them is effected by zero current. Said converters work with trapezium contour of current and voltage. In this case the power at the moment of switching on of commutation transistor is close to zero while at switching off losses are inconsiderable and are determined by the commutation capacity. The trapezium form of both voltage and current brings about to decrease in electromagnetic disturbances due to supply.**

Keywords – **CATV, supply voltage**

I. Introduction

Cable TV and other areas of telecommunications industry entered a stage of introducing new digital infrastructures. The latter are known as "data highways" and could offer a set of multimedia, communications, information and amusement services. Power supply of these systems is an integral part of such infrastructures. It is also necessary to ensure effective and reliable solution of feed from one end to the other.

The form of voltage, its capacity and frequency of net supply is still under development. In other to ensure the necessary power to the units included in the net with a certain level of voltage and lower level of electromagnetic disturbances, it is recommendable to feed voltage with trapezium form, which differs from sinusoidal. Normally a supply voltage of 60-90 V and frequency of 50 (60) Hz is used as it guarantees a good blend of safety, increased load capacity, corrosion, level of electromagnetic disturbances and transmission losses [1-3] without any comment of these properties. In [4] are shown some of the indisputable advantages in terms of minimizing the "scissors" effect of current and the high multiplier of power.

Fig. 1 presents the form of supply voltage with its basic components. Fig 2 shows a simplified chart of the resonance converter of the supply unit employed in the popular type of highway amplifiers which are manufactured in the west. It is evident, however, that with trapezium form of supply voltage the so-called "mild commutation" of powerful transistors

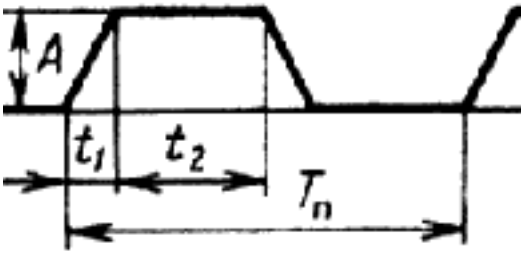


Fig. 1. Supply voltage waveform

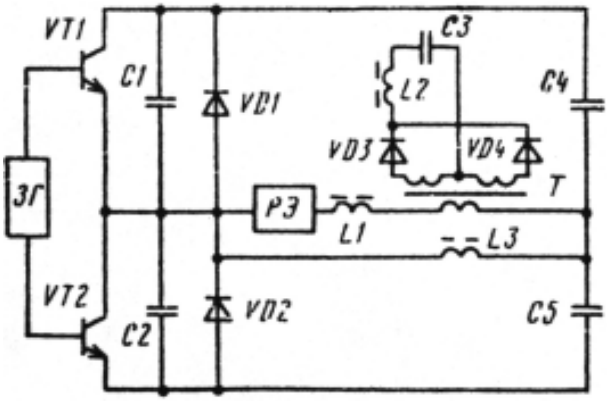


Fig. 2. Resonance converter of the supply unit

will be ensured. The time chart of switching of these transistors in this circuit is shown in Fig. 3 the rate of increment of current has an unchangeable time constant which is determined by the inductance L1 of the primary side. It provides a comparatively quick increment of current up to its functional value. Likewise inductance L2 is used in the same way, but included in the secondary side.

The presence of commutation condensers C1 and C2 as well as inductance's L1 and L3 allows the change of switch-off (cut out) voltage for the power transistors VT1 and VT2, namely, the voltage of the the transistor opened earlier goes up to the level of supply voltage whereas the voltage of the transistor that was closed earlier goes down to zero (Fig. 3) That state is kept in the course of a definite time period until the dissipation of the energy accumulated in the coils of the chokes L1 and L3. Opening of transistor whose collector's voltage reaches zero does not cause commutation losses. It is evident that in this case power losses at the time of starting the transistors are close to zero and at the time of cut-

[1]Kiril Koichev is with the Department of Communications Technology and Equipment, Technical University of Gabrovo, str. "Hadji Dimitar" No 4, 5300 Gabrovo, Bulgaria, E-mail: koichev@tugab.bg

[2]Stanimir Sadinov is with the Department of Communications Technology and Equipment, Technical University of Gabrovo, str. "Hadji Dimitar" No 4, 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

[3]Stefan Nemtsov is with the Department of Communications Technology and Equipment, Technical University of Gabrovo, str. "Hadji Dimitar" No 4, 5300 Gabrovo, Bulgaria, E-mail: stefan@tugab.bg

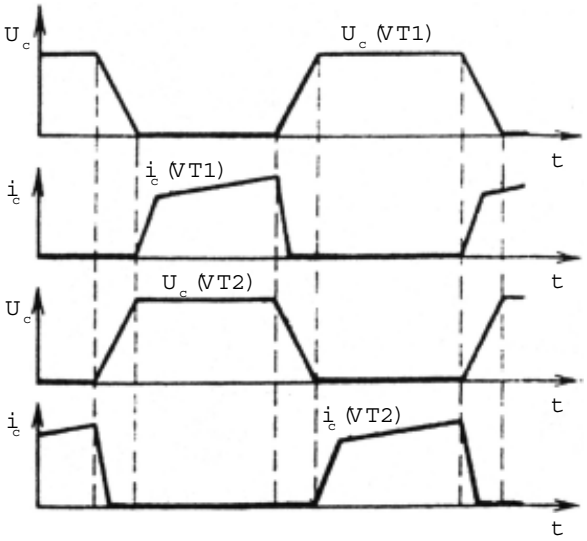


Fig. 3. Time chart of switching of the transistors

out these losses are inconsiderable and are determined by the commutating condensers. Because of the trapezium form of the supply voltage of such type of rectifier unit it is possible to switch off the short circuit mode at the power transformer output in switching the diodes of the output rectifier. Such form of supply allows to considerably reducing electromagnetic disturbances created by the rectifier unit of highway amplifiers that are used in CATV.

Evaluation of noise resistance at the output of the supply unit of the pulse transducer in cable amplifiers can be carried out in the following way. The range of disturbances in rated voltage of the pulse transducer is determined by the expression:

$$U_{\text{out}_i}(S) = K_i(S)U_{\text{in}}(S) \tag{1}$$

where $K_i(S)$ is the transmission function of the converter; $U_{\text{in}}(S)$ is the function of disturbances in the network.

Expression (1) corresponds to that case when at the moment $U_{\text{out}_i}(S)$ all transition processes in the transducer are complete. If they are still going on then the range of disturbances in rated voltage is to be determined by

$$U_{\text{out}}(S) = \sum_{i=1}^{n} U_{\text{out}_i}(S)e^{-t_i s} \tag{2}$$

where t_i corresponds to backward holding of $U_{\text{in}_i}(S)$ in relation to the beginning of reading. Spectrum characteristics of rated voltage in network disturbances impact will be determined by the expression (1):

$$20\lg|U_{\text{out}_i}(j\omega)| = 20\lg|K_i(j\omega)| + 20\lg|U_{\text{in}_i}(j\omega)| \tag{3}$$

where $K_i(j\omega)$ is the transmission function of the converter; $U_{\text{in}_i}(j\omega)$ is the function of network disturbances. Having in mind (2) we can express it:

$$20\lg|U_{\text{out}}(j\omega)| = 20\lg|\sum_{i=1}^{n} U_{\text{out}_i}(j\omega)e^{-t_i j\omega}|. \tag{4}$$

By using expressions (3) and (4) it is possible to chart the spectrum characteristics of network disturbances of rated voltage of pulse voltage converters.

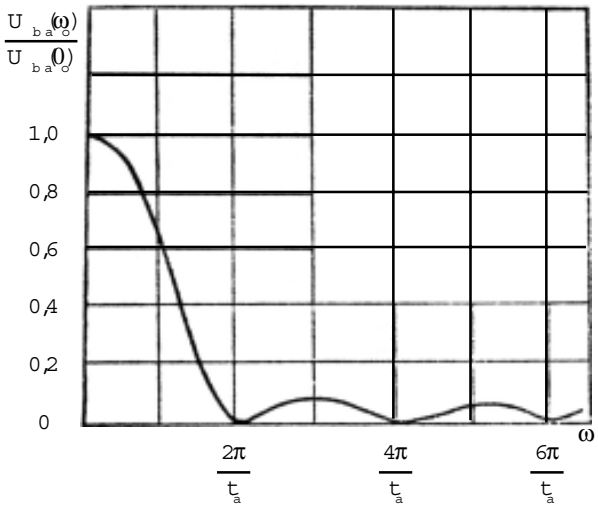


Fig. 4. Spectrum density module

For better evaluation of the parameters of the voltage pulse or the transistor current it is good to use the generalized charts of spectrum characteristics. If we know the particular pulse parameters we can obtain equivalent spectrum density of Niquist bandwidth [5]:

$$U_{\text{out}}(\omega) = U_{\text{out}}(j\omega) = \frac{8A}{t_i\omega^2}\sin\frac{\omega t_i}{8}\sin\frac{3\omega t_i}{8}. \tag{5}$$

For trapezium shaped pulses (Fig. 1) there is a graphic display of the spectrum density module in Fig. 4. This module could be conditionally divided into three areas (low, medium and high frequencies) with two break frequencies f_1 and f_2 and the corresponding amplitudes A_1 and A_2 (see Fig. 5) as follows:

$$\begin{aligned} u_{\text{lf}} &= 126 + 20\lg A(t_1 + t_2) \\ u_{\text{mf}} &= 116 + 20\lg A - 20\lg f \\ u_{\text{hf}} &= 106 + 20\lg(A/t_1) - 40\lg f \end{aligned} \tag{6}$$

Table 1 contains analytical expressions for frequency characteristics obtained from the generalized charts and also the amplitudes and frequencies of transmission with the various parameters of trapezium voltage.

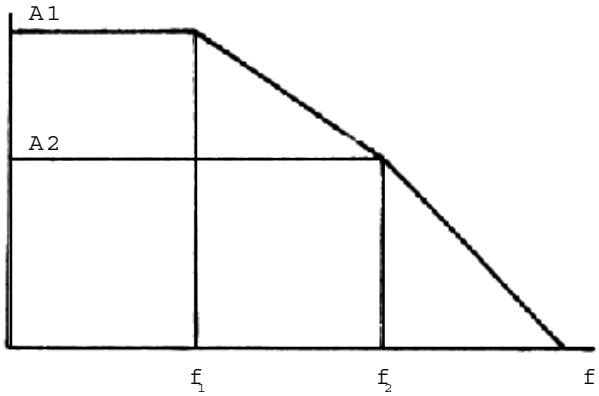


Fig. 5. Spectrum density module three areas

Table 1.

Duration, μs		Amplitude A_1, A_2 $\left[\mathrm{dB}/\frac{\mu\mathrm{V}}{\mathrm{MHz}}\right]$ and refraction frequencies f_1, f_2 [kHz] at amplitude A [V]					
t_1	t_2	90		60		30	
		A_1/f_1	A_2/f_2	A_1/f_1	A_2/f_2	A_1/f_1	A_2/f_2
0.1	40	213.6/7.9	161.5/3200	207.6/7.9	155.4/3200	186.7/7.9	134.6/3200
1		213.8/7.7	181.4/320	207.8/7.8	175.4/320	186.9/7.8	154.6/320
5		214.6/7.08	196/60	208.6/7.1	190/60	187.6/7.1	168.4/60
0.1	18	206.7/17.6	161.5/3200	200.7/17.6	155.4/3200	179.8/17.6	134.6/3200
1.0		207.1/16.8	181.5/320	201.1/16.8	175.5/320	180.2/16.8	154.6/320
0.1	9	200.7/35	161.5/3200	194.7/35	155.4/3200	173.8/35	134.6/3200

II. Conclusion

On the grounds of the results obtained the following conclusions can be drawn:

1. Voltage and current passing through the transistors of the pulse transducers of the power supply unit of the highway cable amplifiers appear to be a major source of electromagnetic disturbances. The level of electromagnetic disturbances depends on the amplitude and shape of the voltage pulses and the current passing through power transistors. The alternating transistor current creates a prevailing magnetic field and the changing potential creates electric field.

2. With the increase of the duration of the fronts of the trapezium shaped supply pulses the frequency range of the spectrum characteristics move toward the side of low frequencies, for instance if the duration of the front is changed from 0.1 to 5 μs then the frequency of refraction f_2 will shift from 3200 to 60 kHz; in other words its change will be approximately 50 times.

3. Increase in duration of supply voltage front causes a decrease of the disturbances' amplitude for high frequencies and also causes a minor increase in lower frequencies. At the same time, however, the increase in the duration of the pulses's front will cause a corresponding increase in the commutating transistors, hence the lower efficiency coefficient

4. With the increase of frequency translation, i.e. with the decrease in the duration of voltage pulses, the amplitude of disturbances is slightly decreased for lower frequencies. For example, with a change of pulse duration from 40 to 9 μs the amplitude of disturbances goes down from 213.6 to 200.7 dB/μV/MHz, but remains unchanged for medium and high frequencies.

5. The increase of the front of the trapezium shaped supply voltage will cause a decrease in the intensity of disturbance creation. If this is accompanied by a solution of the problem for reducing the power losses then the technical parameters of supply units will be boosted considerably.

References

[1] Walter Ciciora, James Farmer, David Large, "Modern Cable Television Technology", Morgan Kaufmann Publishers Inc., 1999.

[2] Mei Qiu, Praveen K. Jain, "Modeling and Performance of aPower distribution System for Hybrid Fiber/Coax Networks", IEEE Transactions on Power Electronics, vol. 14, #2, March 1999.

[3] Kaiser F., T. Osterman, "Broadband network powering issues" in Proc. 18th Int. Telecommunications Energy Conf., October 1996.

[4] "General requirements for powering optical network units infiber-in-the-loop systems", Bellcore Tech. Advisory, TA-NWT-1500, No 1, December 1993.

[5] Thottuvelit V. and R. Kakale, "Analysis and design of broadband power systems" in Proc. 18th Int. Telecommunications Energy Conf., October 1996.

[6] Bichev G. "Multichannel Digital Communication Systems", Sofia, Tehnika, 1980.

Chaotic Signals Generated by Some Circuits – Comparative Study (Correlation Analysis)

Stoitscho V. Manev[1]and Vladimir Iv. Georgiev[2]

Abstract – **In this paper correlation analysis of chaotic signals, generated by some circuits, has been made. These circuits have been developed in practice. The signals, obtained at the outputs of the circuits, have been digitized and analyzed.**

Keywords – **chaotic circuits, chaotic attractors**

I. Introduction

The study of the autocorrelation functions of the chaotic signals is of great interest. This is the first step toward the qualitative and quantitative estimation of the chaotic signals. The determination of the Ljapunov exponents, using an algorithm, based on the algorithm of Eckmann, Kamphorst, Ruelle and Ciliberto (EKRC algorithm), requires the precise estimation of some parameters, concerning the autocorrelation analysis. In the proposed paper some practical developed circuits have been used, in terms of obtaining chaotic signals. Graphics of the autocorrelation functions of the discussed signals have been presented.

II. Classification

In the presented work signals, obtained from the following circuits, have been discussed [3-4]:

1) H-generator.

2) 4-D generator

3) Circuits, based on the canonical realization of Chua's circuit.

These circuits have been developed in practice. The signals, obtained at the outputs of the circuits, have been digitized by means of oscilloscope interface and after that the obtained results have been used as basis for further investigations. One of the main goals of the discussed paper is, to show the relationship between the structure and the parameter values of each of the analyzed variants and the form of the obtained signals. The analysis has been made in time domain.

The precise determination of the Ljapunov's exponents is from great importance for the understanding the dynamical behaviour of the analyzed circuits. The necessary condition for this aim requires analysis, based on the autocorrelation functions (ACF) of the obtained signals.

[1]Stoitscho Velizarov Manev, Dept. of Radiotechnique in Faculty of Communication and Communication Technologies in TU-Sofia, Bulgaria

[2]Vladimir Ivanov Georgiev, Dept.of Theoretical Electrotechnic in TU-Sofia, Bulgaria

III. Experimental Results

The circuits, discussed in [3-4], have been practical developed. Using the lab prototypes, a number of experiments have been carried out.

Most realistic notion about the nature of the generated chaotic signals can be obtained by inspection of the photos from the trajectory pictures, observed on oscilloscope in the phase plane.There are not any restrictions, resulting from the analog - digital transforms.

A. H-generator [3]

In [3] a LC-generator from H-type, designed to produce chaotic signals, has been discussed. For different values of the circuit parameters a variety of chaotic signals has been obtained. The signals, observed in the phase plane, have been presented on Fig. 1 [3].

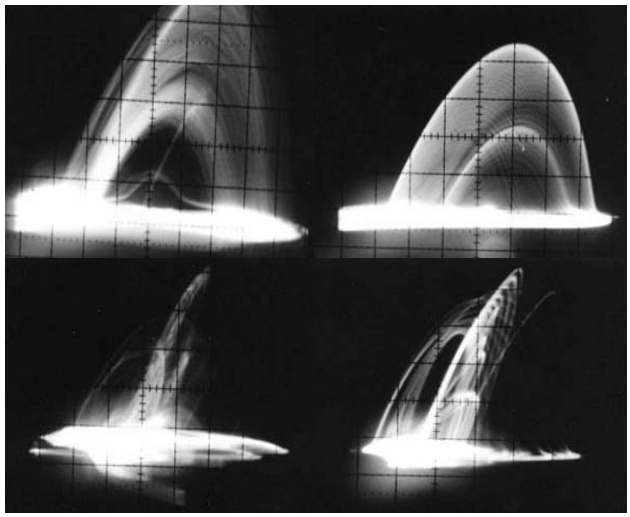


Fig. 1. Photos of output signals, observed in the phase plane[3]

For a fixed set of values of the parameters a chaotic signal has been produced. This signal has been digitized. The

Fig. 2. Time domain presentation

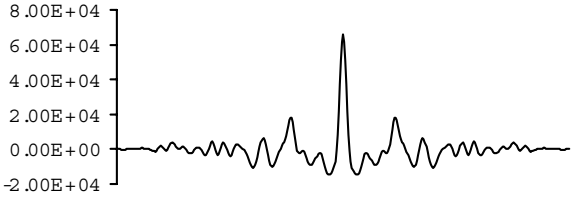


Fig. 3. ACF of the obtained chaotic signal

obtained digital values have been used as basis for further investigations. The time domain presentation and the auto-correlation function (ACF) of the signal have been shown respectively on Figs. 2 and 3.

The value of τ for the discussed signal has been obtained after computations, based on the ACF. The computed value for τ is $\tau = 6$.

Another set of parameter values of the analyzed circuit lead, as expected, to different experimental results (Figs. 4 and 5).

Fig. 4. Time domain presentation

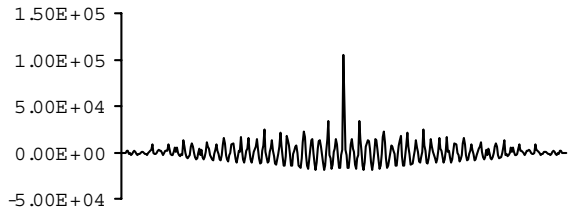


Fig. 5. ACF of the obtained chaotic signal

After computations, based on the ACF, the value of τ for the discussed signal has been obtained: $\tau = 2$.

Another set of values of the parameters of the analyzed circuit causes a corresponding change in the form of the obtained signals, as indicated on Figs. 6 and 7.

Fig. 6. Time domain presentation

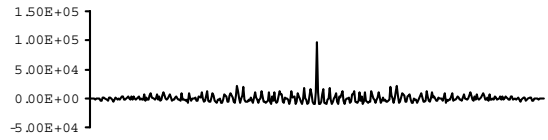


Fig. 7. ACF of the obtained chaotic signal

The value for τ obtained for the discussed signal after some computations is equal to 2.

The presence of close relationship between the set of the values of the parameters and the form of the chaotic signal, analyzed at the output of the circuit, can be shown.

B. Four-dimensional chaotic generator [3]

In [3] a 4-D chaotic generator with modified external driven nonlinearity has been presented. The trajectories in the phase plane have been presented on Fig. 8 [3].

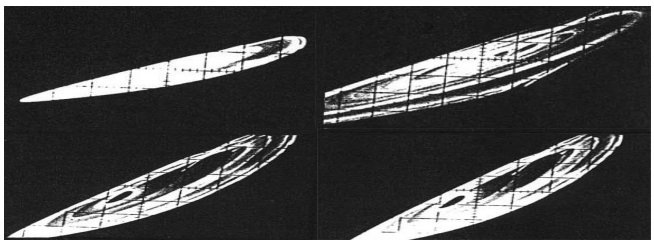


Fig. 8. Photos of output signals observed in the phase plane

By a fixed set of values of the parameters the following chaotic signal has been obtained (Figs. 9, 10).

Fig. 9. Time domain presentation

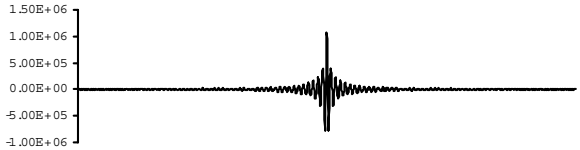


Fig. 10. ACF of the obtained chaotic signal

After some computations $\tau = 4$ has been obtained.

Fig. 11. Time domain presentation

Fig. 12. ACF of the obtained chaotic signal

By different conditions, concerning the set of values of the parameters of the 4-D generator, discussed in [3], a different signal has been obtained (Figs. 11, 12).

The computed value for τ is: $\tau = 9$.

A first version of Chua's circuit. In [4] a version, based on the well known Chua's circuit, has been discussed. The trajectories of different output signals have been presented in the phase plane on Fig. 13 [4].

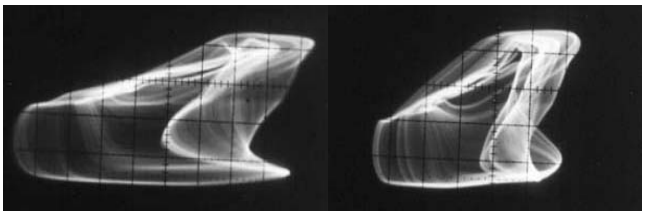


Fig. 13. Photos of output signals observed in the phase plane

For the experiments an appropriate signal has been chosen. The signal presented in time domain has been shown on Fig. 14. The corresponding ACF has been presented on Fig. 15. The value for τ of the discussed signal obtained after computations is: $\tau = 2$.

Fig. 14. Time domain presentation

Fig. 15. ACF of the obtained chaotic signal

A second version of Chua's circuit. In [4] another version, based on the Chua's circuit, has been presented. The circuit, discussed there, includes external driven nonlinearity. Different output signals, observed in the phase plane, have been shown on Fig. 16 [4].

For the calculations an appropriate signal has been chosen. The generated signal has been presented in time domain on Fig. 17.

Fig. 16. Photos of output signals observed in the phase plane

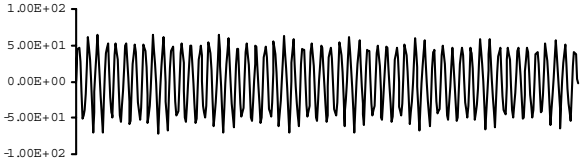


Fig. 17. Time domain presentation

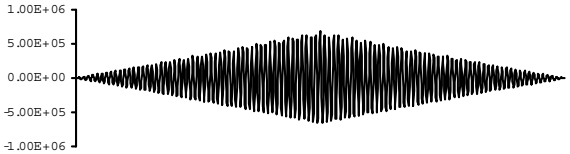


Fig. 18. ACF of the obtained chaotic signal

The corresponding ACF has been shown on Fig. 18. The value for τ, obtained after computations, is: $\tau = 3$.

A third version of Chua's circuit. In an another circuit, presented in [4], a different way of design of the nonlinear element has been chosen.

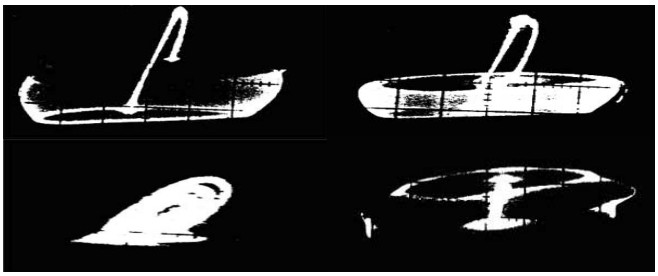


Fig. 19. Photos of output signals observed in the phase plane

The trajectories in the phase plane have been shown on Fig. 19 [4]. A several types of signals, obtained by different sets of values of the parameters of the circuit, analyzed in [4], have been presented. For a fixed set of values of the parameters a chaotic signal has been produced (Figs. 20, 21).

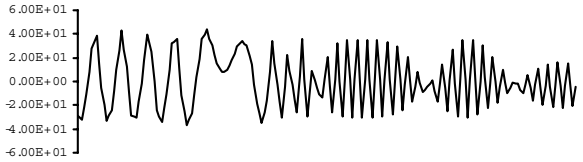


Fig. 20. Time domain presentation

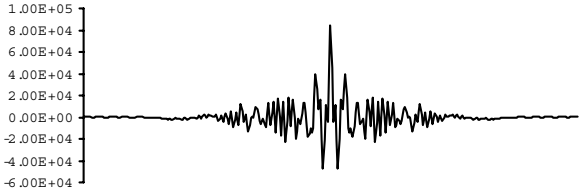


Fig. 21. ACF of the obtained chaotic signal

The value for τ, obtained after computations, is: $\tau = 2$.

Changing the set of values of the parameters, another signal at the output of the discussed circuit, presented in [4], has been obtained (Figs. 22, 23).

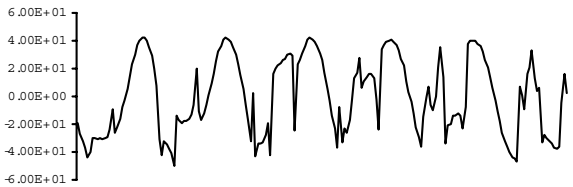


Fig. 22. Time domain presentation

Fig. 23. ACF of the obtained chaotic signal

The value for τ, obtained after computations, is: $\tau = 8$.

The next change in the parameter values leads to a corresponding change in the form of the generated signal (Fig. 24).

Fig. 24. Time domain presentation

The corresponding ACF has been presented on Fig. 25.

Fig. 25. ACF of the obtained chaotic signal

For the discussed signal, by means of computations, based on the ACF, $\tau = 2$ has been obtained.

IV. Conclusions

- Some circuits, designed to generate chaotic signals, have been investigated.
- The discussed circuits have been developed in practice.
- Experiments on the lab prototypes have been made.
- The dependence between the change in the sets of the values of their parameters and the corresponding change in the form of the obtained chaotic signals has been shown.
- Signals, obtained from the discussed circuits, have been digitized and presented in the time domain.
- The corresponding ACF have been shown.
- Based on the obtained ACF, computations, necessary for the determination of the Lyapunov exponents, have been carried out.

References

[1] Stoycho Panchev, “Theory of chaos” (with Examples and Applications), Academic Publishing House “Prof. Marin Drinov”, 1996 (in Bulgarian).

[2] Antonis Karantonis, M. Pagistas, “Comparative study for the calculation of the Lyapunov spectrum from nonlinear experimental signals”, pp. 5428-5444, Phys. Rev. E, Vol. 53, 1996.

[3] Stoitscho V. Manev and Vladimir Iv. Georgiev, “Chaotic Signals in Radiocommunication Generated by some Circuits”, pp. 627-630, ICEST Niš 2002.

[4] Stoitscho V. Manev and Vladimir Iv. Georgiev, “Study of some Circuits in Radio Communication Designed to Generate Chaotic Signals”, pp. 631-634, ICEST Niš 2002.

Chaotic Signals Generated by Some Circuits – Comparative Study (Ljapunov Exponents)

Stoitscho V. Manev[1]and Vladimir Iv. Georgiev[2]

Abstract – **In this paper an algorithm, based on the algorithm of Eckmann, Kamphorst, Ruelle and Ciliberto (EKRC algorithm) for determination of the Ljapunov exponents of chaotic signals, obtained from some circuits, has been applied. These circuits have been developed in practice. The signals, obtained at the outputs of the circuits, have been digitized and analyzed. After the determination of the Ljapunov exponents, conclusions have been made.**

Keywords – **chaotic attractors, Ljapunov exponents**

I. Introduction

It is well known, that the estimation of the Ljapunov exponents is from great importance for the examination of the chaotic behaviour. In the presented paper an algorithm, based on the one of the most popular algorithms, used for determination of the Ljapunov exponents, has been applied. From the obtained results conclusions for the application of the analyzed circuits can be made.

II. Kolmogorov's Entropy and It's Relationship to the Ljapunov's Exponents[1,2]

The entropy is a basic physical parameter. The entropy gives the rate of randomness in an observed system. The Kolmogorov's entropy is one of the most important quantitative characteristics of chaotic motion in phase space with arbitrary high dimension. The Kolmogorov's entropy (K) shows, the rate of chaotic behaviour of the observed system [2,1].

In work [1] H.G.Schuster presents the basic relations, concerning the Kolmogorov's entropy and it's relationship with the Ljapunov's exponents. The following conclusions, concerning the Kolmogorov's entropy, have been presented in [1]:

- The Kolmogorov's entropy determines in time domain the rate of the loss of information about the state of the analyzed dynamical system.
- By one-dimensional maps the Kolmogorov's entropy is directly connected to the Ljapunov's exponent.
- For high-dimensional systems the Kolmogorov's entropy is a measure for the fragments deformation in the phase space.
- The Kolmogorov's entropy is inversely proportional to the time interval, where a prediction for the state of the analyzed system can be made.
- The lower bound of the Kolmogorov's entropy can be obtained directly from the measurement in time domain of one of the components of the chaotic system.

After Schuster the Kolmogorov's entropy is this fundamental quantity, which characterizes the chaotic behaviour and the strange attractor can be treated as a attractor, characterized by positive entropy [1].

For determination of the Kolmogorov's entropy the Ljapunov's spectrum can be used. In [2] has been shown, that, by the assumption, that the Ljapunov's exponents are known, the Kolmogorov's entropy (K) can be determined by means of the following expression [2]:

$$K \leq \sum_{i=1}^{k} \lambda_i = \sum_{\lambda_i > 0} \lambda_i ,$$

where k is the number of the positive Ljapunov exponent with the most little value.

III. Algorithms for Determination of the Ljapunov Exponents [2-5]

The Ljapunov exponents are very important by the examination of the chaotic attractors and by the estimation of the entropy of the analyzed circuits or systems.

There are several well known algorithms for determination of the Ljapunov exponents [2-5]:

- Algorithm of Alan Wolf, Jack B. Swift, Harry L. Swinney, John A.Vastano.
- Algorithm of Sano and Sawada.
- Algorithm of Eckmann, Kamphorst, Ruelle and Ciliberto (EKRC algorithm).

An algorithm for determination of the Ljapunov exponents from a time series has been presented by A. Wolf, J. Swift, H. Swinney and John A.Vastano in their work [3]. The authors have been presented the relation between the dynamic behaviour of the observed system and the Ljapunov's exponents.

In [3] has been shown, that from the point of view of the information theory, from the magnitudes of the Ljapunov's exponents, a quantitative estimate of the attractor's dynamic behaviour can be obtained, or i.e. the exponents measure the rate, at which system processes create or destroy information.

[1]Stoitscho Velizarov Manev, Dept. of Radiotechnique in Faculty of Communication and Communication Technologies in TU-Sofia, Bulgaria

[2]Vladimir Ivanov Georgiev, Dept.of Theoretical Electrotechnic in TU-Sofia, Bulgaria

The relation between the information dimension d_f and the Ljapunov's spectrum has been given by the equation [3]:

$$d_f = j + \frac{\sum_{i=1}^{j} \lambda_i}{|\lambda_{j+1}|},$$

where: $\sum_{i=1}^{j} \lambda_i > 0$ and $\sum_{i=1}^{j+1} \lambda_i < 0$ [3].

Two another well known algorithms for determination of the Ljapunov's exponents are the algorithm of Sano and Sawada and the algorithm of Eckmann, Kamphorst, Ruelle and Ciliberto (EKRC algorithm).

In their work [5] the authors (Eckmann, Kamphorst, Ruelle and Ciliberto) presented and analyzed an algorithm for computing Ljapunov exponents from an experimental time series. The algorithm includes the following steps [5]:

a) reconstructing the dynamics in a finite dimensional space.
b) obtaining the tangent maps to this reconstructed dynamics.
c) deducing the Ljapunov exponents from the tangent maps.

In [4] a comparison between the algorithm of Sano and Sawada and the EKRC algorithm has been made. After the conclusions, made there, the following properties have been observed [4]:

- The EKRC algorithm better approximates the zero and negative exponents in most cases and it seems more promising for the accurate determination of the whole spectrum.
- The EKRC algorithm gives stable Ljapunov exponents for a larger region of parameters and thus more easily determines the Ljapunov spectrum.

On other hand [4]:

- The Sano and Sawada algorithm is easy to implement and can determine the Ljapunov exponents for smaller values of the embedding dimension.

IV. Experimental Results

In what follows by the computation of the Ljapunov's exponents an algorithm, based on the well known EKRC algorithm, has been used. This algorithm has been applied by the investigations, carried out upon several circuits, designed to produce chaotic signals. The obtained values of the Ljapunov's exponents permit a qualitative and quantitative comparison between the considered circuits to be made. Based on the obtained results an estimate about the suitability of the investigated circuits for generating of chaotic signals can be made.

The determination of the time delay is from great importance for the accurate calculation of the Ljapunov exponents. The algorithm, used in the presented paper, is based on analysis of the autocorrelation functions (ACF) of the obtained chaotic signals. The time value, where the ACF for the first time obtains value near to zero, has been chosen as basis for the other necessary calculations [2].

In the presented work the Ljapunov exponents for signals, generated from the following circuits, have been computed:

1) H-generator [7];
2) 4-D generator [7];
3) Circuits, based on the canonical realization of Chua's circuit [8].

These circuits have been developed in practice. The signals, obtained at the outputs of the circuits, have been digitized by means of oscilloscope interface and have been analyzed. The obtained signals in time domain and the related autocorrelation functions have been presented in an another work [6].

Following the methodology, proposed in [2], the value for the time delay (τ) in each case has been determined.

The dynamic behaviour of the observed circuits has been analyzed in 3 or in 4-dimensional phase space. The phase space has been formed after the proposal, given in an example in [2].

An algorithm, based on the EKRC algorithm, has been implemented for determination of the Ljapunov's exponents and the Ljapunov's dimension.

A. H-generator [7]

Different signals have been obtained by the experiments. Some of them, displayed in the phase plane, have been shown on Fig. 1 [7].

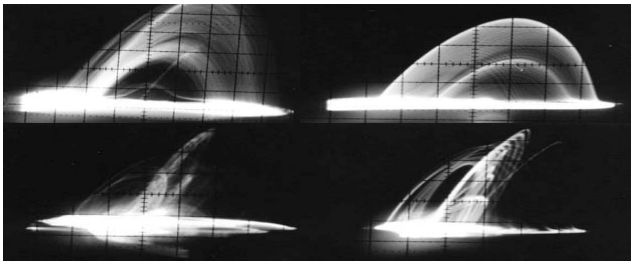


Fig. 1.

The Ljapunov's exponents for different sets of values of the parameters have been obtained:

A) By a first state of the parameter set of the presented in [7] H-generator, a chaotic signal has been obtained. The values of τ, the Ljapunov's exponents and the Ljapunov's dimension have been computed, as follows:

$$\tau = 6, \quad \lambda_1 = 0.0426, \quad \lambda_2 = -0.0137,$$
$$\lambda_3 = -0.0463, \quad \lambda_4 = -0.0864, \quad D_{Ljap} = 4.1175$$

B) By a second state of the parameter set an another chaotic signal has been obtained. The values of τ, the Ljapunov's exponents and the Ljapunov's dimension have been computed, as follows:

$$\tau = 2, \quad \lambda_1 = 0.1623, \quad \lambda_2 = 0.0117,$$
$$\lambda_3 = -0.1538, \quad \lambda_4 = -0.3167, \quad D_{Ljap} = 3.1319$$

C) By a third state of the parameter set a different chaotic signal has been obtained. The values of τ, the Ljapunov's exponents and the Ljapunov's dimension have been computed, as follows:

$$\tau = 2, \quad \lambda_1 = 0.2536, \quad \lambda_2 = 0.0135,$$
$$\lambda_3 = -0.1398, \quad \lambda_4 = -0.3084, \quad D_{Ljap} = 3.9105$$

The values, obtained for the Ljapunov's exponents, show a variation, depending on the set of the values of parameters. As expected, the magnitudes of the Ljapunov's dimension are significant. The results show, that on base of the H-generator, presented in [7], chaotic signals with different properties can be produced.

B. Four-dimensional chaotic generator [7]

In [7] a four-dimensional chaotic generator with modified external driven nonlinearity has been presented. The trajectories in the phase plane have been presented on Fig. 2 [7].

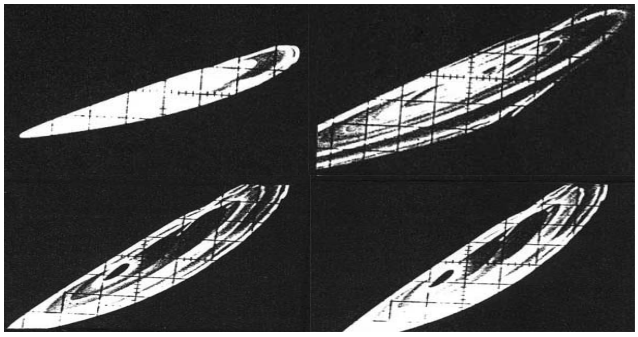


Fig. 2.

A) For a first state of the parameter set of the 4-D generator, presented in [7], the following values for τ, for the Ljapunov's exponents and for the Ljapunov's dimension have been obtained:

$$\tau = 4, \quad \lambda_1 = 0.0642, \quad \lambda_2 = 0.0105,$$
$$\lambda_3 = -0.0399, \quad \lambda_4 = -0.1452, \quad D_{Ljap} = 3.8722$$

B) By an another set of values of the system parameters the following values for τ, for the Ljapunov's exponents and for the Ljapunov's dimension have been obtained:

$$\tau = 9, \quad \lambda_1 = 0.0378, \quad \lambda_2 = 0.0085,$$
$$\lambda_3 = -0.0134, \quad \lambda_4 = -0.0537, \quad D_{Ljap} = 5.4536$$

The values of the Ljapunov's dimensions are significant. This is a logical consequence from the type of the system of differential equations, describing the processes in the 4-D generator, analyzed in [7].

The obtained results show, that the 4-dimensional generator, discussed in [7], is appropriate for generating of chaotic signals.

A first version of Chua's circuit [8]. The signals, concerning the discussed version, presented in [8], have been observed in the phase plane. They have been presented on Fig. 3 [8].

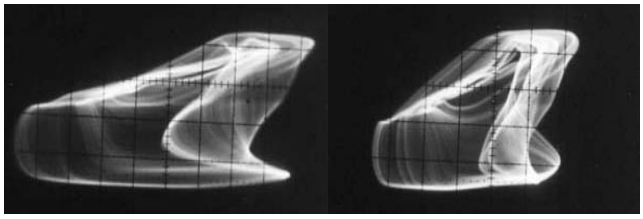


Fig. 3.

For the signal, which characteristic in time domain and the related ACF have been presented in [6], has been computed: $\tau = 2$. This signal has been obtained by a fixed set of values of the parameters. For determination of the Ljapunov exponents an algorithm, based on the EKRC algorithm, has been implemented.

Here the analysis has been conducted for a 3-dimensional system and the number of the corresponding Ljapunov exponents is 3.

The following values have been obtained:

$$\lambda_1 = 0.0998, \quad \lambda_2 = -0.0802, \quad \lambda_3 = -0.3151,$$
$$D_{Ljap} = 2.2435$$

The value of the Ljapunov's dimension is not as significant, as the values, obtained by the H-generator and the 4-dimensional generator, presented respectively in cases A and B. Nevertheless, from the final results becomes obvious, that the first version of Chua's circuit, discussed in [8], is appropriate for generating of chaotic signals.

A second version of Chua's circuit [8]. Another circuit, discussed in [8], includes external driven nonlinearity. By defined conditions (i.e by some set of values of the parameters) the presence of chaotic behaviour of the signal at the output of the circuit has been established.

The obtained signals have been observed in the phase plane and some of them have been presented on Fig. 4 [8].

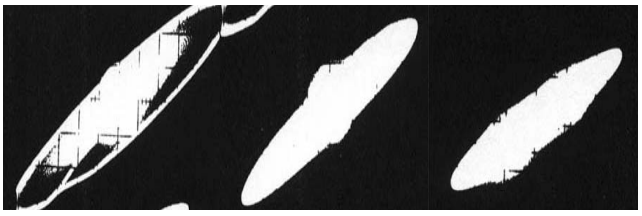


Fig. 4.

The following values for τ, for the Ljapunov's exponents and for the Ljapunov's dimension have been obtained:

$$\tau = 3, \quad \lambda_1 = 0.0885, \quad \lambda_2 = -0.0179, \quad \lambda_3 = -0.2162,$$
$$D_{Ljap} = 5.935$$

The value of the Ljapunov's dimension is significant. This fact is favourable to the implementation of the discussed variant, presented in [8], for generation of signals with chaotic behaviour.

A third version of Chua's circuit [8]. Another version of Chua's circuit, designed to generate chaotic signals, has been

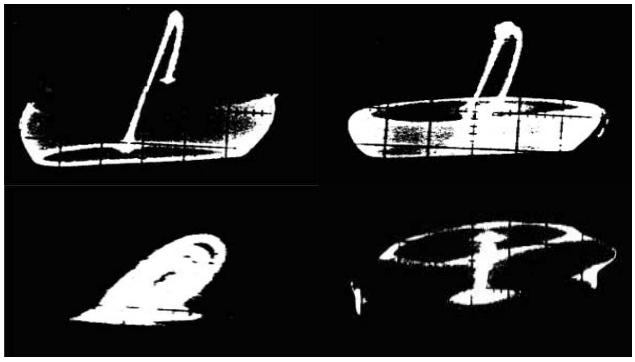


Fig. 5.

presented in [8]. At the output of the circuit a variety of signals have been obtained. Photos of the trajectories, observed in the phase plane, have been shown on Fig. 5 [8].

A) For a first state of the parameter set of the discussed circuit, presented in [8] , the following values for τ, for the Ljapunov's exponents and for the Ljapunov's dimension have been obtained:

$$\tau = 2, \quad \lambda_1 = 0.1645, \quad \lambda_2 = -0.0609, \quad \lambda_3 = -0.3401,$$
$$D_{Ljap} = 3.6997$$

B) By another conditions the following values for τ, for the Ljapunov's exponents and for the Ljapunov's dimension have been obtained:

$$\tau = 8, \quad \lambda_1 = 0.0419, \quad \lambda_2 = -0.0061, \quad \lambda_3 = -0.0411,$$
$$D_{Ljap} = 7.8468$$

C) By different values of the circuit parameters the following values for τ, for the Ljapunov's exponents and for the Ljapunov's dimension have been computed:

$$\tau = 2, \quad \lambda_1 = 0.2333, \quad \lambda_2 = -0.0324, \quad \lambda_3 = -0.3331,$$
$$D_{Ljap} = 8.21$$

Compared to the values of the Ljapunov's dimensions, computed for the first variant of Chua's circuit, the values, obtained here, are more significant.

It is obvious, that the magnitudes of the Ljapunov's dimension change to a considerable extent. The reason for the change in the values is the corresponding change in the set of the values of the parameters in the analyzed circuit. In this way, by means of the discussed circuit, presented in [8], by appropriate choice of the system parameters, different types of signals can be produced.

V. Conclusions

- Signals, obtained from some versions of circuits, designed to generate chaotic signals, have been studied.
- On the lab prototypes of these circuits experiments have been made.
- The signals, obtained at the outputs of the circuits, have been digitized and analyzed.
- By means of the autocorrelation functions, related to the signals, investigated in time domain, for each case the parameter has been computed and the necessary for the space construction calculations have been made.
- The Ljapunov's exponents have been computed by means of an algorithm, based on the EKRC algorithm. The algorithm, used in the presented work, will be improved and will be investigated more properly in future experiments.
- The obtained results have been compared and conclusions about the suitability for the implementation of each variant have been made.

References

[1] Heinz Georg Schuster, "Deterministc Chaos. An Introduction.", Physik-Verlag, Weinheim, 1984, Moscow, "Mir", 1988 (in Russian).

[2] Stoycho Panchev, "Theory of chaos (with Examples and Applications)", Academic Publishing House "Prof. Marin Drinov" , 1996 (in Bulgarian).

[3] Alan Wolf, Jack Swift, Harry Swinney, John A. Vastano, "Determining Lyapunov exponents from a time series", pp. 285-317, *Physica* 16 D, 1985.

[4] Antonis Karantonis, M. Pagistas, "Comparative study for the calculation of the Lyapunov spectrum from nonlinear experimental signals", pp. 5428-5444, Phys. Rev. E, Vol. 53, No.5, May 1996.

[5] J. P. Eckmann and S. Oliffson Kamphorst, D. Ruelle, S. Ciliberto, "Liapunov exponents from time series", pp.4971-4979, Phys. Rev. A, Vol. 34, No.6, Dec. 1986.

[6] Stoitscho V. Manev and Vladimir Iv. Georgiev, "Chaotic signals, generated by some circuits – comparative study (correlation analysis)", ICEST Sofia 2003.

[7] Stoitscho V. Manev and Vladimir Iv. Georgiev, "Chaotic Signals in Radiocommunication Generated by some Circuits", pp. 627-630, ICEST Nish 2002.

[8] Stoitscho V. Manev and Vladimir Iv. Georgiev, "Study of some Circuits in Radio Communication Designed to Generate Chaotic Signals", pp.631-634, ICEST Niš 2002.

Characterization of Blood Glucose Levels in Human as Chaotic Biological Signal

R. Penev[1] and M. Wada[2]

Abstract – **The Characterization method of the chaotic bi-signal in human is investigated. For the measured time series of bio-data, the internal dynamics with a form of vector field by using embedding method is constructed. In order to characterize and recognize such dynamic structure, we will be constructing in next paper the chaotic neural system artificially. Then we characterize the similarity by the synchronized or de-synchronized Responses among the artificial dynamics and the target internal dynamics of bio-signal.**

I. Introduction

Self-monitoring of blood glucose is an effective method to determine of the glucose levels [1-3]. The method is currently used by many diabetics for their treatment and blood glucose control. Typical values of the blood glucose [4,5] is as following scheme

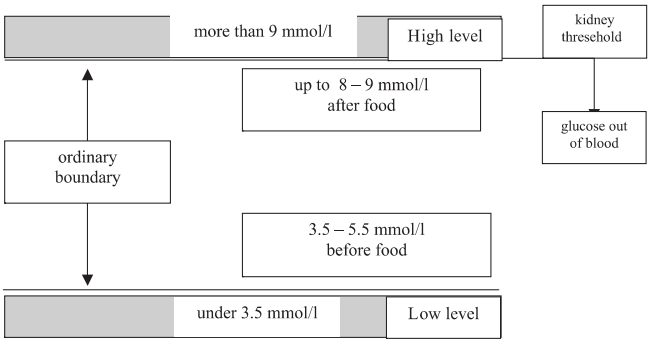


Fig. 1.

Good rectification

- level of blood glucose < 10 mmol/l during twenty-four hours;
- before food < 4 to 7 mmol/l
- 1 hour after food < 10 mmol/l
- 2 hour after food < 8 mmol/l
- level of hemoglobin **A1c** – from 4.5% to 7.5%
- glucoside hemoglobin **HbA 1c** – under 7.5%

Now, measurement of glucose in whole blood !

!!! An exception (See GLUCOWATCH of *Cygnus*)

Each measurements, used in this papers, are conducted by means of glucometer with characteristics described below:

[1]R. Penev, Faculty of Communications, Technical University – Sofia, rpenev@tu-sofia.acad.bg

[2]M. Wada, Hokkaido University, wada@complex.eng.hokudai.ac.jp

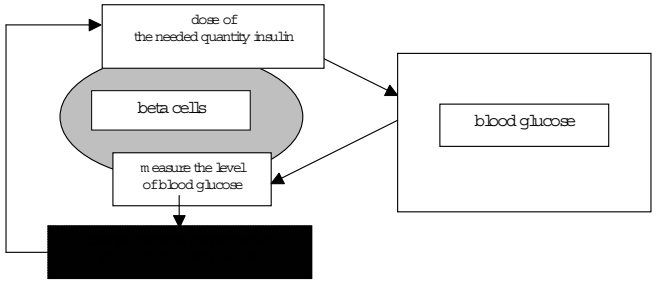


Fig. 2.

GLUCOCARD II (blood glucose test meter) – part from Operating Manual manufactured by: KDK KORPORATION, 57 NISHI AKETA-CHO, HIGASHI-KUJO, MINAMI-KU, KYOTO 601, JAPAN

Test: Glucose in whole blood

Sample size: Approximately 5μl

Measuring Range: 40-500 mg/dl (2.2-27.8 mmol/l)

Measuring Time: 60 seconds

Temperature Compensation: Automatically compensation using a built-in thermo-sensor

Calibration: Automatically selects the appropriate calibration curve by using a CALIBRATION Strip

Principle: The blood sample is drawn into the Test Strip through capillary action. Glucose in the sample reacts with glucose oxidase and potassium ferricyanide in the strip, producing potassium ferrocyanide. Potassium ferrocianyde is produced in proportion to the glucose concentration of the blood sample. Oxidation of the potassium ferrocyanide produces an electrical current which is then converted by the meter to display the glucose concentration.

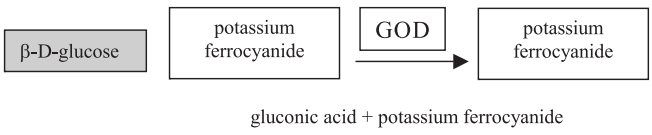


Fig. 3.

Following pages show experimental results from processing of data in environment of Matlab software

II. A Possibly Experimental System (A proposal for scheme of the measurements)

The experimental system is identical as described in [1-3]. In advance it is known that:

- The hemoglobin has a property of adsorption of infrared ray [1-4]
- The hemoglobin contains an information for blood glucose [4,5]
- hemoglobin **A1c** – from 4.5% to 7.5% (good values)
- glucoside hemoglobin **HbA 1c** – under 7.5% (good values)

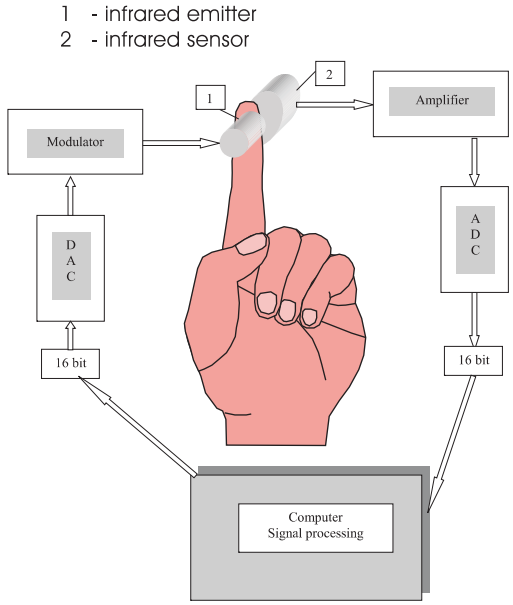


Fig. 4.

Outcome results: Time series as biological signal which contain information for the instantaneous values of levels of blood glucose

Problems: – How we can extract that information?
– What kind modulation will be effective with accordance of pursue goals?

Possible decisions (Application of Chaos Theory)

1. Embedding method
2. Mutual information
3. Correlation dimension
4. Time delay
5. Lyapunov exponent
6. Unstable Periodic Orbits

Images below are by means of [16] Software package NDT (Nonlinear Dynamics Toolbox)

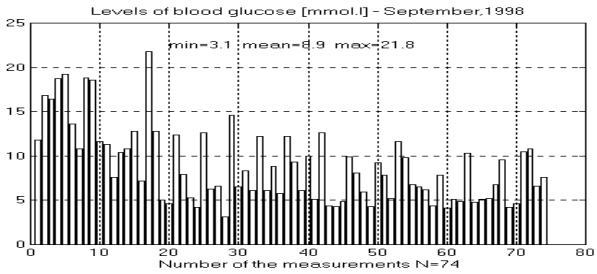


Fig. 5a.

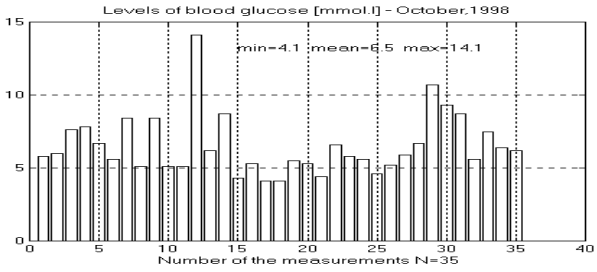


Fig. 5b.

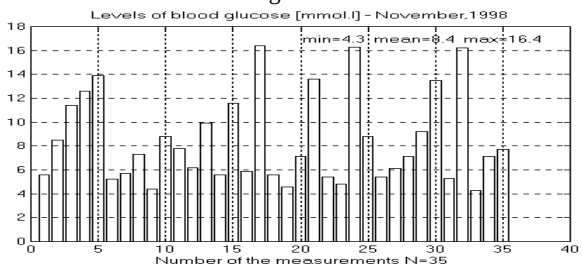


Fig. 5c.

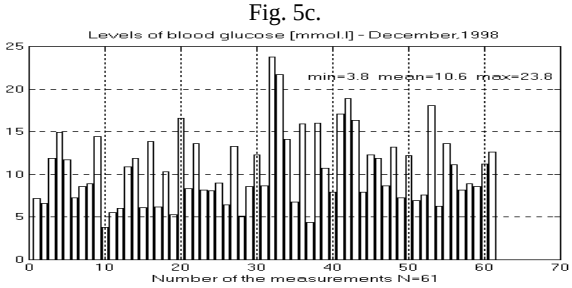


Fig. 5d.

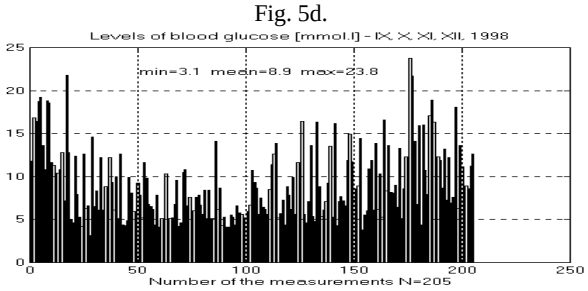


Fig. 5e.

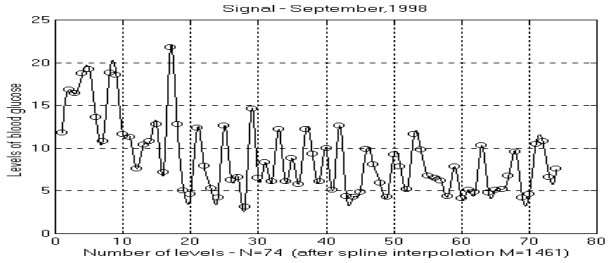


Fig. 6a.

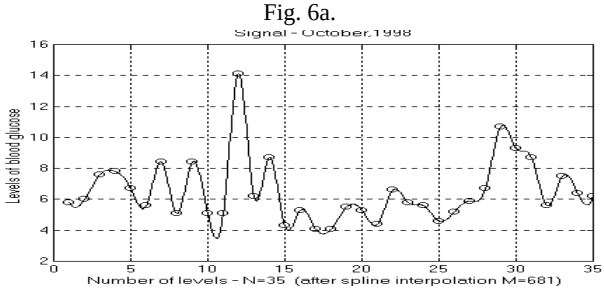


Fig. 6b.

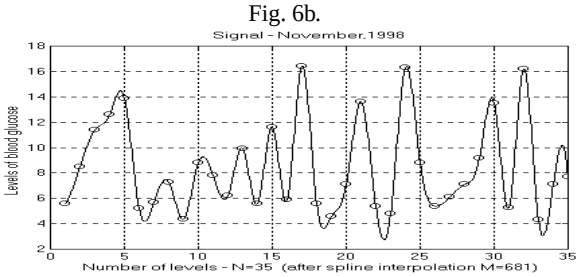


Fig. 6c.

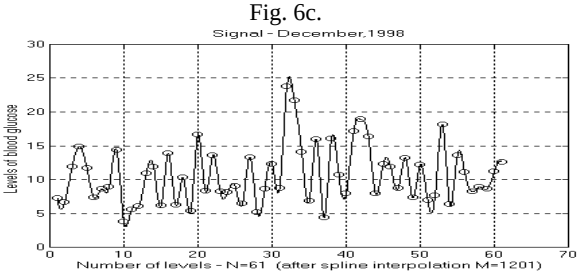


Fig. 6d.

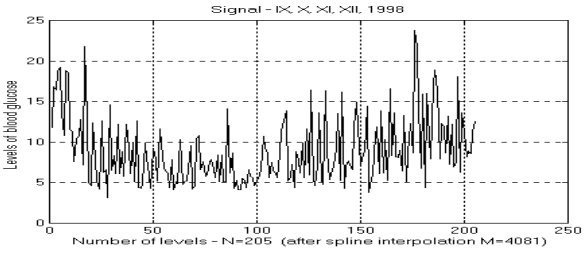


Fig. 6e.

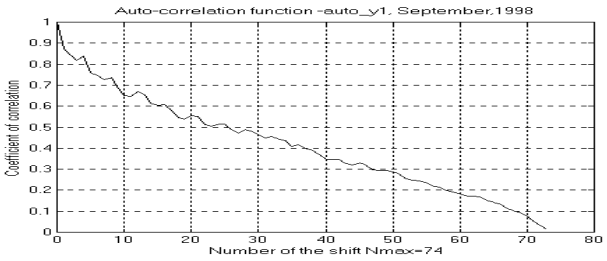


Fig. 7a.

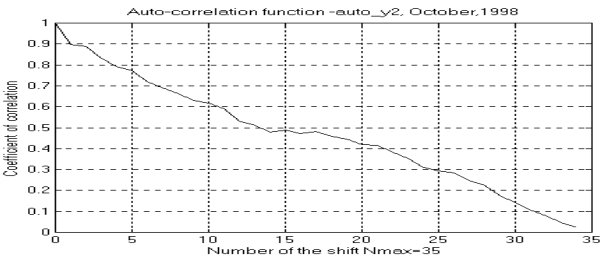


Fig. 7b.

Fig. 7c.

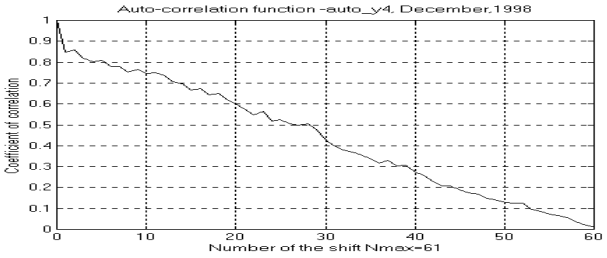


Fig. 7d.

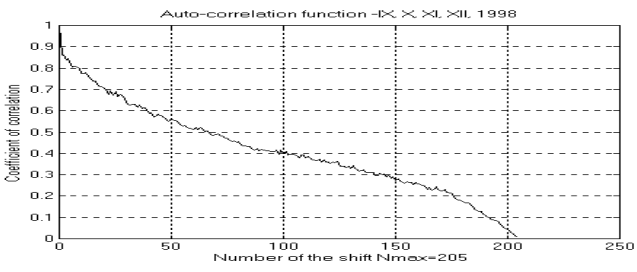


Fig. 7e.

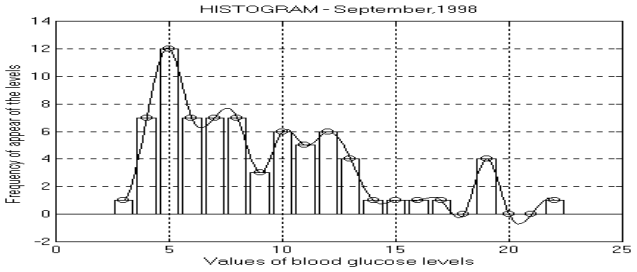


Fig. 8a.

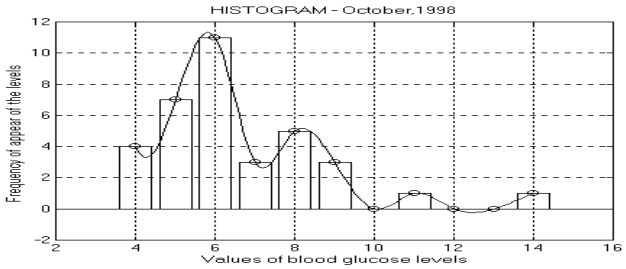


Fig. 8b.

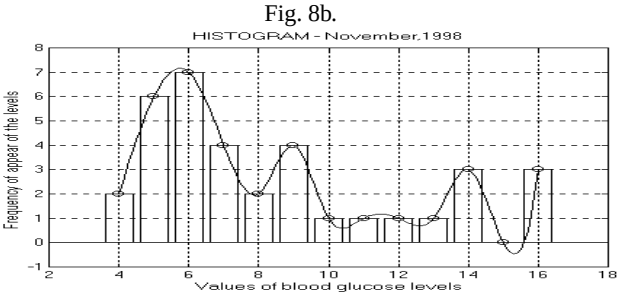


Fig. 8c.

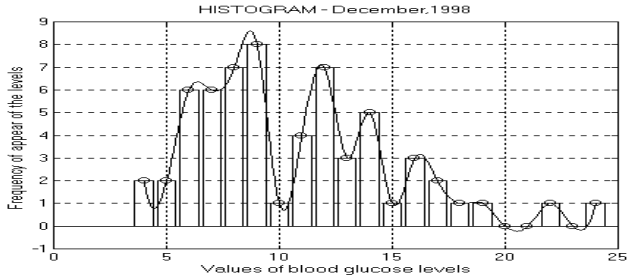


Fig. 8d.

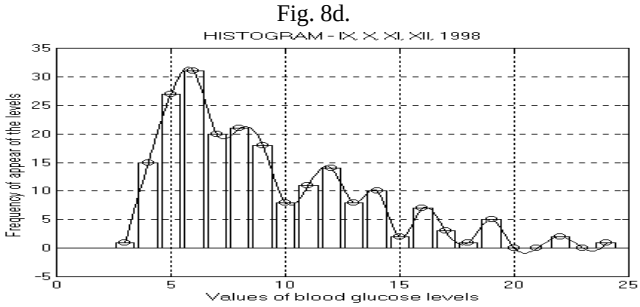


Fig. 8e.

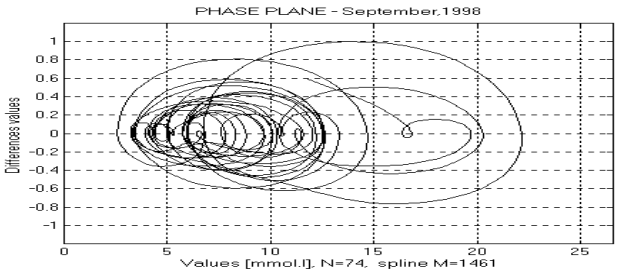


Fig. 9a.

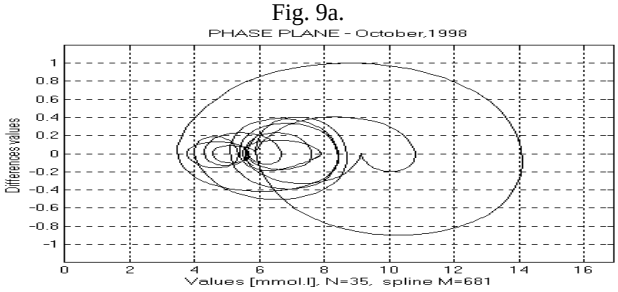


Fig. 9b.

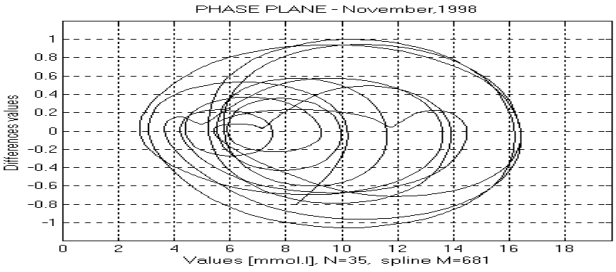


Fig. 9c.

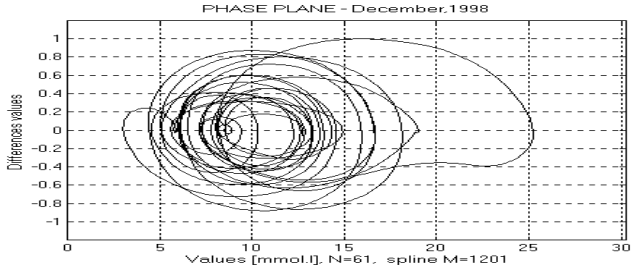


Fig. 9d.

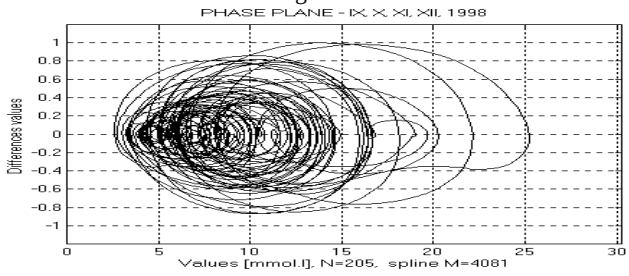


Fig. 9e.

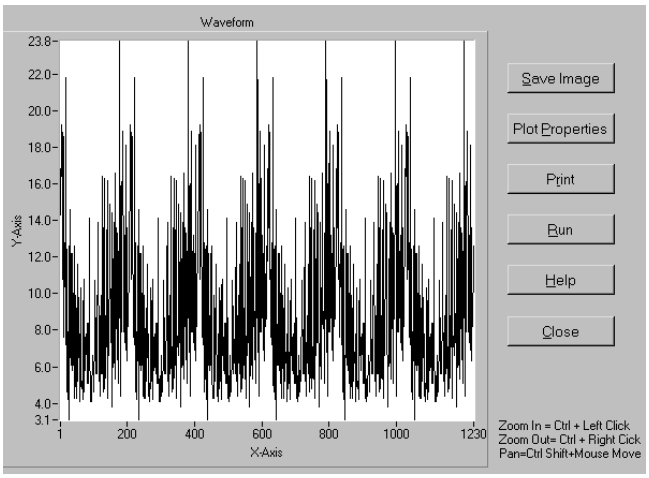


Fig. 10.

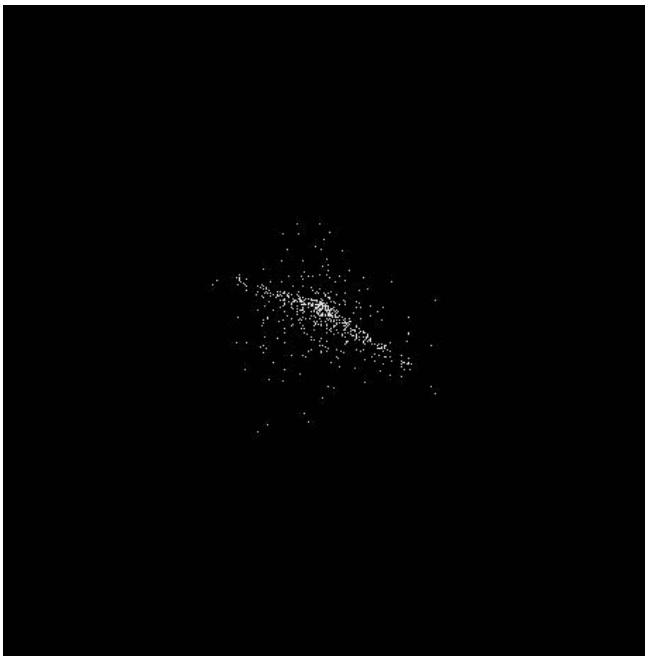


Fig. 11. 3-D plot of embedding data

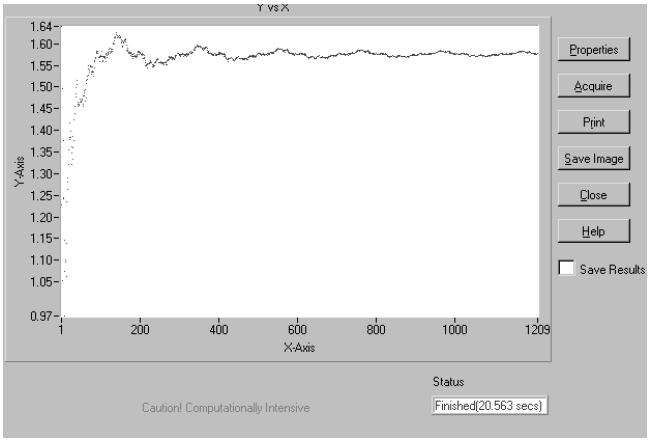


Fig. 12. Dominate Lyapunov exp.

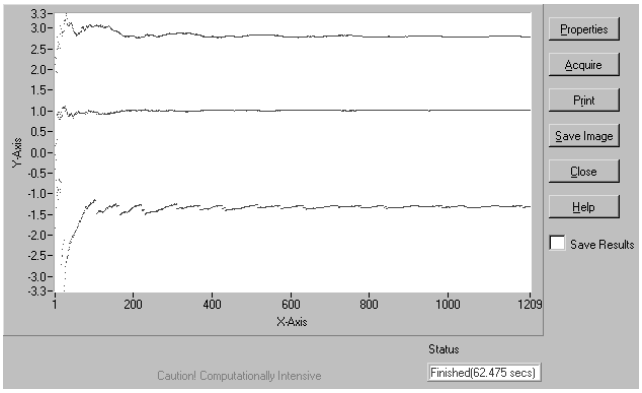


Fig. 13. Lyapunov spectrum

III. Conclusion

We found chaotic property in time series of data of human blood glucose levels. See positive Lyapunov spectrum and Lyapunov exponent. This is meaning that tools of analysis of chaotic signals are appropriate for this bio signal.

References

[1] Yamaguchi, S. Miami, H. Watanabe, M. Wada, “Extraction of a Dynamic Structure of Biological Signals as Stochastic Automata for a Human-Machine Interface”, *Proceedings of 6th IEEE Intetnational Workshop on Robot and Human Communication*, pp.438-441, 1997.

[2] Yamaguchi, S. Mikami, H. Watanabe, M. Wada, “Characteization of Biological Internal Dynamics by the Synchronization of Coupled Chaotic System”, *Intelligent Autonomous Systems*, IOS Press, pp.670-677, 1998.

[3] A. Yamaguchi, S. Mikami, M. Wada, “Identification of biological internal dynamics using a structure of unstable periodic orbits”, *Proceedings of International Conference on Systems, Man and Cybernetics Conference’99*, Tokyo, 1999.

[4] Standl Eberhard, Hellmut Mehnert, Das große THRIAS-Handbuch für Diabetiker, Verlag, 1988.

[5] 5. Пенев Р. П., Нелинейни комуникационни системи, изд. на ТУ–София, 1998. Penev R. P., Nonlinear comunication systems, (in Bulgarian), Printed Office – Technical University of Sofia, 1998

[6] Hale Jack K., Huseyin Kocak, Dynamics and Bifurcation, Springer- Verlag,1991.

[7] Parrcer S. Tomas, Leon O. Chua, Practical Numerical Algorithm for Chaotic Systems, Springer-Verlag, 1989.

[8] Randall R. B., B.Tech, Frequency analysis, Printed in Denmark, Larsen & Son, 1987.

[9] Thompson J.M.T, H.B. Stewart, Nonlinear Dynamics and Chaos, John Wiley and Sons, 1991.

[10] ТИИ Э Р (перевод с английского). Хаотические системы, Тематический выпуск, том 75, 1987. (IEEE translated into Russian) Chaotic systems

[11] IEEE Circuits and Systems-I: Fundamental theory and applications, Special issue on chaos in nonlinear electronic circuits. Part A. Tutorials and reviews, volume 40, October, 1993

[12] IEEE Circuits and Systems-I: Fundamental theory and applications, Special issue on chaos in nonlinear electronic circuits. Part B. Bifurcation and chaos, volume 40, November, 1993

[13] IEEE Circuits and Systems-II: Analog and digital signal processing, Special issue on chaos in nonlinear electronic circuits. Part C. Application, volume 40, October, 1993

[14] Зарубежная радиоэлектроника, Тематический выпуск. Проблемы динамического хаоса, No 10, 1997 (in Russian) Problems Dynamics Chaos

[15] IEEE Circuits and Systems-I: Fundamental theory and applications, Special issue on chaos synchronization and control:theory and application, volume 44, October, 1997

[16] Software package NDT (Nonlinear Dynamics Toolbox), created by the Scientific Visualization Lab and Applied Chaos Lab of Georgia Tech.

Evaluation of Dispersion Characteristics of Conventional Single-Mode Fibers

Dimitar Slavov[1]

Abstract – **The use of Chromatic dispersion, dispersion slope and relative dispersion slope as dispersion characteristics are discussed. The results for two types of single-mode fibers in S+C band are presented.**

Keywords – **Chromatic dispersion, Single-mode fibers, Dispersion slope, RDS, Dispersion compensation**

I. Introduction

The chromatic dispersion is the most significant limitation factor of the fiber length for high-speed optical transmission systems. This limitation is more relevant for DWDM systems with high count of channels. The diverse types of fibers have significant differences in their dispersion characteristics. The new NZDS optical fibers have small values of chromatic dispersion in C and L bands [1]. At the same time, most of the optical networks today, use conventional single-mode fibers (G.652). These fibers were designed originally for single channel operation at 1310 nm. They have dispersion of 15-18 ps/nm.km in the 1550 nm window. At the datasheets can only be found dispersion values at 1300 an 1550 nm, as well values for dispersion slope (S_0) at zero dispersion wavelength (λ_0). When increasing the needs of transmission capacity, the operation of these fibers in C as well in S band (1460-1530) nm becomes very important [1].

II. Chromatic Dispersion Characteristics

The broadening of the optical pulse, travelling within a single-mode fiber can be calculated as follows:

$$\Delta t_{out} = \sqrt{(\Delta t_{in})^2 + (DL\Delta\lambda_{mod})^2 + (DL\Delta\lambda_{sour})^2} \quad (1)$$

where: Δt_{in} – length of input pulse; D – chromatic dispersion; L – fiber length; $\Delta\lambda_{mod}$ – modulation spectral broadening; $\Delta\lambda_{sour}$ – source spectral width;

Usually $\Delta t_{in} \ll \Delta t_{out}$, then the pulse broadening will be approximately:

$$\Delta t_{out} \approx |D|\Delta\lambda_{eff} L \quad (2)$$

where

$$\Delta\lambda_{eff} = \sqrt{\Delta\lambda_{mod}^2 + \Delta\lambda_{sour}^2}.$$

At small values of D (near zero dispersion wavelength) and at high bit rates, the dispersion high-order components must be taken into account:

$$\beta_2(\omega) = \beta_2 + \beta_3(\omega-\omega_0) + \frac{\beta_4}{2}(\omega-\omega_0)^2 + \frac{\beta_5}{6}(\omega-\omega_0)^3 \quad (3)$$

where β_2 – dispersion value of the fiber at the reference frequency ω_0; β_3, β_4 and β_5 – high order dispersion coefficients. β_2 [ps^2/km], correspond to chromatic dispersion $D(\lambda)$ [ps/nm.km], and β_3 [ps^3/km] – to dispersion slope $S(\lambda)$ [ps/nm^2.km] respectively [4].

The dispersion coefficients β_4 and β_5 must be taken into account at small values of the dispersion and ultra high bit rates (40 Gbit/s and more) [2]. The first two coefficients (β_2 and β_3) only can be used for evaluation of the dispersion at transmission speeds up to 10 Gbit/s:

$$\begin{aligned} D(\lambda) &= \frac{dt_g}{d\lambda} \text{ [ps/nm.km]} \\ S(\lambda) &= \frac{dD(\lambda)}{d\lambda} \text{ [ps}^2\text{/nm.km]}. \end{aligned} \quad (4)$$

The relative dispersion slope (RDS) of the fiber is a good assessment of the possibilities for dispersion compensation. RDS is the ratio between the dispersion slope (S) and the dispersion (D):

$$\text{RDS} = \frac{S}{D} \text{ [1/nm]}. \quad (5)$$

The first condition to get precise broadband compensation is that the dispersion compensating fiber (DCF) should compensate dispersion and dispersion slope of the transmission fiber. This can be achieved by matching the RDS of DCF to that of the transmission fiber.

III. Results

Two typical G.652 single-mode fibers whose characteristics conform to the characteristics of SMF28 (Corning Inc.) and SMF (Alcatel) were modeled with FiberCad.

Dispersion slope and RDS are calculated with the obtained values of D in (1460-1560) nm wavelength band. Fig. 1 shows the chromatic dispersion in the above band for both fibers.

The chromatic dispersion for 1460 and 1550 nm, determined with FiberCAD simulation, matches the values given in data sheets by the manufactures.

The calculated dispersion slope for both of fibers are shown on Fig. 2.

The dispersion slope in the S-band remains less then 0.07 ps/nm^2.km. As expected, this value decreases with increasing of wavelength. If the RDS is known it is possible to determine the usable bandwidth of the fiber. It is defined as

[1]Dimitar Slavov, Dept. of Telecommunications, Technical University of Varna, Studentska 1, 9010 Varna, Bulgaria, E-mail: slavov@ms3.tu-varna.acad.bg

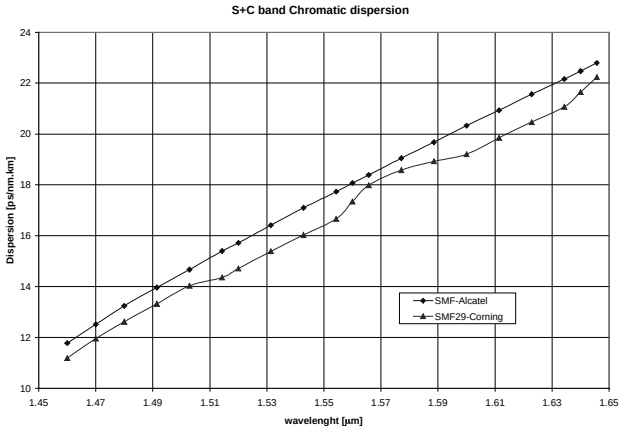


Fig. 1. Dispersion changes in S and C band

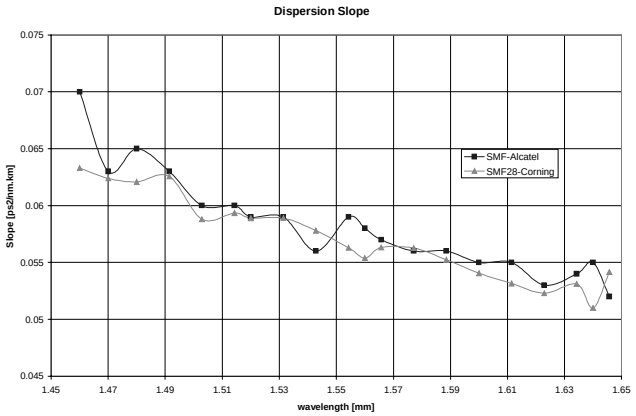


Fig. 2. Dispersion changes in S and C band

the maximal wavelength band in which the residual dispersion after compensation is less than 1 ps/nm [6].

As we can see in [6] the usable bandwidth and RDS are inverse proportional. We have determined the RDS for both fibers as Fig. 3 shows.

The values of RDS for the whole wavelength range are between (0.0022 and 0.0059) 1/nm. The calculated usable bandwidth vs. RDS of the transmission fiber is shown on fig. 4.

IV. Conclusion

The obtained results show that the G.652 SMF have relative small values of chromatic dispersion in S-band. Dispersion slope is approximately the same as that of the DCF. This allows the use of G.652 SMF at these wavelengths even in multi channel applications. In C-band the dispersion is grater, but dispersion slope has smaller values, which allows better dispersion compensation. Dispersion compensating fibers can be used for dispersion compensation in both C and S bands. The obtained results for dispersion and dispersion slope of conventional single mode fibers at different wavelengths in S and C band ca be used to find the appropriate DCF and compensations scheme.

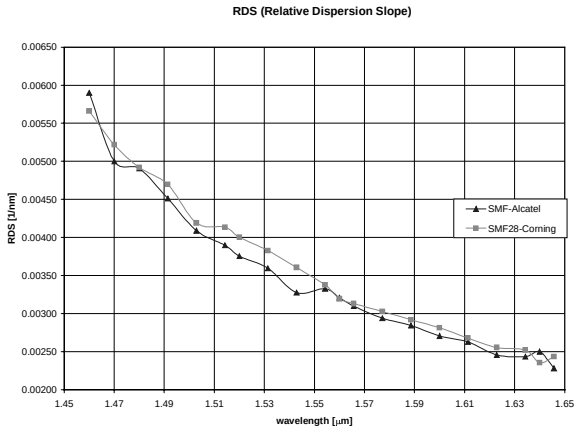


Fig. 3. The RDS changes in 1460-1560 wavelength band

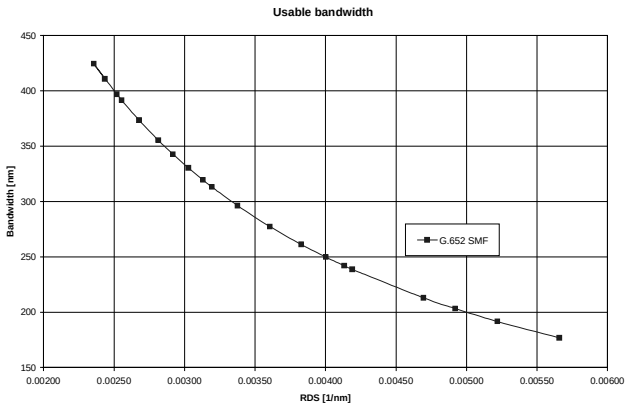


Fig. 4. Bandwidth vs. RDS for G.652 optical fiber

References

[1] Y. Danziger, D. Askegard, An Innovative Approach to Chromatic Dispersion Management that enables optical Networking in Long-haul high-speed transmission systems, *Optcal Networks Magazine* Vo.2 Issue 1, Jan/Feb 2001

[2] G. Novak, Exact chromatic dispersion theory of optical fibers, *Optik* No.10 1999, Urban & Fischer Verlag

[3] Lars Gruener-Nielesen, Bent Edvold, Status and future promises for dispersion compensating fibers, *Proc. of OFC 2002*, May 2002.

[4] Fiber_CAD, Optical Fiber Design Software, Technical Documentation, Optiwave Corporation 2000

[5] Marie Wandel, Poul Kristensen et al, Dispersion compensating fibers for non-zero dispersion fibers, *Proc. of OFC 2002*, May 2002

[6] Jacob Rathje, Lars Gruener-Nielsen, Relationship between Relative Dispersion Slope of transmission fiber and the usable bandwidth after dispersion compensating, *Proc. ECOC 2002*

Link Range of Free Space Laser Communication System

Erwin Ferdinandov[1], Tsvetan Mitsev[2]

Abstract – **In this paper a model analytical description of a Free Space Laser Communication System the type ground-to-space is constructed. It is used for quantitative determination of the maximum range between transmitter and receiver depending on the bit error rate (BER). The laser radiation extinction in the atmosphere is accounted for by means of the visibility.**

Keywords – **Laser Communication, Wireless Communication, Free Space Optics, Ground-to-Space Optical Communication, Atmospheric Attenuation**

I. Introduction

Contemporary development of laser physics and technology offers new possibilities for the use of free space laser communication systems (FSLCS) [1]. The development of FSLCS of the type ground-to-space (or ground-to-space LCS) is of special interest [2-5]. This is due to two reasons: (I) these systems, in contrast to FSLCS of the type ground-to-ground (point-to-point), have no alternative in fiber optic communication systems (FOCS); (II) ground-to-space LCS are directly with present and future free space reclamation.

In this paper we attempt to construct an analytical model of ground-to-space LCS, and to connect parameters of structural links and characteristics of free space with the quantitative indices of the system as a whole.

II. Analytical Model of Surface-To-Space LCS

We assume that the transmitter part of the system is constructed on the basis of single mode Nd^{3+}:YAG laser excited by semiconductor lasers. This gives us reason to adopt Gaussian-amplitude and synphase distribution of the optical field in the aperture of the transmitter's aerial. Of course, we must take into account the unavoidable and very often considerable diversions of the real distribution from the shown theoretical idealization which leads to a substantial increase of the laser beam divergence. For this purpose, we correct the current radius of the Gaussian laser beam with the experimentally determined radius of the beam at distance Z from the transmitter's aperture, introducing their ratio $K(Z) > 1$.

On the basis of diffraction spatial structure of the Gaussian laser beam and accounting for the energy losses in the transmitter's and receiver's aerials, losses from the extinction in the earth's atmosphere and losses due to the use of pulse code modulation (PCM) we obtain the expression for the mean signal optical flux in the aperture of the photo-detector, namely:

$$\Phi_S = \frac{\tau_t \tau_r \tau_m}{2\sqrt{2}(1-e^{-2})}\Big[1 - \exp\Big(-\frac{k^2\rho_0^2 R_r^2}{2K^2(Z)Z^2}\Big)\Big]\Phi_L, \quad (1)$$

where: Φ_L – optical flux in the laser output aperture (before PCM), $k = 2\pi/\lambda$, R_r – radius of the receiver aerial aperture, λ_0 – distance from the center of transmitter aerial aperture on which the optical field decreases e times (or initial radius of Gaussian laser beam), τ_t, τ_r, τ_m – transparencies of transmitter aerial, of receiver aerial, and of free space, respectively.

For $\lambda = 0.53$ mm (second harmonic of Nd^{3+}:YAG laser) it is possible to neglect the absorption of laser radiation in the atmospheric aerosols and atmospheric gases ($\tau_{aer}^{(a)} = \tau_{mol}^{a)} = 1$) and to assume that the extinction is only due to corresponding scattering, i.e.:

$$\tau_m = \tau_{aer}^{(s)} \tau_{mol}^{(s)}. \quad (2)$$

The determination of $\tau_{mol}^{(s)}$ is accomplished on the basis of Relay theory of scattering. For standard atmosphere and for $\lambda = 0.53$ mm we have $\tau_{mol}^{(s)} = 0,9$.

To find we use the Elterman's model, according to which for laser beam propagation through the entire atmosphere we have

$$\tau_{aer}^{(s)}(S_m) = \exp\Big[-\frac{1}{b(S_m)}\alpha_{aer}(0,S_m)\Big]. \quad (3)$$

The quantity

$$\alpha_{aer}(0,S_m)\ [\mathrm{km}^{-1}] = \frac{3.92}{S_m\ [\mathrm{km}]}\Big(\frac{\lambda\ [\mu\mathrm{m}]}{0.55}\Big)^{-0.585 S_m\ [\mathrm{km}]^{\frac{1}{3}}} \quad (4)$$

in Eq. (3) is the volume coefficient of aerosol scattering at ground level, S_m is the visibility at ground level, b is a coefficient determined by the curve in of Fig. 1.

The height H is connected with the distance Z with the relation

$$H = Z\cos\psi, \quad (5)$$

where ψ is the zenith angle of free space channel. We further suppose that a photo-multiplyer (PMP) is used as an optical radiation detector.

Analyzing the operation of a photo detector we obtain an expression for signal-to-noise ratio (SNR), namely:

$$SNR = \frac{S_{\mathrm{C}}^2 \Phi_{\mathrm{S}}^2}{2eN\Delta f(\sqrt{2}S_{\mathrm{C}}\Phi_{\mathrm{S}} + i_{\mathrm{B}} + i_{\mathrm{D}})}, \quad (6)$$

where $S_{\mathrm{C}} = e\eta\lambda/hc$ is cathode sensitivity of PMP (η is its quantum efficiency; e, electron charge; h, Plank's constant, N, noise coefficient from amplifying; Δf, signal frequency-spectrum bandwidth; i_{B} and i_{D}, mean values of background

[1]Erwin Ferdinandov is with the Faculty of Communication Techniques and Technologies, Kliment Ohridski blvd. 8, 1797 Sofia, Bulgaria.

[2]Tsvetan Mitsev is with the Faculty of Communication Techniques and Technologies, Kliment Ohridski blvd. 8, 1797 Sofia, Bulgaria, E-mail:Mitzev@vmei.acad.bg.

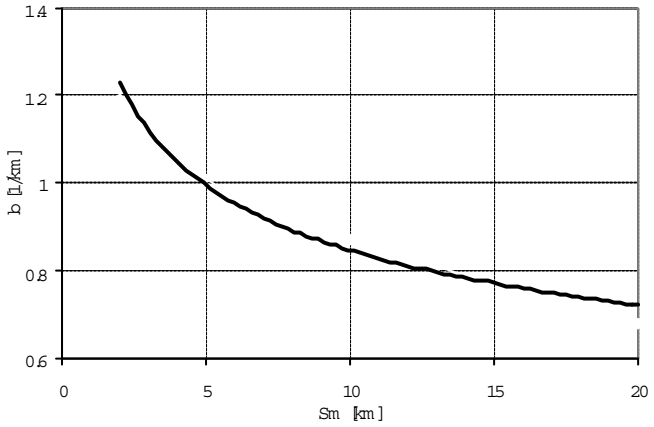


Fig. 1. Dependence of coefficient b on visibility

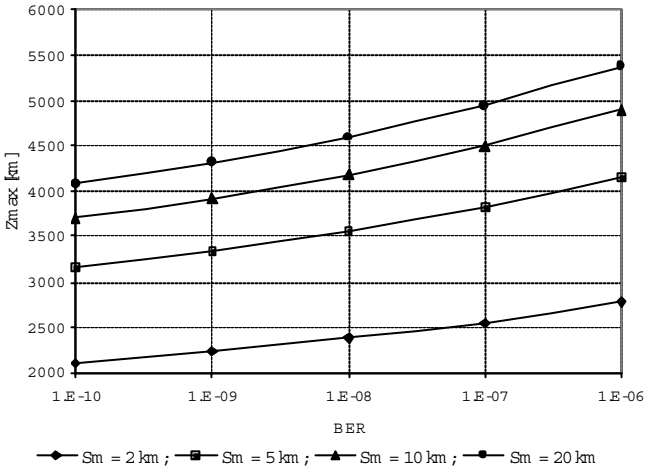


Fig. 2. Dependence of range limit value on the bit error rate

and dark currents in the cathode circuit of PMP, respectively. The values of the currents are calculated with the well known relations

$$i_{\mathrm{B}} = \frac{\pi^2 e\eta\lambda\tau_r}{hc} L_{\lambda,\mathrm{B}} \frac{R_{\mathrm{PMP}}^2 R_r^2}{f^2} (\Delta\lambda)_{\mathrm{F}} \tag{7}$$

and

$$i_{\mathrm{D}} = \frac{i_{\mathrm{Da}}}{G}, \tag{8}$$

where $L_{\lambda,\mathrm{B}}$ is the spectral density of background brightness; R_{PMP}, radius of input aperture of PMP; f, equivalent focal distance of receiver aerial; $(\Delta\lambda)_{\mathrm{F}}$, interference filter optical bandwidth; i_{Da}, anode dark current of PMP; G, current gain coefficient of PMP.

The connection between BER and SNR is given by the equation

$$BER = \frac{1}{\sqrt{\pi SNR}} \exp\Big(-\frac{1}{4}SNR\Big). \tag{9}$$

The calculation of the dependence $Z_{\max} = Z_{\max}(BER)$ is performed by substituting (1) in (6) and (6) in (9) with the subsequent solution of the resulting relation with respect to Z. Fixing the value of BER transforms Z to $Z_{\max}$.

III. Calculations

As example the dependence of $Z_{\max}$ on BER with parameter S_m is carried out with the following values: $\Phi_L = 1$ W; $\rho_0 = 2$ cm; $K(Z) \approx K = 10$; $\tau_t = 0.8$; $R_r = 20$ cm; $\tau_r = 0.4$; $f = 2$ m; $(\Delta\lambda)_{\mathrm{F}} = 10$ Å; $\Delta f = 100$ MHz (information capacity 200 Mbit/s); $\psi = 0$; $R_{\mathrm{PMP}} = 3$ mm; $\eta = 0.1$; $N = 1.5$; $G = 10^7$; $i_{\mathrm{Da}} = 10$ nA; $L_{\lambda,\mathrm{B}} = 10^{-3}$ [W/m^2.sr.Å]. The results are shown in Fig. 2 for $S_m = 2$ km, 5 km, 10 km, 20 km.

IV. Conclusion

As one can see on Fig. 2, in the most important from a practical viewpoint interval of BER (from 10^{-10} to 10^{-6}) the decrease in information losses (decrease of BER) leads to an acceptable decrease of the system link range. The values of the realizable link range demonstrate the great potential of ground-to-space LCS. One can also see the strong influence of the atmospheric transparency on $Z_{\max}$ for a given BER value. For a relatively clean atmosphere, $S_m > 10$ km, the S_m value is not of substantial importance. However, for low visibility ($S_m < 5$ km), $Z_{\max}$ drops rather quickly with S_m.

References

[1] J. Gowar, *Optical Communication Systems*, London, Prentice/Hall International, Inc., 1984.

[2] I. Kim, R. Stieger, J. Koontz, C. Moursund, M. Barclay, P. Adhikari, J. Schuster, E. Korevaar, R. Ruigrok and C. DeCusatis, "Wireless Optical Transmission of Fast Ethernet, FDDI, ATM, and ESCON Protocol Data Using the TerraLink Laser Communication System", *Optical Engineering*, vol. 37, no. 12, pp. 3143-3155, 1998.

[3] R. Strickland, M. Lavan, E. Woodbridge and V. Chan, "Effects of Fog on the Bit-Error Rate of a Free-Space Laser Communication System", *Applied Optics*, vol. 38, no. 3, pp. 424-431, 1999.

[4] D. Begley, "Laser Cross-Link Systems and Technology", *IEEE Communication Magazine, Free Space Laser Communications*, pp. 126-132, August 2000.

[5] K. Wilson, M. Enoch, "Optical Communications for Deep Space Missions", *IEEE Communication Magazine, Free Space Laser Communications*, pp. 126-132, August 2000.

Modeling of a Loaded Cylindrical Metallic Cavity with Real Excitation Using 3-D TLM Method

B. Milovanović[1], A. Marinčić[2], J. Joković[3] and A. Atanasković[4]

Abstract – **For the example of the cylindrical metallic cavity with circular cross section loaded with a lossy dielectric slab placed on the bottom of the cavity, the real excitation modeling, using TLM wire node, are presented. As an excitation form, a straight wire conductor is used, according to the wanted type of mode in the cavity. Water at a temperature of 15°C is used as a dielectric layer. The modeling process is described and the obtained TLM numerical results in frequency range $f = [1.5 - 3]$ GHz are compared with the experimental ones. Comparing numerical results and experimental ones, an excellent agreement is observed. Also, in order to investigate the influence of real probe length to the resonant frequencies of modes, TLM results with real probe are compared with results calculated by using the theoretical approach, that is TLM method with impulse excitation, and the appropriate conclusions are given.**

Keywords – **TLM method, microwave applicator, cavity, real excitation, wire node, lossy dielectric sample, resonant frequency**

I. Introduction

Cylindrical metallic cavities represent a configuration very suitable for good modeling of some practical heating and drying applicators. The knowledge of the mode tuning behavior under loading condition has important significance and would help in designing these applicators. For this reason, some researches of the cylindrical cavities, based on using the different approaches, were presented by a number of authors [1-3]. Also, some experimental work has been done in order to investigate the mode tuning behavior experimentally [1,2].

TLM (Transmission-Line Modeling) method is a general, electromagnetic based numerical method that has been applied very successfully in the area of cylindrical metallic cavities modeling [3-5]. In all this applications, an impulse excitation was used to establish desired field distribution in the modeled cavity. However, this way of enhancing the wanted TE or TM mode is different from the experimental case where a small probe inside the cavity is used as an excitation. This difference in the cavity excitation causes that the TLM results in the case of impulse excitation being different from the experimental ones. With some recent improvements in TLM method, it is possible to model a small probe inside the cavity using TLM wire node [6] and to investigate the influence of the real excitation to the resonant frequencies of the cavity.

In practice, depending on the position and the mode of excitation (waveguide, capacitive probe, inductive loop or slots), the number of modes will be different from theoretical case. For instance, placing the coaxial cable in the middle of cavity height will not generate modes with even-mode numbers in z-plane. From the remaining odd-mode numbers some modes will not be excited, depending on whether they have an electric field component in the direction of the source electric field. The resulting electric field distribution will then be given by the sum of the modes excited in the cavity. Another problem is identification of the precise modes. Although the S_{11} plots give the number of modes in the cavity, they do not indicate exactly which modes are present. This situation is made worse when many modes are present. The probe presence also tends to shift the modes and sometimes split degenerate modes.

The goal of this paper is to describe the possibilities of TLM method for modeling of loaded microwave applicator with real excitation probe. The applicator is represented in the form of a cylindrical metallic cavity loaded with a homogeneous lossy dielectric sample placed on the bottom of the cavity. As the microwave applicator is often used for drying of wet material, which as a dominant element within itself have water, as a dielectric layer, water, at a temperature 15°C, is used.

TLM method is applied to the cavity with dimensions $a =$ 7 cm and $h = 14.24$ cm, loaded with a homogeneous lossy dielectric sample with thickness $t = 3$ cm, placed on the bottom of the cavity. As an excitation form straight wire conductor loaded in the cavity is used. Excitation probe is placed in the height $l = 7.24$ cm (slightly different from $h/2$) from bottom on the cavity, in the r direction. The probe length is variable in order to investigate the influence of the real excitation presence to the resonant frequencies of the cavity. Obtained TLM results for resonant frequencies in the case of cavity with real excitation are compared with results calculated by using the theoretical approach, that is TLM with impulse excitation. Also, in order to verify TLM method the obtained numerical results of resonant frequencies for TE_{111} and TE_{211} modes in frequency range $f = [1.5 - 3]$ GHz are compared with the experimental ones. Experimental set up for resonant frequencies measurement is shown on the Fig. 1.

[1]Bratislav Milovanović is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Nis, Yugoslavia, E-mail: bata@elfak.ni.ac.yu

[2]Aleksandar Marinčić, SANU member, is with the Faculty of Electrical Engineering, Bulevar Kralja Aleksandra 73, 11000 Beograd, Yugoslavia, E-mail: marmarij@eunet.yu

[3]Jugoslav Joković is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Nis, Yugoslavia, E-mail: jugoslav@elfak.ni.ac.yu

[4]Aleksandar Atanasković is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Nis, Yugoslavia, E-mail: beli@elfak.ni.ac.yu

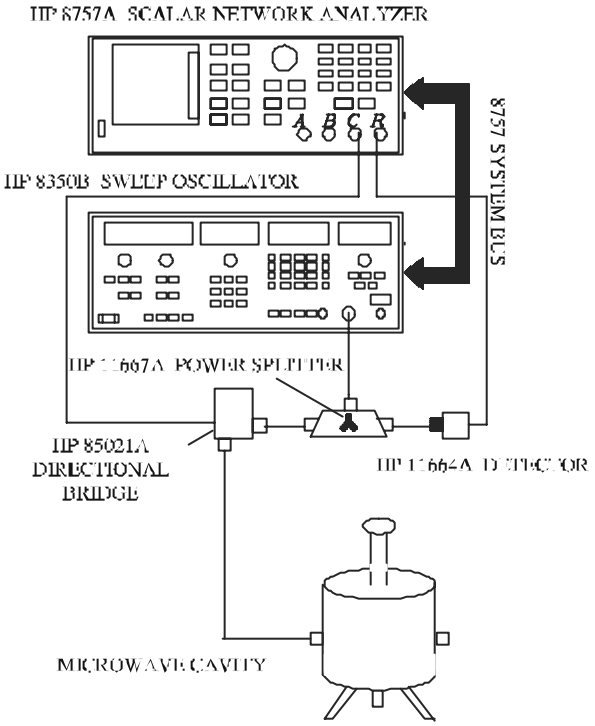


Fig. 1. Experimental set up for resonant frequency of the cylindrical metallic cavity with circular cross-section measurement

II. Problem Modeling

In TLM method, an electromagnetic (EM) field distribution in three dimensions, for a specified mode of oscillation in a microwave cylindrical cavity, is modeled by filling the field space with a network of transmission lines and exciting a particular field component in the mesh by voltage source placed on the excitation probe. EM properties of a medium in the cavity are modeled by using a network of interconnected nodes, a typical structure being the symmetrical condensed node (SCN), which is shown in Fig. 2. To operate at a higher time-step, a hybrid symmetrical condensed node (HSCN) [7] is used. An efficient computational algorithm of scattering properties, based on enforcing continuity of the electric and magnetic fields and conservation of charge and magnetic flux [8] is implemented to speed up the simulation process.

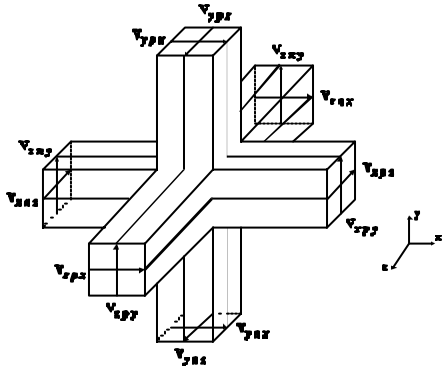


Fig. 2. Symmetrical condensed node

For accurate modeling of this problem, a finer mesh within the dielectric layer and cells with arbitrary aspect ratio suitable for modeling of particular geometrical features, are applied.

Losses can be incorporated in a TLM model by introducing loss stubs into the scattering points (i.e. nodes). The loss stubs may be viewed as infinitely long, or equivalently, as terminated (matched) by their own characteristic impedance. The matched stubs can be used to model both 'electrical' and 'magnetic' losses. In the HSCN, the presence of matched stubs is incorporated directly into the scattering matrix. Given the effective electrical conductivity σ_e, loss 'electrical' element for the 3-D time-domain TLM method is defined as [8]:

$$G_e = \sigma_e f(\Delta x, \Delta y, \Delta z) \tag{1}$$

where: Δx, Δy and Δz are dimensions of TLM node in the x, y and z directions respectively. Complex permittivity is related to effective electrical conductivity as:

$$\varepsilon^* = \varepsilon_0 \varepsilon_r^* = \varepsilon_0 \varepsilon_r' - j\sigma_e/\omega. \tag{2}$$

III. TLM Wire Node

In TLM wire node, wire structures are considered as new elements that increase the capacitance and inductance of the medium in which they are placed. Thus, an appropriate wire network needs to be interposed over the existing TLM network to model the required deficit of electromagnetic parameters of the medium. In order to achieve consistency with the rest of the TLM model, it is most suitable to form wire networks by using TLM link and stub lines (Fig. 3) with characteristic impedances, denoted as Z_{wy} and Z_{wsy}, respectively.

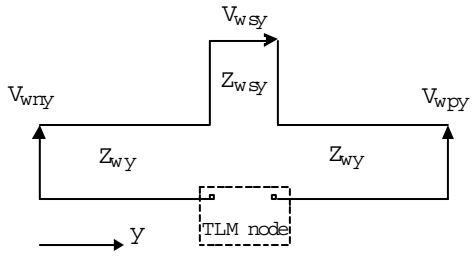


Fig. 3. Wire network

An interface between the wire network and the rest of TLM network must be devised to simulate coupling between the electromagnetic field and the wire. In order to model wire junction and bends, wire network segments pass through the center of the TLM node In that case, coupling between the field and wire coincides with the scattering event in the node which makes the scattering matrix calculation, for the nodes containing a segment of wire network, more complex. Because of that, a simple and elegant approach is developed [6], which solves interfacing between arbitrary complex wire network and arbitrary complex TLM nodes without a modification of the scattering procedure.

IV. Numerical Analysis

The numerical results, which illustrate the effect of the real excitation probe on the resonant frequency, are presented for a cavity with circular cross-section. Dimensions of the investigated cavity are chosen to be $a = 7$ cm and $h = 14.24$ cm, starting from the example from [3]. Cavity is loaded with dielectric sample placed on the bottom of the cavity. The thickness of dielectric layer is $t = 3$ cm. Permittivity of hypothetical lossy homogeneous dielectric sample is equal to that of water at a temperature $15°$C ($\varepsilon_r = 77 - j5$).

For modeling of this cavity non-uniform TLM mesh with $45 \times 45 \times Nz$ nodes was used. The real excitation in form of small straight wire conductor is modeled by using TLM wire node. The excitation probe is placed on the height $l = 7.24$ cm from bottom on the cavity (slightly different from $h/2$), in the r direction (Fig. 4.). In this way, it is possible to excite modes having r-component of the electrical field in the cavity.

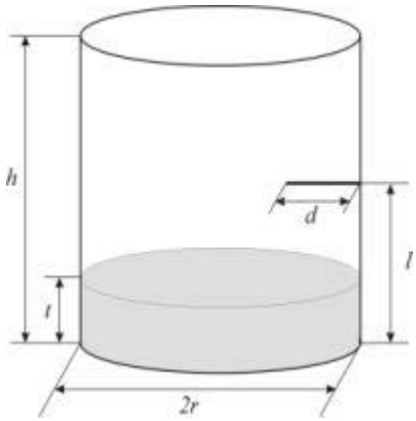


Fig. 4. Real excitation loaded in a metallic cavity with circular cross section ($a = 7$ cm, $h = 14.24$ cm, $t = 3$ cm, $l = 7.24$ cm)

The radius of the excitation probe is $r = 0.5$ mm and length d is variable in order to investigate the influence of the real excitation presence to the resonant frequencies of the cavity. Excitation probe is connected with voltage source $V_{source} = 1$ V, $R_{source} = 50\ \Omega$. The resonant frequencies are determined from the reflection characteristic (S_{11} plot).

The obtained TLM numerical results and experimental results of resonant frequencies for modes in the frequency range $f = [1.5 - 3]$ GHz, versus length of the real excitation probe d, are shown in Table 1. To the aim the of illustrating the agreement between experimental and numerical TLM results and dependence of resonant frequencies of probe length, obtained results are shown in the Fig. 5. The circle symbols indicate the results obtained by using TLM method with real excitation and triangle indicate experimental results. The straight lines present the values of resonant frequencies calculated by using TLM method with impulse excitation. Also, quarter-wavelength curve is presented in order to identify areas of capacitive and inductive character of probe impedance.

As it can be seen from Table 1 and Fig. 5, in comparison

Table 1. The resonant frequencies versus probe length, calculated by using TLM method and experimentally, respectively

d [cm]	Resonant frequencies f_{res} [MHz]							
	TE_{114} Theoretical value = 1809 MHz		TM_{015} Theoretical value = 2151 MHz		TE_{215} Theoretical value = 2463 MHz		TM_{116} Theoretical value = 2928 MHz	
	TLM	exp.	TLM	exp.	TLM	exp.	TLM	exp.
2	1807	1805	2137	2138	2441	2414	2923	2931
3	1789	1776	2136	2128	2571	2504	2953	2952
4	1949	1891	2155	2143	2542	2504	2959	2958
5	1890	1880	2163	2146	2522	2505	2959	2963
6	1891	1875	2160	2145	2532	2504	2959	2965

with results calculated by using theoretical approach where an impulse excitation was used, the obtained TLM numerical results in the case of applying real excitation show a much better agreement with experimental ones, which indicates good TLM modeling of the real excitation probe.

The Fig.5. shows that the values of resonant frequencies for both TE and TM modes considerable depend on the real probe length d. The results calculated by using TLM method and experimental ones, where a probe inside the cavity is used as an excitation, are strongly deviate from the results calculated by using the theoretical approach where an impulse excitation was used to establish desired field distribution in the modeled cavity. In the area of capacitive character of probe impedances ($d < \lambda/4$), due to increasing of wire conductor length the values of resonant frequencies shift to lower frequencies. In inductive area ($d > \lambda/4$) results of resonant frequencies have higher values than in the case applying TLM method with impulse excitation. Also, due to in-

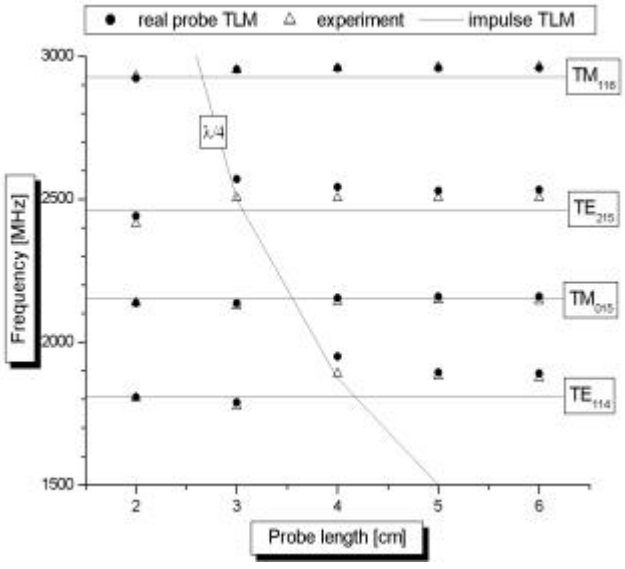


Fig. 5. Resonant frequencies of excited modes in frequency range $f = [1.5 - 3]$ GHz versus probe length

creasing probe length resonant frequencies decrease and tend toward theoretical values.

To the aim of illustrating the good agreement between experimental and numerical TLM result, in the Figs. 6. and 7. are shown S_{11} plots (reflection characteristic) for the probe length $d = 5$ cm, obtained experimentally and by using TLM method, respectively.

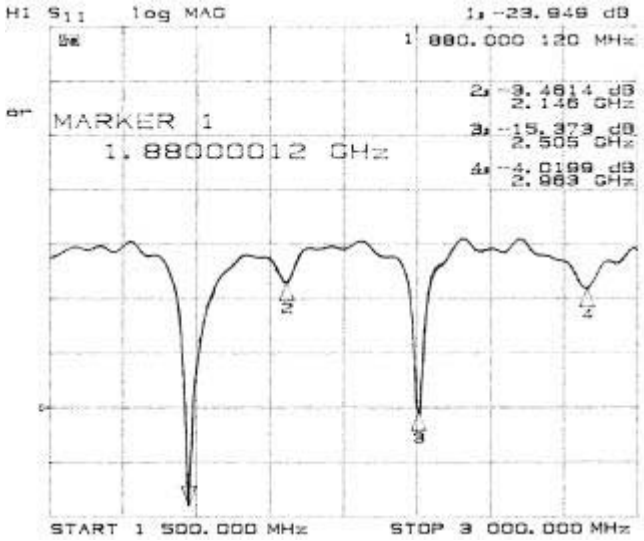


Fig. 6. S_{11} plot in frequency range $f = [1.5 - 3]$ GHz for the probe length $d = 5$ cm, obtained experimentally

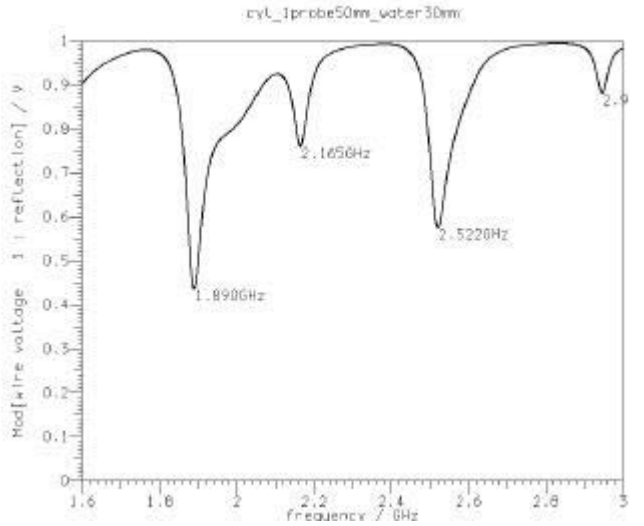


Fig. 7. Voltage reflection in frequency range $f = [1.5 - 3]$ GHz for the probe length $d = 5$ cm, obtained by using TLM method

V. Conclusion

In this paper, real excitation probe in a loaded cylindrical metallic cavity is modeled by using TLM method and influence of real probe presence to the resonant frequencies is analyzed. TLM numerical technique has been implemented in the appropriate software and applied to the problem of determining resonant frequencies as important information in the microwave applicator design.

In comparison with results calculated by using the theoretical approach where an impulse excitation was used, the obtained TLM numerical results in the case of real excitation show a much better agreement with experimental ones, which indicates good TLM modeling of the real excitation.

Also, the influence of the probe length to the resonant frequencies of modes in the frequency range $f = [1.5 - 3]$ GHz are investigated. The obtained results where a probe inside the cavity is used as an excitation show that values of resonant frequencies depend on length of wire conductor. This dependence is related with character of probe impedances.

In this paper, for the first time, real excitation in a loaded cylindrical metallic cavity with lossy dielectric sample is modeled by using TLM method. According to previously showed results a general conclusion can be derived that TLM approach gives valid result. Therefore it is expected that these resonant structures can be successfully modeled by TLM method, independently of probe position and dimensions and location of dielectric sample in the cavity.

References

[1] A. Baysar, J.L. Kuester and S. El-Ghazaly, "Theoretical and Experimental Investigations of Resonance Frequencies in a Microwave Heated Fluidized Bed Reactor", *IEEE MTT-S Digest*, pp.1573-1576, 1992.

[2] S. Ivkovic, B. Milovanovic, A. Marincic and N. Doncov, "Theoretical and Experimental Investigations of Resonance Frequencies in Loaded Cylindrical Microwave Cavity", *Proc. of the Third IEEE TELSIKS'97 Conference*, Niš, Yugoslavia, 306-309, 1997.

[3] B. Milovanovic, N. Doncov, V. Trenkic and V. Nikolic, "3-D TLM Modelling of the Circular Cylindrical Cavity Loaded by Lossy Dielectric Sample of Various Geometric Shapes", *Proc. of the Third International Workshop on TLM Modelling*, Nica, France 187-195, 1997.

[4] N. Doncov, "Microwave structures analysis using 3-D TLM method", M.Sc thesis, Faculty of Electronic Engineering, University of Nis, 1999.

[5] B. Milovanovic, N. Doncov, A. Atanaskovic, "Tunnel Type Microwave Applicator Modelling using TLM Method", *Problems in Modern Applied Mathematics, A Series of Reference Books and Textbooks: Mathematics and Computers in Science and Engineering*, WSES Press, pp.327-332, 2000.

[6] V. Trenkic, A.J. Wlodarczyk, R.A. Scaramuzza, "Modelling of Coupling Between Transient Electromagnetic Field and Complex Wire Structures", *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, Vol.12, No.4, pp.257-273, 1999.

[7] C. Christopoulos, "The Transmission-Line Modelling Method", *IEE/OUP Press*, 1995.

[8] V. Trenkic, "The Development and Characterization of Advanced Nodes for TLM Method", Ph.D. Thesis, University of Nottingham, 1995.

Implementation of Microwave Transistor Neural Noise Models into Standard Microwave Simulators

Vera Marković, Zlatica Marinković[1]

Abstract – **Authors of this paper have proposed efficient transistor noise models based on neural networks. In this paper, implementation of proposed neural noise models into standard microwave simulators, such as Libra or ADS is presented. For a specified bias point, a microwave transistor can be described within the simulator by an $s2p$ file containing transistor signal and noise data. Using proposed neural models such $s2p$ file can be generated and assigned to the two-port circuit representing the transistor.**

Keywords – **Neural network, microwave transistor, noise parameters, S-parameters, microwave simulator**

I. Introduction

In the last decade, neural networks have been widely applied in many areas such as pattern recognition, speech and image processing, control process applications, etc. Neural network based methods belong to the class of so-called "black box" methods since only sources and responses are considered [1]. Since a neural network itself has no any knowledge about the problem to be solved, it has to be trained. In the training process, known input-output combinations related to the problem are presented to the neural network, and network parameters are adjusted. Once trained, neural network is expected to generate correct responses for all inputs even they are presented to the network during the training or not. Neural network ability to give correct response for the inputs not included in the training set, called generalization, is the strongest motive for the further research activities in the filed of neural networks.

As highly nonlinear structures, neural networks are able to model nonlinear relations between different data sets. Owing to this ability, they have been applied in a wide area of problems. Especially, they are interesting for problems not fully mathematically described. Once trained they can predict response with quite a good accuracy, even for input values not presented in the training process, without changes in their structure and without additional knowledge of considered problem. Neural models are simpler than physically based ones but retain the similar accuracy. They require less time for response providing; therefore using of neural models can make simulation and optimization processes less time-consuming, shifting computation efforts from on-line optimization to off-line training.

Recently, neural networks have been applied in the microwave area [2]. Neural models of passive components are presented in [3-5]. There are some neural models that refer to the important parts of many modern communication systems – microwave transistors. Very good results are obtaining in noise modeling of microwave transistors, considering transistor bias condition as well, [6-9]. In this paper, implementation of the developed neural models into the standard microwave simulators, such as Libra or ADS, will be presented.

[1]Vera Marković and Zlatica Marinković are with the Faculty of Electronic Engineering, Beogradska 14, 18 000 Niš, Serbia and Montenegro e-mail: [vera,zlatica]@elfak.ni.ac.yu

II. Transistor Signal And Noise Characteristics

Microwave transistors are usually represented as a two-port circuit characterized by its scattering ($[S]$) matrix, that contains four complex scattering parameters, $S_{ij}, i, j = 1, 2$ described by their magnitudes and phases. It is common that the manufacturers provide S-parameters' data at certain number of frequencies from the specified frequency range, where these data are related to one or a few bias conditions. S-parameters define so-called signal performance of the transistor.

In addition to the signal performance, transistor noise performance is of a great importance for the low noise applications. Any two-port noisy component can be characterized by a noise figure F, which is a measure of the degradation of the signal-to-noise ratio between input and output of the component, [10], and can be expressed as

$$F = F_{\min} + \frac{4R_n|\Gamma_g - \Gamma_{opt}|^2}{Z_0(-|\Gamma_g|^2)|1 + \Gamma_{opt}|^2}, \qquad (1)$$

where $F_{\min}$ is a minimum noise figure, R_n is an equivalent noise resistance, Γ_{opt} is the optimum reflection coefficient, and finally, Z_0 is normalizing impedance. The optimum reflection coefficient refers to the optimum source impedance that results in minimum noise figure, $F = F_{\min}$. The set of four noise parameters: $F_{\min}$, magnitude and angle of Γ_{opt} and R_n describe inherent behavior of the component, independent of a connected circuit and are not a direct or physics-based representation of the noise produced by the device, but play an important role in describing the performance of the noise figure as a function of the generator reflection coefficient.

III. Multilayer Perceptron (MLP) Neural Network

The basic idea of neural network application in microwave transistor noise modeling is developing of appropriate noise

models that can predict transistor noise parameters accurately in a wide frequency range for all bias points from the operating range. As a first step, transistor noise parameters dependence on biases and frequency is modeled using multilayer perceptron network - MLP. A standard MLP neural network is shown in Fig. 1. [1].

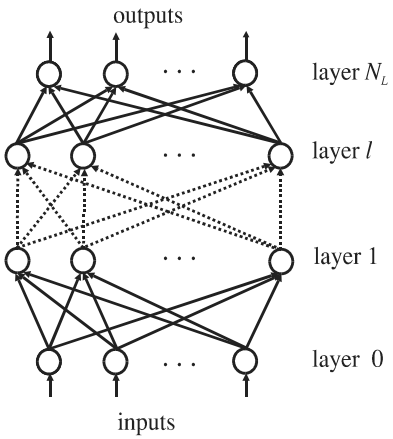


Fig. 1. MLP neural network

This network consists of neurons (circles) grouped into the layers. The input signal is presented to the neurons from the input layer. Each neuron from one layer is connected to all neurons from the next layer. The output layer neurons represent outputs of the network. The layers not directly connected to the outside environment are hidden layers. Neurons are characterized by their activation functions. Here, a linear function for input and output layer and a sigmoid function for hidden layers are chosen. The connections between neurons are characterized by weighting factors.

Input vectors are presented to the input layer and fed through the network that then yields the output vector. Network training is a process of adjusting of network parameters (activation function thresholds and connection weights) in order to minimize the difference between a network response and reference values. This process is iterative and it proceeds until errors are lower than the prescribed goals or until the maximum number of epochs (epoch – the whole training set processing) is reached. Here, for training purposes, *Levenberg-Marquardt* algorithm (a modification of "backpropagation" algorithm) is used.

IV. Transistor Modeling Using Neural Networks

MLP networks are applied with the aim to model the HEMT transistor noise parameters dependence on frequency and bias conditions (dc drain-to-source voltage and dc drain-to-source current). The used MLP network structure has four layers (i.e. two hidden layers). There are three neurons in the input layer:

- dc drain-to-source voltage V_{dc},
- dc drain-to-source current $I + dc$ and
- frequency f.

The output layer consists of four neurons corresponding to:

- minimum noise figure,
- magnitude of optimum reflection coefficient,
- angle of optimum reflection coefficient and
- normalized equivalent noise resistance (50 Ω normalizing impedance).

This approach is presented in Fig. 2 and denoted by "bf" (the mark is related to the network inputs – bias and frequency).

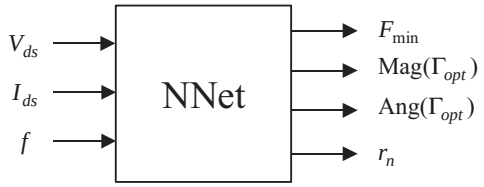


Fig. 2. Neural model for noise parameters dependence on bias conditions and frequency (bf approach)

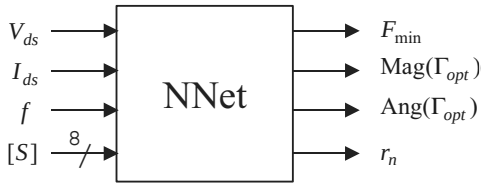


Fig. 3. Neural model for noise parameters dependence on bias conditions, frequency (sbf approach)

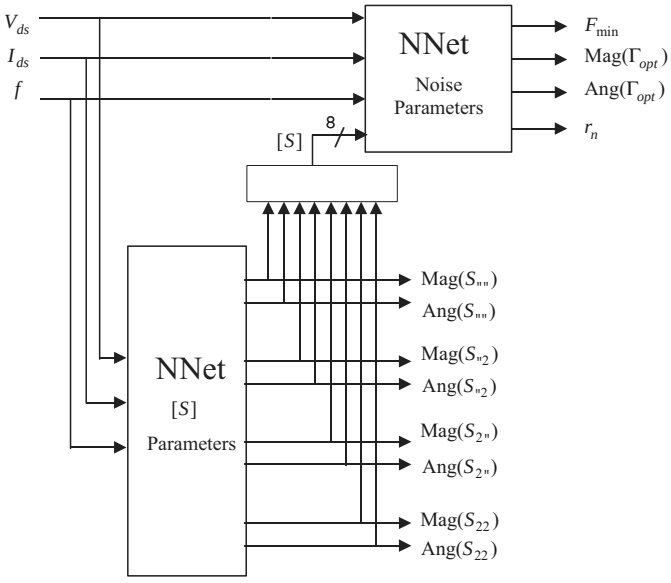


Fig. 4. Neural model for accurate noise parameters prediction

Obtained models are able to predict noise parameters for a given bias point even in the case of the bias point not presented in the training process, without additional computation or change in the network structure.

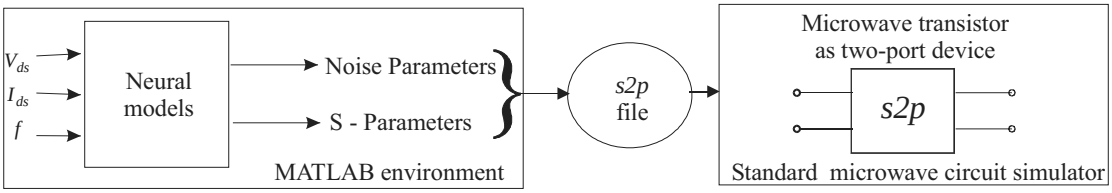


Fig. 5. $s2p$ file generating using transistor neural noise models

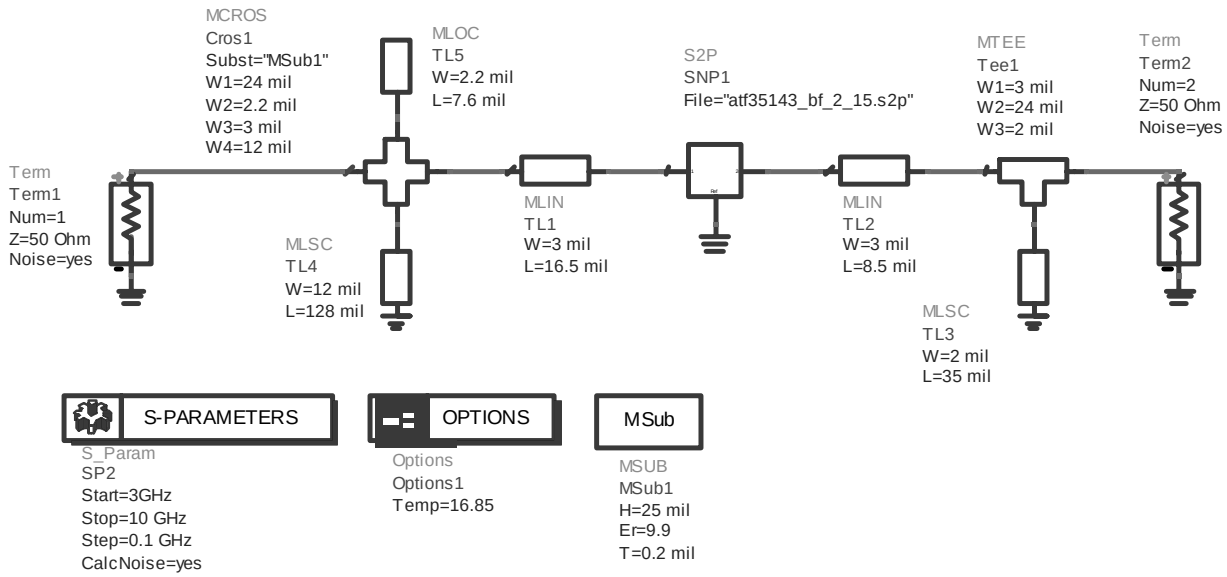


Fig. 6. One-stage microwave amplifier

Further, in order to improve the modeling, transistor scattering parameters are introduced as additional inputs of the neural network, as it is shown in Fig. 3 [6].

Although S-parameters easier to be measured than noise parameters much time can be saved using neural models of S-parameters as well. At that way, all noise parameters can be predicted with high accuracy without additional measuring of S-parameters or their determination by simulation. This approach is presented in Fig. 4 [9].

V. Neural Model Implementation Into The Standard Circuit Simulator

In the microwave community several powerful software packages for analyzing, optimization and design of microwave circuits are popular and widely used, like Libra [11], ADS [12], etc. These software packages are often called microwave simulators. Within microwave simulators, one transistor can be represented as a two-port circuit described by using so-called $s2p$ file. This file contains table values of S-parameters and, in the case of low-noise transistors, the noise parameters as well. $s2p$ file is formed according specified syntax. The data is related to one specified combination of biases, and organized in the following way: after text header there are magnitudes and arguments of four S-parameters (S_{11}, S_{12}, S_{21} and S_{22}) at a number of frequency points. These data are followed by noise parameters' values: minimum noise figure $F_{\min}$, magnitude and angle of optimum reflection coefficient Γ_{opt} and, finally, normalized equivalent noise resistance (where the normalization resistance value is 50 Ω). The noise parameters are given at either the same frequencies or the different frequencies as S-parameters are given at.

The basic idea is forming the $s2p$ file according to the existing syntax using data generated by the transistor neural models of S- and noise parameters for the specified bias conditions. Then, this file is assigned to the two-pot circuit that describes the transistor in the microwave simulator.

In this way, extraction of elements of transistor equivalent circuit and/or model parameters, that is necessary in the existing transistor signal and noise models, is avoided. Here, on-line optimization is shifted in off-line training of neural networks. Once trained, neural models provide signal and noise data prediction practically instantaneously. Therefore, it should compute neural models' responses for specified bias conditions and form $s2p$ file, Fig. 5. This could be done efficiently in the environment used for neural network training, like MATLAB program package environment used here.

VI. Modeling Example

The procedure described above can be illustrated by an example of implementation of transistor neural models in ADS, the standard software tool for microwave design. Noise performance analysis for a microwave amplifier realized as hy-

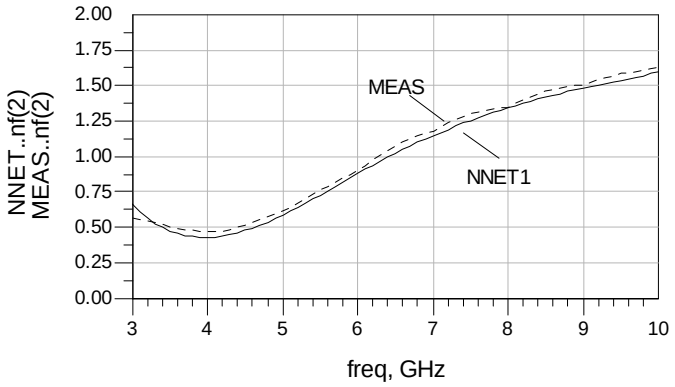


Fig. 7. Noise figure of the microwave amplifier vs. frequency

brid microwave circuit is performed. Schematic design of this amplifier in ADS simulator is shown in Fig. 6. This is a simple one-stage amplifier consisting of active component and input and output matching network realized in microstrip technology. As active component, a microwave pHEMT transistor *Hewlett Packard* ATF35143 is used. First, appropriate signal and noise models for this transistor have been developed, [9]. The models are trained using the data from the manufacturer's catalogue over the (0.5-10) GHz frequency range and for different bias conditions. The both of developed neural models, sp_10_10 model for S-parameters and bf_10_10 model for noise parameters, have ten neurons in each of two hidden layers. Using data generated at the different bias conditions, standard $s2p$ files corresponding to these bias conditions are formed. Further, within ADS, these $s2p$ files can be simply assigned to two-port circuit corresponding the transistor in the amplifier circuit, and amplifier performance analysis (including noise) could be done. Also, optimization of matching networks could be performed if necessary.

In order to verify the proposed approach, not only $s2p$ files generated using neural models but also corresponding $s2p$ files given by the manufacturer are used, and the amplifier analyses simulation results are compared.

It is very interesting to compare results for transistor biases different form the biases used for the training of neural models. As an illustration, Fig. 7 shows noise figure at the amplifier output at the bias point (2 V, 15 mA). The data related to this bias point were not used in the training of neural models. Noise figure obtained in the case of $s2p$ file generated using transistor neural models is denoted by solid line (NNET), and in the case of $s2p$ file given by the manufacturer by doted line (MEAS). It could be seen that these two lines are very close, verifying the proposed procedure.

VII. Conclusion

Using available data of S-parameters and noise parameters of a microwave transistor in operating frequency range and for different bias conditions, neural networks that model dependences of these parameters on biases and frequency could be trained. The trained neural models can predict S- and noise parameters in the whole operating range without their changes. This is their main advantage comparing to most of the existing transistor signal and noise models, which are valid for only one bias condition. Within microwave simulator a microwave transistor can be represented as two-port circuit with assigned file containing information about S- and noise parameters for several frequencies. Since this file can be obtain very easily for any transistor bias conditions using transistor neural models, the proposed approach provides very efficiently representing of the transistor in the whole operating range.

Acknowledgement

This work was supported in part by Ministry of Science, Technologies and Development of Republic of Serbia, under project 1351: "*The development of physically based noise models for micro- and millimeter wave range*".

References

[1] S. Haykin, *Neural networks*, New York, IEEE, 1994.

[2] Q. J. Zhang, K. C. Gupta, *Neural Networks for RF and Microwave Design*, Artech House, 2000.

[3] F. Wang, Q. J. Zhang, "Knowledge-Based Neural Models for Microwave Design", *IEEE Trans., Microwave Theory Tech.*, vol. 45, no. 12, 1997, pp. 2333-2343.

[4] P. M. Watson, K. C. Gupta, "Design and optimization of CWP circuits using EM-ANN models for CPW components", *IEEE Trans., Microwave Theory Tech.*, vol. 45, no. 12 pp. 2515-2523, 1997.

[5] B. Milovanovic, Z. Stankovic, S. Ivkovic, "Modelling of the Cylindrical Metallic Cavity with Circular Cross-section using Neural Networks", *MELECON'2000 Conference Proceedings*, Vol.II, Cyprus 2000, pp. 449-452.

[6] V. Marković, Z. Marinković, "Neural Models of Microwave Transistor Noise Parameters Based on Bias Conditions and S-parameters ", *Proceedings of the Conference TELSIKS 2001*, Nis, September 2001, pp. 683-686.

[7] O. Pronić, V. Marković, Z. Marinković, "Noise Modeling of Packaged HEMTs by using neural model of Noise Wave Temperatures", *11th Conference on Microwave Technique COMITE 2001*, September 18-19, Paradubice, Czech Republic, pp. 163-166

[8] B. Milovanović, V. Marković, Z. Marinković, Z. Stanković, "Microwave circuits modeling using neural networks – overview of the results achieved at the Faculty of Electronic Engineering in Nis" – invited paper, *Neurel 2002*, Beograd, Septembar 2002, pp. 177-184

[9] V. Marković, Z. Marinković, "Signal and Noise Neural Models of pHEMTs", *NEUREL 2002*, Belgrade, 2002, pp. 177-184

[10] D. Pozar, *Microwave Engineering*, J. Wiley & Sons, Inc., 1998.

[11] Touchstone and Libra Users Manual, EEsof, Inc. 1990.

[12] "Advanced Design Systems-version 1.5", Agilent EEsof EDA, 2000.

Microwave Transistors Noise Modeling Using Noise Wave Temperatures

Olivera R. Pronić, Vera V. Marković

Abstract – **In this paper we propose the noise models of MESFET / HEMT transistors based on the wave representation of transistor intrinsic circuit. The noise wave temperatures are introduced as empirical parameters of the models. Besides the noise model based on the constant noise wave temperatures, the noise model based on the frequency dependent noise wave temperatures is also developed and the comparative analysis is done. These frequency dependences are modeled using second order polynomial regression. The results for transistor noise parameters obtained by the proposed procedures are verified by the comparison to experimental data.**

Keywords – **MESFET, HEMT, noise model, wave approach**

I. Introduction

The noise models of microwave transistors are based mostly on the well-known fact that a linear noisy two-port may be represented by a noiseless two-port and two additional noise sources, [1]. These noise sources are usually equivalent voltage and/or current sources.

In the last decade the Pospieszalski's approach to noise modeling of MESFETs / HEMTs has gained much attention in microwave community, [2]. The noise model he proposed is based on H representation of transistor intrinsic circuit with two uncorrelated noise sources, the voltage noise source at the gate side and the current noise source at the drain side. However, it has been found, [3], that in some cases the inaccuracy in transistor noise modeling caused by this assumption is not negligible. Therefore, the model including the correlation between noise sources has been developed and implemented into a standard microwave circuit simulator, [4].

In the microwave frequency region, a treatment of noise in terms of waves is more appropriate since it allows the use of scattering parameters for noise computations, [5]. It has been shown, [6], that the wave approach is useful for both noise modeling and measurement of microwave FETs. Using a similar approach, the new extraction formulas for the noise wave sources in the noise equivalent circuit of MESFETs / HEMTs, where the correlation between noise sources is included, are proposed in [7]. The noise parameter characteristics obtained by using that procedure are in better agreement with the measurements than the existing model, [6].

The noise wave modeling procedures of MESFETs, HEMTs and dual-gate MESFETs based on T representation of transistor intrinsic circuit are presented by the authors in previous papers, [8], [9]. Three noise wave temperatures are introduced as empirical parameters of those models. These temperatures, being constant over the whole frequency range, are obtained on the basis of some experimental noise data by applying standard optimization procedures.

However, it is shown that the noise wave temperatures are frequency dependent. Therefore, the microwave FETs' noise model based on variable noise wave temperatures will be also presented here. Thus, two different procedures for the noise wave modeling of packaged MESFETs / HEMTs are considered and compared in this paper. The verification of presented procedures is done by comparison with measured data.

II. Noise Modeling

We used a MESFET / HEMT small-signal equivalent circuit as shown in Fig. 1. This equivalent circuit represents the packaged devices very well.

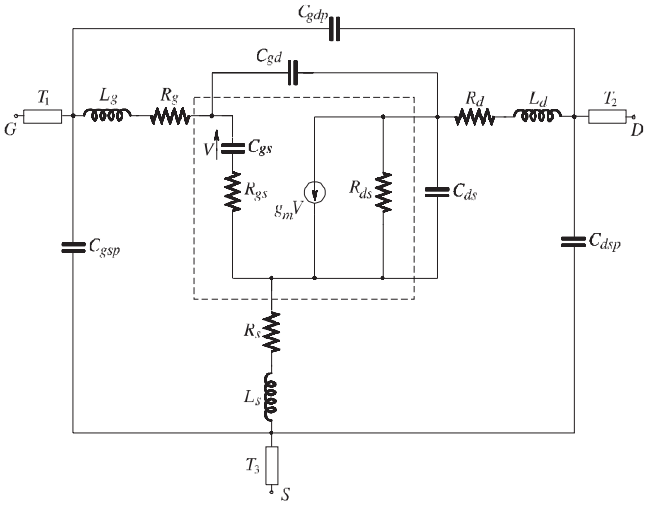


Fig. 1. A small-signal equivalent circuit for a MESFET / HEMT package ($g_m = g_{m0}e^{-j\omega\tau}$)

T representation of the intrinsic circuit with two correlated noise sources is considered. The intrinsic part of the circuit (denoted by a dashed line in Fig. 1), can be represented by a noiseless two-port defined by transfer scattering parameters, $[T]$, and two noise wave sources a_n and b_n referred to the input, as shown in Fig. 2.

The linear matrix equation describing this noisy two-port

[1]Olivera R. Pronić and Vera V. Marković are with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mails: oljap, vera@elfak.ni.ac.yu

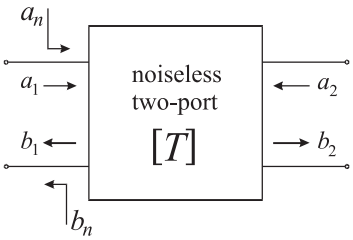


Fig. 2. T representation of a noisy two-port

is:

$$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} b_2 \\ a_2 \end{bmatrix} + \begin{bmatrix} a_n \\ b_n \end{bmatrix}, \quad (1)$$

where a_i and b_i, $i = 1, 2$, are incident and output waves at the i-th port.

The elements of the noise source vector are correlated and characterized by the correlation matrix C_T given by

$$C_T = \begin{bmatrix} \langle |a_n|^2 \rangle & \langle -a_n b_n^* \rangle \\ \langle -b_n a_n^* \rangle & \langle |b_n|^2 \rangle \end{bmatrix} \quad (2)$$

where the brackets $\langle\rangle$ indicate time average of the quantity inside and * indicates complex conjugation.

It is very convenient to use the noise wave temperatures as empirical noise model parameters, [10]. In this way, the correlation matrix C_T can be expressed by

$$C_T = k \triangle f \begin{bmatrix} T_a & |T_c| e^{j\varphi_c} \\ |T_c| e^{-j\varphi_c} & T_b \end{bmatrix}, \quad (3)$$

where k is the Boltzmann's constant and $\triangle f$ is the noise bandwidth (it is assumed that $\triangle f$=1Hz). In this way the noise performance of a two-port network is completely characterized by two real temperatures T_a and T_b and a complex correlation temperature $T_c = |T_c| e^{j\omega\tau_c}$.

Using this representation of noise, the noise wave temperatures can be expressed in term of intrinsic circuit noise parameters - minimum noise figure F_{min}, optimum reflection coefficient $\Gamma_{opt} = |\Gamma_{opt}| e^{j\varphi_{opt}}$ and noise resistance R_n, as

$$T_a = T_0 (F_{min} - 1) + \frac{4 R_n T_0 |\Gamma_{opt}|^2}{Z_0 |1 + \Gamma_{opt}|^2}, \quad (4)$$

$$T_b = \frac{4 R_n T_0}{Z_0 |1 + \Gamma_{opt}|^2} - T_0 (F_{min} - 1), \quad (5)$$

$$T_c = \frac{4 R_n T_0 \Gamma_{opt}}{Z_0 |1 + \Gamma_{opt}|^2}, \quad (6)$$

where Z_0 is the normalization impedance (Z_0=50Ω) and T_0 is standard reference temperature (T_0=290K).

III. Numerical Example

The numerical results presented in this paper are related to small-signal and noise modeling of Siemens HEMT packaged device, type CFY65A. All simulations are performed with microwave circuit simulator Libra, [11].

The parameters of the proposed MESFET / HEMT noise models are three noise wave temperatures. Generally, besides S parameter data, the noise parameters measured at several frequency points are needed to predict the transistor noise parameters over the whole operating frequency range.

The noise modeling procedures are performed in the following way:

At the beginning, the small-signal equivalent circuit elements are extracted from the scattering parameters measurements. The extracted values are given in Table 1. The transmission lines segments (T_1, T_2 and T_3) are characterized by the characteristic impedances (Z_1, Z_2 and Z_3) and electrical lengths (e_1, e_2 and e_3) at frequency f=10 GHz.

Table 1. Equivalent circuit element values

intrinsic circuit elements	*values*	*parasitics*	*values*
C_{gs} (pF)	0.176	C_{ds} (pF)	0.046
R_{gs} (Ω)	3.29	C_{gd} (pF)	0.028
g_m (mS)	37.4	R_g (Ω)	0.75
τ (ps)	2.82	L_g (nH)	0.112
R_{ds} (Ω)	591.95	R_d (Ω)	0.52
		L_d (nH)	0.262
		R_s (Ω)	0.34
		L_s (nH)	0.001
		C_{gdp} (pF)	0.011
		C_{gsp} (pF)	0.01
		C_{dsp} (pF)	0.099
		Z_1 (Ω)	51.21
		Z_2 (Ω)	68.62
		Z_3 (Ω)	31.8
		e_1 (°)	35.78
		e_2 (°)	24.49
		e_3 (°)	14.1

In order to determine all model parameters, after the extraction of the small-signal equivalent circuit elements, it is also necessary to determine the noise temperatures.

The intrinsic circuit noise parameters needed for the noise wave temperatures calculation can be obtained by applying the deembeding procedure in Libra. The deembedding of device parasitics is done by adding parasitic elements with negative values and in reverse order to the complete circuit. In that way we got the noise parameters of an intrinsic circuit. After that the noise wave temperatures could be calculated by applying Eqs. (4)-(6).

Since the calculated values of the noise temperatures vary with the frequency, it is useful to model this frequency dependence by some mathematical relationships. Here, a polynomial regression model of the second order is stated to model

the frequency dependence of the noise wave temperatures,

$$T_i = a_i + b_i f + c_i f^2, \quad i = a, b, c\,. \tag{7}$$

The parameters for polynomial fit of the noise wave temperatures are given in Table 2.

Table 2. Parameters for polynomial fit

	a	b	c
T_a	-3.69	31.67	-1.70
T_b	-214.95	111.03	-6.79
$\|T_c\|$	-78.19	60.26	-3.68
φ_c	4.88	8.52	-0.25

The noise wave temperatures could be converted to standard noise parameters with the aim to perform a comparison with the measured data. Equations for the conversion between these parameter sets are:

$$\Gamma_{opt} = \left(\frac{T_a + T_b}{2|T_c|} - \sqrt{\left(\frac{T_a + T_b}{2|T_c|} \right)^2 - 1} \right) e^{j\omega\tau_c}\,, \tag{8}$$

$$R_n = Z_0 \frac{|T_c|}{4T_0|\Gamma_{opt}|} \left[1 + 2|\Gamma_{opt}| \cos\phi_{opt} + |\Gamma_{opt}|^2 \right]\,, \tag{9}$$

$$F_{min} = 1 + \frac{T_a - T_b}{2|T_0|} + \frac{1}{2T_0} \sqrt{(T_a + T_b)^2 - 4|T_c|^2}\,. \tag{10}$$

The frequency dependence of standard noise parameters of the intrinsic circuit is obtained by replacing the parameters for polynomial fit of the noise wave temperatures in Eqs. (8)-(10). In order to obtain the noise parameters of the complete transistor model it is necessary to include all remaining elements of the equivalent circuit that represent the parasitics (as shown in Fig. 1).

The second approach is based on approximation that the noise wave temperature values are constant over the whole frequency range. The constant noise wave temperatures can be extracted by using the optimization capabilities of a powerful microwave circuit simulators like Libra, ADS, etc, in the following way: First, the expressions for the intrinsic circuit noise parameters (Eqs. (8)-(10)) are programmed using the "equation" capability of the circuit simulator and assigned to the intrinsic circuit by the corresponding statement. After that, all parasitics are connected and the topology of the entire transistor is described. Finally, all small-signal circuit elements and the noise wave temperatures are optimized with the aim that the complete model fits the measured S parameters and noise parameters as well as possible.

The frequency dependences of the transistor noise parameters obtained by the proposed models are presented in Figs. 3-6. The characteristics for the minimum noise figure are presented in Fig. 3. Real and imaginary parts of the optimum reflection coefficient are shown in Fig. 4 and Fig. 5, respectively. Finally, equivalent noise resistance, normalized with respect to 50 Ω, is presented in Fig. 6.

The curves obtained by using the first approach - fitting by the polynomial regression - are denoted by MOD1. The

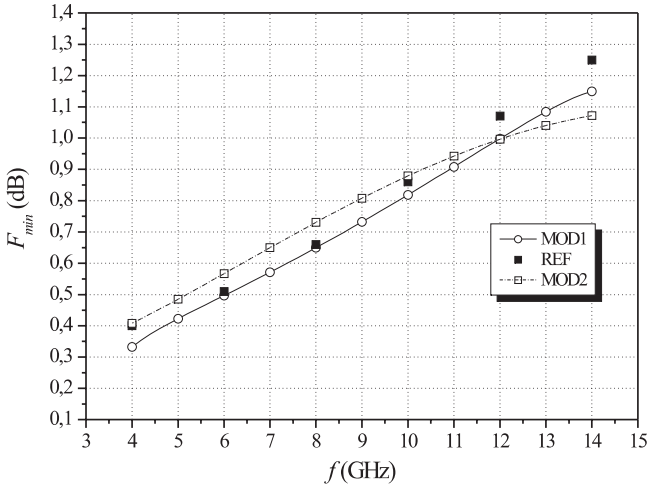


Fig. 3. Minimum noise figure

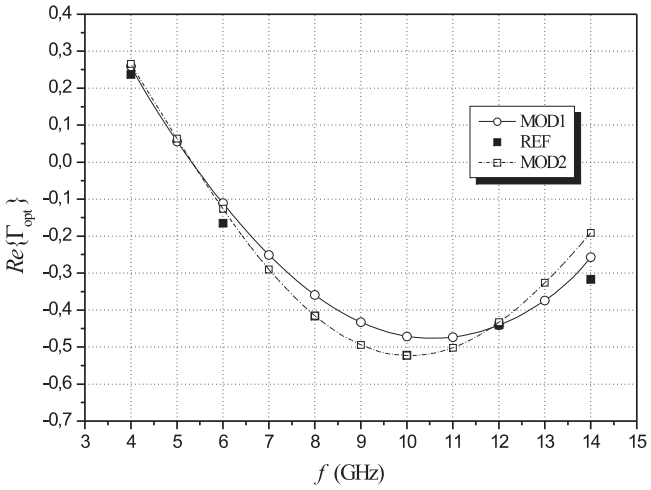


Fig. 4. Real part of the optimum reflection coefficient

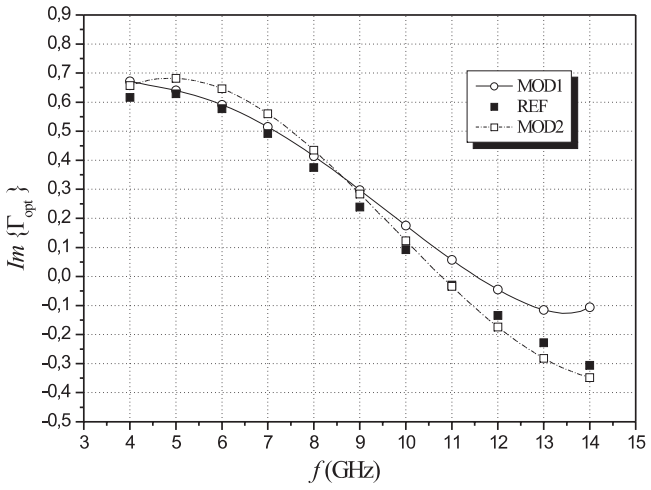


Fig. 5. Imaginary part of the optimum reflection coefficient

characteristics based on the second approach - using of constant noise wave temperatures and optimization procedure, are denoted by MOD2. The referent values, based on the data measured by manufacturer, are denoted by REF.

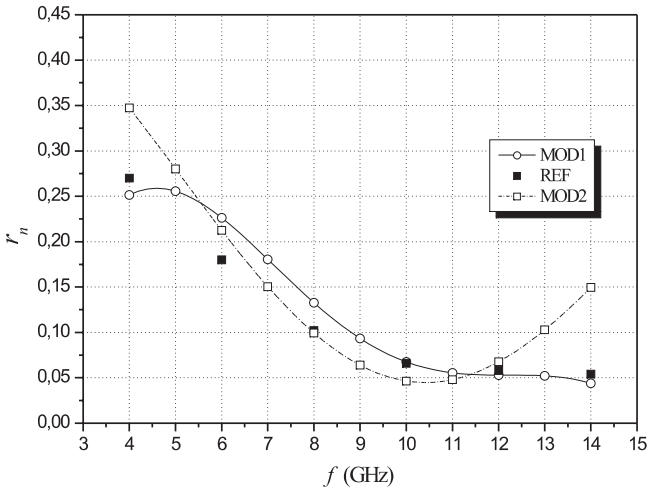


Fig. 6. Normalized equivalent noise resistance

Generally, a good agreement between measured and modeled parameters can be observed in both cases. However, as expected, the better modeling of referent data is achieved when the first approach, MOD1, is applied (especially for the noise resistance).

Several other microwave FETs in packaged form have been analyzed and modeled with similar results.

IV. Conclusion

The modeling of packaged MESFET / HEMT devices presented in this paper is based on a noise wave representation of transistor intrinsic circuit. Two noise models, one based on the frequency dependent, and the other based on the constant noise wave temperatures are presented. The frequency dependences of noise wave temperatures are modeled using second order polynomial regression. In that way frequency extrapolation of noise parameters is enabled and on-line optimization in circuit simulator is omitted. The example of packed HEMT noise modeling is presented. A good agreement with the measured noise parameters is observed in both cases, but the model including frequency dependant noise temperatures enables at some degree more accurately prediction of noise parameters, in comparison to the model with constant noise wave temperatures.

References

[1] J. A. Dobrowolski, *Introduction to Computer Methods for Microwave Circuit Analysis and Design*, London, Artech House, 1991.

[2] M. W. Pospieszalski, "Modeling of noise parameters of MESFET's and MODFET's and their frequency and temperature dependence", *IEEE Trans. Microwave Theory Tech.*, vol.MTT-37, pp. 1340-1350, September 1989.

[3] J.H. Han, "A New Extraction Method for Noise Sources and Correlation Coefficient in MESFET", *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-44, No. 3, pp. 487-490, March 1996.

[4] V. Marković, B. Milovanović, N. Maleš-Ilić, "MESFET Noise Modeling Based on Three Equivalent Temperatures", *Conference Proceedings EUMC'97*, pp. 966-971, Jerusalem, Israel, September 1997.

[5] R. P. Hecken, "Analysis of liner noisy two-ports using scattering waves", *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29, pp. 997-1004, October 1981.

[6] S.W. Wedge and D.B. Rutledge, "Wave techniques for noise modeling and measurement", *IEEE Trans. Microwave Theory Tech.*, vol.MTT-40, pp. 2004-2012, November 1992.

[7] O. Pronić, V. Marković, N. Maleš-Ilić: "The wave approach to noise modeling of microwave transistors by including the correlation effect", *Microwave and Optical Technology Letters*, Vol.28, Issue 6, pp. 426-430, March 2001.

[8] O. Pronić, V. Marković, N. Maleš-Ilić: "MESFET Noise Modeling Based on Noise Wave Temperatures", *TELSIKS'99, Conference Proceedings*, Vol.2, pp. 407-410, Niš, Yugoslavia, 1999.

[9] O. Pronić, V. Marković, "A wave approach to signal and noise modeling of dual-gate MESFET", *AEÜ- Archiv für Elektronik und Übertragungstechnik (International Journal of Electronics and Communications)*, Vol.56, No.1, pp. 61-64, January 2002.

[10] R.P. Meys, "A wave approach to the noise properties of linear microwave devices", *IEEE Trans. Microwave Theory Tech.*, vol.MTT-26, pp. 34-37, January 1978.

[11] *Touchstone and Libra Users Manual*, EEsof, Inc. 1990.

GENESYS-Compatible Models for Two-Dimensional Circuit Analysis in Frequency Domain

Miodrag V. Gmitrović[1], Saša M. Gmitrović and Biljana P. Stojanović[1]

Abstract – **In this paper new multi-port models convenient for two-dimensional ($2D$) electrical circuit analysis are suggested. These models are implemented in the known program *GENESYS*. Microwave transmission lines can be efficiently analyzed by the suggested procedure. A microstrip lowpass filter is analyzed. The results obtained by $2D$ analysis are compared with the ones obtained by $1D$ analysis and $3D$ electromagnetic analysis.**

I. Introduction

Electrical circuits containing microwave lines can be modelled by one-dimensional ($1D$) circuits composed of transmission lines. Analysis of such circuits in frequency domain is simple and very fast. Sometimes, the results of analysis differ very much from the desired exact results. This is happening due to appearance of discontinuities that are results of different physical dimensions and shapes of line. Because of that, more exact, but also more complex methods, based on two-dimensional ($2D$) and three-dimensional ($3D$) approaches to the analysis of circuits containing microwave lines, are used. For that purpose, a number of approaches is developed, and some of them are given in the papers [1-7]. Also, a number of software packages is developed, such as *ADS* [9], *GENESYS* [10-11], *FAMIL* [3-4], and etc.

DC and frequency analyses of complex circuits can be done with software package *GENESYS* [10-12]. It unites many simulators, such as synthesis of active and passive filters, fast linear $1D$ analysis of active and passive circuits, $3D$ electromagnetic analysis, nonlinear harmonic analysis, spectral analysis of systems, drawing of circuit schemes and layouts, and etc. Its platform is projected in the manner that different modules can work together, it can analyse many different circuits at the same time, and it can easily count different output parameters and clearly show them on different graphics.

An approach to the analysis of $2D$ circuits in the software package *GENESYS*, based on new models for multi-port networks, is proposed in this paper. Existent models of multi-port networks are here modified, so they can be simple and uncomplicated used for analysis of $2D$ electrical circuits. The approach has general character, and the verification is done on microstrip lowpass filter which is modelled with $2D$ complex circuit composed of lumped elements. The same filter is then analysed as $1D$ circuit with transmission lines, and $3D$ electromagnetic analysis is also done.

[1]Miodrag V. Gmitrović and Biljana P. Stojanović are with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš,Serbia and Montenegro, E-mails: [gmitrovic, bilja]@elfak.ni.ac.yu

II. New Multi-Port Models

A $2D$ electrical circuit with lumped LC elements, which is very suitable for modelling of microwave transmission lines [1-4], is shown in Fig.1. In the case of uniform microwave lines, complex network consists of large number of simple networks, as shown in Fig.2a), which inductances and capacitances can be obtained by relations given in the reference [3]

$$L_1 = L_2 = \frac{Z_c d\sqrt{\varepsilon_r^{eff}}}{2c_0} \cdot \frac{nu}{nk}\,, \tag{1}$$

$$L_3 = L_4 = \frac{Z_c w^2\sqrt{\varepsilon_r^{eff}}}{2c_0 d} \cdot \frac{nk}{nu}\,, \tag{2}$$

$$C_1 = \frac{Z_c\sqrt{\varepsilon_r^{eff}}}{c_0 Z_c} \cdot \frac{1}{nk \times nu}\,. \tag{3}$$

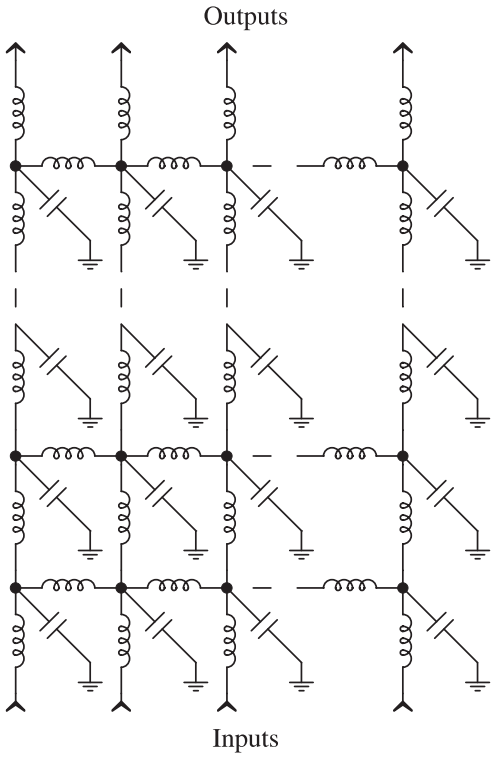


Fig. 1. A $2D$ electrical network.

The parameters in the previously given relations are: d - length and w - width of line, nu - number of $2D$ network inputs, nk - number of cascade-connected networks, c_0 - speed of the light in vacuum, Z_c - line characteristic impedance

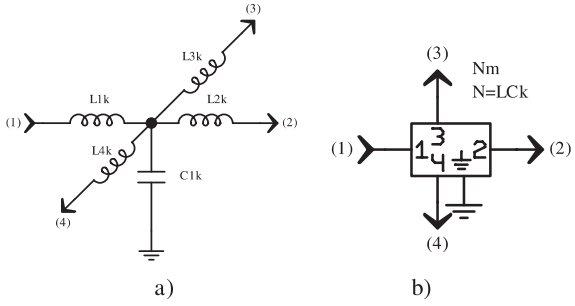


Fig. 2. a) LC node and b) its equivalent 4-port network.

and ε_r^{eff} - effective dielectric constant. Formulas for and are known and given in the paper [8]. In the program *GENESYS*, one part of the analysed network can be replaced with multi-port network. For example, a part of LC network shown in Fig. 1, with 5 inputs and 5 outputs, can be replaced with network which symbol is given in the Fig. 3a). Now, whole network can be represented by a large number of such constituent networks, which are connected between them. How all ports in this symbol are putted at one side, connection of constituent networks is very difficult and whole network is very confused. In order to simplify replacement of $2D$ electrical network, which is very complex in its nature, with equivalent multi-port networks, new models are suggested. New symbols for the case of network with 10 ports are given in Fig. 3b) and c). Network shown in Fig. 3c) is used only in the case of cascade connection, ports 1-5 and 6-10. Network shown in Fig. 3b) is used for cascade connection, ports 1-4 and 6-8, and for side connection, ports 9 and 10. Now, networks of such shapes can be connected very easily. The resulting equivalent network has simple shape; it is also very clear and easy for controlling.

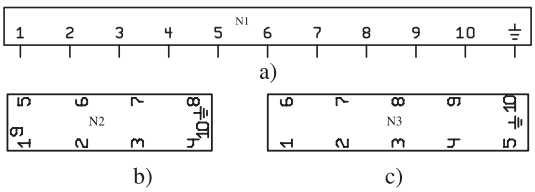


Fig. 3. a) Original symbol of 10-port network, b) and c) new equivalent symbols.

III. Forming of Networks by Using New Models

Procedure of forming equivalent network by using new models of multi-port networks will be explained on an example of microstrip lowpass filter, which layout is shown in Fig. 4. It is a 7th order filter with a cutoff frequency of 900 MHz and 50 Ohm terminations, which is taken from the reference [10] and analysed in the paper [4]. The filter parameters are: $d1 = d9 = 2540\mu$m, $d2 = d8 = 13385.9\mu$m, $d3 = d5 = d7 = 2873.93\mu$m, $d4 = d6 = 30022.8\mu$m, $w_1 = w_9 = 932.59\mu$m, $w_2 = w_4 = w_6 = w_8 = 465.19\mu$m, $w_3 = w_7 = 15679.79\mu$m, $w_5 = 18016.59\mu$m, ε_r=6.0, h=635 and t =18.03μm. Chosen segmentation for the lines 1 and 9 is nu=4 and nk=8, for the lines 2 and 8 is nu=2 and

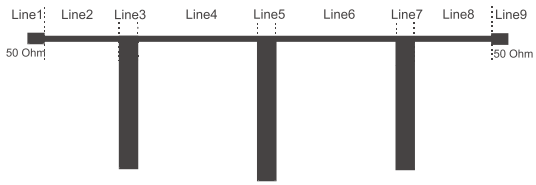


Fig. 4. A layout of microstrip lowpass filter.

nk=40, for the lines 4 and 6 is nu=2 and nk=100, for the lines 3 and 7 is nu=45 and nk=9, and for the line 5 segmentation is nu=54 and nk=9.

At the beginning, numerical values of the physical dimensions of microstrip filter, data about substrate and data about a number of line segments in transverse direction, nu, and in longitudinal direction, nk, are assigned in the *Equations* option by using program statements which are accessible in the program *GENESYS*. Then, the characteristic impedance, Z_c, and effective dielectric constant, ε_r^{eff}, are counted by using program statements for the formulas given in the reference

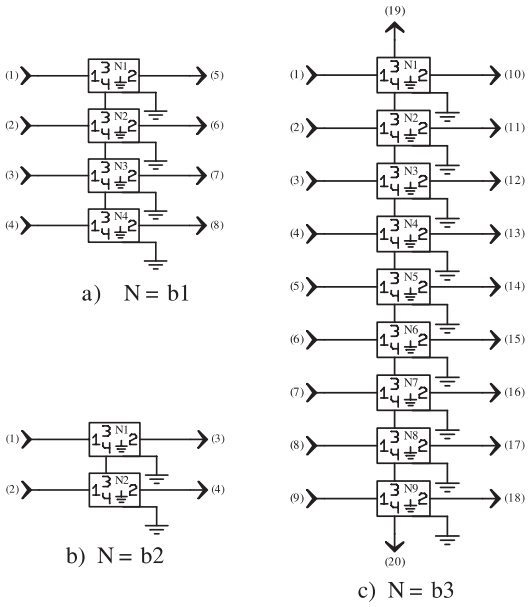


Fig. 5. a) 8-port network, b) 4-port network and c) 20-port network.

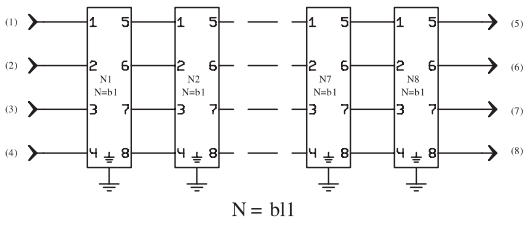


Fig. 6. Cascade connection of eight 8-port networks.

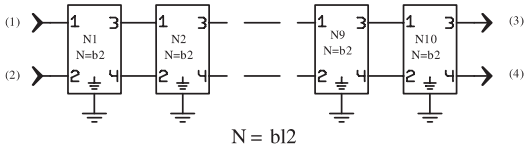


Fig. 7. Connection of ten 4-port networks.

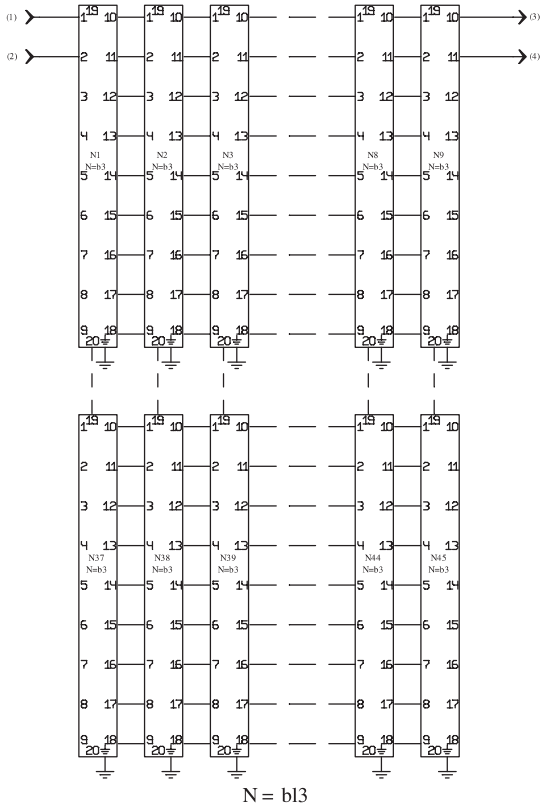


Fig. 8. Connection of forty five 20-port networks.

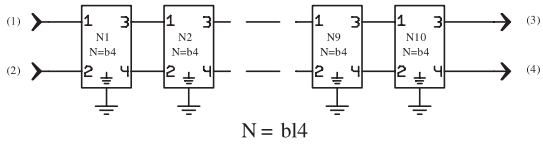


Fig. 9. Connection of ten 4-port networks.

[8]. The inductance and capacitance values shown in the network given in Fig. 2a), for the lines 1 to 5, are counted by using the relations (1-3).

Now, it walks up to drawing the needed multi-port electrical circuits and networks. Because of the symmetrical analysed filter, only the networks for the approximation of the lines 1 to 5 are formed. Five separated electrical schemes, $N = LCk,\ k = 1, .., 5$, are drawn in the *Workspace*. Their inductance and capacitance values are automatically taken from the option *Equations*, and they depend on the physical dimensions of the corresponding lines and the segmentations in transverse and longitudinal directions. These networks are replaced with 4-port networks given in the Fig. 2b) and then, next multi-port networks are drawn, Fig. 5. The first line is simulated by network given in Fig. 6, where the network assigned by $N = b1$ is given in the Fig. 5a). The second one is simulated by network shown in the Fig. 11, where the network assigned by $N = bl2$ is shown in the Fig. 7. The network assigned with $N = b2$ is given in Fig. 5b). Because the third line is very wide, for its simulation are used 20-port networks as shown in Fig. 8, where the network assigned by $N = b3$ is given in Fig. 5c). The fourth and fifth networks are simulated as the second and third ones, respectively. The

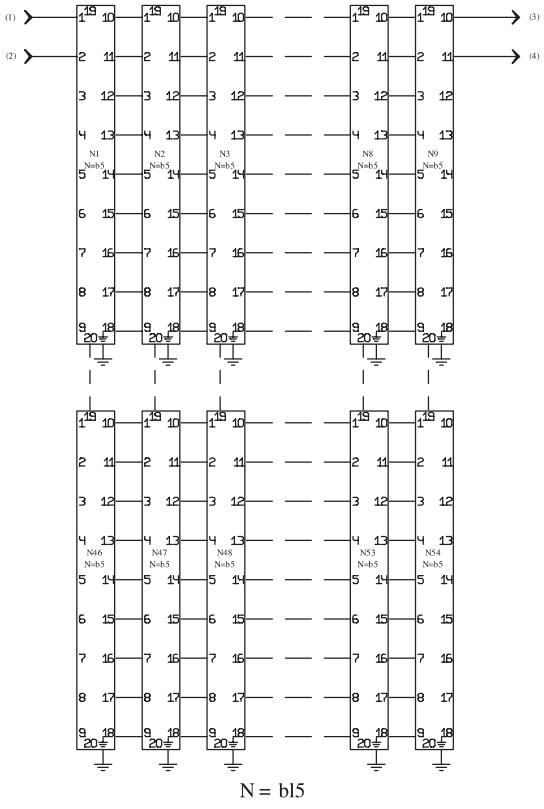


Fig. 10. Connection of fifty four 20-port networks.

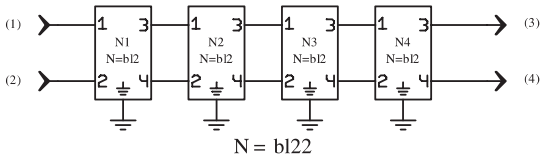


Fig. 11. Cascade connection of four 4-port networks.

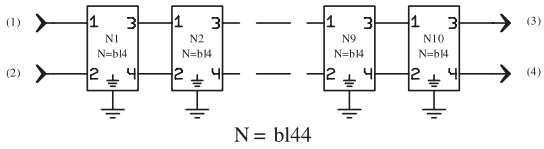


Fig. 12. Cascade connection of ten 4-port networks.

complete network used for simulation of microstrip lowpass filter is shown in Fig. 13. The input lines are simulated by 3-port networks, and terminated impedances are $Z_{cl} = 50\Omega$.

IV. Analysis Results

In the software package *GENESYS*, microstrip lowpass filter is analyzed by using three different procedures: 1) simulation of the filter by $1D$ circuit with transmission lines, 2) $3D$ electromagnetic simulation of the filter and 3) simulation of the filter by $2D$ circuit consisting of lumped LC elements according to the suggested procedure. Response results of the transmission parameter S_{21} are shown at the same graphic in Fig. 14. From this graphic it can be concluded that results obtained by suggested procedure are nearer to the val-

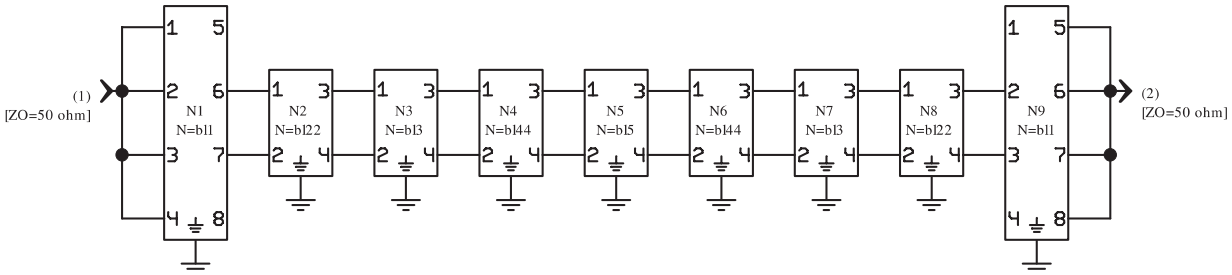


Fig. 13. Microstrip lowpass filter simulated by cascade connection of nine complex multi-port networks.

ues obtained by $3D$ electromagnetic simulation in whole frequency band than to the ones obtained by $1D$ simulation. The same microstrip filter is analyzed in the paper [4] by using *ETS* (Equivalent Thevenin Source) method incorporated in the software package *FAMIL*. Counted values are in full agreement with the ones counted by procedure suggested in this paper.

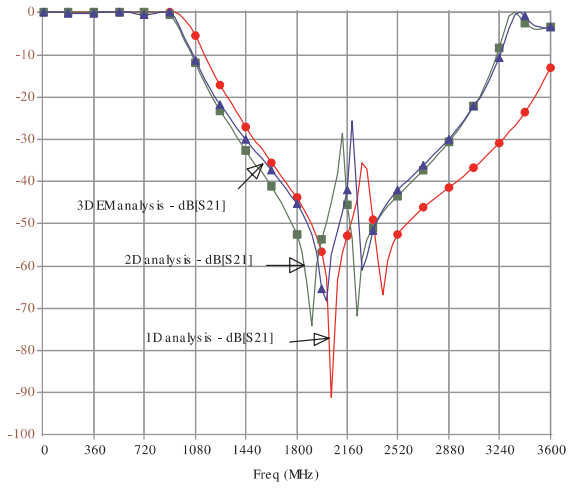


Fig. 14. Frequency response of the microstrip lowpass filter.

V. Conclusion

In this paper, simple procedure for frequency analysis of complex $2D$ circuits with lumped LC elements by using program *GENESYS*, is given. Electrical circuits, formed in this way, are using very successfully for modelling of microwave transmission lines. In order to analyze such circuits by program *GENESYS*, new models of the multi-port networks are formed. The schedule of forming electrical circuits and multi-port networks is shown on example of microstrip lowpass filter. The analysis results obtained by filter simulation with $2D$ circuit with lumped LC elements are in better agreement with results of $3D$ electromagnetic simulation than with the ones obtained by filter simulation with $1D$ circuit with transmission lines.

The suggested procedure is simple, has general character and can be applied to microwave circuits with both uniform and nonuniform transmission lines of complex configuration. Choosing different segmentation solves the discontinuity problem, i.e. changes of line widths.

Acknowledgment

This work has been supported by the Ministry of Science, Technologies and Development of Republic of Serbia.

References

[1] B. Stojanović and M. Gmitrović, "Analysis of 2D Lumped Circuit by Equivalent Thevenin Source Method", *14th Intern. Conf. on Microwaves, Radar and Wireless Communications - MIKON 2002*, May 20-22, Gdansk, Poland, pp. 549-552.

[2] M. Gmitrović and B. Stojanović, "Analysis of Cascade-Connected Planar Transmission Lines by ETS Method", *XXXVII Intern. Scientific Conf. on Information, Communication and Energy Systems and Technologies - ICEST 2002*, October 1-4, Niš, Yugoslavia, Volume 1, pp. 317-320.

[3] B. Stojanović and M. Gmitrović, "Planar Transmission Line Analysis", *XVLI Yugoslav Conference - ETRAN, June 3-6, 2002*, Banja Vrućica - Teslić, Republic of Srpska, pp. 241-244.

[4] B. Stojanović and M. Gmitrović, "Analysis of Cascade-Connected Transmission Lines with Different Widths by ETS Method", *X Telecommunication Forum - TELFOR 2002*, November 26-28, Belgrade, Yugoslavia, pp. 595-598.

[5] J. H. Thompson, T. R. Apel, "Simplified Microstrip Discontinuity Modeling Using the Transmission Line Matrix Method Interfaced to Microwave CAD", *Microwave Journal*, July 1990, pp. 79-88.

[6] J. D. Geest, T. Dhaene, N. Fache and D. De Zutter, "Adaptive CAD-Model Building Algorithm for General Planar Microwave Structures", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 9, September 1999, pp. 1801-1808.

[7] W. K. Gwarek, "Analysis of Arbitrarily Shaped Two-Dimensional Microwave Circuits by Finite-Difference Time-Domain Method", *IEEE Transactions on Microwave Theory and Techniques*, Vol.36, No.4, April 1998, pp. 738-744.

[8] D. M. Pozar, *Microwave Engineering*, New York: John Wiley & Sons, 1998, pp. 160-163.

[9] Advanced Designing System ADS, *Agilent Technologies*, User's Manual, 2000.

[10] R. W., Rhea, *HF Filter Design and Computer Simulation*, Noble Publishing Corporation, USA, 1994.

[11] RF and Microwave Design Software GENESYS, Eagleware Corporation, 635 Pinnacle Court, Norcross, GA 30071, 2001.

[12] S. M. Gmitrović, "Analysis and Optimization of RF and Microwave Circuits in the Program Package *GENESYS*", *Final exam*, Faculty of Electronic Engineering, Niš, 2003.

[13] M. V. Gmitrović, S. M. Gmitrović and B. P. Stojanović, "Frequency Analysis of 2D Electrical Circuits by Program *GENESYS*", *XVLII Conference - ETRAN*, June 3-6, 2003, Budva, Serbia and Montenegro, accepted paper.

Modeling of the Propagation Curves from ITU-R P.370-7 Recommendation using Neural Approach

Bratislav Milovanović, Zoran Stanković and Nebojša Vasić[1]

Abstract – **In this paper modeling of the propagation curves from ITU-R P.370-7 recommendation will be done using neural model based on multilayer perceptron network. In order to increase accuracy and generalization ability of neural models a special two-phase training process procedure has been developed and applied. This way enables fast and accurate computing the field strength level at reception points for any given effective antenna height from 37.5 m to 1200 m and range from 10 km to 650 km.**

Keywords – **ITU-R P.370-7 recommendation, neural network, neural model, electromagnetic field strength.**

I. Introduction

Design of any telecommunication system requires efficiently prediction of the electromagnetic field strength level. Great number of global and local parameters such as: configuration of the nearby terrain, obstacles and their attributes, climate zone, refraction index, multipath etc. affect the propagation of the electromagnetic waves. Therefore, it is almost impossible to develop a universal algorithm, meaning the reasonably short computing time, which will be able to solve the problem for the macro cell, micro cell and indoor propagation. To produce optimized algorithm almost every method omits one or more of the mentioned parameters.

A propagation model is a set of mathematical expressions, diagrams and algorithms by which signal strength level or path loss can be adequately represented or calculated. Generally, the propagation models can be either statistic (also called empirical), or deterministic, or a combination of these two, also known as a pseudo-deterministic.

Empirical models are derived from the extensive terrain measurements and statistic data analysis. The diagrams and tables of corrections that are result of the previous statistic analyze then give the correlation between terrain parameters and signal strength level. ITU-R method is typical statistic method. This method is based on reading the field strength level on the given diagrams in ITU-R P.370-7 recommendation [1,2]. Diagrams show field strength level in dBμV/m as a function of distance and effective height (Fig. 1).

The main problem is the inaccuracy of the reading due to visual errors and the existence of the only six discrete effective height curves. So the interpolation procedure must be the part of the immoderate algorithm, described in [2]. In practice it needs to calculate the field strength level in numerous

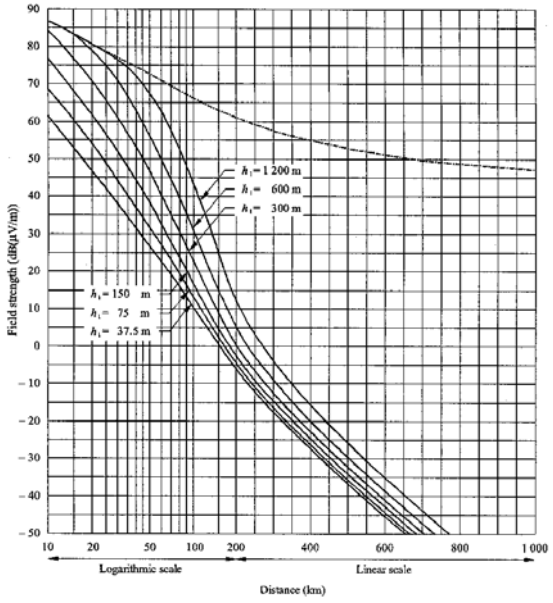


Fig. 1. Field strength (dB(μV/m)) for half wave dipole, which radiates 1 kW e.r.p. [1], land, 50% of the time, 50% of locations, frequency 450-1000 MHz.

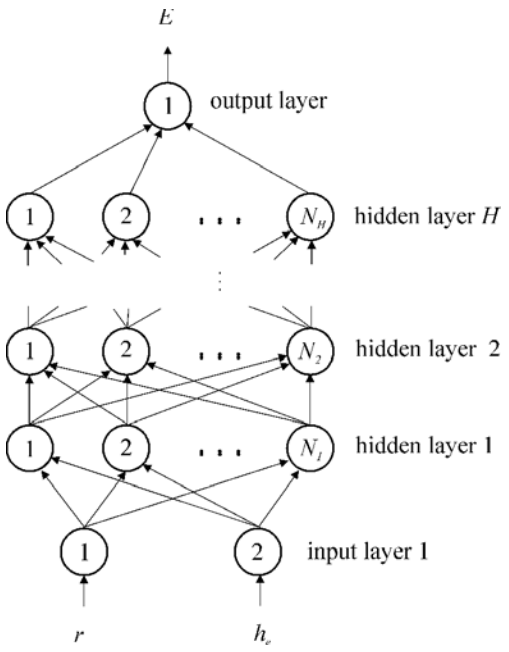


Fig. 2. Neural model architecture used for prediction of the electric field level E as a function of effective height h_e and distance between receivers r.

[1]Bratislav Milovanović, Zoran Stanković and Nebojša Vasić are with Faculty of Electronic Enginnering, Beogradska 14, 18 000 Nis, Serbia and Montenegro, E-mail: [bata, zoran, nvasic]@elfak.ni.ac.yu

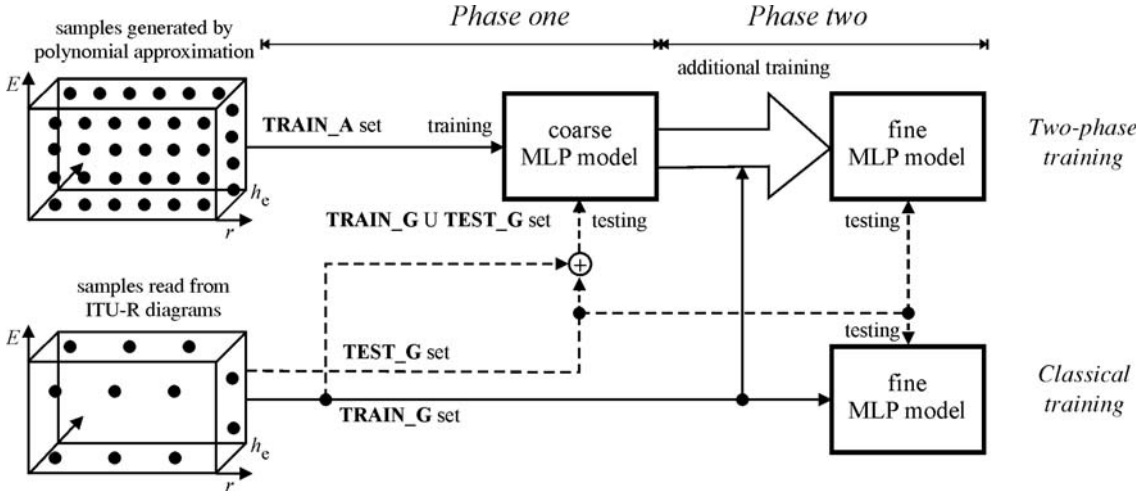


Fig. 3. Two-phase and classical training of the propagation curve's MLP neural models.

points for each azimuth in the sector of interest so the method becomes thorough job.

As one of the alternatives to the visual reading is modeling of the propagation curves by neural networks. First neural models for this purpose are based on the multilayer perceptron network (MLP-Multi Layer Perceptron) [3-8] and has been introduced in [7,8] showing good accuracy and fast computation. Further accuracy improvement and generalization abilities of the MLP neural models of the propagation curves, especially in ranges slightly out of the defined input parameters range, can be achieved using two phase training procedure that will be presented in this paper.

II. Neural Model of the Propagation Curves from ITU-R P.370-7 Recommendation

The statement of the problem is as follows: find the field strength level function as a function of distance and effective transmitter's height

$$E = f(h_e, r) \tag{1}$$

According to the above MLP neural model will have two neurons in the input layer and one neuron in the output layer (Fig. 2). Number of hidden layers is great than one but the results obtained in [7,8] showed that the optimum is two hidden layers. This conclusion was decisive which model to choose and thus we will deal only with models with two hidden layers. Activation functions in the hidden layers are of tang-hyperbolic type, but at the output neuron they are linear. In an agreement with this notation for MLP model is M$n - l_1 - l_2 - ... - l_n$ where n is the total number of neural layers in the given model and a $l_1, l_2, ..., l_n$ are the numbers of neuron in the hidden layers, respectively.

The method applied in training procedure of the MLP neural models of the propagation curves consists of two phases (Fig 3): "coarse" training using the large set of the input variables and "fine" training using the smaller set of input variables defined by ITU-R P.370-7 recommendation. Coarse training phase was conducted in the large set of inputs - 10000 samples generated by approximation function from

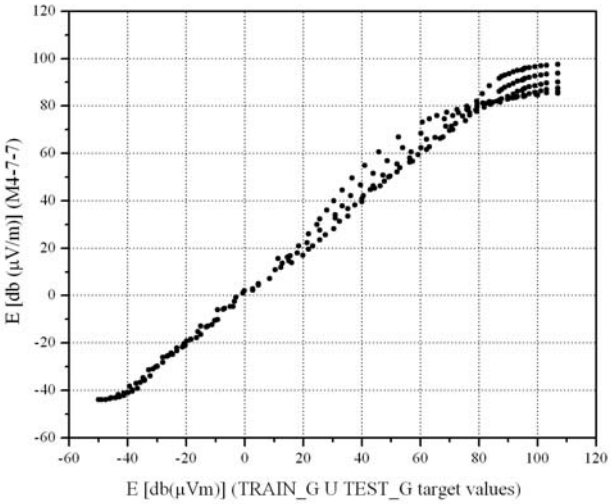


Fig. 4. M4-7-7 correlation diagrams after first training phase.

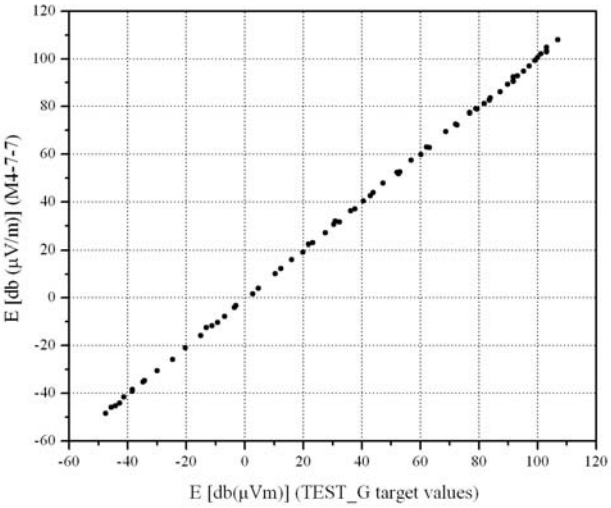


Fig. 5. M4-7-7 correlation diagrams after second training phase.

Table 1. Testing results of the propagation curve's MLP neural models

MLP model	Two-phase training (phase one) (TRAIN_G U TEST_G set)		Two-phase training (phase two) (TEST_G set)		Classic training (TEST_G set)	
	WCE [%]	ATE [%]	WCE [%]	ATE [%]	WCE [%]	ATE [%]
M4-5-5	18.50	2.85	1.65	0.45	39.21	1.51
M4-6-6	17.35	2.70	1.97	0.48	3.70	0.57
M4-7-7	14.95	2.48	1.60	0.45	9.82	3.69
M4-8-2	16.94	2.75	2.74	0.61	18.93	2.29
M4-8-5	19.25	2.58	3.21	0.84	5.62	1.03
M4-8-8	18.29	2.62	12.94	0.77	4.25	0.92
M4-9-5	18.19	2.30	4.13	1.14	18.56	5.57
M4-9-8	17.91	2.39	3.34	1.03	7.60	2.08
M4-9-9	17.83	2.57	2.41	0.57	8.68	1.10
M4-10-4	14.83	2.22	1.93	0.55	3.68	1.04
M4-10-9	16.06	2.48	2.17	0.57	5.90	1.05
M4-10-10	17.30	2.31	3.46	0.91	3.99	1.21

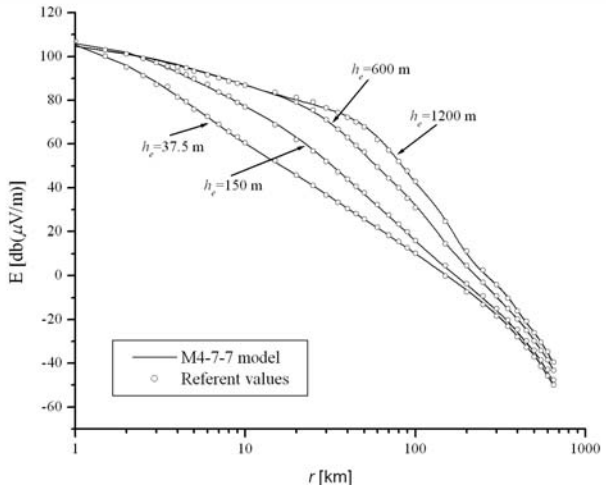


Fig. 6. Comparison of the M4-7-7 output curves to referent values that haven't been included in training set (two-phase training).

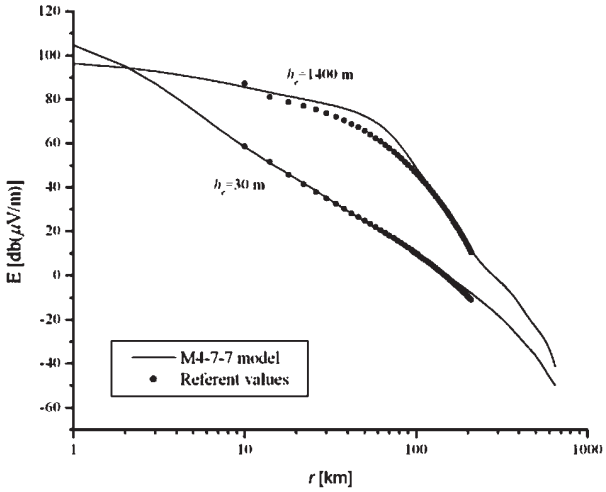


Fig. 7. Comparison of the M4-7-7 output curves to values obtained by polynomial approximation application for the cases of h_e=30 m and h_e=1400 m that are not belonging to a range of training neural models (two-phase training).

[2] in the range of 30 m<h<1400 m and 1 km<r<650 km. In this phase we applied Quasi-Newton training method with required maximum absolute error of 10^{-3}. Despite the lower efficiency of Quasi-Newton method compared to the Levenberg-Marquardt method it was optimal solution thanks to its applicability on large sets of inputs. Neural models that showed the best results in testing procedure were picked up for the second phase. This way the chosen models from the first phase have the a priori knowledge and increase the accuracy in the second phase. In order to test MLP model at the first training phase (cycle) we used set of 228 samples directly read from the ITU-R diagrams (TRAIN_G ∪ TEST_G). This set was split then in two subsets: one for training in the second cycle (155 samples - TRAIN_G) and the other for testing (73 samples - TEST_G). Training set was quite smaller compared to the previous set. In this phase Levenberg-Marquardt training method was applied showing the well known efficiency when dealing with the smaller set of training samples.

Simultaneously using the same set of input values form the second phase (cycle) training was done directly on classic way for the purpose of comparing with the new herein described training approach (Fig. 3). Testing of both models (new two cycle phase approach) and classic (direct) was carried on the same set of 73 samples and the results are shown in Table 1. Average absolute error (ATE) and maximum absolute error (WCE) present mean values of the three best trainings of the same MLP model. Analyzing the results shown in the Table 1 one can observe the advantages of the two phase training method: decrease of the maximum test error and consequently decrease of the average test error. Test correlation diagrams for M4-7-7 model which have the smallest errors are shown in Fig. 4 and Fig. 5 (after first and second phase respectively)

III. Simulation Results

Neural model based on two phase training method, which has shown the minimum error during testing, was used then for simulation of the field strength level as a function of effective antenna height and distance from transmitter's point. Fig. 6 shows the propagation curves for effective antenna height values of 37.5, 150, 600, and 1200 m obtained by M4-7-7 model with respect to the reference values read from the dia-

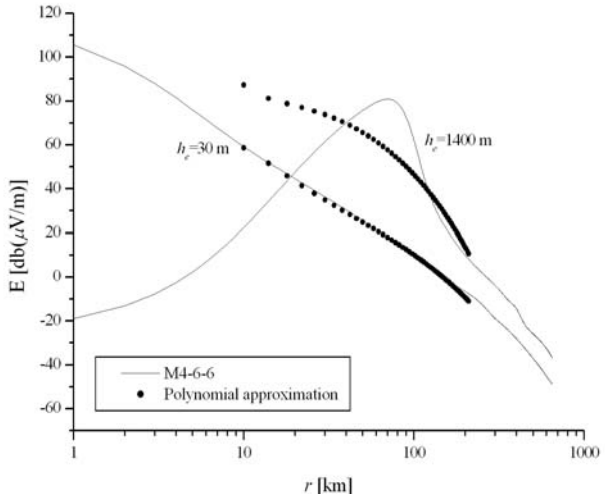


Fig. 8. Comparison of the M4-6-6 output curves to values obtained by polynomial approximation application for the cases of h_e=30 m and h_e=1400 m that are not belonging to a range of training neural models (classical training).

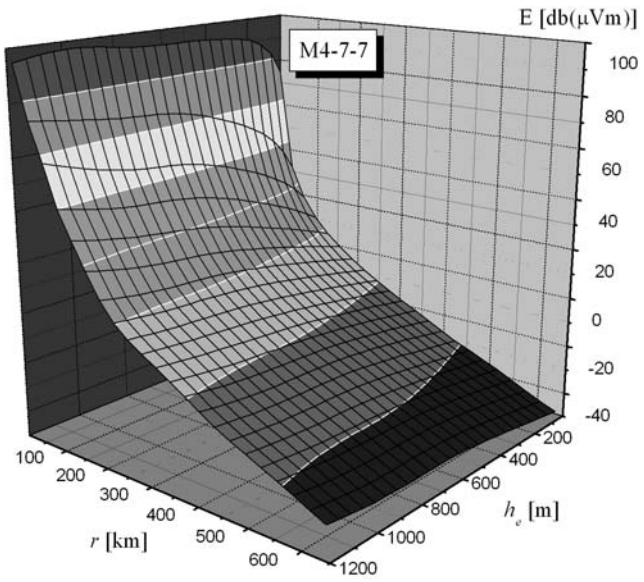


Fig. 9. Three dimensional presentation of the field strength level as a function of effective antenna height and distance, generated by M4-7-7 model.

gram. We can see the excellent fitting of the values obtained by M4-7-7 model in the reference values. This signifies the high accuracy of the two phase model.

Furthermore in order to explore the extrapolation the same neural model was used to generate the propagation curves for h_e=30 m and h_e=1400 m (Fig. 7). Due to inability to get the values from the diagram these curves were then compared to the approximate values. For the same effective heights (30 m and 1400 m) propagation curves were generated using M4-6-6 model that has been passed the standard (direct) training procedure and showed minimum error (Fig 8). It can be seen that M4-7-7 model is better than M4-6-6 especially in area where extrapolation is a must.

M4-7-7 model was used to generate complete three dimensional presentation of the field strength level as a function of effective antenna height and distance. Also, simulation was done in 15640 points producing the three dimensional pattern on Fig. 7. To do that it took only 5 seconds on PIII 450 MHz hardware platform with 128 MB RAM.

IV. Conclusion

Probably the most reliable statistic method for field strength prediction in radio and TV systems is ITU-R method. Reading the diagrams often makes this method effortful, so the adequate neural model has great practical utility. The results presented in this paper prove that neural model based on MLP network is accurate alternative to the classic reading the diagrams. Excellent agreement between referent values (the diagram's values) and values obtained using MLP neural model so as fast computation time (only 5 seconds last computation in 15000 points) are the main advantages of the MLP models developed with two phase training method. In addition two phase method requires less training samples than direct method to get the same target accuracy thanks to knowledge gained in the first training phase. This implies that pre-knowledge improves the accuracy and fast learning at the second training phase. The greatest accuracy increase is in the extrapolating area - slightly out of the range of input effective height values, which is very useful in practice. Finally, thanks to its features MLP model can be embedded in software for prediction of the field strength level.

References

[1] Recommendation ITU-R P.370-7, VHF and UHF Propagation Curves for the Frequency Range from 30 MHz to 1 000 MHz, Broadcasting services.

[2] Bratislav Milovanović, Nebojsa Vasić, Vladan Stanković, Aleksandar Atanasković, Jugoslav Joković, "Approximation of the Propagation Curves from the Recommendation ITU-R P. 370-7", TELSIKS 2001 Conference Proceedings, Niš 2001, pp.699-702.

[3] S. Haykin, *Neural Networks*, New York, IEEE, 1994.

[4] J. Hertz, A. Krogh and R. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, 1991.

[5] B. Milovanovic, Z. Stankovic and V. Stankovic, "Loaded Cylindrical Metallic Cavities Modeling using Neural Networks", TELSIKS'99 Conference Proceedings, Nis 1999, pp.214-217.

[6] B. Milovanovic, Z. Stankovic, S. Ivkovic, "Modelling of the Cylindrical Metallic Cavity Loaded by Lossy Dielectric Slab Using Neural Networks", NEUREL 2000 Conference Proceedings, Beograd 2000, pp. 141-145.

[7] Zoran Stanković, Bratislav Milovanović, Jelena Jovković, Jelena Antonijević, " Modeling of the ITU-R P.370-7 Propagation Curves by Neural Network", ICEST 2002 Conference Proceedings, Niš 2002, pp. 103-106.

[8] Zoran Stanković, Bratislav Milovanović, Jelena Antonijević, Jelena Jovković, "Neural Model of the Propagation Curves from ITU-R P.370-7", NEUREL 2002 Conference Proceedings, Beograd, Septembar 2002, str. 191-196

Antenna Miniaturization Using Fractal Geometry

Aleksandar Atanasković, Bratislav Milovanović[1]

Abstract – **The expected benefit of using a fractal as a dipole antenna is to miniaturize the total height of the antenna at resonance, where resonance means having no imaginary component in the input impedance. In this paper three types of fractals are investigated as dipole wire antennas. They include two planar structures, Koch curve and a fractal tree, and a three dimensional fractal tree. These three types of fractals are compared among each other and to a straight dipole. The starting structure for each of these fractal geometries is straight dipole that is resonant in the PCS band, at 1900 MHz. In the simulation, the antenna height is held constant and the frequency is swept. It can be seen that the resonant frequency decreases as the number of fractal iterations increases. The decrease in resonant frequency can correlate to a miniaturized antenna, if the resonant frequency would be held fixed.**

Keywords – **Fractal antenna, antenna miniaturization, Koch curve, fractal tree**

I. Introduction

Although fractals are mainly discussed in mathematical forums, they exist in all parts of nature. For example Mandelbrot [1] discusses the basics of fractal theory as applied to the characteristics of a coastline (Fig. 1.). The length of a coastline depends on the size of the measuring yardstick. As the yardstick we use to measure every turn and detail decreases in length, the coastline perimeter increases exponentially. As the view of a coastline is brought closer, we discover that within the coastline there lie miniature bays and peninsulas. As we examine the coastline on a rescaled map, we discover that each of the bays and peninsulas contain sub-bays and sub-peninsulas. There is a self-similar trait observed as we look at the coastline at various resolutions. The number of microscopic structures begin to approach infinity. In fact, because of the immense number of irregularities, the physical length of a coastline is virtually infinite. Self similarity (seen in the coast example above) is defined by structures that look the same at variable magnifications. This recurring self-similarity is one of the many attributes of many fractals. Much like the coastline described above, any small part in a self-similar fractal is going to look exactly like the fractal as a whole.

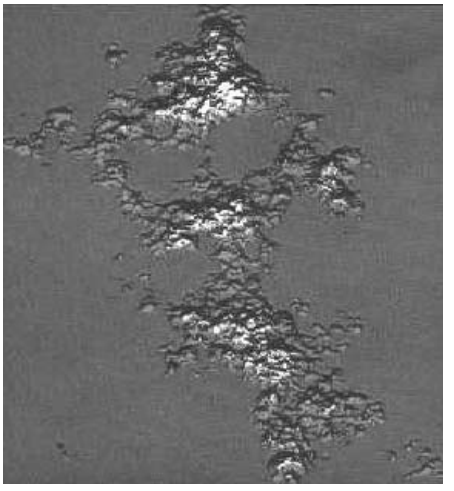


Fig. 1. Fractal-generated coastline

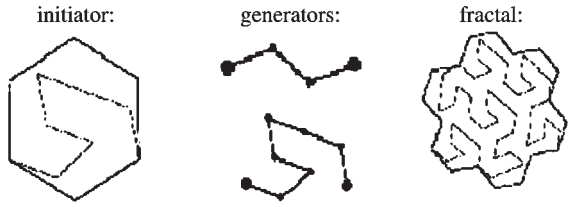


Fig. 2. Initiator/generator fractal

Another type of figure uses a generator/initiator relationship system of construction. This construction begins by placing an ”initiator”, which will be the base format for the figure. The initiator is then divided into a collection of lines upon which the generator(s) will be placed. Fig. 2. shows the initiator and its first stage of growth where the lines are replaced, or added to, by one of the two generators. Once the generators replace the lines belonging to the initiator, the generator may repeat ”n” number of time, or a different generator may begin growth upon the one already in place.

II. Koch Curve

The first fractal shape that is investigated as a dipole antenna is Koch curve [2,3,6]. The geometry of how this antenna could be used as a dipole is shown in Fig. 3.

A Koch curve is generated by replacing the middle third of each straight section with a bent section of wire that spans the original third. Each iteration adds length to the total curve. This can be seen from the figure depicting the generating process (Fig. 3.). Each iteration results in a total length that is 4/3 times the original geometry. However, the original overall height of the fractal does not change from one iteration to the next. Therefore, if the process is carried out for an infinite number of times, the curve would have an infinite length while the overall height would not change.

The starting structure that is used is a half of a resonant PCS dipole, which is 3.75 cm in length. The overall length

[1]Authors are with Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: beli@elfak.ni.ac.yu

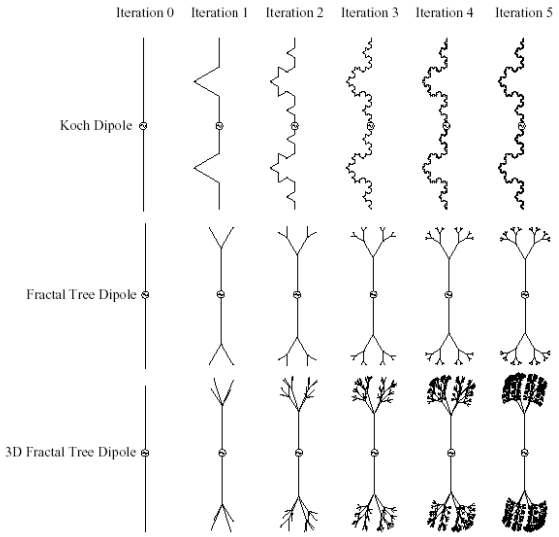


Fig. 3. The various fractal geometries are configured as dipoles, including a Koch fractal, a fractal tree and a three dimensional fractal tree. The starting size of the geometries are identical PCS band dipoles.

of the resonant dipole is 7.5 cm, which is slightly smaller than $\lambda/2$ at 1900 MHz.

The total length of the Koch curve is given by:

$$l_{Koch} = l\left(\frac{4}{3}\right)^{n}, \qquad (1)$$

where n is the number of iteration and h is the height of the straight starting generator.

These fractals are analyzed as resonant dipole antennas using WIPL-D software [4]. The input match, compared to 50Ω, of the fractal dipoles and straight dipoles as calculated are shown in Fig. 4. It can be seen how the resonant frequency drops as the number of generating iterations for the fractal is increased. Also, it is interesting to note that the resonant frequency approaches an asymptotic limit. This limit gives an insight into where the resonance of an ideal Koch fractal curve as a dipole would lie, if such a structure were manufacturable. The simulated input impedance plots are shown in Fig. 5.

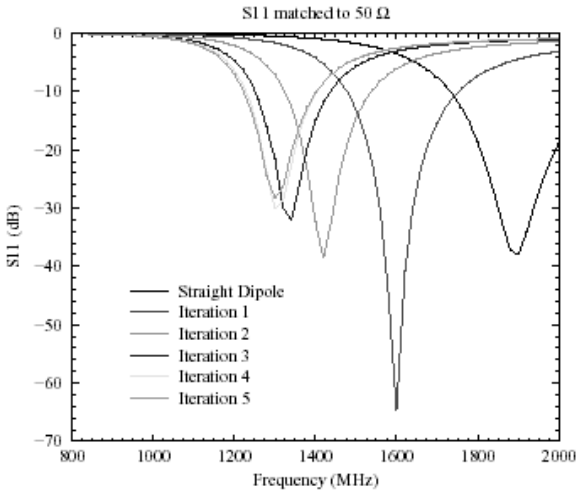


Fig. 4. Simulated input match of the straight dipole and the first five iteration for the Koch dipole antennas matched to 50 Ω

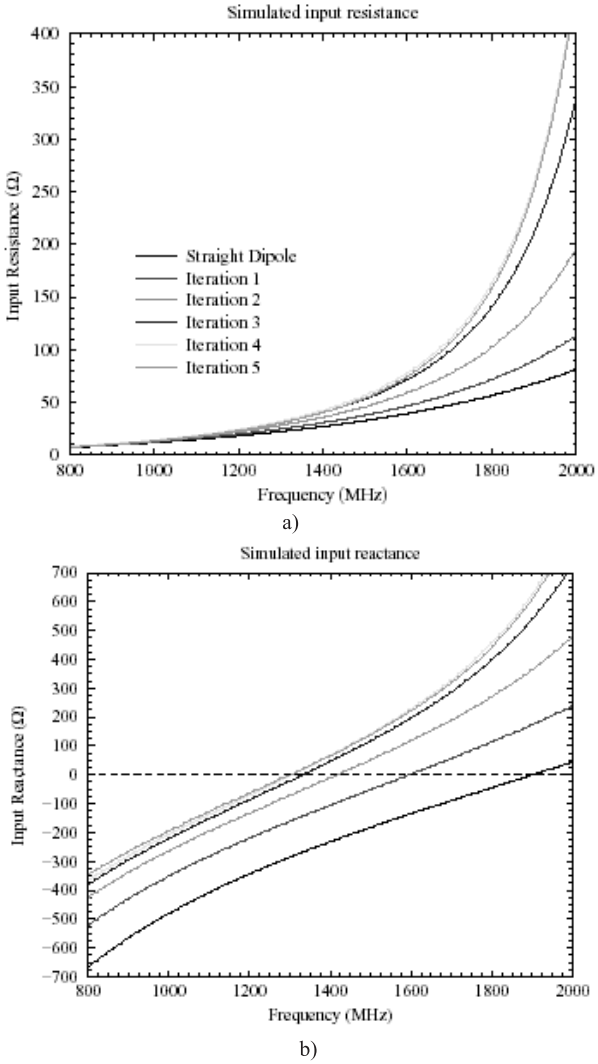


Fig. 5. Simulated input impedance for the first five fractal iterations of Koch dipoles plus a straight dipole for comparison. a) input resistance b) input reactance

III. Fractal Tree

Another type of fractal that can be utilized as a dipole is a fractal tree. The geometry of how the fractal is used is shown in Fig. 3. This deterministic fractal is a simple model of branching found in nature. Again, the goal of using this type of fractal is to reduce the height of a resonant dipole antenna.

The fractal is generated by applying an iterative sequence to the starting structure. The fractal is started with a simple monopole. The top segment of this monopole is then split at a pre-determined angle, $\theta = 60°$, to form the first two branches. As the iterative process continues, the end segment of each branch splits into two more branches. The total electrical length of the conductor, l, remains constant throughout

the iterative process. The total electrical length can be defined as the shortest length from base of the fractal to any other end. The lengths of each straight section in the first five iterations are shown in Table 1. It can be seen from the section lengths that the total conductor length, l, always adds up to unity for each iteration.

Table 1. Length of each straight section of the fractal tree and 3D fractal tree for the first five iterations

Iteration	0	1	2	3	4	5
	1	1/3	1/7	1/15	1/31	1/63
		2/3	2/7	2/15	2/31	2/63
			4/7	4/15	4/31	4/63
				8/15	8/31	8/63
					16/31	16/63
						32/63

The first five iteration plus a straight dipole were analyzed. In the previous section describing the Koch dipole antenna, the overall height was maintained from iteration to iteration. For the tree fractal, the total length of the conductor path is maintained among iterations. The subsection size for each iteration of the antenna is the same.

The input match, compared to 50Ω, of the fractal dipoles as calculated are shown in Fig. 6.

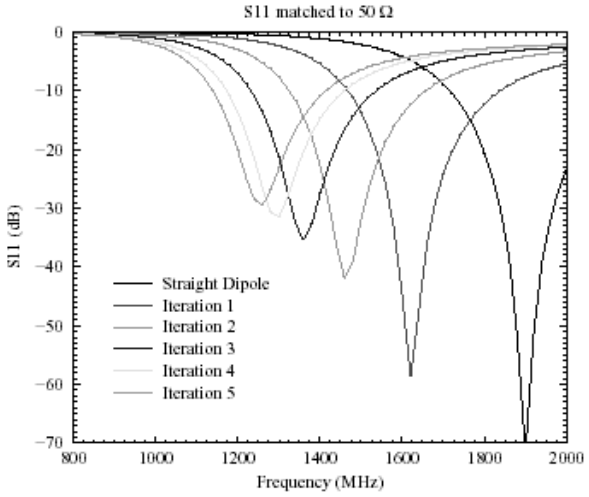


Fig. 6. Simulated input impedance matched to 50Ω for the first five iterations of a fractal tree dipole with a split angle of 60° and for a straight dipole

It can be seen that the resonant frequency drops as the fractal iteration is increased. The ratio of miniaturization versus the fractal iteration is very similar to that of the Koch dipole. As the fractal iteration increased, the resonant frequency decreases in a saturating manner. At each iteration the extra number of branches top loads the antenna. Even though the electrical length of a single conductor path from the generator port of the antenna to the top of a branch is identical for all antennas, there are more branches after each iteration. This adds more conduction paths at the top of antenna serving as a top-loaded device. This, in turn, lowers the resonant frequency at every iteration. It can be seen that the top loadings effect diminishes as the number of iterations is increased. The length of wire that branches out during each iteration is almost half as small as the previous iteration, thus the effect it has on the input characteristics of the antenna diminishes.

IV. Three Dimensional Fractal Tree

A three dimensional fractal tree has a similar geometry as the fractal tree. However, instead of branching in one plane, the fractal branches out in three dimensions. The resulting antenna exhibits similar benefits as the two dimensional case to a greater degree. The geometry of how this type of fractal can be utilized as a dipole is shown in Fig. 3.

The three dimensional fractal tree is generated in a similar fashion as the two dimensional case. The top of a straight monopole is split into four branches. The branches split off at a set angle in two orthogonal planes. The angle used in this case is 60°. The resulting four branches then split in a similar manner. The ratio of the sizes of each of the branches at each iteration is outlined in Table 1. For the purpose of studying this fractal as an antenna, the first five iteration are used. As before, this shows us the trends of the benefits of using a fractal within the computational limitations of the simulations. The fractal generated is mirrored at the base. These antennas are simulated in a dipole configuration.

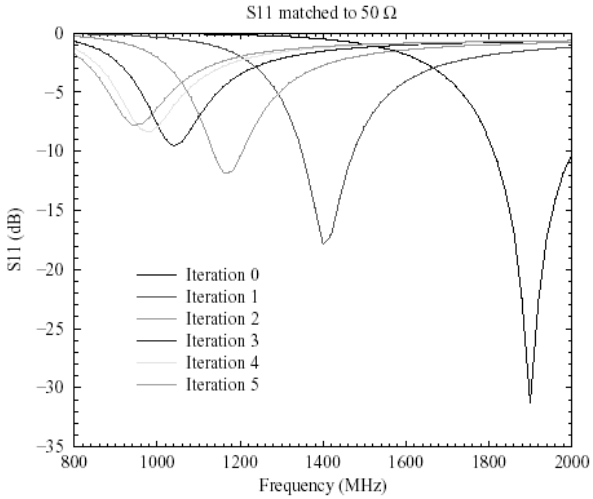


Fig. 7. Input match for various iterations of a three dimensional fractal tree matched to 50 Ω

The simulated input match for the antennas is shown in Fig. 7. It can be seen how the resonant frequency decreases as the fractal iteration is increased. In a similar fashion as the previous fractal dipoles studied, the input resistance decreases as the fractal iteration is increased, resulting in a poorer input match.

V. Fractal Dipole Comparison

The benefits of the various fractal geometries can be compared. All of the dipoles that are compared have the same starting height. The starting geometry is a resonant dipole

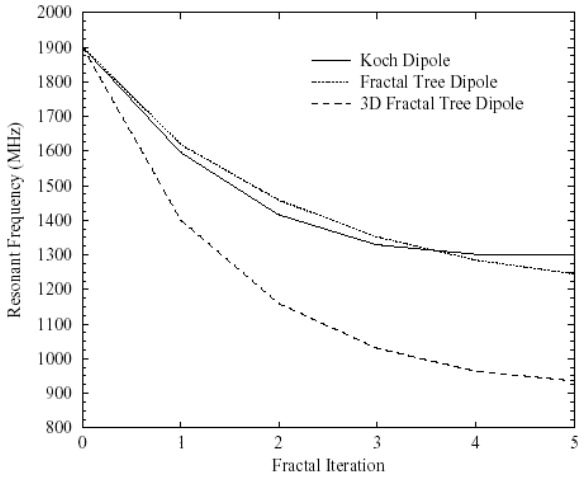


Fig. 8. The resonant frequency for each of the fractal antennas versus the number of iterations for a Koch tree, a fractal tree, and a 3D fractal tree in a dipole configuration as simulated with WIPL

that is 7.5 cm in length, resonant in the PCS band at 1900 MHz. The relative geometry of all of the compared dipoles is shown in Fig. 3.

The benefits of using a fractal geometry are dependent on the type of fractal that is chosen. A comparison of the miniaturization of the antennas by increasing the number of generating iterations is depicted graphically in Fig. 8.

It can be seen that the miniaturization benefits of both two dimensional structures, the Koch fractal and the fractal tree, are very similar. The benefits of the three dimensional fractal tree, however, is more pronounced.

Even thought the three dimensional fractal miniaturizes the antenna at resonance to a greater degree than the other fractals, the input resistance is lowered by a significant amount, as well.

It can be seen from Fig. 9. that the input resistance of the Koch and fractal tree dipoles drops to near 30 Ω at resonance for the fifth iteration. Likewise, the input resistance of the three dimensional fractal tree drops to 20 Ω due to the increased amount of conducting branches. This would decrease the match to a 50 Ω feed line. The fractal geometry chosen for a particular application would have to weight the trade-off between increased miniaturization versus input resistance.

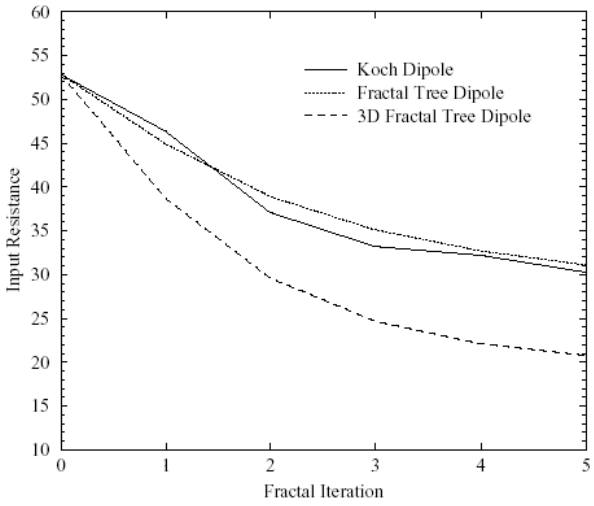


Fig. 9. Simulated input resistance versus the number of generating iterations for three fractal antennas

It can be seen from the plots of the simulated input match for the various dipoles that they are all narrow band antennas. The simulated 3 dB bandwidth of the dipole antenna is about 2.4%. This can be compared with the 3 dB bandwidth of the simulated fractals generated from the highest number of iterations, which have the lowest resonant frequency. The simulated bandwidth for the highest iteration of the Koch dipole is around 3.1%. For the fifth iteration of the fractal tree dipole, the simulated bandwidth is 4.2%. The simulated bandwidth of the fifth iteration of the three dimensional fractal tree is 12.7%, but only has a -7.75 dB input match at resonance.

VI. Conclusion

Fractal dipole antennas have shown the possibility to miniaturize antennas and to improve input matching. There are three distinct advantages which are reached by using fractal antennas. First, fractal geometries can be implemented to miniaturize dipole antenna. Also, designing with fractal geometries can overcome limitations to improve the input resistance of antenna that are typically hard to match to feeding transmission lines. Furthermore, the self-similar nature in the fractal geometry can be utilized for operating a fractal antenna at various frequencies.

Acknowledgement

This work was supported in part by Ministry of Science, Technologies and Development of Republic Serbia, under project ”Development of broadband wireless distribution systems”, contract number IT.1.15.0186.A.

References

[1] Mandelbrot, B.B., *The Fractal Geometry of Nature*, New York, W. H. Freeman and Company, 1983.

[2] C. Puente, J. Romeu, R. Pous, J. Ramis, A. Hijazo, ”Small but long koch fractal monopole”, *Electronic Letters*, 34(1):9-10, January, 1998.

[3] C. Puente-Baliarda, J. Romeu, A. Cardana, ”The Koch Monopole: A Small Fractal Antenna”, *IEEE Transaction Antennas and Propagation*, Vol. 48, No. 11, 2000, pp. 1773-1781

[4] B.M.Kolundzija, J.S.Ognjanovic, T.P.Sarkar, *WIPL-D: Electromagnetic Modeling of Metallic and Dielectric Structures*, Artech House, Inc., Boston, 2000.

[5] N. Cohen, (Unsigned), *Fractal Antenna White Paper*, Fractal Antenna Systems, Inc., 1999.

[6] J.P. Gianvittorio, Y. Rahmat-Samii, ”Fractal Antennas: A Novel Antenna Miniaturization Technique and Applications”, *IEEE Antennas and Propagation Magazine*, Vol. 44, No. 1, February 2002, pp. 20-36

[7] J.J. Carr, *Practical Antenna Handbook*, 2nd edition, TAB Books, New York, 1994

[8] W.L. Weeks, *Antenna Engineering*, McGraw-Hill Book Company, New York, 1969

Research of the Substrate Characteristics' Influence on the Bandwidth of Rectangular Microstrip Resonator Antennas

Nicola Dodov[1] and Ralitsa Stoyanova[2]

Abstract – **In this paper it is made a research of the frequency bandwidth of a rectangular microstrip resonator antenna depending on relative dielectric constant (ε_r) and the height (h) of the used substrate. The results give the possibility to be taken definite engineer decisions when projecting microstrip antenna.**

Keywords – **Microtrip resonator antenna, Frequency bandwidth, Quality factor, Dielectric constant.**

I. Introduction

The idea for microstrip resonator antennas (patch antennas) originates from 1953 but they receive a significant development and application after 1973. That specific kind of antennas has an application in many different areas such as wireless and mobile radio communications and they can also be used for connecting spaceships, air crafts, for security systems, etc. The convenience of these antennas comes of their low cost, easy technological implementation, small weight and aerodynamic profile. Because of their small dimensions and narrow bandwidth, single elements often are projected in large scanning arrays and the number of the elements can reach some thousands. Also the antenna array can be an array with signal processing, which makes these antennas more attractive.

Microstrip resonator antennas as all the others have many advantages, but some disadvantages aren't absent. The main advantages, from technological point of view come from the fact that they obtain mechanical robust and they can be easy mounted over different surfaces. On another side, by varying with the resonators' dimensions and forms and with proper choice of the feeding method, it can be achieved appropriate working mode, optimal pattern and frequency bandwidth. This type of antennas are very flexible when choosing the resonant frequency, polarization, pattern, amplification coefficient and these characteristics can be corrected.

The microstrip antennas' disadvantages come from the fact that the frequency bandwidth is very narrow and limited (exert influence when the antenna is used for scanning), because of the high quality factor Q. Usually thin substrates are used, which makes impossible high power feeding. Also there is spurious radiation, coming from the feeder and low level of polarization purity. Because of the losses in the dielectric material, the efficiency is low. The problem with the narrow bandwidth is solved by different methods such as extension of the substrate height, which leads to efficiency increase, but it also leads to increase of the surface waves.

A basic part of the microstrip antenna is the dielectric substrate and it has to be chosen and measured off very carefully. To a higher extend by substrate choice depends the characteristics and the application of the antenna. In this aspect that papers has the aim to research the frequency bandwidth dependence when varying the substrate relative dielectric constant ε_r and thickness h. The results from this analysis can be considered when taking engineer decisions in microstrip antennas design.

II. Theory

It is well known that the frequency bandwidth B of one structure in higher extend depends on the quality factor Q of this structure. Higher the quality factor is, the frequency characteristic becomes thinner and the bandwidth decreases.

According to the model for defining these two basic characteristics of the microstrip resonator antennas given by [1] it is clearly shown that Q and B depend on a number of factors. The total quality factor Q_t represents the antenna losses and according to [1] the formula is:

$$\frac{1}{Q_t} = \frac{1}{Q_{rad}} + \frac{1}{Q_c} + \frac{1}{Q_d} + \frac{1}{Q_{SW}} , \qquad (1)$$

where Q_{rad} is a quality factor, due to radiation losses; Q_c – quality factor, due to conduction losses; Q_d – quality factor, due to dielectric losses; Q_{SW} – quality factor, due to surface waves.

For every single quality factor above there is a definite expression. The quality factor, which expresses the losses caused by radiation, according to [1], is written as:

$$Q_{rad} = \frac{2 \cdot \omega_0 \cdot \varepsilon_r}{h \cdot G_t / l} \cdot K , \qquad (2)$$

where ω_0 is the circus frequency, on which the resonator works; ε_r – substrate relative dielectric constant; h – substrate height (thickness); G_t/l – total conductance per unit length of the radiating aperture;

$$K = \frac{\iint\limits_{area} |E|^2 dA}{\oint\limits_{perimeter} |E|^2 dl} , \qquad (3)$$

[1]Nicola Dodov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ndodov@vmei.acad.bg

[2]Ralitsa Stoyanova is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: r_stojanova@softhome.net

where E is the electric field.

The quality factor Q_c, due to the conduction losses, according to [1], can be expressed as:

$$Q_c = h \cdot \sqrt{\pi \cdot f_0 \cdot \mu \cdot \sigma} \,, \tag{4}$$

where f_0 is the resonant frequency of the resonator; μ – absolute magnetic constant of the dielectric material; σ – conductivity of the conductors assosiated with the patch and ground plane.

The quality factor Q_d due to the conductance losses, it is given by [1] as:

$$Q_d = \frac{1}{tg\delta} \,, \tag{5}$$

where $tg\delta$ is the loss tangent of the substrate material.

The frequency bandwidth B is inversely proportional to the total quality factor Q_t of the antenna and according to [1] it is expressed by:

$$\frac{B}{f_0} = \frac{1}{Q_t} \,. \tag{6}$$

With the account of the impedance matching at the input terminals of the antenna, according to [1], Eq. 6 modifies to:

$$\frac{B}{f_0} = \frac{VSWR - 1}{Q_t \cdot \sqrt{VSWR}} \,, \tag{7}$$

where $VSWR$ is voltage standing wave ratio.

The exposition that was made shows that the method for defining the frequency bandwidth by quality factor Q_t is very complicated. That's why a simple expression is used – Stutzman formula, which according to [2], is given by:

$$B = 3.77 \frac{\varepsilon_r - 1}{{\varepsilon_r}^2} \cdot \frac{W}{L} \cdot \frac{h}{\lambda_0} \,, \tag{8}$$

where W is the width of the patch; L – length of the patch; λ_0 – central wavelength of the resonator.

Fig. 1 shows the dimensions and the profile of the patch.

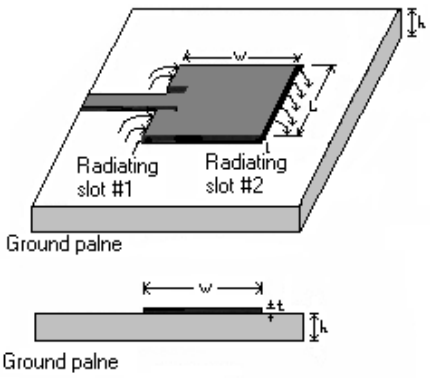


Fig. 1. Microstrip resonator.

According to the transmission line model, described in [1], the dimension W (width of the patch) is a function of the frequency and the dielectric constant. According to [1] and [3] the expression is:

$$W = \frac{C}{2 \cdot f_0} \sqrt{\frac{2}{\varepsilon_r + 1}} \,, \tag{9}$$

where C is the light speed.

When the length L is determined, the influence of the "edge effect" is taken in account and according to [1] and [3], can be written as:

$$L = \frac{C}{2 \cdot f_0 \sqrt{\varepsilon_{reff}}} - 2 \cdot \triangle L \,, \tag{10}$$

where ε_{reff} is the effective relative dielectric constant; L – extended incremental length as a result of the "edge effect".

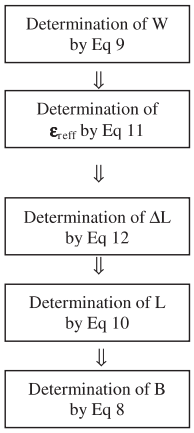
The effective relative dielectric constant reff, according to [1] and [3], it is given by:

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{\sqrt{1 + 12 \cdot h/W}} \,. \tag{11}$$

On the other hand the extended incremental length $\triangle L$, according [1] and [3], can be written as:

$$\triangle L = 0.412 \cdot h \cdot \frac{(\varepsilon_{reff} + 0.3)\,(W/h + 0.264)}{(\varepsilon_{reff} - 0.258)\,(W/h + 0.8)} \,. \tag{12}$$

The aim of the results, received with the help of simulation analysis, is to show how the frequency bandwidth B changes when different kinds of substrates with various height h and relative dielectric constant ε_r are used. The initial parameters are: resonant frequency f_0=10GHz, ε_r and h that are varying in definite limits. The algorithm is as follows:



The results can be explained with the physical processes in the microstrip line and according to [4] they are as follows:

1. As the substrate height h increases, the quality factor Q becomes lower and therefore the frequency bandwidth becomes larger.

2. As the dielectric constant ε_r increases, the electric field concentrates deeper into the microstrip line. As the substrate height h increases, the quasitransverse electromagnetic wave concentration becomes weaker.

3. As the dielectric constant ε_r becomes larger, the influence of the height h decreases.

These fundamental conclusions can be used for an explanation of the results.

1) Fig. 2 shows that the frequency bandwidth decreases with the increase of ε_r and also that for higher values of ε_r the alteration rate decreases.

2) The increase of the bandwidth increases proportional to the increase of the substrate height h.

3) Fig. 3 shows that with the increase of the substrate height h, the bandwidth B increases linearly.

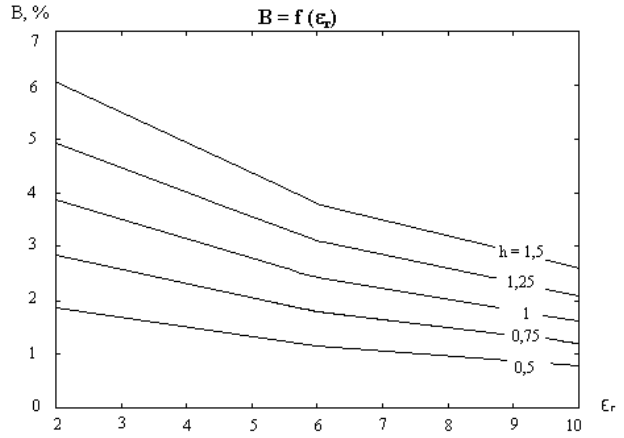


Fig. 2. Results from the research of the bandwidth (B,%) as a function of the substrate relative dielectric constant (ε_r) and a parameter – height (h,mm).

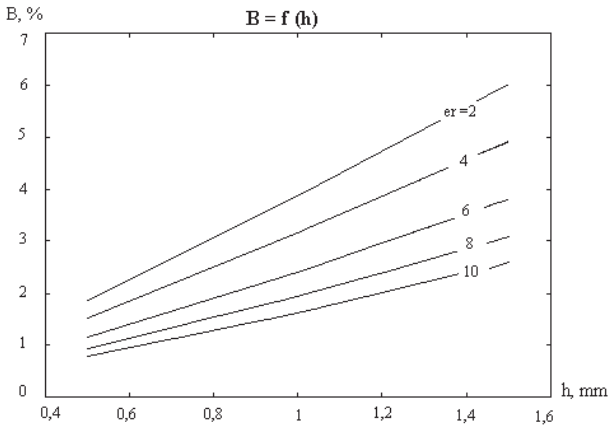


Fig. 3. Results from the research of the bandwidth (B,%) as a function of the height (h,mm) and a parameter – substrate relative dielectric constant (ε_r).

III. Recommendations

The following recommendations can be made on the base of Figs. 2 and 3:

1. When it is necessary to expand the resonator bandwidth, there has to be used substrates with lower values of ε_r and higher thickness.

2. When the frequency is resonant and a narrow frequency bandwidth is necessary and it is desirable to use thin substrates. The value of the ε_r depends on the frequency band and the technological requirements.

3. Microstrip resonator antennas are with narrow bandwidths. The standard width is around $(1 \div 2)$%.

References

[1] Balanis C. A. *Antenna Theory (Analysis and Design)*, John Wiley and Sons, Inc., New York, 1997

[2] Warren Stutzman, Gary Thiele, *Antenna Theory and design*, John Wiley and Sons, Inc., 1998

[3] Drabowitch S. A. Papiernik. *Modern Antennas*, Chapman and Hall, London, 1998

[4] Kai Fong, Wel Chen, *Advances in Microstrip and Printed Antennas*, John Wiley and Sons, Inc., New York, 1997

[5] Hansen R. C. *Phased Array Antennas*, John Wiley and Sons, Inc., New York, 1998

[6] Mailloux R. J. *Phased Array Antenna Handbook*, Artech House, Boston, London 1994

[7] Ramesh G. B. Prakash, *Microstrip Antenna Design Handbook*, Artech House, Boston, London, 2001

Evaluation of Edge Effects in Measuring of a Rectangular Microstrip Resonator Antenna

Nicola Dodov[1] and Mincho Pankov[2]

Abstract – **It is well known that in measuring of a microstrip resonator antenna it is necessary to pay attention on the "edge effect", which reflects in determining the resonator length (L). In this paper a research is made of the varying of the extended incremental length ($\triangle L$) depending on the substrate characteristics - relative dielectric constant (ε_r) and the height (h).**

Keywords – **Edge effect, Relative dielectric constant, Height, Extended incremental length.**

I. Introduction

Microstrip resonator antennas are used in spacecraft and aircraft, in satellite connections, where small weight, cost, performance technology and aerodynamic profile are very important. They have other applications in mobile and wireless communications where a great amplification, scanning possibility and low spurious radiation are required.

The main advantages of the microstrip antennas are:
– varying with the form of the antenna;
– possessing mechanical robust;
– easy correction of the characteristics;
– easy matching;
– a possibility for polarizations regulation;
– working on two polarization with one structure;
– small weight and dimensions.

The disadvantages of the microstrip resonator antennas are as follows:
– not resistant on high power (power limited to 1 or 2 W);
– narrow frequency bandwidth (about 1 to 2%);
– bad polarization purity;
– low efficiency;
– spurious feed radiation.

These disadvantages can be overcome by using different geometric shapes, shields, materials and different feeding methods.

For microstrip antennas' dielectric substrates are used dielectric materials with low dielectric constant, low losses and good mechanical characteristics.

The aim of the present papers is to research the dependence of the edge effect when varying the substrate dielectric constant (ε_r) and the height (h), which is very important when projecting such type of antennas.

[1]Nicola Dodov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ndodov@vmei.acad.bg

[2]Mincho Pankov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: m_pankov@softhome.net

II. Theory

The edge effect is a result of the fringing of the electric lines at the end of the microstrip resonator, as shown in Fig. 1.

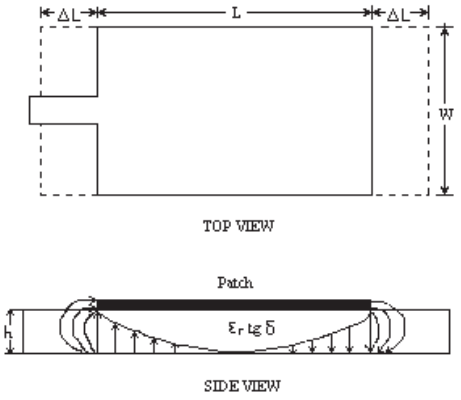


Fig. 1. Microstrip resonator.

The cause for that electric field configuration is the implementation of the boundary conditions in electrodynamics, namely that the electric lines are always open lines and they start and finish perpendicularly to the metal surface. It is clearly shown in Fig. 1 that the E vector lines at the edge and out of the edge of the microstrip resonator can not finish at the top surface of the metallic strip without fringing their shape. This fringing is the reason the effective resonator dimensions which define resonant frequency to be bigger than the real technical dimension (L) with an extended incremental length $\triangle L$.

The edge effect exhibition is due to the fact that the patch dimensions are finite along his length and width. The extended incremental length $\triangle L$, according to [1], depends on the effective dielectric constant of the dielectric substrate and the height and is given with the following formula:

$$\triangle L = 0.412 \cdot h \cdot \frac{(\varepsilon_{reff} + 0.3)\,(W/h + 0.264)}{(\varepsilon_{reff} - 0.258)\,(W/h + 0.8)}\,, \qquad (1)$$

where h is the height of the substrate; ε_{reff} – effective dielectric constant; W – width of the patch.

The patch width, according to [1], is a function of the wavelength and the dielectric constant and it is given with the following expression:

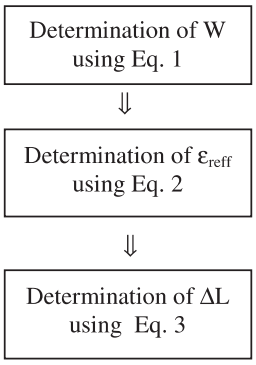
$$W = \frac{\lambda_0}{2}\sqrt{\frac{2}{\varepsilon_r + 1}}\,, \qquad (2)$$

where λ_0 is the central resonant wavelength, which the resonator works in; ε_r relative dielectric constant of the substrate.

The effective dielectric constant, according to [1], depends of the substrate dielectric constant and height and the width of the patch:

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2\sqrt{1 + 12 \cdot \frac{h}{W}}} . \qquad (3)$$

The simulation analysis, which was made, has the aim to show the variation of the dimension $\triangle L$ as function of the given parameters ε_r, h and the resonant frequency f_r = 10 GHz. The algorithm of the research is given bellow:



III. Results and Conclusions

The results can be explained with the physical processes in the microstrip line and according to [3] they are as follows:

1. As the dielectric constant ε_r increases, the electric field concentrates more in the microstrip line.

2. As the substrate height h increases, the quasitransverse electromagnetic wave concentration becomes weaker.

3. As the substrate height h increases, the quality factor Q becomes lower and subsequently the frequency bandwidth becomes larger.

4. As the dielectric constant ε_r becomes larger, the influence of the height h decreases.

The results can be explained with the physical processes in the microstrip line and according to [3] they are as follows:

1) Fig. 2 shows that with the increase of ε_r, the effective extended incremental length $\triangle L$ decreases and the decrease rate gets lower with the increase of ε_r.

2) It can be seen in Fig. 3 that the increase of the extended incremental length $\triangle L$ as a function of the height h, is linear. But as higher ε_r is, lower the increase rate of $\triangle L$ is lower and ε_r influence decreases.

The conclusions that were made correspond completely with the well-known statement, according to [6], that the higher ε_r is, the quasi-stationary field is more stuck around the microstrip line. When ε_r value is very high, an effect of "saturation" takes place and it is connected with the constant delay coefficient $\kappa = V_\Phi / C$, which is near to the value $1/\sqrt{\varepsilon_r}$.

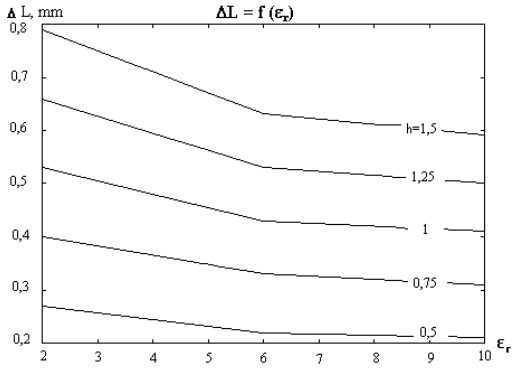


Fig. 2. Results from the research of the extended incremental length ($\triangle L$, mm) as a function of the substrate relative dielectric constant (ε_r) with parameter - height (h, mm).

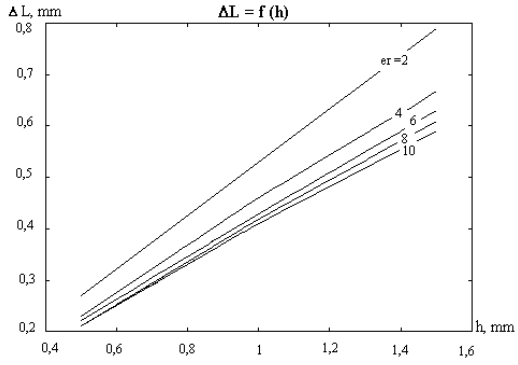


Fig. 3. Results from the research of the extended incremental length ($\triangle L$, mm) as a function of the substrate height (h, mm) with parameter - dielectric constant (ε_r).

IV. Recommendations

The following recommendations can be made on the base of Figs. 2 and 3:

1.]When it is necessary in design to achieve lower value of the extended incremental length $\triangle L$, there has to be used small substrate thickness h. In this case it is not necessary to use relative dielectric constant ε_r with a high value.

2. A smaller value of the effective extended incremental length $\triangle L$ can also be achieved with a high value of ε_r, when it is required by the frequency band and the microwave technology, for instance using of a ceramic substrate.

References

[1] Balanis C. A. *Antenna Theory (Analysis and Design)*, John Wiley and Sons, Inc., New York, 1972.

[2] Drabowitch S. A. A. Papiernik *Modern Antennas*, Chapman and Hall, London, 1998

[3] Kai Fong, Wel Chen, *Advances in Microstrip and Printed Antennas*, John Wiley and Sons, Inc., New York, 1997

[4] Hansen R. C. *Phased Array Antennas*, John Wiley and Sons, Inc., New York, 1998

[5] Mailloux R. J. *Phased Array Antenna Handbook*, Artech House, Boston, London 1994

[6] Ramesh G. B. Prakash, *Microstrip Antenna Design Handbook*, Artech House, Boston, London, 2001

Influence of Tropospheric Duct Parameters Changes on Microwave Path Loss

I. Sirkova[1] and M. Mikhalev[2]

Abstract – **This work studies the influence of essential evaporation and surface-based ducts parameters on microwave path loss in the context of a communications link. Discussed is the need to use range dependent refractivity profiles in order to increase the accuracy in path loss prediction especially in coastal regions.**

Keywords – **Tropospheric ducts, microwave propagation modeling, wireless network planning.**

I. Introduction

This work follows the investigation made on tropospheric ducting and its possible effects on wireless communications systems design reported in [1]. Coastal areas worldwide are known to be especially ”rich” in super refractive layers and ducts that affect microwave propagation [2]. In this report the attention is focused on a) evaporation duct, due to evaporation from sea surface, and b) surface-based duct, caused, for instance, by advection. Ducts due to evaporation are practically almost present at lower latitudes [3], their depths increasing during the summer months and during the daytime. Even in moderate latitudes evaporation duct is not an occasional event. In [4] are reported refractometer measurements data accomplished in years 1973-1976 at the north part of Bleak sea indicating super refraction and ducting conditions in the layer 0-40 m above sea level during 20-25% of the time with maximum in July. Advection ducts arise when warm air from a dry landmass moves over the cooler sea water. Surface ducts of such nature appear about 15% of the time worldwide [5]. Advection may reinforce a preexisting evaporation duct and increase its depth. Even though stable formations, ducts suffer seasonal and diurnal variations [6] especially in the coastal zone where the sharp contrast between land and sea contributes to temporal and spatial variability. This leads to highly variable propagation conditions and thus affects significantly radio com-munications links performance.

This report studies the influence of the changes of the duct parameters on path loss assessment. For range independent refractivity models path loss calculations is made using the parabolic equation (PE) electromagnetic field propagation model based on finite element numerical scheme as described in [1]. When range dependent refractivity profiles are applied the Advanced Propagation Model (APM) routines of the SPAWAR Systems Center, San Diego, CA, USA, are used. This code is based on the radio physical optics model and the Terrain Parabolic Equation Model and makes essentially use of the split-step Fourier PE method [7]. Horizontally polarized Gaussian beam antenna with frequency 2 GHz, 2.5 GHz and 5.8 GHz is used and smooth perfectly conducting underlying surface is assumed. The limit values of the duct parameters have been chosen following the values reported in [6] and [8].

[1]I. Sirkova is with the Institute of electronics, Bulgarian Academy of Sciences, blvd. ”Tzarigradsko chaussee” 72, 1784 Sofia, Bulgaria, E-mail: irina@ie.bas.bg

[2]M. Mikhalev is with the Institute of electronics, Bulgarian Academy of Sciences, blvd. ”Tzarigradsko chaussee” 72, 1784 Sofia, Bulgaria, E-mail: matam@ie.bas.bg

II. Results and Discussion

The evaporation duct is modeled using the log-linear model [9]:

$$M(z) = M_0 + 0.13\left[z - z_d \ln\left(\frac{z+z_0}{z_0}\right)\right], \qquad (1)$$

where M is modified refractivity, z is altitude in m, M_0 is the value of modified refractivity at the sea surface, z_d is evaporation duct height in m, and z_0 is the aerodynamic roughness parameter assumed here to be 1.5x10^{-4} m. When the electromagnetic field is calculated in a single frequency the parameter M_0 can be set to an arbitrary constant without affecting the interference pattern in height, thus the evaporation duct model is entirely governed by z_d. The refractivity in the case of surface-based ducts is modeled by bilinear model with important parameters duct height z_d and M-deficit $dM = M(z_d) - M_0$ (once again, the offset of the profile is not important). The slope above the inversion is set to 0.118 M-units/m (standard troposphere). This simple model is rough but allows pointing out the influence of basic parameters.

To illustrate the influence of z_d in the case of evaporation duct a series of duct height measurements taken over a 100-minutes period by means of a series of atmospheric sensors [10] is used, Table 1. During this 100-minutes time interval z_d has suffered significant changes. For the evaporation duct from Table 1 two frequencies, f=2.5 GHz and f=5.8 GHz, and two links are investigated over distances of r=3 km, 5 km and 10 km: 1) the first link has transmitter height z_t=30 m

Table 1. Evaporation duct height variation with time

z_d, m	13.3	9	6.8	10.8	9.6	9.6	10
Time	5	10	15	20	25	30	35
z_d, m	7.5	10	11	11.5	14.2	10	16
Time	40	45	50	55	60	65	70
Z_d, m	21	18	24	21	15.2	17.5	
Time	75	80	85	90	95	100	

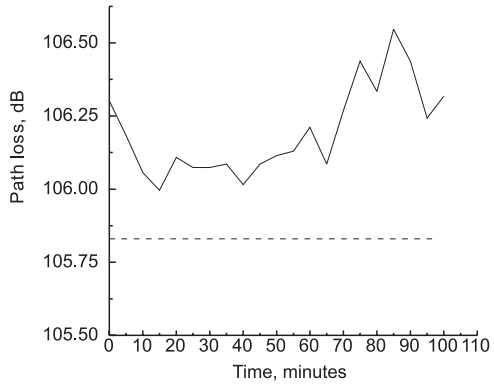


Fig. 1. Path loss vs time for f=2.5 GHz, r=3 km, link 1)

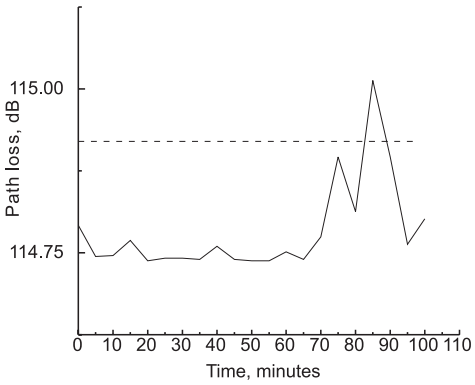


Fig. 5. Path loss vs time for f=2.5 GHz, r=10 km, link 1)

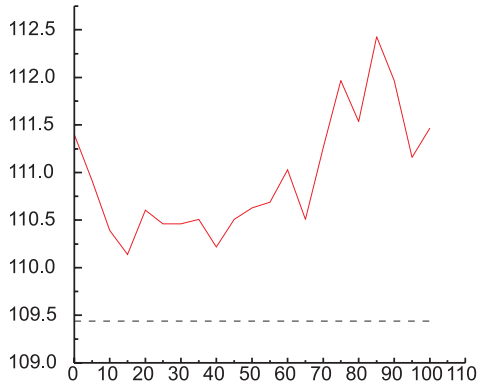


Fig. 2. Path loss vs time for f=2.5 GHz, r=3 km, link 2)

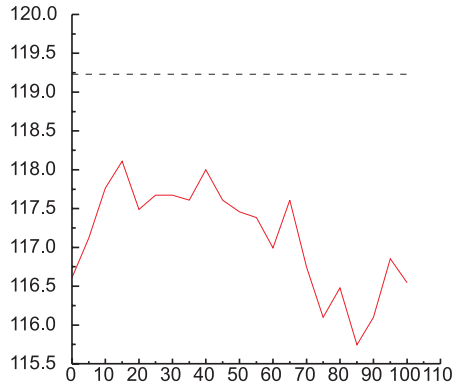


Fig. 6. Path loss vs time for f=2.5 GHz, r=10 km, link 2)

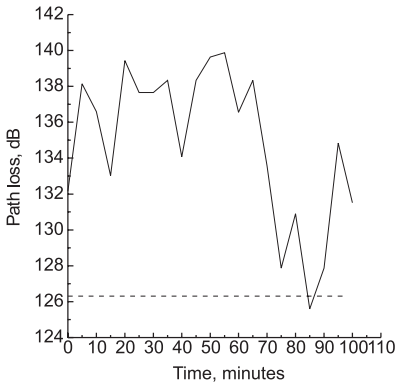


Fig. 3. Path loss vs time for f=2.5 GHz, r=5 km, link 1)

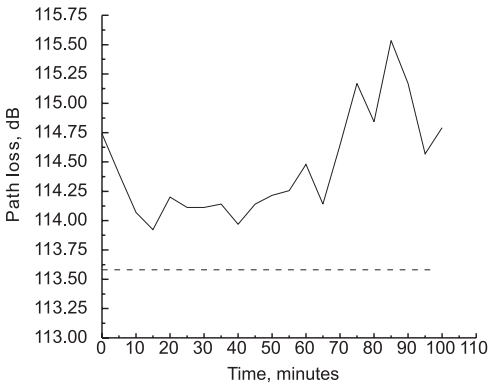


Fig. 7. Path loss vs time for f=5.8 GHz, r=3 km, link 1)

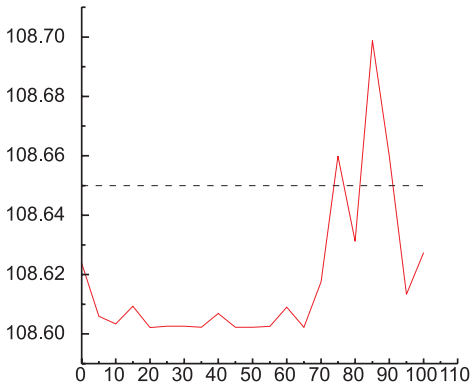


Fig. 4. Path loss vs time for f=2.5 GHz, r=5 km, link 2)

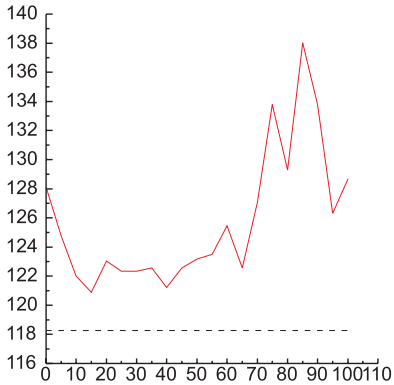


Fig. 8. Path loss vs time for f=5.8 GHz, r=3 km, link 2)

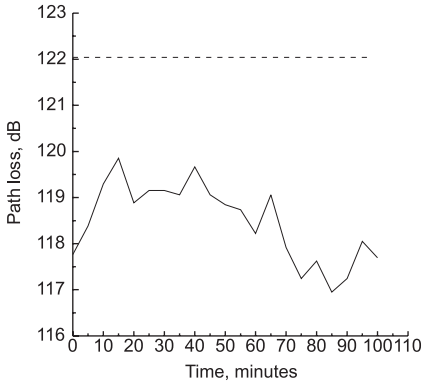


Fig. 9. Path loss vs time for f=5.8 GHz, r=5 km, link 1)

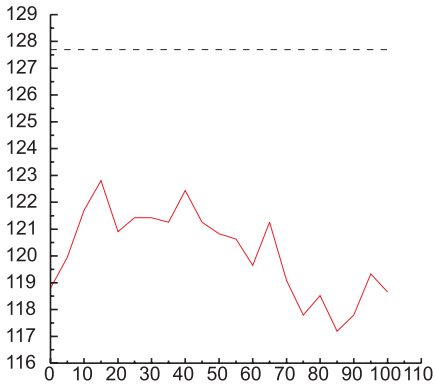


Fig. 10. Path loss vs time for f=5.8 GHz, r=5 km, link 2)

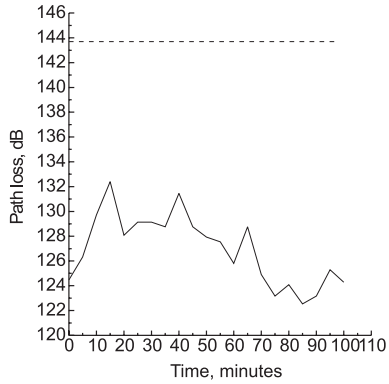


Fig. 11. Path loss vs time for f=5.8 GHz, r=10 km, link 1)

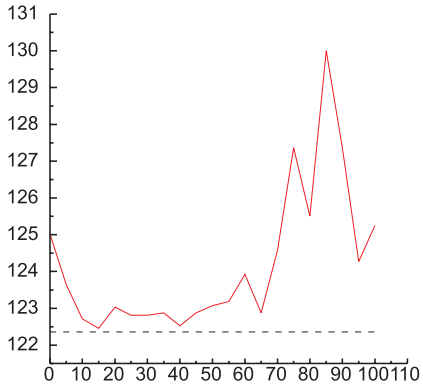


Fig. 12. Path loss vs time for f=5.8 GHz, r=10 km, link 2)

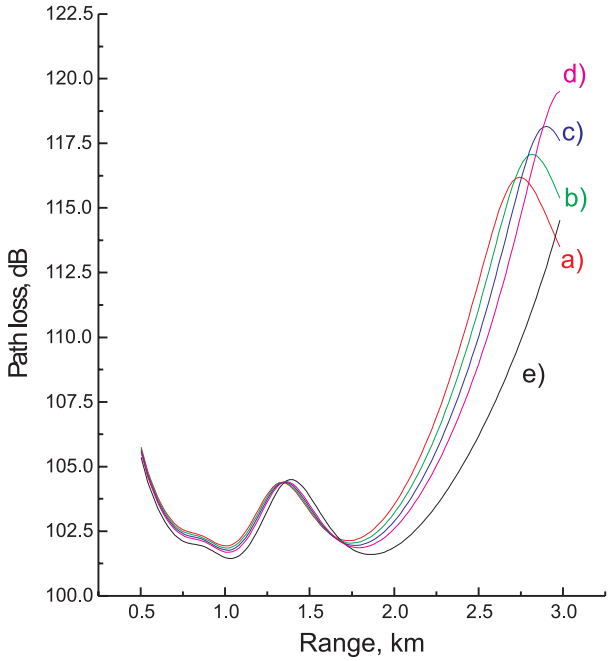


Fig. 13. Surface-based duct, influence of dM.

(z_t is always above the duct), receiver height z_r=10 m; 2) the second link has z_t=15 m (for this link z_t is submerged in the duct during 1/3 of the time), z_r=10 m. For both links z_r is within the duct during 2/3 of the time. For all cases antenna beam-width=2° and tilt=0° are used. Dashed line indicates standard troposphere path loss.

Figs. 3, 8, 12, 6, 9, 10, 11 show strong and moderate influence of the changes in z_d on path loss. In all these figures path loss values differ markedly from its value for standard troposphere. In Figs. 1, 2, 4, 5 and 7 the influence of z_d changes is quasi-negligible, but even here path loss values differ (except in Figs. 4 and 5) from the case of standard troposphere. Path loss decrease in longer distances due to ducting as well as its fluctuations may cause interference and, in an worse case, reception from an unwanted link and lack of signal from the wanted link. Comparison of Figs. 9 and 10 shows the duct decreased the path loss for both links below the value for standard troposphere but for link 2) this decrease is more important. From Figs. 3 and 4 it is clear that for all z_d (except at 85th minute) the path loss for link 1) is significantly increased above the value for standard troposphere whereas the influence of the duct on link 2) is negligible. Suppose, link 1) is the desired link and link 2) is the unwanted one: the above mentioned examples will aggravate the suppression of link 2) signal. To avoid such cases additional interference reduction techniques could be used or anomalous propagation conditions should be accounted for when link budget is calculated.

Figs. 13 and 14 illustrate the influence of the changes of the surface-based duct parameters z_d and dM. Fig. 13 shows path loss for surface-based duct with z_d=50 m and changing dM: a) dM=10 M-units; b) dM=20 M-units; c) dM=30 M-units; d) dM=40 M-units; e) dM=70 M-units. The other parameters are: frequency f=2 GHz, transmitter height z_t=20 m, receiver height z_r=10, beamwidth 1° (no

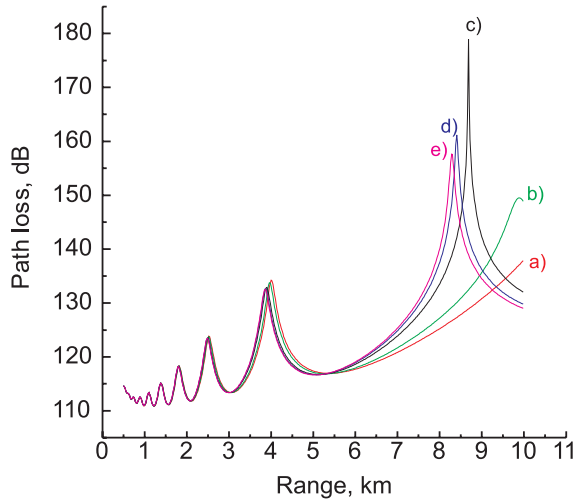


Fig. 14. Influence of z_d: a) z_d=50 m; b) z_d=60 m; c) z_d=100 m; d) z_d=130 m; e) z_d=150 m, dM=10.

tilt).

The path loss vs range for f=5.8 GHz, z_t=20 m, z_r=10 m and changing z_d (dM=10 M-units) is shown in Fig. 14: a) z_d=50 m; b) z_d=60 m; c) z_d=100 m; d) z_d=130 m; e) z_d=150 m. This Figure (as well as Fig. 15) is computed for antenna beam-width=1°, tilt=0°. Clearly seen are the differences in path loss provoked by the changes in z_d even in the case of week ducts.

Fig. 15 refers to the case of range dependent ducts. It is well known [6] that refractivity profiles over sea and over land differ. Thus, a more realistic description of the propagation conditions along a mixed land-sea path will be the use of two (or more) M-profiles: one at the transmitter site and another close to the receiver. Fig. 15 shows path loss for f=2 GHz, z_t=25 m, z_r=10 m and: a) range independent surface-based duct with z_d=100 m, dM=60 M-units over the entire distance of 10 km; b) surface-based duct with z_d=100 m, dM=60 M-units at the transmitter and standard troposphere at distance of 2 km; c) surface-based duct with z_d=100 m, dM=60 M-units at the transmitter and standard troposphere at distance of 5 km. As it is seen from Fig. 15, in coastal regions the influence of the horizontal changes of refractivity could not be neglected.

III. Conclusion

This report presents simulation results on the influence of the essential evaporation and surface-based ducts parameters on the path loss for a microwave link. Shown is the need to use range dependent refractivity profiles in order to increase the accuracy in path loss prediction especially in coastal areas. In these regions communications systems designed without accounting for the refraction and ducting could potentially suffer interference from each other. The correct preliminary assessment of the expected path loss using in situ refractivity data and more precise propagation prediction models will decrease the cost of the planning links and improve their performance.

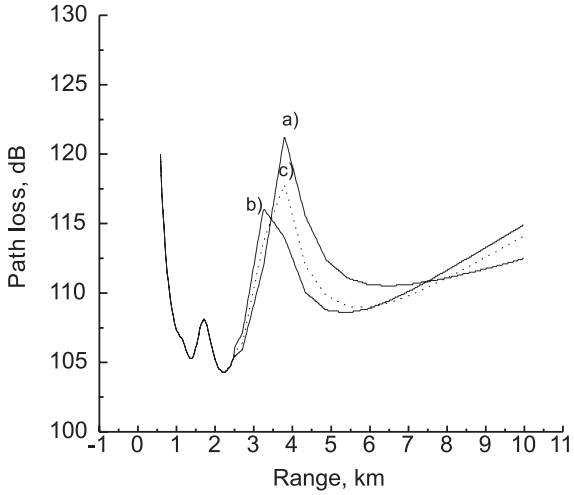


Fig. 15. Range dependent duct, f=2 GHz.

Acknowledgements

The authors are grateful to the SPAWAR Systems Center, San Diego, for making available the APM code.

References

[1] I. Sirkova, M. Mikhalev, "Influence of Tropospheric Ducting on Microwave Propagation in Short Distances", *XXXVIII ICEST*, Sofia, Bulgaria, 2003.

[2] Turton, J. D., D. A. Bennetts, and S. F. G. Farmer, "An introduction to radio ducting", *Meteor. Mag.*, vol. 117, pp. 245-254, 1988.

[3] K. D. Anderson and R. A. Paulus, "Rough evaporation duct (RED) experiment", *Battlespace Atmospherics and Cloud Impact on Military Operations, BACIMO 2000*, Fort Collins, CO, 25-27 April 2000.

[4] N. A. Dorfman, V. A. Kabanov, F. V. Kivva and I. S. Tourgenev, "Refractive index statistical characteristics in above the see layer", *Izv. Acad. Sci. SSSR Fizika Atmosferi i Okeana*, vol. 14, pp. 549-553, 1978 (in russian).

[5] P. Gerstoft, D. F. Gingras, L. T. Rogers and W. S. Hodgkiss, "Estimation of Radio Refractivity Structure Using Matched-Field Array Processing", *IEEE Trans.*, vol. AP-48, no. 3, pp. 345-356, 2000.

[6] B. W. Atkinson, J.-G. Li, and R. S. Plant, "Numerical Modeling of the Propagation Environment in the Atmospheric Boundary Layer over the Persian Gulf", *J. Appl. Meteorology*, vol. 40, pp. 586-603, 2001.

[7] A. E. Barrios "A Terrain Parabolic Equation Model for Propagation in the Troposphere", *IEEE Trans.*, vol. AP-42, pp. 90-98, 1994.

[8] L. T. Rogers, "Likelihood estimation of tropospheric duct parameters from horizontal propagation measurements", *Radio Sci.*, vol. 32, pp. 79-92, 1997.

[9] R. A., Paulus and K. D. Anderson, "Applications of an Evaporation Duct Climatology in the litoral", *Battlespace Atmospherics and Cloud Impact on Military Operations, BACIMO 2000*, Fort Collins, CO, 25-27 April 2000.

[10] A. Kerans, A. Kulessa, G. Woods and J. Hermann, "Evaporation Duct Statistics Around Australia and the West Pacific", *Proc. AP2000*, Davos, Switzerland, April 2000.

Influence of Tropospheric Ducting on Microwave Propagation in Short Distances

I. Sirkova[1] and M. Mikhalev[2]

Abstract – **Tropospheric ducting effects are normally considered to be a long-range phenomenon and propagation prediction models used in cells planning and channel characteristics assessment usually do not account for them. In this work it is shown that trapping layers, existing significant percentage of time in certain regions, can affect the expected signal level by shifting the location of the interference maximums in comparison to the propagation under standard troposphere conditions.**

Keywords – **Tropospheric ducts, microwave propagation modeling, parabolic equation.**

I. Introduction

The increasing demand for more services and better quality in the mobile communications poses higher requirements to the propagation prediction models applied in network planning tools. In addition, the UMTS radio network is known to be more sensitive to the propagation environment than is the GSM network [1]. The improvement in coverage and interference assessment/prediction optimizes the base stations planning and decreases the cost of system deployment and exploitation.

This work investigates the influence of tropospheric ducting on microwave propagation in short distances. Ducting effects are normally considered to be a long-range phenomenon, [2], leading to signal enhancement near and beyond the radio horizon. Thus, the classical and even more sophisticated propagation prediction models used in cells planning and channel characteristics assessment usually do not account for tropospheric super refraction and ducting. But, as reported in [3], ducts can affect the propagation in short ranges in two ways: they provoke a shift of the location of the last interference maximum (in terms of path loss) and decrease of the signal level near the last interference minimum. In this work calculations of path loss versus range for some common cases of surface-based ducts are compared to the path loss obtained assuming standard troposphere. The calculations are performed using the parabolic equation (PE) method in conjunction with a finite element based numerical scheme [4].

[1]I. Sirkova is with the Institute of electronics, Bulgarian Academy of Sciences, blvd. "Tzarigradsko chaussee" 72, 1784 Sofia, Bulgaria, E-mail: irina@ie.bas.bg

[2]M. Mikhalev is with the Institute of electronics, Bulgarian Academy of Sciences, blvd. "Tzarigradsko chaussee" 72, 1784 Sofia, Bulgaria, E-mail: matam@ie.bas.bg

II. Method Description

The PE approximation to the wave equation and its application to the tropospheric propagation problems are well documented [5-7] and here only brief description of the method is given.

As paraxial approximation, PE assumes the problem has some preferred propagation direction, say, the x-axis in a Cartesian coordinate system, and transforms the scalar wave equation in a 3D PE:

$$\frac{\partial u(x,y,z)}{\partial x} = \frac{i}{2k}\left(\frac{\partial^2 u(x,y,z)}{\partial z^2} + \frac{\partial^2 u(x,y,z)}{\partial z^2}\right) + \frac{ik}{2}\left(n^2(x,y,z)-1\right)t(x,y,z)\,, \quad (1)$$

where k is the free-space wave number, n is the refractive index of the troposphere, $u(x,y,z)$ is the reduced function, [6], related to a field component E as $E(x,y,z) = u(x,y,z)exp(ikx)$. Equation (1) accounts only for forward propagating field and is very accurate at angles within 15° of the direction of x-axis, [8].

The advantage of equation (1) is that it can be easily marched in range: the solution at range $x + \triangle x$ is obtained from that in range x, provided the field is known on an initial plane and adequate boundary conditions on the outer boundaries of the integration domain are given. To solve (1) a finite-element based numerical scheme allowing easier boundary conditions implementation, [7], is used. Due to its simplicity, the 2D form of (1) is the most widely used and has been adopted here.

III. Results and Discussion

In order to point out the influence of the ducting all other propagation mechanisms are ignored and a smooth perfectly conducting underlying surface is assumed. Horizontally polarized Gaussian beam antenna with 2 GHz frequency is used. Studied are different surface-based ducts and positions of the transmitting antenna in respect of the trapping layer. Piece-wise linear range independent profiles for the modified refractivity $M(z)$ ($M(z) = 10^6(m(z)-1)$, $m = n + z/a_e$, where a_e is the Earth radius) are used with small, moderate and strong M-deficits. Results for path loss calculations are compared to those, obtained for standard troposphere conditions. Standard troposphere is characterized by a modified refractivity gradient that increases monotonically at a rate of 0.118 M-units per m.

Fig. 1 shows path loss for h_t=20 m, h_r=10 m, beam-width=2° (without tilt), the red curve a) referring to a surface-based duct with thickness Z_d=50 m and M-deficit $\triangle M$=10 M-units; the black curve b) is for standard troposphere conditions. Fig. 2 shows path loss for the same antennas heights but for beamwidth=4° (without tilt), Z_d=100 m and $\triangle M$=30 M-units. On Fig. 3 are shown the results for a 70 m surface-based duct formed by a trapping layer between h=50 m and h=70 m with $\triangle M$=30 M-units, h_r=10 m, beam-width=4° (without tilt) and different h_t: a) h_t=20 m (the curves are shifted of -30 dB from their real position); b) h_t=30 m; c) h_t=50 m (the curves are shifted of +30 dB); d) h_t=60 m (the curves are shifted of +60 dB). Fig. 4 reports

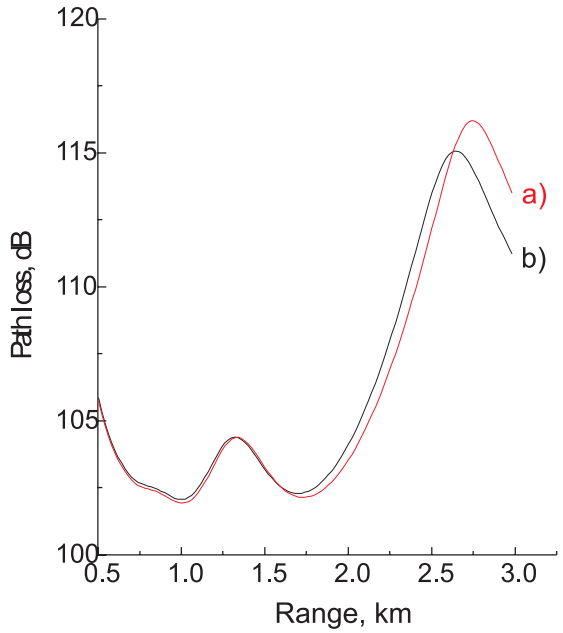


Fig. 1. Path loss for 50 m surface based duct (red curve) and for standard troposphere (black curve), beamwidth=2°, h_t=20 m, h_r=10 m.

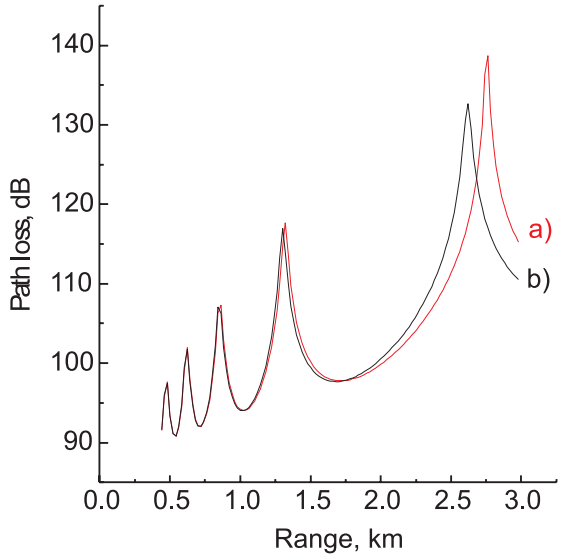


Fig. 2. Path loss for 100 m surface based duct (red curve) and for standard troposphere (black curve), beamwidth=4°, h_t=20 m, h_r=10 m.

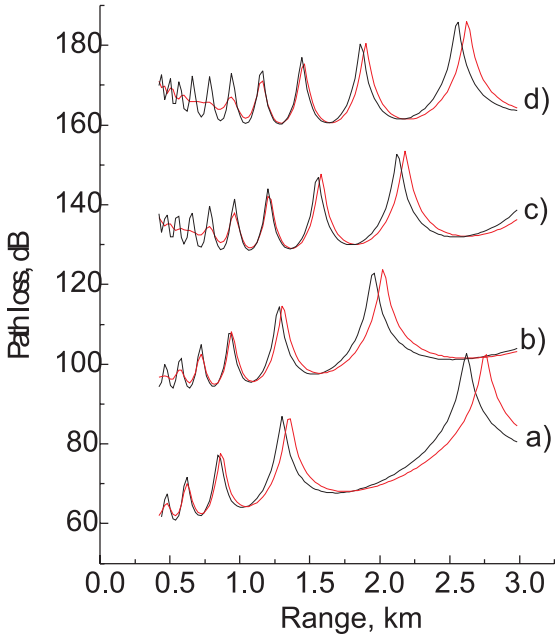


Fig. 3. Path loss for 70 m surface based duct (red curves) and for standard troposphere (black curves), for beamwidth=4°, h_r=10 m and: a) h_t=20 m, b) h_t=30 m, c) h_t=50 m, d) h_t=60 m.

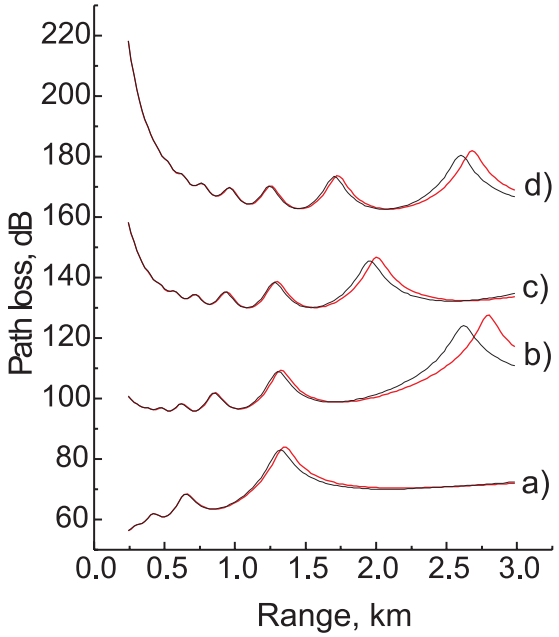


Fig. 4. Path loss for 25 m surface-based duct (red curves) and for standard troposphere (black curves), for beamwidth=2°, h_r=10 m and: a) h_t=10 m, b) h_t=20 m, c) h_t=30 m, d) h_t=40 m.

the results for 25 m surface-based duct ($\triangle M$=10 M-units), h_r=10 m, beam-width=20 (without tilt) and: a) h_t=10 m (shift of -30 dB); b) h_t=20 m, c) h_t=30 m (shift of +30 dB); d) h_t=60 m (shift of +60 dB). The red curves in Figs. 3 and 4 indicate path loss under ducting, the black curves – under standard tropospheric conditions.

From Figs. 1 to 4 it is clearly seen that there is a shift of the location not only of the last interference maximum. As an illustration, in Tables 1 and 2 are given the locations of the

path loss (PL) maximums, their values and the differences in dB for these locations between the cases with duct and the standard troposphere. Table 1 refers to the right maximums of Figs. 1 and 2, Table 2 is for the curves d) from Fig. 3 (starting from the right). The shift of the maximums locations is up to 140 m and 100 m for Figs. 1 and 2, respectively. The shifts for Fig. 3 are of 60, 40 and 20 m respectively. There is significant difference in path loss (the column △, dB in the Tables) for one and the same place under ducting and under standard troposphere conditions.

Table 1. Path loss maximums and their locations for Figs. 1 and 2

		PL_{max}, dB	Distance, km	', dB
Fig. 1	Duct	138.68	2.762	21.62
	Standard trop.	132.64	2.622	14.94
Fig. 2	Duct	116.19	2.742	1.76
	Standard trop.	115.07	2.642	0.19

Table 2. Path loss maximums and their locations for curves d) from Fig. 3

		PL_{max}, dB	Distance, dn	', dB
Duct	1^{st} max	125.97	2.622	10.96
	2^{nd} max	120.52	1.902	6.76
	3^{rd} max	115.38	1.462	2.06
Standard trop. .	1^{st} max	125.74	2.562	10.45
	2^{nd} max	120.36	1.862	7.19
	3^{rd} max	117.01	1.442	2.81

The decrease of the signal level near the interference minimums has also been investigated. As it is seen from the reported Figures, for the studied cases it is negligible (less than 1 dB).

As far as ducting is known to be present about 15% of the time all over the world [3], and even more often in maritime environment, the duct effects inclusion in radio propagation modeling could improve the coverage and interference prediction.

References

[1] M. Coinchon, A. Salovaara and J. Wagen, "The impact of radio propagation predictions on urban UMTS planning", *COST 273 TD(01)041*, Bologna, Italy, 15-17 Oct. 2001.

[2] H. V. Hitney, J. H. Richter, R. A. Pappert, K. D. Anderson and G. B. Baumgartner, Jr., "Tropospheric radio propagation assessment", *Proc. IEEE*, vol. 73, pp. 265-283, 1985.

[3] K. D. Anderson, "Radar detection of low-altitude targets in a maritime environment", *IEEE Trans. Antennas Propagat.*, vol. 43, no. 6, pp. 609-613, 1995.

[4] I. Sirkova and H. E. Hernandez-Figueroa, "Local transparent boundary condition applied to the modeling of tropospheric ducting propagation", *J. Microw. Opt. Technol. Lett.*, vol. 21, no. 5, pp. 343-346, 1999.

[5] J. R. Kuttler and G. D. Dockery, "Theoretical description of the parabolic approximation/Fourier split-step method of representing electromagnetic propagation in the troposphere", *Radio Sci.*, vol. 26, pp. 381-393, 1991.

[6] K. G. Craig and M. F. Levy, "Parabolic equation modeling of the effects of multipath and ducting on radar systems", *IEE Proc.-F*, vol. 138, pp. 153-162, 1991.

[7] I. Sirkova, "On transparent boundary conditions application to the tropospheric ducting propagation modeling", *J. Applied Electromagnetism*, vol. 3, no. 1, pp. 59-78, 2000.

[8] Ch. Zelley and C. Constantinou, "A 3-dimensional parabolic equation applied to the VHF/UHF propagation over irregular terrain", *IEEE Trans. Antennas Propagat.*, vol. AP-47, no. 10, pp. 1586-1596, 1999.

Overview of User Location in Cellular Networks

Vladimir A. Kyovtorov[1]

Abstract – **An additional application for wireless location is presented -the concept of user location in cellular networks. Different radiolocational and triangulation methods for subscriber location are represented. A brief comparison between these contrivances is made.**

Keywords – **subscriber location, wireless location, TOA, OTD, E-OTD, AOA, A-GPS, ICEST 2003.**

I. Introduction

The contemporary telecommunication market is the one of the most modern and dynamically developing market structures. Unremittingly new methods to increase the potentialities of the wireless network are searched. The number of clients grows progressively. It becomes necessaries to additional services. One new service is on the way to give new appearance of the network - subscriber location service. Recently, in a few tested areas, rental cars equipped with location devices and map displays have aided visitors in an unfamiliar territory. Taxi and delivery drivers have utilized location technology in Tokyo to navigate the myriad of streets [1]. Fleet operators use location technology to improve product delivery times and the efficiency of the fleet management process. In cellular telephone networks, location technology could be used for radio resource and mobility management. The Federal Communications Commission (FCC) requires that starting October 1, 2001, all wireless carriers be able to provide the position (or location) of an emergency 911 caller to the appropriate Public Safety Answering Point (PSAP) [2]. Location technologies not requiring new, modified, or upgraded mobile stations (MS) must determine the caller's longitude and latitude within 100 meters for 67% of the emergency calls, and 300 meters for 95% of the calls. If new, modified, or upgraded handsets are required, the requirements are more stringent: 50 meters for 67% of the calls, and 150 meters for 95% of the calls. For successfully resolving this problem many radio location and navigations methods are used.

II. Overview of Existing Location Systems

Location technologies [4] fall into two major categories: **network-based solutions** and **handset (mobile station MS)-based solutions**. It depends on network environment there are some differences and summaries. It is normal to exist hybrid systems, systems that use many different signals (synchronization signal, information, SMS signal, .. etc.) in the different existing world networks [5].

[1]Vladimir A. Kyovtorov is with the Faculty of Communications and Communications Technologies, Technical University, Sofia, Bulgaria, E-mail: vladimir_ak@yahoo.com

The most common system used for cellular location if so named **"Cell-ID based positioning system"** location. The entire service area of a mobile phone network consists of a mosaic or honeycomb of overlapping radio cells, each centered on a base station where the radio antenna is installed. The size of the cells varies according to the intensity of user traffic, and each one is uniquely identifiable by its cell ID. Cell-based location does not require any modifications to the users mobile phones nor to the network. The accuracy is depended on the size of the cell in large cities and conurbations, where the cells are very small and highly concentrated (so-called pico-cells), it is possible to identify a users location to within about 300 meters, In more thinly populated areas and in more isolated regions, a single cell is theoretically capable of covering a radius of up to 35 kilometers. Cell-ID based positioning system have most mobile networks including the most European countries and the Asia Pacific countries [6].

Other radiolocation technology is Positioning using Signal Strength (**SS**). The measurement employs a well-known mathematical model describing the path loss attenuation with distance [7]. This model is reflected in Eq. 1 - Path Loss Attenuation Model:

$$Pl(d) = Pl(d_0) + 10n\log(d/d_0) + X_\sigma \qquad (1)$$

In this equation, $Pl(d)$ (is the path loss as a function of the distance between a transmitter and a receiver. $Pl(d_0)$ is the path loss at a known reference distance d_0, n is 2 for free space and usually higher for wireless channels, and X_σ a zero mean Gaussian random variable reflecting the variation in average received power. X_σ of special interest in this model since it describes the influences of shadowing in an ideal environment (free space and Line-Of-Sight (LOS) propagation) the transmitter lies on a circle centered at the receiver [6].

Some of several fundamental approaches for implementing a radiolocation system including those based on signal-strength are angle of arrival (AOA) and time of arrival (TOA)[8,1]. It is important to note that line-of-sight (LOS) propagation is necessary for accurate location estimates. Angle of Arrival (AOA) techniques estimate the location of a mobile station (MS) by using directive antennas or antenna arrays to measure the AOA at several base stations (BS) of a signal that is transmitted by the MS [8]. Simple geometric relationships are presented in Fig. 1. Consider the error due to multipath propagation, but do not consider angle estimation errors. It is assumed that the MS uses an omnidirectional antenna, so that. In the absence of an LOS signal component, the antenna array will lock on to a reflected signal that may not be coming from the direction of the MS. Even if an LOS component is present, multipath will still interfere with the angle measurement. The accuracy of the AOA method diminishes with increasing distance between the MS and BS due

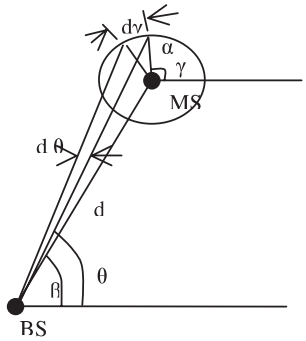


Fig. 1. MSBS geometry assuming a ring of scatterers for macrocells.

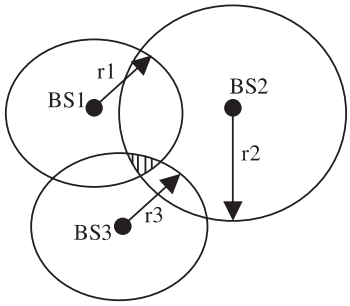


Fig. 2. The location of the MS is constrained to the intersection area (shaded region) of circles of radius $c(\tau_i - \tau)$ centred at each BS.

to fundamental limitations of the devices used to measure the arrival angles as well as changing scattering characteristics. For macrocells, scattering objects are primarily within a small distance of the MS, since the BSs are usually elevated well above the local terrain [8,10]. Consequently, the signals arrive with a relatively narrow AOA spread at the BSs (Fig. 1 [1]). For microcells, the BSs may be placed below rooftop level. Consequently, the BSs will often be surrounded by local scatterers such that the signals arrive at the BSs with a large AOA spread. Thus, while the AOA approach is useful for macrocells, it may be impractical for microcells.

Other systems to measurement position of an object are time based systems. It uses a Time-Of-Arrival (TOA) method. The position of an object is found by measuring the propagation time of a signal travelling from a mobile station to a fixed transceiver or vice versa. Geometrically, this provides a circular locus centered at the transceiver. In order to be able to resolve this location ambiguity for a two-dimensional environment, two more measurements have to be made. A TOA system has some disadvantages. Firstly, it requires all transceivers (either at the network-side or the object itself) to have precisely synchronized clocks. Secondly, multipath propagation caused by signal reflections has a strong influence on the accuracy of the location estimate and can be overcome only by employing sophisticated techniques that screen out unwanted portions of the original signal [1,11]. On the basis of that is development some methods that can increase the timing measurements [12,13] – Fourth Order Cumulating Estimation of Signal Parameters via Rotational Invariance Techniques (FOC-ESPIRT) approaches [13]. The approaches based on higher-order cumulates are powerful in suppressing the Gaussian noise. The Generalized Successive Interference Cancellation (GSIC) is a computationally effective approach when the multipath channel is long and sparse [12].

The idea behind the Time-Difference-Of-Arrival (TDOA) method is to determine the relative position of a transceiver by examining the difference in time rather then the absolute arrival time. A straightforward method of TDOA estimation is to form the cross-correlation between signals received at a pair of BSs. Suppose that the signal $d(t)$ is received at $\mathrm{BS_A}$ corrupted by noise n_A such that $s_A(t) = d(t) + n_A(t)$ The same signal is received at $\mathrm{BS_B}$ with a delay of D and also corrupted by noise $n_B(t)$, giving $s_B(t) = d(t-D)+n_B(t))$.

The cross-correlation function of these signals is:

$$C_{A,B}(\tau) = \frac{1}{T}\int_0^{\infty} S_A(t)S_B(t+\tau)\mathrm{d}t \qquad (2)$$

The TDOA estimate is the value τ that maximizes $C_{A,B}$. This approach requires the analog signals $s_A(t)$ and $s_B(t)$ to be digitalized and transmitted to a common processing site. Also, a strict time reference is required at each BS. In the IS-95A CDMA standard, all BSs are referenced to a systemwide time that uses the GPS time scale [1]. Because of properties of the PN codes, the TOA estimates can be derived from the pseudo-noise (PN) code acquisition and tracking algorithms employed in Spread Spectrum (SS) receivers [15]. Two approaches are generally used to calculate the location of an MS from TOA or TDOA estimates. One approach uses a geometric interpretation to calculate the intersection of circles (Fig. 2) or hyperbolas (Fig. 3), depending on whether TOA or TDOA is used. This approach becomes difficult if the hyperbolas or circles do not intersect at a point due to time measurement errors. A second approach calculates the position using a nonlinear least-squares (NL-LS) solution [1], which is a more statistically justifiable approach. The algorithm assumes that the MS, located at (x_0, y_0), transmits its sequence at time τ_0. The N BS receivers located at coordinates (x_1, y_1), (x_2, y_2),...(x_N, y_N) receive the sequence at times τ_1, τ_2,...τ_N. As a performance measure, we consider the function[1]:

$$f_i(x) = c(\tau_i - \tau) - \sqrt{(x_i - x)^2 + (y_i - y)^2}\,, \qquad (3)$$

where c is the speed of light, and $x = (x, y, \tau)^T$. This function is formed for each BS receiver, $I = 1, ...N$, and all the $f_i(x)$ could be zero with the proper choice of x, y, and τ. However, the measured values of the arrival τ_i times are generally in error due to multipath and other impairments, and non-LOS propagation introduces errors into the range estimates that are derived from the arrival times.

In a GSM network a (Base Station) BS sends synchronization information (i.e. training sequence in the synchronization burst transmitted on the synchronization channel) to a mobile station on how to advance their frame timing to ensure correct frame synchronization with the serving BS. After two consecutive inter-cell handovers, three timing advances are known to the network and therefore provide sufficient information to compute the mobile phones position. Additional handovers may contribute to the improvement of

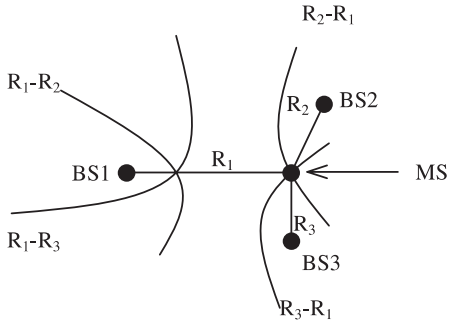


Fig. 3. Principals of Time-Difference-Of-Arrival (TDOA) (Hyperbolical geometric interpretation)

the location estimate. The positioning method is referred to as **Timing Advance** (TA) and can be classified as a TOA technique, and the nature of the architecture is network-based. The system can be implemented without modifications to the current generation of GSM mobile stations [6].

The **Observed Time Difference Positioning (OTD)** employs TDOA measurements in order to determine the position of a mobile station. Through TA, the latter has information about the propagation delays to different BS and can therefore detect the time difference between them. The system can either work mobile-based or network based depending on whether the GSM terminal calculates its position from two independent measurements or if the information is passed on to the network. For the latter, the network computes the location estimate by solving the hyperbolic Eq. 3 presented in the preceding chapter. A major drawback of the positioning concepts based on GSM signaling aspects is their limited accuracy. The evaluation of positioning architectures based on TAs and OTDs resulted in a resolution f 554 m given LOS propagation [5] and may be further degraded by multipath propagation effects. The given accuracy figure is a direct result of the limited bit resolution of TA and OTD measurements. Since the this accuracy specification will likely not meet the requirements for emergency location reporting in the U.S., alternative approaches have to be undertaken. This could be achieved, for example, by overlaying an autonomous TDOA positioning system. The method to measure the Observed Time Difference (OTD) on a terminal has been used since radar systems were first invented over 50 years ago and has been successfully adapted to E-OTD by the major standardization bodies for cellular networks. E-OTD method can be related to both network based and terminal based positioning techniques. If the BSs are not synchronized, such a system operates by placing fixed reference measuring points. E-OTD seems to be a promising candidate for enabling positioning at current and future mobile network systems (in UMTS networks this technology is called OTDOA (Observed Time Difference of Arrival) [9].

An obvious idea is to use the Global Positioning System (GPS) [14,16] and incorporate GPS receivers into mobile phones, especially considering substantial increase in GPS accuracy after the May 2000 removal of deliberately introduced errors through Selective Availability (SA) [4,9]. This can successfully incorporate with Handset-based techniques. Assisted-GPS (A-GPS) technology overcomes the downsides of the conventional GPS solution, and achieves high location accuracy at reasonable cost [14,4]. The assistance to the mobile phone trying to determine its own location comes from the network over the air-interface, and this distributed approach leads to performance levels that exceed those of conventional GPS. What makes this technology work so well is that the wireless network, using its own GPS receivers, as well as an estimate of the mobiles location down to cell/sector, can predict with great accuracy the GPS signal the handset will receive and convey that information to the mobile. With this assistance the size of the search space is greatly reduced, and the time-to-firstfix (TTFF)[4, 14] shortened from minutes to a second or less. In addition, an A-GPS receiver in the handset can detect and demodulate signals that are order of magnitude weaker than those required by conventional GPS receivers. Only a partial GPS receiver is required in the handset to achieve this functionality, but legacy terminals cannot be used and new handsets are required for the technology to operate. The A-GPS technology concept is shown in Fig. 4. The main system components are a wireless handset with partial GPS receiver; AGPS server with reference GPS receiver that can ”see”the same satellites as the handset (DGPS service can be used as well); and wireless network infrastructure, that is, base stations and a mobile switching center (MSC)[4].

The classification in Table 1[4] gives short overview and description in location technologies. It is made according to where signals are measured, since subsequent calculations for location determination can be done anywhere in the system. (In the Table, BS denotes a wireless base station, MS a mobile handset, PDE is position-determination equipment, RTD is the real time difference, and LMU stands for Location Measurement Unit).

III. Conclusion

The subscriber location is a very interesting area in the additional application of radio communication networks. There are many ”pure” radio navigational technics utilized. Most of the ideas in subscriber location could be applied successfully for bistatic radar systems, which are based on the existing radio systems and subsystems.

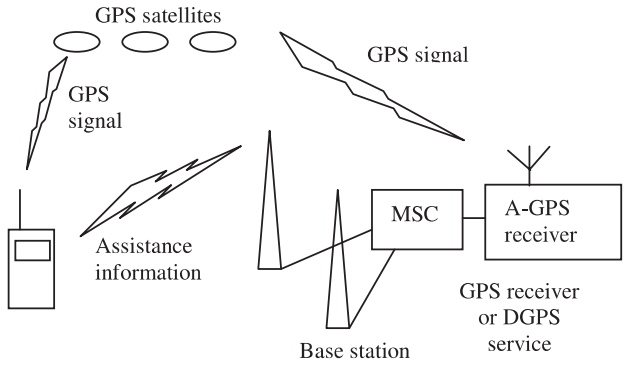


Fig. 4. Assisted-GPS concept

Table 1. Location technologies short description

MS-Based location	Global Positioning System (GPS)	Built into the handset is a full-fledged GPS receiver that operates as a standalone device (see description in the Section on GPS).
	Assisted GPS (A-GPS)	A partial GPS receiver in the handset is aided in its operation by the wireless network (see description in the Section on A-GPS).
	Observed Time Difference (OTD)	MS monitors signals from at least three BSs and observes time differences of arrivals. BS locations are known and fixed and position is calculated by triangulation after subtracting OTD from RTD. LMU required if BSs not synchronized (e.g. GPS at each site). Added network traffic is minimal (data are gathered at MS,) but handsets need firmware upgrade. In GSM, method is known as Enhanced OTD (E-OTD).
Network-Based Location	Time of Arrival (TOA)	MS signal is received by at least three BSs, which measure TOA independently and send data to PDE where location is calculated. No handset modifications are needed, but network requires upgrade and signaling is increased.
	Time Difference of Arrival (TDOA)	At least three BSs monitor MS signals using dedicated location receivers. Location estimate based on apparent arrival time differences between pairs of sites. Strict time synchronism among all BSs is a must. Time difference measurements less accurate in analog and digital narrowband systems (AMPS, TDMA,) leading to inferior performance.
	Angle of Arrival (AOA)	Special antenna arrays and location receivers in BS determine AOA of the MS signal. Location is found at the intersection of apparent arrival directions. At least two BSs needed, but three or more used to increase accuracy and reduce uncertainty due to multipath. No modifications to MS are needed.
	Multipath Fingerprinting	MS location found by matching the multipath-produced "fingerprint" of the signal received by one or more BSs with location/fingerprint database. Technique requires continuous database management and updating.
	Timing Advance (TA)	During link establishment MS aligns its frame/slot times with the serving BS, and uses this as a measure of its distance to BS (TDMA or GSM). Using network-enforced handoff, at least three measurements with different BSs are made and location is determined via triangulation. Sequential measurements make the method unreliable when MS is moving. No modifications to handsets and minor changes in BS software.
Network/MS-Based Location	Enhanced Forward Link Triangulation (E-FLT)	Solution unique to CDMA. Primarily based on TDOA using forward-link signals received by MS. Performance enhanced by complementary methods, including pattern matching of RF characteristics, statistical modeling, round trip delay measurements, and AOA.
Hybrid-Type	TDOA & Received Signal Strength (RSS)	Combine highly accurate with highly robust methods, and use multiple inputs to improve both the robustness and coverage. E.g., in A-FLT/A-GPS, A-FLT can extend coverage deep indoors where not enough GPS satellites are visible; in TDOA/AOA, AOA enables operation even when only two BSs can receive the MS signal.
	TDOA & AOA	
	A-FLT & A-GPS	
	E-OTD & A-GPS	

Acknowledgements

I wish to sincerely thank associate prof. Veselin Demirev Ph.D. from the Faculty of Communications and Communication Technologies at the Technical University in Sofia, Bulgaria.

References

[1] J. Caffery, Jr. and G. Stber, *"Subscriber Location in CDMA Cellular Networks"*, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL.47 ,NO.2, MAY1998.

[2] Relevant FCC documents, can be found at www.fcc.gov/e911/.

[3] M. Cherniakov, K. Kubik, *Secondary applications of wireless technology (SAWT)*, 2000 European Conference on Wireless Technology Paris 2000.

[4] G. Djuknic, R. Richton, *Geolocation and Assisted-GPS*, Bell Laboratories, Lucent Technologies, 2000.

[5] C. Drane, M. Macnaughtan, C. Scott, *"Positioning GSM Telephones"*, IEEE Communications Magazine, April 1998.

[6] J.Kriegl, *"Location in Cellular Networks"*, Institute for Applied Information Processing and Communications, University of technology Gratz, Austria, 2000.

[7] J.Andersen, T. Rappaport, S. Yoshida, *"Propagation Measurements and Models for Wireless Communication Channels"*, IEEE Communications Magazine, January 1995.

[8] J. Caffery, Jr. and G. Stber, *"Overview of Radiolocation in CDMA Cellular Systems"*, IEEE Communications Magazine, April 1998.

[9] Openvawe, *"Overview in Location Technologies"*, November 2002.

[10] G. Wlfle, R. Hoppe, D. Zimmermann, F. Landstorfer, *"Enhanced Localization Technique within Urban and Indoor Environments based on Accurate and Fast Propagation Models"*, Institut fr Hochfrequenztechnik, University of Stuttgart, Stuttgart, Germany,2002.

[11] W. Geary, *"Location Determination Technologies for Cellular Enhanced 9-1-1 Service"*, West Virginia University,1999

[12] S. Kim, T. Pals, R.Iltis, H. Lee, *"CDMA Sparse Channel Estimation Using a GSIC/AM Algorithm for Radiolocation"*, University of California, Santa Barbara, CA93106.

[13] L.Ying, Y. Liang, S. Wang, *"Location Parameters Estimation In Mobile Communication Systems"*, Institute of Information Science and Engineering, JiLin University of Technology, 2000.

[14] E. Kaplan, *"Understanding GPS Principals and Applications"*, Artech Huose, Boston 1996.

[15] J. Lee, L. Miller, *CDMA Systems Engineering Handbook*, Artech House, Hardcover, November 1998.

[16] http://www.trueposition.com

Microwave Autonomous Angular Position Finding System for Middle Range Unmanned Aerial Vehicle

Vladimir Smiljaković, Zoran Golubičić, Predrag Manojlović, Zoran Živanović[1]

Abstract – **Development and realization of autonomous microwave angular position finding system of unmanned aerial vehicle (UAV) in one (azimuthal) plane relative to ground control station (GCS) is described in this paper. Functioning of the system is based on microcontroller supported microwave monopulse receiver of original construction and is field proven.**

Keywords – **microwave link, monopulse receiver, UAV (unmanned aerial vehicle), autonomous angular position finding**

I. Introduction

System for remote guidance of unmanned aerial vehicle has great importance for its successful and safe flight, as well as fulfilling missions objective, because it has to guarantee integrity of UAV and operatorss complete control of it during the whole flight [1]. To obtain that task, it is necessary to have good quality of communication in both directions: from GCS to UAV (commands) and vice versa (telemetry and mission equipment data). One of the basic data needed by operator to make proper decisions during the UAVs mission is knowing reliably UAVs position, ie its real trajectory, so he can compare its actual to wanted position or to deliver command towards UAV to get to certain position. One and only exception from request of having good communication is when UAVs mission in some zones is under cover and in some parts of trajectory radio silence is a must. Usually, in such cases UAV performs preprogrammed trajectory and activity. Even in those cases in come back phase of flight: approaching GCS, preparing for landing and landing, communication between UAV and GCS is reestablished with finding its exact position as one of the most important data for safe landing as the most critical phase of flight.

II. UAVs Position Finding Systems

There are global position finding methods of UAV, all of them based on network of transmitters covering complete Earth, and autonomous methods - local and independent from anything outside of UAV-GCS system. Common to both methods is their aim: to define position of an object (in this case UAV) in space as precisely, quickly, reliably and cost effectively as possible. For that task three coordinates that describe UAVs position have to be found, independently of chosen reference system.

[1]All authors are with the Institute of microwave technique and electronics IMTEL, Bulevar Mihajla Pupina 165 b, 11 070 Novi Beograd, Srbija, E-mail insimtel@Eunet.yu, E-mail of the first author smiljac@insimtel.com

Basically there are two kinds of global systems: the first one consisting of transmitters at Earths satellite network such as GPS as representative of space based and the second like OMEGA and LORAN-C as representatives of older, Earth surface based global systems. In global systems usually latitude, longitude and height above sea level is defined as a result of activity.

As opposite approach, functioning of autonomous RPV position finding systems do not depend of any equipment outside RPV-GCS complex itself. Of course, its quality and operation range are limited by characteristics of the chosen method of position finding (inertial-gyroscope, primary or secondary radar, Doppler radar, compass, terrain recognition) and properties of telecommand-telemetry system. When autonomous position finding systems are concerned, usually results of position defining algorithm are: azimuth refering to the GCS, slant distance between GCS and UAV and height of UAV above sea level barometric (or above terrain radar altimeter or elevation of UAV). As it is known from radar technique, distance measurement relative error is about three orders of magnitude less than angle measurement relative error [2], unless special methods are applied, while the complexity of appropriate methods is inversely proportional. This is the reason of great importance of angle measurement algorithm in position finding process as a part of autonomous position finding system.

Choice between applying global or autonomous navigation system usually depends primarily on intended operational range and possibility of direct communication between GCS and UAV.

In the case of using already existing telemetry and telecommand chanels (equipment) in UAV-GCS communication system for position finding such as in this particular case, characteristics of that equipment together with chosen operating frequencies of telecommand and telemetry (due to known dependency of effects of atmosphere on characteristics of electromagnetic waves propagation) limit range of operation. At the same time they make possible reliable position finding of UAV during all phases of flight, which is described in more detail in [3].

Angular position finding method based on monopulse receiver principle applied in above mentioned UAV-GCS complex deserves more detailed description because of its elegance, reliability and at the same time very good cost over performance relationship [3].

Communication system of UAV-GCS consists of ground equipment signal receiver) and airborne part (telecommand receiver and appropriate antenna systems both on UAV and

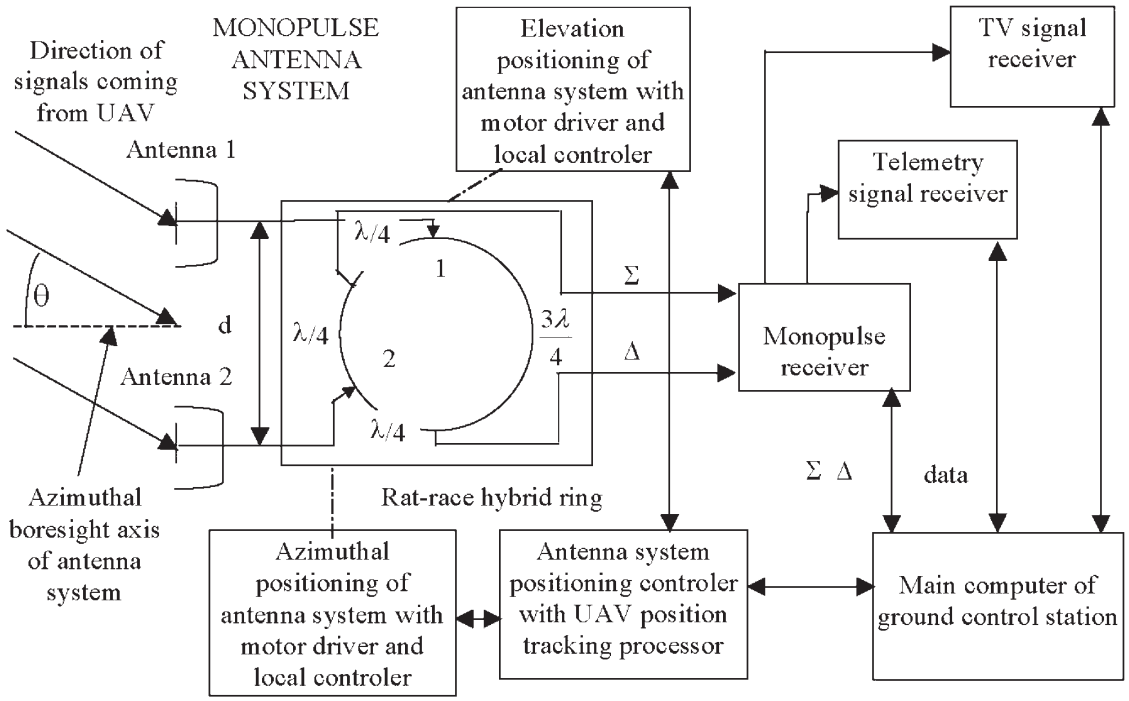


Fig. 1. Realized monopulse receiver without reference channel

GCS ([1], [3]). It connects GCS main computer and UAVs control computer, performing at the same time additional function - finding UAVs position: distance and azimuth angle while basic communication function is performed. Third coordinate height above sea level is obtained from barometer altimeter in UAV and sent to GCS via telemetry message. The distance finding system is based on application of secondary radar principle [4], and realization is described in detail in [5].

III. Monopulse Receiver

There are three basic classes of detection of angle θ č between boresight axis of monopulse antenna system and direction of arrival of signal from UAV: amplitude of arriving signal detection and comparison, phase of arriving signal detection and comparison and sum-and-difference of arriving signals [6]. Monopulse receiver configuration of the third class which is the most insensible to interference signal is depicted at Fig. 1. As it can be seen signals from two antennas (ie antenna arrays) are fed to special microwave unit - hybrid ring known as ”rat-race” in dielectric substrate technique or ”magic-T” in waveguide technique, which has two outputs. Because of the physical shape of this unit (distances between ports - two inputs and two outputs) it has special properties. At the one output of that unit signal is equal to the difference (designated as $\triangle$) of the input antenna signals, while at the other output signal is equal to the sum of the antenna signals (designated as Σ). It is known in literature (for example [6] where its functioning is elaborated in detail, as well as mathematical model) that angle θ defined at the beginning of this section depends on above defined signals $\triangle$ and Σ. Those signals are, generally speaking, complex signals, ie have components in phase - denoted by subscript I, and in quadrature – denoted by subscript Q, referring to the phase of the local oscillator (LO) as the part of the demodulator in the receiver.

Existence of components of the signals which are not in phase with appropriate basic signal are a measure of uncoherence of these signals with signal of the local oscillator and imperfections of parameters of receiving chains in $\triangle$ and Σ branches of monopulse receiver (as a consequence of differences of parameters characteristics which are unequal changed with change of frequency or temperature for example). It means that in the case of the coherent receiving process signals $\triangle$ and Σ signals have the same phase relationship to the incoming signal as to the signal of the receivers local oscillator. In one plane (in this case azimuthal) dependance between angle of incidence of the incoming signal and boresight axis θ and electrical signals in receiver is described by the following equation:

$$\theta = \mathrm{Im}\{\triangle/\Sigma\} = (\triangle_Q \Sigma_I - \triangle_I)/({\Sigma_Q}^2 + {\Sigma_I}^2)\,. \qquad (1)$$

When received signal and local oscillator are strictly in phase (quadrature), some of their components disappear (their projections on appropriate axes are zero) so expression becomes simply:

$$\theta = \triangle_Q/\Sigma_I\,. \qquad (2)$$

Using this fact monopulse receiver with proper operating mode (signals $\triangle$ and Σ in quadrature) is used for incidence angle measurement. Using normalization of the signals by Σ_I, it is enough to measure amplitude of $\triangle_Q$ signal to define angular position of the incidence signal according to the boresight axis, and sign of the angle is defined by the sign of $\triangle_Q/\Sigma_I$ ratio. Of course, this assumption is valid only for relatively small angles, because phase difference ϕ of inci-

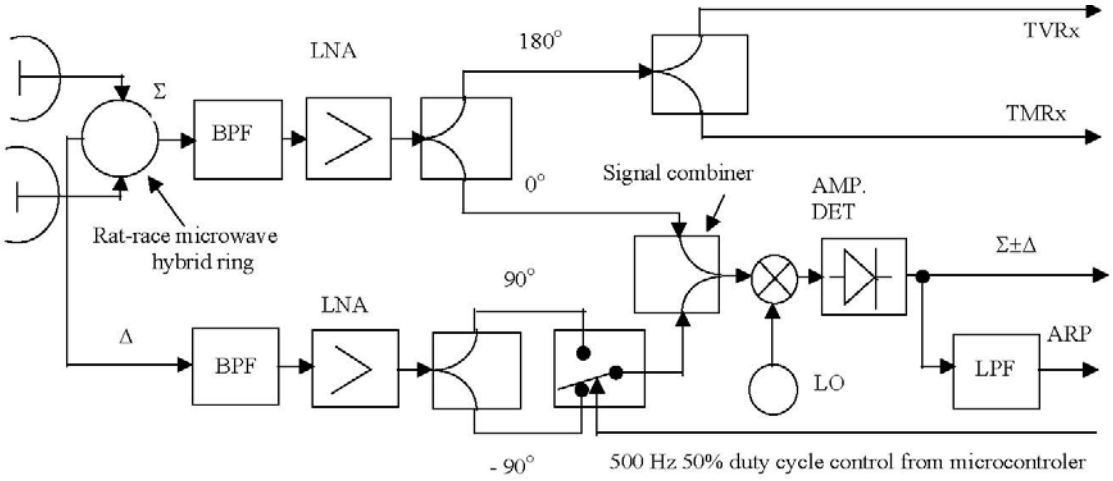


Fig. 2. Realized monopulse receiver without reference channel

dent signals coming to two antennas forming monopulse receiving system is sinusoidal ie periodical function of incident angle θ defined at the beginning of this section:

$$\phi = (2\pi \mathrm{d}/\lambda)\sin\theta\,, \tag{3}$$

where d is the distance between centers of the antennas and λ is wavelength of incoming signal. This relation shows that phase difference is equal to 0 not only when θ is 0, but every time $\theta = k\pi$ (radians).

IV. Realized Monopulse Receiver without Reference Signal Channel

Because of fluctuations of receiver branches parameters additional reference channel is usually used to suppress their unwanted effects, and additional circuitry adds complexity to the receiver [7]. This increases price and efforts in maintenance, so it is motive to introduce another, new type of monopulse receiver, originally developed and realized. This new monopulse receiver is depicted on Fig. 2.

As can be seen from the picture, sum signal is splited after amplification in two parts. One part is with reversed phase and it is further splitted in two in phase components which are fed to the receiver of TV signals (as signal of mission equipment) and to the receiver of the telemetry signals. Another sum signal from the first splitter is fed to signal combiner.

Difference signal is after amplification fed to another signal splitter with two outputs. One output is difference signal phase shifted by 90° in advance, and the other output is the same signal only with phase delayed by the same amount. These two output signals are fed as inputs to multiplexer with periodical command signal (periode 2 ms, duty factor 50%), generated by microprocessor supported monopulse receiver controller, that choses which one of the two inputs is fed to the output. In that way, bearing in mind that $\triangle$ signal originally at the place of generation is in quadrature with Σ signal, at the output of multiplexer is alternation of difference signal which is in phase with sum signal and difference signal which is in phase opposite to sum signal. This output signal of the multiplexer is fed to earlyer mentioned combiner, so output of the combiner is alternatively once $\Sigma + \triangle$ and $\Sigma - \triangle$.

After mixing of the output of the combiner signal with the receivers local oscillator signal, at the output of the mixer we get alternation of sum+difference and sum-difference signals in baseband, which are fed to input of analog to digital converter synchronously with the above mentioned periodical command signal of multiplexer.

Intensity of the phase difference signal ie intensity of the incident angle of the received signal is obtained by simple arithmetic operation of summing the digital representatives of $\Sigma + \triangle$ and $\Sigma - \triangle$ signals and dividing the result by two in microprocessor. The sign of the phase difference $\triangle$ ie the indication is the incident angle of the incoming signal at the left or the right side of boresight axis of antenna system is obtained by comparing the phase of periodic multiplex command signal and the phase of $\triangle$ signal (difference of input antenna signals).

Data describing position of the UAV (distance, azimuthal angle) microprocessor controller of the monopulse receiver sends to main computer of the ground control station every 100 ms. During one period of the multiplexer command signal (2 ms), several measurements (AD conversions) are performed and the results are averaged in microcontroller to suppress noise, and only after filtering are sent to main computer.

Having in mind dynamics of flying object and electromechanical piedestal for antenna system, this velocity of obtaining data about UAVs position is quite satisfying. The whole applied algorithm of finding angular position by use of simple monopulse microprocessor supported receiver is laboratory and field tested and compared to other, more complex algorithms (with reference channel) also developed and realized by IMTEL institute and gave comparative advantages.

V. Conclusion

Simple, yet effective, reliable and low cost monopulse microwave receiver is designed and realized for autonomous angular position determination of unmanned aerial vehicle. Thanks to microprocessor supported functioning and special configuration, common used receivers reference channel is avoided as well as cumbersome and time consuming trimming of receiver parameters. The design is laboratory and field tested as a part of UAV-GCS complex with good results.

References

[1] Slobodan Tirnanić, *Bespilotne letilice* (in Serbian), Vojnoizdavački zavod, Beograd, Yugoslavia, 2001

[2] George Biernson, *Optimal Radar Tracking Systems*, New York, USA, John Wiley & Sons, 1990.

[3] V.Smiljaković, Z.Golubičić, .Simić, D.Obradović, S.Dragaš, M.Mikavica, "Telemetry System of Light Aerial Vehicle "Raven", *Proceedings of TELSIKS99 Conference*, pp. 604-607, Niš, Yugoslavia, October 13-15. 1999.

[4] Michael Stevens: *Secondary Surveillance Radar*, Boston Ma. USA, Artech House Inc, 1988.

[5] V.Smiljaković, Z.Golubičić, P.Manojlović, Z.Živanović, "Autonomous Distance Finding Microwave System for Middle Range Remotely Piloted Vehicle", *Proceedings of ICEST2002 Conference*, pp. 229-232, Niš, Yugoslavia, October 2002

[6] D.H. Rhodes: *Introduction to Monopulse*, Dedham Ma. USA, Artech House Inc, 1980.

[7] Z.Golubičić, V.Smiljaković, P.Manojlović, "Kompenzacija faznih nejednakosti u mikrotalasnom BPSK monoimpulsnom prijemniku primenom mikroračunara" (in Serbian), *Zbornik radova XXXIX Konferencije ETRAN*, knjiga 2, strana 410-412, Zlatibor (Yugoslavia), 6-9. Juni 1995.

Automatic Radar Processing Using OSCA CFAR Detector

Slavcho Lishkov[1], Rumen I. Arnaudov[2], Rossen G. Miletiev[3]

Abstract – **The algorithm for an automated radar processing is analyzed, so the CFAR detector based on order statistics and cell-averaging is examined. By reason of that the expressions of the false alarm rate, the detection probabilities and measure ADT under the Swelling II assumption are calculated and are compared with the analogous parameters to the well known CFAR detectors.**

Keywords – **order-statistics, CFAR, ADT, radar processing**

I. Introduction

The main problem at the radar signal processing is the coherent primary detection of the received signal. The signal processing automation is connected with the development of the algorithms, which work at the wide range of an alteration of the signal statistical characteristics. The detection decision of the radiolocation signals requires the algorithm design, which maximizes the detection probability (P_D) at a constant false alarm probability (P_{f_a}). On the basis on this requirement a lot of CFAR detectors are designed. Constant false alarm rate (CFAR) algorithms are used to detect the targets in noise and clutter backgrounds whose mean power are unknown. Finn and Johnson [1] proposed the well-known CA-CFAR detector (cell-averaging constant false alarm rate). If the outputs of the reference cells are statistically independent and identically distributed random variables from the same population as the cell under test when there is no target, the detection performance of the CA-CFAR detector is optimal. However, the detection performance of the CA-CFAR detector is seriously degraded when the background environment is nonhomogeneous (*Rayleigh* or *Weibull* distributed). To improve the resolution of closely-spaced targets, Trunk [2] Hansen and Sawyers [3,4] proposed the SO-CFAR (smallest-of) and GO-CFAR (greatest-of) detectors respectively, but SO-CFAR processor exhibits severe degradation in false alarm rate control, with respect to the CA and GO-CFAR detectors in the presence of a clutter distribution edge effect. A new class of order statistic (OS) CFAR detectors is firstly introduced by Rohling [5] for multiple target situations.

reduce the processing time of the OS-CFAR detector in ordering magnitudes of the cell in the reference window, some kinds of modified OS-CFAR detectors, such as OSGO, OSCA and OSSO, are examined [6-9]. The best balance between the detection losses and processing time is possessed by the OSCA CFAR detector [9-11].

[1]Slavcho Lishkov is with the Faculty of Communications and Communication Technologies, Technical University of Sofia, Kliment Ohridski 8, Sofia, Bulgaria

[2]Rumen I. Arnaudov is with the Faculty of Communications and Communication Technologies, Technical University of Sofia, Kliment Ohridski 8, Sofia, Bulgaria, E-mail: RA@vmei.acad.bg

[3]Rossen G. Miletiev is with the Faculty of Communications and Communication Technologies, Technical University of Sofia, Kliment Ohridski 8, Sofia, Bulgaria, E-mail: miletiev@yahoo.com

II. System Description

Due to the best balance of the OS-CFAR detection performance, one of its modified variants is analyzed. The block diagram of OSCA-CFAR detector is shown at Fig. 1. The function of the automatic censoring structure is to censure the target echoes from the reference sliding window. The system collects $M + N$ reference cells and implements the adaptive threshold to estimate the noise background. In the generalized order-statistic cell-averaging CFAR detector the estimation of the noise level in the cell under test is the sum of the outputs of the leading window the k-th order statistics and the lagging window the l-th order statistics. The estimated noise level is multiplied with the threshold parameter level $T(k,l)$ and the corrected noise level is compared with the reference cell value at the comparator to take the decision about the target presence. The scaling factor T is represented at Table 1 according to k and l values at $N = M = 16$, $P_{f_a} = 10^{-6}$. When the leading cell number M is equal to the lagging cell number N, there is $T(k,l) = T(l,k)$. This is also true for the all type of OS-CFAR detectors.

We assume that the noise in the test cell is Rayleigh envelope distributed and that the target returns are fluctuated according to the Swelling II model. The system implements an adaptive threshold test:

$$V \underset{H_0}{\overset{H_1}{\geq}} TZ\,, \tag{1}$$

where Z – the final background noise estimation, T – threshold parameter to control the desired probability of false alarm, V – the cell-under-test variable.

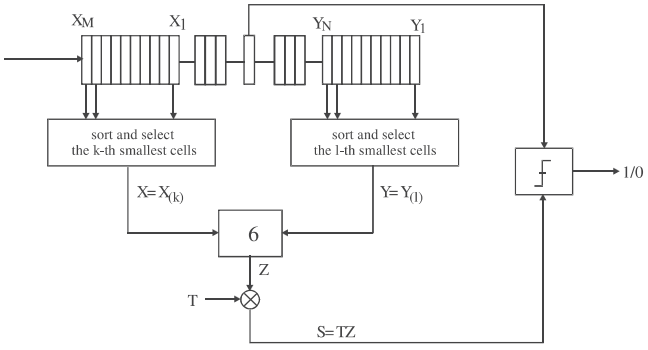


Fig. 1. OSCA-CFAR detector block diagram

A binary hypothesis testing paradigm under the Swerling II assumption is [7]:

$$V = \begin{cases} \frac{1}{\mu}\exp\left(-\frac{\nu}{\mu}\right), \ H_0 \\ \frac{1}{b\mu}\exp\left(-\frac{\nu}{b\mu}\right), \ H_1, \ \nu > 0 \end{cases}, \quad (2)$$

where $b = 1 + S$, S – per pulse average SNR.

For any CFAR detector employing Eq. (1) the detection probability is expressed by the equation [12]:

$$P_D = \Pr[V \geq TZ] = -\int\limits_{-C} \frac{d(u)}{i2\pi} u^{-1} h(u) d(-Tu), \quad (3)$$

where $h(u) = E\left(e^{-u\nu}\right)$, $d(u) = E\left(e^{-uz}\right)$ – moment generating functions (MGF)

The PDF (probability density function) of Z defined by Eq. (1) is given by [13]:

$$f_z(Z) = \int\limits_0^x f_X(x) f_Y(z-x) dx, \ z > 0. \quad (4)$$

The MGF of the noise level estimation is given for the homogeneous environment by [7]:

$$d(u) = d_X(u) d_Y(u), \quad (5)$$

where

$$d_Y(u) = k\binom{M}{k}\int\limits_0^\infty e^{-uz}[\exp(-z)]^{M-k+1}[1-\exp(-z)]^{k-1}dz$$
$$= k\binom{M}{k}\frac{\Gamma(M-k+1+u)\Gamma(k)}{\Gamma(M+u+1)}$$

and

$$d_Y(u) = l\binom{N}{l}\int\limits_0^\infty e^{-uz}[\exp(-z)]^{N-l+1}[1-\exp(-z)]^{l-1}dz$$
$$= l\binom{N}{l}\frac{\Gamma(N-l+1+u)\Gamma(l)}{\Gamma(N+u+1)}$$

Therefore, the detection probability and false alarm probability is calculated by the expressions:

$$P_{f_a} = (P_{f_a})_1 (P_{f_a})_2 \quad (6)$$
$$(P_{f_a})_1 = k\binom{M}{k}\frac{\Gamma(M-k+1+T)\Gamma(k)}{\Gamma(M+1+T)}$$
$$(P_{f_a})_2 = l\binom{N}{l}\frac{\Gamma(N-l+1+T)\Gamma(l)}{\Gamma(N+1+T)}$$
$$P_D = (P_D)_1 (P_D)_2, \quad (7)$$
$$(P_D)_1 = k\binom{M}{k}\frac{\Gamma\left(M-k+1+\frac{T}{1+S}\right)\Gamma(k)}{\Gamma\left(M+1+\frac{T}{1+S}\right)}$$
$$(P_D)_2 = l\binom{N}{l}\frac{\Gamma\left(N-l+1+\frac{T}{1+S}\right)\Gamma(l)}{\Gamma\left(N+1+\frac{T}{1+S}\right)},$$

where S – signal-to-noise ratio.

It is known that ADT (average detection threshold) is an alternative measure to compute the loss of detection performance in a CFAR processors. For given values of P_{f_a}, M and N, the ADT is independent of the detection probability. The ADT for the OSCA-CFAR detector is calculated by the equation [7]:

$$ADT = T\left(\sum_{i=1}^{k}\frac{1}{M-k+i} + \sum_{j=1}^{k}\frac{1}{N-l+j}\right). \quad (8)$$

The average detection threshold value is represented at Table 2 according to k and l values at $N = M = 16$, $P_{f_a} = 10^{-6}$

III. Numerical example

The OSCA-CFAR algorithm is analyzed using MATLAB® routine. The detection performance and detection losses are analyzed and compared with the analogue parameters of the other type of CFAR detectors. The scaling factor T is represented at Table 1 according to k and l values at $N = M = 16$, $P_{f_a} = 10^{-6}$. When the leading cell number is equal to the lagging cell number N, there is $T(k,l) = T(l,k)$.

Table 1. Scaling factor T according to k and l values at $N = M = 16$, $P_{f_a} = 10^{-6}$

1	10	11	12	13	14	15
k						
10	10.885					
11	9.8432	8.9641				
12	8.8647	8.1294	7.4214			
13	7.9316	7.3222	6.7278	6.1381		
14	7.0188	6.5222	6.0315	5.5383	5.0299	
15	6.0892	5.6964	5.3029	4.9019	4.4826	4.0239

The detection performance is the main characteristic, which is defined by the dependence of the detection probability P_D from the signal to noise ratio (S) at the fixed value of the false alarm probability P_{f_a}. It is estimated from the equation (6) and (7) and is shown at Fig. 2.

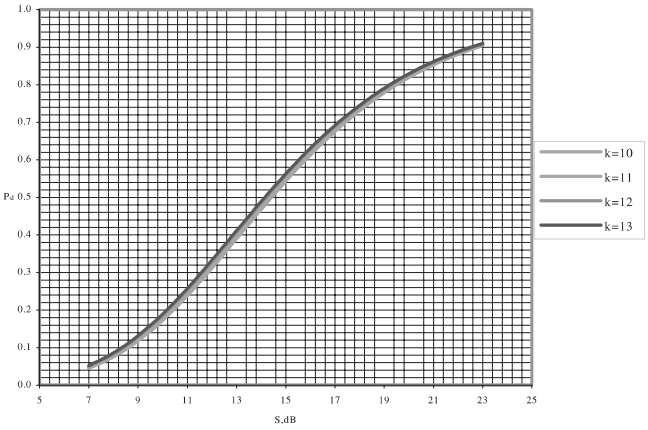


Fig. 2. The detection probability P_D according to k and l values at $N = M = 16$, $P_{f_a} = 10^{-6}$

The represented analyze shows that the detection probability values increase upon k and l augmentation and depend on the small degree from k and l values. But if k is near close to M or l is near close to N, the detection performance may be significantly degraded by the influence of the interfering target in the leading or lagging windows. At $k = M$ and $l = N$ OSCA-CFAR detector is identical with the CA-CFAR detector.

The next analyzed parameter of the CFAR detector is ADT value. It is calculated by the equation (8) and is represented at Table 2 according to k and l values at $N = M = 16,\ P_{f_a} = 10^{-6}$.

Table 2. The average threshold average value according to k and l values at $N = M = 16,\ P_{f_a} = 10^{-6}$

l	10	11	12	13	14	15
k						
10	20.261					
11	19.961	19.674				
12	19.751	19.468	19.257			
13	19.655	19.365	19.139	18.996		
14	19.733	19.424	19.169	18.986	18.919	
15	20.164	19.812	19.504	19.266	19.102	19.160

The shown results represent the low detection losses of OSCA-CFAR detector in relation to the other CFAR detector types. Regarding to that parameter the analyzed CFAR processor is compared to CA-CFAR one and exceeds all type of CFAR detectors [10].

These results define the intermediate position of the OSCA-CFAR detector in relation to CA-CFAR and OS-CFAR detectors. It combines the low detection losses and computing effectiveness of the CA-CFAR detector with the detection behavior of the OS-CFAR detector in the nonhomogeneous background multiple target situations.

IV. Conclusion

The represented CFAR algorithm for the automated radar processing allows a detection of the target returns in the homogeneous and nonhomogeneous backgrounds with an optimal detection performance. The favorable detection performance defines OSCA-CFAR detector as an optimal CFAR detector in relation to the detection losses, processing time and detection conduct in the nonhomogeneous background and interfering targets in the leading or lagging windows.

References

[1] Finn, H.M. ; Johnson, R.S. - Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates, *RCA Rev.*, 1968, pp.414-463

[2] Trunk, G.V. - Range resolution of targets using automatic detectors, *IEEE Trans.*, 1978, AES-14, (5), pp.750-755

[3] Hansen, V.G. - Constant false-alarm rate processing in search radars, *Proceedings of International Radar Conference*, London, 1973, pp.325-333

[4] Hansen, V.G.; Sawyers, J.H. - Detestability loss due to greatest of selection in a cell-averaging CFAR, *IEEE Trans.*, 1978, AEC-14, (1), pp.750-755

[5] Rohling, H. - Radar CFAR tresholding in clutter and multiple target situations, *IEEE Trans.*, 1983, AES-19, (4), pp.608-621

[6] Elias-Fuste, A.R.; de Mercado, M.G.G.; de los Reyes Davo, E - Analysis of some modified ordered statistic CFAR: OSGO and OSSO CFAR, *IEEE Transactions on Aerospace and Electronic Systems*, Volume: 26 Issue: 1 , Jan 1990 pp: 197 -202

[7] You He - Performance of some generalised modified order statistics CFAR detectors with automatic censoring technique in multiple target situations, *IEE Proceedings - Radar*, Sonar and Navigation, Volume: 141 Issue: 4, Aug 1994, pp: 205 - 212

[8] Kyung-Tae Jung; Hyung-Myung Kim - Performance analysis of generalized modified order statistics CFAR detectors, *Proceedings of Time-Frequency and Time-Scale Analysis the IEEE-SP International Symposium on 1998*, 6-9 Oct 1998, pp: 521 -524

[9] He You; Guan Jian; Peng Yingning; Lu Dajin - A new CFAR detector based on ordered statistics and cell averaging, *CIE International Conference of Radar Proceedings 1996*, 8-10 Oct 1996, pp: 106 -108

[10] Novak, L.M.; Hesse, S.R. - On the performance of order-statistics CFAR detectors, *Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, 1991, 4-6 Nov 1991, vol.2, pp: 835 -840

[11] Shor, M.; Levanon, N. - Performances of order statistics CFAR, *IEEE Transactions on Aerospace and Electronic Systems*, Volume: 27 Issue: 2 , Mar 1991, pp: 214 -224

[12] Ritcey, J.A.; Himes, J.L. - Performance of max family of order-statistic CFAR detectors, *IEEE Trans.*, 1991, AES-27, (1), pp.48-57

[13] Papoulis, A. - Probability, random variables, and stochastic processes, *McGraw-Hill*, New York, 1984

Use of Space Correlation of Satellite Move in GPS

Marin S. Marinov[1] and Georgi V. Stanchev[2]

Abstract – **This paper sets out an approach for improving the estimation accuracy of the coordinates of an object that uses the global positioning system GPS.**

Keywords – **GPS, distance estimation, least squares. .**

I. Introduction

The satellite-based navigation system GPS is increasingly used as personal equipment, providing position information, thus it exceeds the bounds of the truly professional equipment. In the case of this new application the user defines its coordinate mostly as a fixed point. The alteration of the distance between the user and the GPS satellites is only determined by the satellite orbital movement and the Earth rotation.

This paper presents an approach where data processing of the distance to the satellite groups takes place, taking into consideration the parameters, known in advance, of the satellite movement in their orbits. This approach is developed for an immobile or slow-moving object. In order to estimate the distance, the least squares approach is used, aiming at an increase in the accuracy while specifying the coordinates of the object.

II. Distance Alteration Model

Each of the GPS satellites transmits ephemeris data, some of the data is related to the orbital parameters. Apart from this parameter type, orbital perturbation corrections are transmitted too. This data provide the necessary precision within the range of an hour. The basic parameters, which are being transmitted, are set out in Table 1.

These parameters can help defining the satellite coordinates, in Earth-Centered-Earth-Fixed coordinate system (ECEF), using the following algorithm Eqs. (1) to (17), [3]:

- Computed mean motion (rad/s)

$$n_0 = \sqrt{\mu A^{-3}} \tag{1}$$

- Time from ephemeris reference epoch (s)

$$t_k = t - t_{oe} \tag{2}$$

- Corrected mean motion (rad/s)

$$n = n_0 + \Delta n \tag{3}$$

Table 1. Ephemeris parameter definition

Param.	Definition
M_0	Mean anomaly at reference time
Δn	Mean motion difference from computed value
e	Eccentricity
$(A)^{1/2}$	Square root of the semi-major axis
Ω_0	Longitude of ascending node of orbit plane at weekly epoch
i_0	Inclination angle at reference time
ω	Argument of perigee
$\dot{\Omega}$	Rate of right ascension
IDOT	Rate of inclination angle
C_{uc}	Amplitude of the Cosine harmonic correction term to the argument of latitude
C_{us}	Amplitude of the Sine harmonic correction term to the argument of latitude
C_{rc}	Amplitude of the Cosine harmonic correction term to the orbit radius
C_{rs}	Amplitude of the Sine harmonic correction term to the orbit radius
C_{ic}	Amplitude of the Cosine harmonic correction term to the angle of inclination
C_{is}	Amplitude of the Sine harmonic correction term to the angle of inclination
t_{oe}	Reference time Ephemeris
IODE	Issue of Data (Ephemeris)

- Mean anomaly (rad)

$$M_k = M_0 + nt_k \tag{4}$$

- Kepler's formula (solved by iteration)

$$E_k = M_k + e \sin E_k \tag{5}$$

- True anomaly (rad)

$$v_k = arctg \left[\sqrt{1-e^2} \frac{\sin E_k}{\cos E_k - e} \right] \tag{6}$$

- Eccentric anomaly (rad)

$$E_k = \arccos \left(\frac{e + \cos v_k}{1 + e \cos v_k} \right) \tag{7}$$

- Argument of latitude (rad)

$$\Phi_k = v_k + \omega \tag{8}$$

- Argument of latitude correction (rad)

$$\delta u_k = C_{us} \sin 2\Phi_k + C_{uc} \cos 2\Phi_k \tag{9}$$

- Radius correction (m)

$$\delta r_k = C_{rc} \cos 2\Phi_k + C_{rs} \sin 2\Phi_k \tag{10}$$

- Correction to inclination (rad)

$$\delta i_k = C_{ic} \cos 2\Phi_k + C_{is} \sin 2\Phi_k \tag{11}$$

[1]Marin S. Marinov, PhD is with the Aviation Faculty, National Military University, 5856 Dolna Mitropolia, Bulgaria, e-mail: mmarinov2000@yahoo.com

[2]Georgi V. Stanchev, PhD is with the Aviation Faculty, National Military University, 5856 Dolna Mitropolia, Bulgaria, e-mail: gstanchev@af-acad.bg

- Corrected argument of latitude (rad)

$$u_k = \Phi_k + \delta u_k \tag{12}$$

- Corrected radius (m)

$$r_k = A\,(1 - e\cos E_k) + \delta r_k \tag{13}$$

- Corrected inclination (rad)

$$i_k = i_0 + \delta i_k + (IDOT)\ t_k \tag{14}$$

- Position in orbital plane (m)

$$\left.\begin{array}{l} x'_k = r_k \cos u_k \\ y'_k = r_k \sin u_k \end{array}\right\} \tag{15}$$

- Corrected longitude of ascending node (rad)

$$\Omega_k = \Omega_0 + \left(\dot{\Omega} - \dot{\Omega}_e\right)\ t_k - \dot{\Omega}_e t_{oe} \tag{16}$$

- ECEF coordinates (m)

$$\left.\begin{array}{l} x_k = x'_k \cos\Omega_k - y'_k \cos i_k \sin\Omega_k \\ y_k = x'_k \sin\Omega_k + y'_k \cos i_k \cos\Omega_k \\ z_k = y'_k \sin i_k \end{array}\right\} \quad . \tag{17}$$

As Marinov, Stanchev [4,5] claim, the change in the satellite coordinates for short periods of time (up to a second) can be approximated with a great precision (probable error - less than 10 cm.), by means of linear function. Thus it's possible to simplify the calculation operations when calculating the distance between GPS receiver and a satellite.

The alteration of the satellite coordinates is set by the following expressions:

$$\left.\begin{array}{l} x_k(t) = x_k(t_0) + \alpha_{k,x} t \\ y_k(t) = y_k(t_0) + \alpha_{k,y} t \\ z_k(t) = z_k(t_0) + \alpha_{k,z} t \end{array}\right\} t_0 \leqslant t \leqslant t_0 + T\ , \tag{18}$$

where k is the satellite number, T is the processing interval, and t_0 is the initial moment of the processing interval. The coefficients $\alpha_{k,x}$, $\alpha_{k,y}$ and $\alpha_{k,z}$are calculated by the ephemeris parameters through the approximation of the alteration of the coordinates to a segment.

The distances between the user and the satellite are set by the relations:

$$D_k(t) = \sqrt{[x_k(t) - x_0]^2 + [y_k(t) - y_0]^2 + [z_k(t) - z_0]^2}\ , \tag{19}$$

where x_0, y_0, z_0 are the user's exact coordinates.

After replacing Eq. (18) in Eq. (19) one can get the expression:

$$D_k(t) = \\ = D_k(t_0)\left\langle 1 - \frac{2}{D_k^2(t_0)} \left\{ \begin{array}{l} \alpha_{k,x}[x_k(t_0) - x_0] \\ +\alpha_{k,y}[y_k(t_0) - y_0] \\ +\alpha_{k,z}[z_k(t_0) - z_0] \end{array} \right\} t + \frac{\alpha_{k,x}^2 + \alpha_{k,y}^2 + \alpha_{k,z}^2}{D_k^2(t_0)} t^2 \right\rangle^{\frac{1}{2}} \tag{20}$$

The square root in Eq. (20) is represented as a sequence:

$$D_k(t) = D_k(t_0) + \frac{\alpha_{k,x}^2 + \alpha_{k,y}^2 + \alpha_{k,z}^2}{2D_k(t_0)}\ t^2 + \\ + \frac{\alpha_{k,x}[x_k(t_0) - x_0] + \alpha_{k,y}[y_k(t_0) - y_0] + \alpha_{k,z}[z_k(t_0) - z_0]}{-D_k(t_0)}\ t. \tag{21}$$

Estimating the distance to the satellites, errors always occur due to different factors [3]. These errors are according to the Gaussian distribution with zero mean. In order to apply the least squares approach, Eq. (21) is reduced to the matrix form:

$$\begin{array}{l} D_k(t) = \mathbf{H}_k(t).\lambda_k \\ \mathbf{H}_k(t) = [1,\ t,\ \frac{\alpha_{k,x}^2 + \alpha_{k,y}^2 + \alpha_{k,z}^2}{2}\ t^2] \\ \lambda_k = [D_k(t_0),\ \beta_k,\ \gamma_k]^T \end{array}\ , \tag{22}$$

where β_k and γ_k are distance model coefficients, estimated by the least squares approach.

The measured values of the distances to the satellites are sum of the true values and errors of different nature:

$$R_k(t) = D_k(t) + \Delta D_k(t). \tag{23}$$

The observation equation, to which we could apply the least squares algorithm, can be drawn from Eqs. (22) and (23)[1]:

$$\mathbf{Z}_{k,n} = \mathbf{H}_{k,n}\lambda_k + \mathbf{\Delta D}_{k,n}; \\ \mathbf{H}_{k,n} = \begin{bmatrix} \mathbf{H}_k(t_0) \\ \mathbf{H}_k(t_0 + \Delta t) \\ \vdots \\ \mathbf{H}_k(t_0 + n.\Delta t) \\ \vdots \\ \mathbf{H}_k(t_0 + N.\Delta t) \end{bmatrix}; \\ \mathbf{\Delta D}_{k,n} = \begin{bmatrix} \mathbf{\Delta D}_k(t_0) \\ \mathbf{\Delta D}_k(t_0 + \Delta t) \\ \vdots \\ \mathbf{\Delta D}_k(t_0 + n.\Delta t) \\ \vdots \\ \mathbf{\Delta D}_k(t_0 + N.\Delta t) \end{bmatrix}; \tag{24}$$

$$0 \leqslant n \leqslant N; N.\Delta t = T; \Delta t = 2.10^{-2} s.$$

The parameter Δt^{-1} is the rate of measuring the distance in the GPS receivers.

The estimation, according to the standard least squares approach, is given through the equation

$$\hat{\lambda}_k = (\mathbf{H}_{k,n}\mathbf{K}_k^{-1}\mathbf{H}_{k,n})^{-1}\mathbf{H}_{k,n}^T\mathbf{K}_k^{-1}\mathbf{Z}_{k,n} \tag{25}$$

where $\mathbf{K}_k$ is covariance matrix of $\Delta\mathbf{D}_{k,n}$. The elements of this covariance matrix are of the type $\exp\{-n\Delta t/t_{corr}\}$.

The algorithm given by Eq. (26) is non-recursive but it can be turned into a recursive one [1,2].

III. Research Results

A model of the distance estimation errors is created for the needs of this research work. For the satellite positioning and movement, true ephemeris data is used in the simulation, and using the algorithms mentioned above the measured and estimated distances to the satellites are modelled. The results are obtained for different error variances.

Fig. 1 shows the results how efficient the suggested algorithm about distance estimation to the satellite is, with error

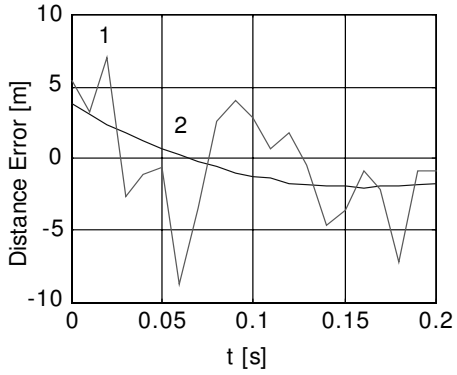


Fig. 1. Distance error for $\sigma^2 = 9\text{m}^2$ and T=200 ms.

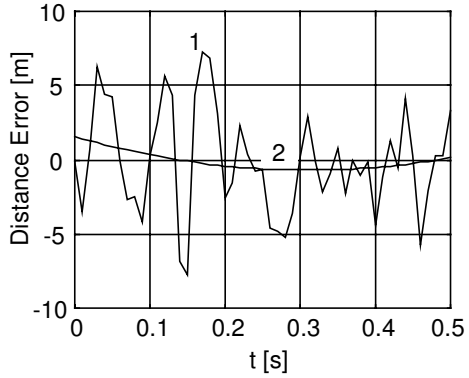


Fig. 2. Distance error for $\sigma^2 = 9\text{m}^2$ and T=500 ms.

variance 9 m^2, geometrical delusion of position GDOP=2.45 and processing interval T=200 ms. Fig. 2 presents the results when the processing interval is 500 ms.

It's obvious that the error in estimating the distance is changing from -8 to +7 meters for the observed values (curve 1), while after applying the suggested algorithm, for the estimated values (curve 2) the error is within the bounds of -2 to +4 meters.

The error in estimating the distance is changing from -7 to +7 meters for the observed values (curve 1), while after apply-ing the suggested algorithm, for the estimated values (curve 2) the error is within the bounds of -1 to +2 meters.

Fig. 3 and Fig. 4 show the results with the error variances 100 m^2.

It's obvious that the error in estimating the distance is changing from -16 to +38 meters for the observed values (curve 1), while after applying the suggested algorithm, for the estimated values (curve 2) the error is within the bounds of -5 to +12 meters.

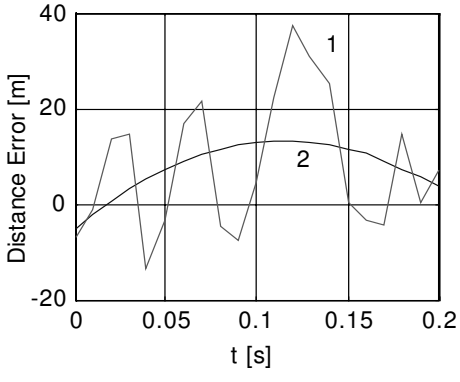


Fig. 3. Distance error for $\sigma^2 = 100\text{m}^2$ and T=200 ms.

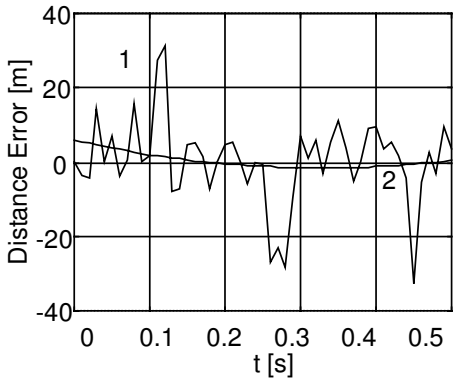


Fig. 4. Distance error for $\sigma^2 = 100\text{m}^2$ and T=200 ms.

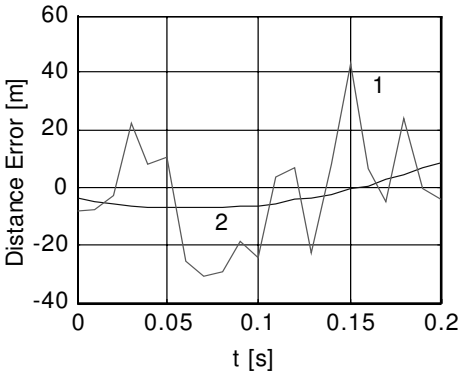


Fig. 5. Distance error for $\sigma^2 = 400\text{m}^2$ and T=200 ms.

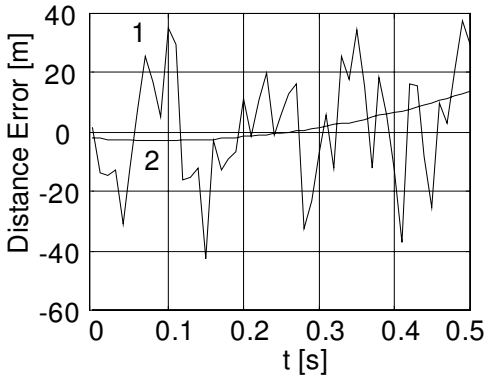


Fig. 6. Distance error for $\sigma^2 = 400\text{m}^2$ and T=500 ms.

It's clear that the error in estimating the distance is changing from -32 to +30 meters for the observed values (curve 1), while for the estimated values, after applying the suggested algorithm, the error is within the bounds of -2 to +7 meters (curve 2).

Fig. 5 and Fig. 6 show the results with the error variances 400 m^2.

It's obvious that the error in estimating the distance is changing from -30 to +42 meters for the observed values (curve 1), while for the estimated values, after applying the suggested algorithm, the error is within the bounds of -8 +10 meters (curve 2).

It's clear that the error in estimating the distance is changing from -42 to +39 meters for the observed values (curve 1), while for the estimated values, after applying the suggested algorithm, the error is within the bounds of -3 to +11 meters (curve 2).

IV. Conclusion

The attained results prove that the accuracy in estimating the distance increases if the suggested algorithm is used. The improvement of the accuracy in estimating the distances leads to improvement of the accuracy in estimating the geographical coordinates of the user. Implementing the algorithm in its recursive form would further more reduce the requirements to the processors.

References

[1] A. P. Sage, J. L. Melse, *Estimation Theory with Application to Communication and Control*, N. Y. McGraw-Hill, 1972.

[2] C. F. N. Cowan, P. M. Grand, *Adaptive Filters*, Prentice-Hall, Englewood Cliffs, 1985.

[3] E. D. Kaplan, *Understanding GPS: Principles and Applications*, Artech House Publishers, Boston, 1996.

[4] G. Stanchev, M. Marinov, "Study of Frequency Shifts of Signals in Uplink Channel of Aircraft Satellite Communication Systems", *Proceeding of Scientific Conference'02*, 25th-28th April 2002., the Air Force Academy of Bulgaria, D. Mitropolia, 2002, vol. 3, pp.235-240.

[5] M. Marinov, G. Stanchev, "A Possibility for Accuracy Increasing in Satellite Navigation Systems, Proceeding of Scientific Conference'01, 12th-13th April 2001., the Air Force Academy of Bulgaria, D. Mitropolia, 2001, Vol.3, pp.140-145.

Use of Joint Space Correlation of Satellite Constellation Move in GPS

Marin S. Marinov[1]and Georgi V. Stanchev[2]

Abstract – **The paper sets out an algorithm, which aims at improving the accuracy in fixing the position of an object using the global positioning system GPS.**

Keywords – **GPS; position accuracy; least squares.**

I. Introduction

In the satellite based navigation system GPS, fixing the user's position is accomplished by means of estimating the distances to the four satellites [3]. The user's coordinates are determined from system of equations, connecting these distances with its position.

One way for improving the accuracy of fixing the coordinates is using of additional processing of the initial estimates of these coordinates.

This paper presents an approach for processing the coordinates data according to the results, which were obtained by means of the algorithm suggested in [6]. This algorithm is developed for an immobile or slow-moving object.

A comparison between the results obtained by this approach and true data obtained by GPS receiver has been made.

II. Fixing the Coordinates

Marinov, Stanchev's study [6] gives quadratic model about the distance alteration between the user and the GPS satellites:

$$\begin{aligned} &0.5[D_1(t) - D_2(t) + R_2(t) - R_1(t)] = \\ &=[x_1(t)-x_2(t)]x+[y_1(t)-y_2(t)]y+[z_1(t)-z_2(t)]z \\ &0.5[D_1(t) - D_3(t) + R_3(t) - R_1(t)] = \\ &=[x_1(t)-x_3(t)]x+[y_1(t)-y_3(t)]y+[z_1(t)-z_3(t)]z \\ &0.5[D_1(t) - D_4(t) + R_4(t) - R_1(t)] = \\ &=[x_1(t)-x_4(t)]x+[y_1(t)-y_4(t)]y+[z_1(t)-z_4(t)]z \end{aligned} \quad (1)$$

where $D_i(t)$ is the distance between the user and the i th satellite. $R_i(t)$ is the distance from the i th satellite to the origin of the coordinate system.

The satellite's coordinates are x_i, y_i and z_i. The index specifies the satellite's number.

The distance estimation is done using the approach suggested by [6].

After solving the equation system, the result is the object's coordinates x, y, z which have an error. During the processing interval the error is changing with time in a second degree curve, therefore the estimation of the coordinates will have the same character.

Additional processing is needed for improving the accuracy of the estimation of the coordinates. The research shows that the distances to the satellites, whose calculations are based on the estimated coordinates, are different from those on the first processing stage. Setting the estimated coordinates, where the difference is minimum, is used as an algorithm for additional processing. On the other hand, the research work reveals that minimum difference is always on one and the same coordinates regardless of the satellite being used.

III. Research Results

The study is based on different variances of the distances measured by the GPS receiver, striking an average of 1000 realization for each of the different variances.

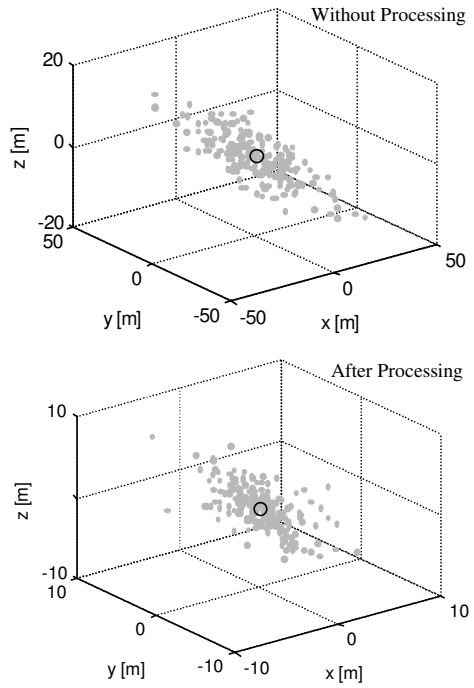


Fig. 1. Coordinate errors for $\sigma^2 = 9\text{m}^2$ and T =200 ms.

[1]Marin S. Marinov, PhD is with the Aviation Faculty, National Military University, 5856 Dolna Mitropolia, Bulgaria, e-mail: mmarinov2000@yahoo.com

[2]Georgi V. Stanchev, PhD is with the Aviation Faculty, National Military University, 5856 Dolna Mitropolia, Bulgaria, e-mail: gstanchev@af-acad.bg

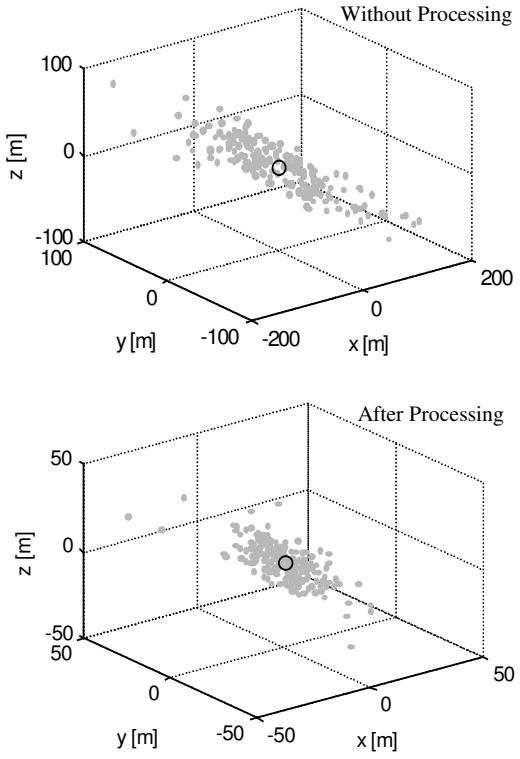


Fig. 2. Coordinate errors for $\sigma^2 = 144\text{m}^2$ and T =200 ms.

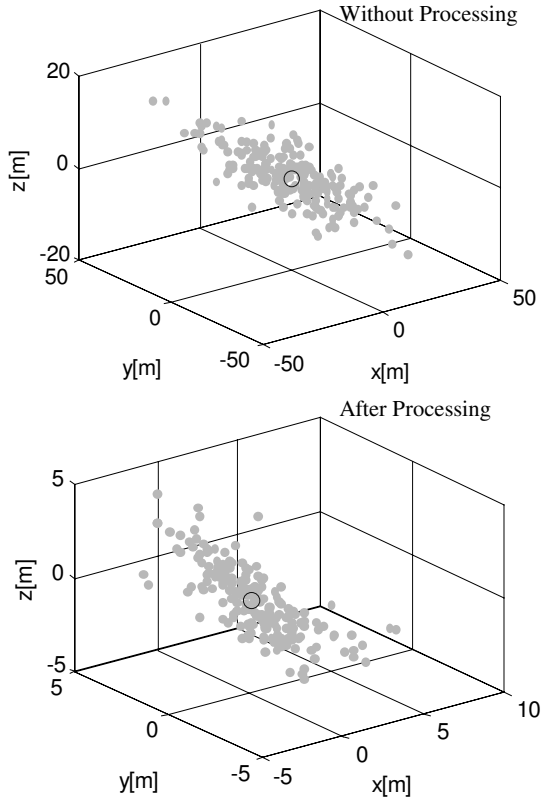


Fig. 3. Coordinate errors for $\sigma^2 = 9\text{m}^2$ and T =500 ms.

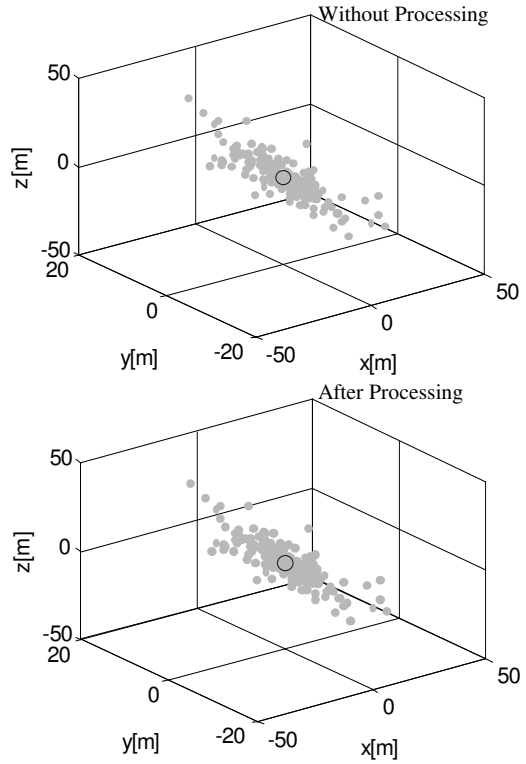


Fig. 4. Coordinate errors for $\sigma^2 = 144\text{m}^2$ and T =500 ms.

Fig. 1 shows the results for the error in fixing the position with variance 9 m^2 and processing interval 200 ms.

It is clear that the coordinate errors are greater in case of no additional processing, rather than the presented algorithm is used. The results prove that the improvement of accuracy is rather than several times when after applying the suggested algorithm. The coordinate errors are no more than 10 meters for all 10000 realizations, when the variance is 9 m^2 and time processing interval is 200 ms.

Fig. 2 shows the results for the error in fixing the position with variance 144 m^2 and processing interval 200 ms.

It is obvious that the coordinate errors are greater in case of no additional processing, rather than the presented algorithm is used. The results prove that the improvement of accuracy is rather than several times when after applying the suggested algorithm. The coordinate errors are no more than 40 meters for all 10000 realizations, when the variance is 144 m^2 and time processing interval is 200 ms.

Fig. 3 shows the results for the error in fixing the position with variance 9 m^2 and processing interval 500 ms.

The results prove that the accuracy is better if time processing interval is 500 ms. The coordinate errors are no more than 8 meters for all 10000 realizations, when the variance is 9 m^2 and time processing interval is 500 ms.

Fig. 4 shows the results for the error in fixing the position with variance 144 m^2 and processing interval 500 ms.

The results prove that the accuracy is better if time processing interval is 500 ms. The coordinate errors are no more than 30 meters for all 10000 realizations, when the variance

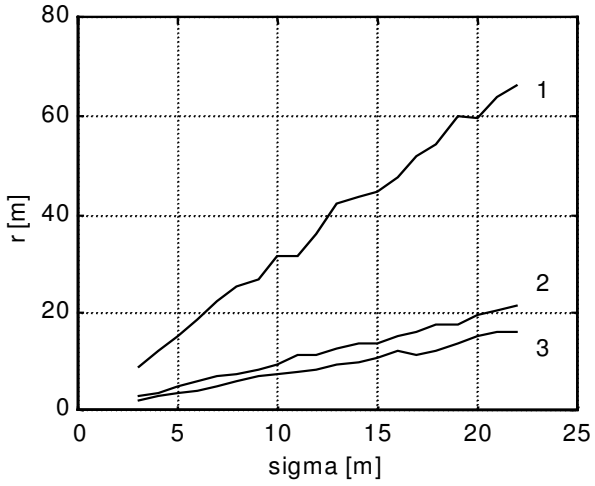


Fig. 5. The averaged error for T=200 ms.

is 144 m^2 and time processing interval is 500 ms.

An averaging of 10000 realizations for each value of the variance is made. The efficiency of the algorithm is valued by the distance between true position and estimated position. The comparison of suggested algorithm with common averaging of coordinates and without any processing of data is done.

Fig. 5 shows the results of the comparison when interval of processing is 200 ms.

The values of errors are the smallest when the proposed algorithm is applied. The distance between true position and estimated position is from 8.77 to 66.38 meters when an additional coordinate processing is not applied (curve 1). The values of the distance are from 2.77 to 20.94 meters in the case of the common averaging (curve 2). In the case of suggested algorithm (curve 3) that distance is from 2.11 to 16.44 meters. The averaged improvement of position accuracy is about 32 % in relation to the common averaging and 321 % in comparison with the case of no data processing.

Fig. 6 shows the results of the comparison when interval of processing is 300 ms.

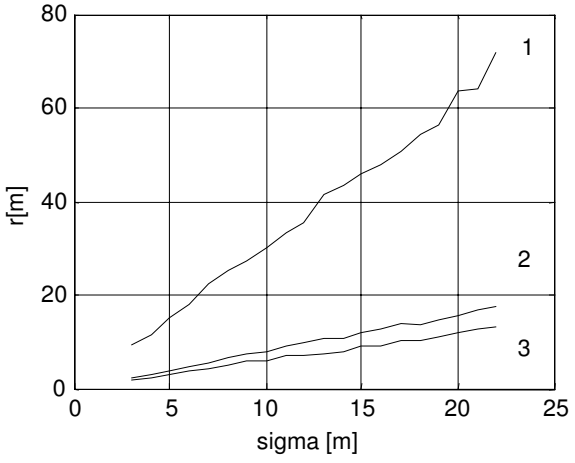


Fig. 6. The averaged error for T=300 ms.

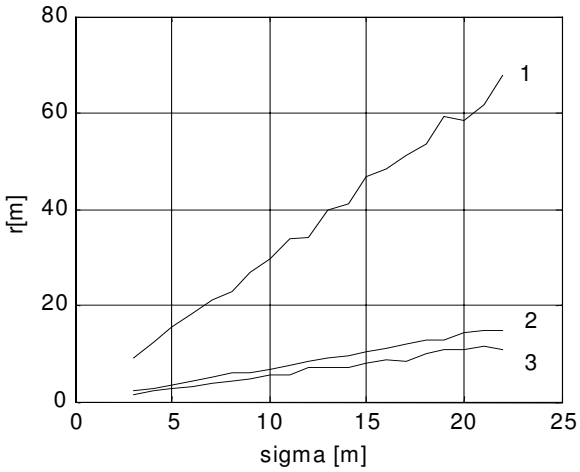


Fig. 7. The averaged error for T=400 ms.

The values of errors are the smallest when the proposed algorithm is applied. The distance between true position and estimated position is from 9.3 to 71.82 meters when an additional coordinate processing is not applied (curve 1). The values of the distance are from 2.42 to 17.48 meters in the case of the common averaging (curve 2). In the case of suggested algorithm (curve 3) that distance is from 1.93 to 13.06 meters. The averaged improvement of position accuracy is about 32.75 % in relation to the common averaging and 411 % in comparison with the case of no data processing.

Fig. 7 shows the results of the comparison when interval of processing is 400 ms.

The values of errors are the smallest when the proposed algorithm is applied. The distance between true position and estimated position is from 9.13 to 67.86 meters when an additional coordinate processing is not applied (curve 1). The values of the distance are from 2.09 to 14.82 meters in the case of the common averaging (curve 2). In the case of suggested algorithm (curve 3) that distance is from 1.49 to 10.88

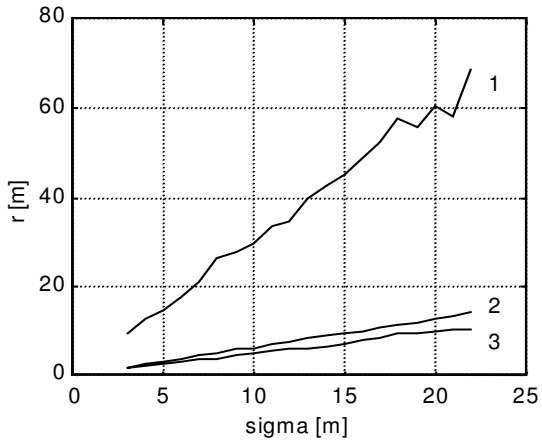


Fig. 8. The averaged error for T=500 ms.

meters. The averaged improvement of position accuracy is about 30.88% in relation to the common averaging and 473% in comparison with the case of no data processing.

Fig. 8 shows the results of the comparison when interval of processing is 500 ms.

The values of errors are the smallest when the proposed algorithm is applied. The distance between true position and estimated position is from 9.32 to 68.5 meters when an additional coordinate processing is not applied (curve 1). The values of the distance are from 1.83 to 14.29 meters in the case of the common averaging (curve 2). In the case of suggested algorithm (curve 3) that distance is from 1.42 to 10.25 meters. The averaged improvement of position accuracy is about 29.33 % in relation to the common averaging and 530 % in comparison with the case of no data processing.

IV. Conclusion

These results reveal that applying the suggested approach there's significant accuracy increase in fixing the user's position. A comparison between the suggested approach and the method about striking average in time shows that there is greater accuracy if the first approach is used. An increase in the processing interval leads to additional accuracy increase. The very nature of this approach implies further optimization of its application.

Acknowledgements

I wish to sincerely thank **associate prof. Ph.D.Veselin Demirev**, from the Faculty of Communications and Communications Technologies, Technical University, Sofia, Bulgaria.

References

[1] A. P. Sage, J. L. Melse, *Estimation Theory with Application to Communication and Control*, N. Y. McGraw-Hill, 1972.

[2] C. F. N. Cowan, P. M. Grand, *Adaptive Filters*, Prentice-Hall, Englewood Cliffs, 1985.

[3] E. D. Kaplan, *Understanding GPS: Principles and Applications*, Artech House Publishers, Boston, 1996.

[4] G. Stanchev, M. Marinov, "Study of Frequency Shifts of Signals in Uplink Channel of Aircraft Satellite Communication Systems",*Proceeding of Scientific Conference'02*, 25th-28th April 2002., the Air Force Academy of Bulgaria, D. Mitropolia, 2002, vol. 3, pp.235-240.

[5] M. Marinov, G. Stanchev, "A Possibility for Accuracy Increasing in Satellite Navigation Systems, Proceeding of Scientific Conference'01, 12th-13th April 2001., the Air Force Academy of Bulgaria, D. Mitropolia, 2001, Vol.3, pp.140-145.

[6] M. Marinov, G. Stanchev, "Use of Space Correlation of Satellite Move in GPS", *Information, Communication and Energy Systems and Technologies*, Sofia, Bulgaria, 2003.

Analysis of the Possibilities for of Secondary Applications of CDMA Cellular Systems

Vladimir A. Kyovtorov[1]

Abstract – **The paper deals a new promising application of the existing broadcasting and mobile communication systems - the secondary applications of the transmitted RF signals for radiolocation purposes. Particular emphasis is given on the CDMA technology. The properties of the different CDMA IS-95 standard signals are analyzed in order to define the best of them for this secondary application. Matlab simulations of the autocorrelation properties of the downlink pilot channel signal are given. The range resolution properties as well as the conditions for un-ambiguity distance measurements are discussed too.**

Keywords – **SAWT, radiolocation, CDMA, secondary applications, cellular applications, ICEST 2003.**

I. Introduction

On the background of existing global communication and navigation networks appeared the concept of Secondary Application for Wireless Technology – (SAWT). The research in this area is connected with using radio communication and electromagnetic signals, which cover a large frequency band from KHz to tens GHz. The situations, which could be observed, are various. For instance many different powerful ground-located radio transmitters, satellite transmitters or wireless telecommunication networks can be used for bistatic radiolocation. There is a comparatively simple situation when the receiver receives signals that have a direct and a reflected component. Analyzing the angle of arrival (AOE) and time of arrival (TOA) appropriately a definite radiolocation scene could be composed, which could give us a concept of some of the target's parameters. Investigation of these problems could extend and develop the wireless services.

II. Short Analysis Examples and Results

The characteristic of the information, which radiolocation device receives, depends on the structure and the properties of the used signal. According to its purpose the signal has to allow optimum realization of:

- Sufficient signal energy.
- Maximum detection of distance and velocity
- High noise resistance /from artificial and natural sources/

[1]Vladimir A. Kyovtorov is with the Faculty of Communications and Communications Technologies, Technical University, Sofia, Bulgaria, E-mail: vladimir_ak@yahoo.com

In the contemporary communication networks signals correspond to some basic radiolocation requirements:

- The communication codes have good autocorrelation function (ACF). At its hand the good ACF ensures the correct recognition of the reflecting signal and is a precondition for good range detection.
- These codes have a good cross correlation function, which allows it low cross cannel interference. On the other side thanks to the good division we can do radiolocation statistical analysis of the different data from different channels. This allows us to increase radiolocation measurements.
- Optimum frequencies that can provide necessary frequency band.

The analysis of the opportunity for secondary application of wireless signals is connected with the radiolocation evaluation of the signal. Various algorithms for detection of the signal are used. Also different methods to evaluate Angle of Arrival (AOA) and Time of Arrival (TOA) of the direct and reflected received signals are used [1]. Powerful means in the radiolocation analysis is the ambiguity function of radar signals. It describes the complex envelope curve on the entry of radio locator like a function of target range and velocity. It is completely defined with the Eq. (1)[2]:

$$\chi(\tau,\Phi)=\int_{-\infty}^{+\infty}u(t)u(t+\tau)e^{-j2\pi\Phi t}\mathrm{d}t\,, \qquad (1)$$

where u(t) – complex modulating function; τ – time delay of the signal is connected with range $\Re$; Φ – Doppler frequency difference connected with radial velocity V.

Let Φ = const., then we obtain an expression for the function of the ambiguity function by range.

$$\chi(\tau,\Phi)=\chi(\tau)=\int_{-\infty}^{+\infty}u(t)u(t+\tau)dt \qquad (2)$$

In this situation, range ambiguity function coincides with Autocorrelation function (ACF) of radiolocation signal. If we express Eq. (2) with digital signals we will obtain the well-known formula – ACF of a digital signal: [3,4]

$$P(u(t))=\sum_{i=1}^{n}Pu_i(t)\delta(u-u_i(t)) \qquad (3)$$

By synthesizing and analyzing the ambiguity function of contemporaneous digital signals we can to obtain notion of

radiolocation properties of these signals. We can also estimate range and velocity with certain accuracy of the radiolocation measurements with conventional digital communication signals [4]. As the ACF has a determined period [3,5], we can also define a parameter – range of synonymous evaluation. This parameter is connected to the period of repetition of the ACF, which at eventual radiolocation range measurement will lead to ambiguity of distance detection at the moment of delay, which equals a period of ACF.

Consider the simple geometrical equation (Fig. 1) [4], which gives a classic situation of a bistatic radiolocation. At point A we have a source of radio waves, which is a part of radio communication network. At point B there is a target, which reflects a radio communication signal, and at point C there is a receiver. The receiver could evaluate with known

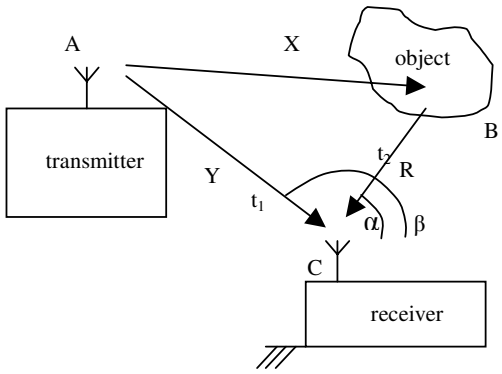


Fig. 1. Block diagram of wireless system

methods the time delay of both signals, which gives an idea of traversed distance, t1 and t2 respectively. We assume that the two angles (α and β) can be obtained quite exactly with the help of a course and bearing indicator, for example. The covered distance will be as follows:

$$AC = Y = \frac{c}{t1} \tag{4}$$

$$d = ABC = X + R = \frac{c}{t2}\,, \tag{5}$$

where c =3.10^{8}m/s.

From the known cosine theorem we can obtain an arithmetical expression of range R:

$$X^2 = R^2 + Y^2 - 2YR\cos(\beta - \alpha) \tag{6}$$

from Eq. (5)

$$X = d - R \tag{7}$$

replace in Eq. (6)

$$(d - R)^2 = R^2 + Y^2 - 2Y(d - R)\cos(\beta - \alpha) \tag{8}$$

After a suitable transformation we can obtain the final expression for estimating the unknown distance to the target:

$$R = \frac{d^2 - Y^2}{2d - 2\cos(\beta - \alpha)} \tag{9}$$

Consider the well-known communication cell system based on the CDMA principals – IS-95A[6]. The planned structure of the network is made with the purpose of assuring maximal coverage of a defined area, which on the other side guarantees reliable receiving in the whole range of the communication network. The structure of the network is mapped in the Fig. 2[6].

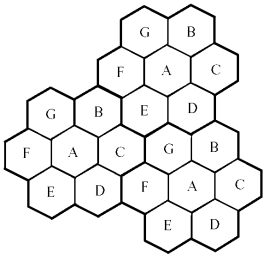


Fig. 2. Cellular structure

The question is could the situation from the Fig. 1 be obtained on the structure of communication network from Fig. 2[6]. We reduced the analysis of the problem to evaluation of the ambiguity function of a communication signal in this network. We attempt to explain this signal as s radiolocation signal in the bistatic radar system. The standard IS-95A is based on the spread spectrum signals (SSS)[3]. It was found that these signals could be used as radiolocation signals, because they are synthesized with a low cross correlation function with the purpose of low influence between particular channels [6]. On the other hand signals used for spreading of informational signals (PN code) are m-sequence [6], which guaranteed very good ACF. If this signal can be used for radiolocation observing, it has to correspond to certain conditions.

- Uninterruptedness of the radio transmittion
- Easily detectable
- According to bistatig reflecting surface of the target the signal has to have enough energy to cover some measurement range

On this base we have to exclude the signals for Uplink, because they are low powered and are of chance character. For this reason we have to exclude Traffic and Paging channels from Downlink channel. We concentrate on the Pilot signal. It is uninterrupted and consists of simple signals. It is used primarily as a coherent phase reference for demodulation the other channels. For this reason, IS-95 requires that the chip timing and carrier phase of each Downlink channel be in very close agreement. The pilot signal are constant logical 0, it is modulated at Walsh chip rate of 1.2288 Mcps by H_0, the 0-th row of the 64x64 Hadamart matrix, which is the Walsh sequence consisting of 64 zeros-thus, in effect, it is not modulated at all. The two distinct short PN codes – I and Q (Eq. (10) and Eq. (11)) are maximal length sequence generated by 15-stage shift registers and lengthened by the insertion of one chip per period in a specific location in the PN sequence. Thus, these PN codes have periods equal to the normal sequence length of 2^{15}-1 = 32767 plus one chip,

or 32768 chips [6]. The first channel architecture is shown in Fig. 3 [6]. The PN-code spreading is followed by classic QPSK modulation of the radio frequency carrier. This signal is perfect for secondary radiolocation because it is more powerful signal than the others. This is the reason for the easy detection and also it has a short period of repetition, which allows fast synchronization.

$$f_1(x) = 1 + x^2 + x^6 + x^7 + x^8 + x^{10} + x^{15} \quad (10)$$

$$f_2(x) = 1 + x^3 + x^4 + x^5 + x^9 + x^{10} + x^{11} + x^{12} + x^{15} \quad (11)$$

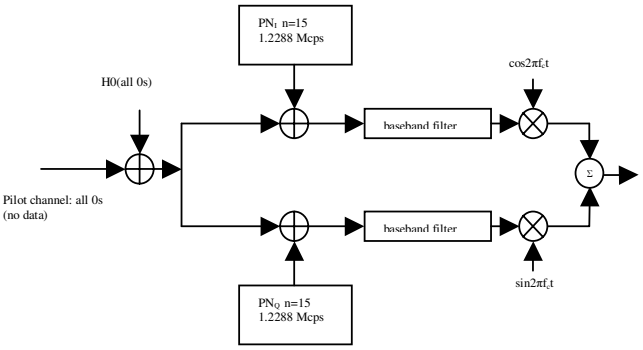


Fig. 3. Pilot channel modulation

With the assistance of MATLAB [8] we did visualization of ACF of the two PN sequences (Fig. 4). From the figure we could see that both functions have small side lobes (Fig. 6) which level is $1/N$ (N is the length of the m-sequence). Consider the range of range of un-ambiguity (Fig. 4), which can be achieved by using such a signal for bistatic radiolocation. The period of repetition of both functions could be evaluated with:

$$P = \frac{N}{Cr}\,, \quad (12)$$

where N is the size of the PN code (N=2^n-1, n-power of the equation of the PN code) and C_r – chip rates (chip/s) [6]. It turns out that the period of repetition is 26.66 ms, which at its hand defines the range of un-ambiguity:

$$S = t * c = 26,66 * 10^{-3} * 3 * 10^8 \approx 7980km\,, \quad (13)$$

where c=3.10^8 m/s.

Some words about the range resolution. Having in mind the structure of spreading PN code and theory for these sequences could be build on the following scene Fig. 5. In this figure are shown the main peak (main lobe) of the function of the autocorrelation of both codes.

The base width equals to 2T (T-the period of repetition of the code) [3]:

$$T_r = \frac{1}{C_r} = \frac{1}{1.2288 * 10^6} \approx 0.8\mu s \quad (14)$$

As seen from the figure, range resolution is different depending on the level of ACF it works on. The closer to the top the higher the accuracy of the range estimation. Factors that can influence the level are phase jitter and thermal noise. This is a problem that impedes the conventional usage of the

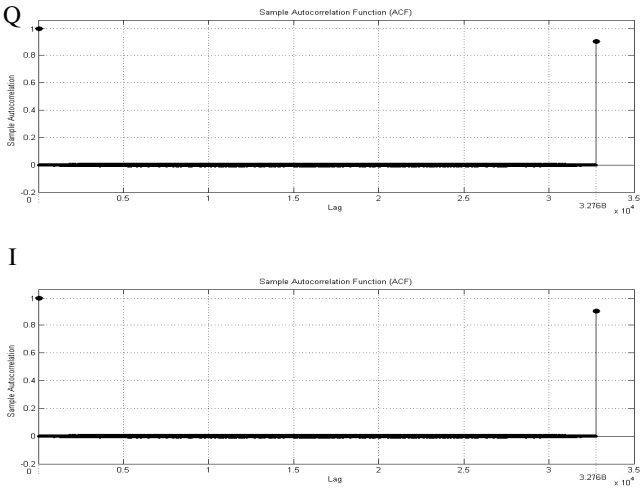


Fig. 4. ACF of the I and Q PN codes

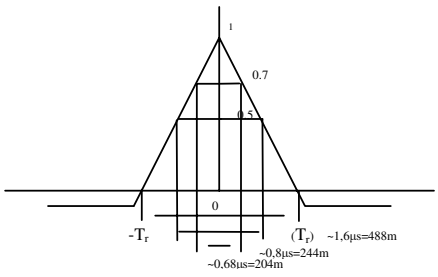


Fig. 5. Estimation of range resolution of the PN I and Q codes

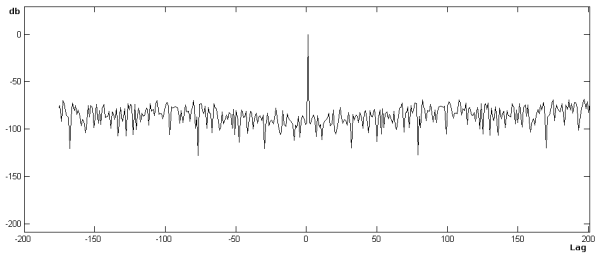


Fig. 6. Main lobe of the ACF of the PN I and Q codes

mobile network – it affects the synchronization of the receivers, so that it is being fought effectively. In Fig. 6 are shown the real situation of the main lobe of the ACF. From figure could be seen that the side lobes have very low level (less than 50 dB). That concludes good signal/noise (S/N) ratio and small probability of un-ambiguous range.

This consideration is presented in using the PN signal in the base band situation. There are possibilities to evaluation of the signal in the radio frequency range. In those situations observing of the phase shifting and the Doppler frequency could give us information about the target velocity.

III. Conclusion

The Secondary Applications of Wireless Technology is an promising aspect of using of wireless communications. The adoption of modern high quality digital radio communication systems and Spread Spectrum Signals will make these scientific researches fascinating.

Acknowledgements

I wish to sincerely thank **associate prof. Ph.D.Veselin Demirev**, from the Faculty of Communications and Communications Technologies, Technical University, Sofia, Bulgaria.

References

[1] J. Caffery, G. StÜber, *Overview of Radiolocation in CDMA Cellular Systems*, Georgia Institute of Technology, IEEE Communication Magazine, April 1998.

[2] М. Сколник, *Введение в технику радиолокационныіх систем*, Мир, М. 1965.

[3] G. Cooper, C. M$_c$Gillom, *Modern Communications and Spread Spectrum,* M$_c$Graw-Hill Book Co., 1996.

[4] M. Cherniakov, K. Kubik, *Secondary applications of wireless technology (SAWT)*,

[5] 2000 European Conference on Wireless Technology – Paris 2000.

[6] S. Haykin, *Communication systems*, M$_c$master university, John Wiley &Sons, Inc.Book, 1994.

[7] J. Lee, L. Miller, *CDMA Systems Engineering Handbook*, Artech House, Hardcover, November 1998.

[8] D. Masters, P. Axelrad, *University of Colorado, Boulder* Zavorotny V., *NOAA Environmental Technology Laboratory* Katzberg S.J., *NASA Langley Research Center* Lalezari F., *Ball Aerospace Corporation* "A Passive GPS Bistatic Radar Altimeter for Aircraft Navigation" Presented at ION GPS 2001, Salt Lake City.

[9] The MathWorks, Inc., http://www.mathworks.com/

The Influence of the Prediction Order onto the Accuracy of a BPNN Tracking Filter

Mimi D. Daneva[1]

Abstract – **In this paper an investigation of the accuracy of a BPNN tracking filter with respect to the model (prediction) order is presented. The performances of some BPNN filters of different order are evaluated and compared with standard recursive Kalman filter. Simulated and recorded real radar data are used.**

Keywords – **Radar data processing, neural networks.**

I. Introduction

The basic operations for radar data processing (RDP) are measurement formation, correlation and association of the measurements to the existent target trajectories, and track filtering and prediction. During the measurement formation stage, it is accepted to work in inertial coordinate system, so the measured kinematic parameters of the aircraft are naturally obtained in polar coordinates, and next, if is necessary, they are transformed in other coordinate system, more convenient for future work [1]. Different methods for track filtering and prediction are used for estimation of the current and future kinematic parameters (position, velocity, and acceleration) of the aircraft. For tracking on multiple targets, the achievement of high accuracy of the tracking filter is very important quality of the algorithm for filtering and prediction. This accuracy influences onto the quality of performance of the algorithm for measurement-to-track classification (data association) and plays a key role for correct classification decisions making. Two main types of tracking algorithms are used in practice: probabilistic and heuristic [2]. The first, such as recursive Kalman filter (KF), fixed coefficients filters, etc., are based on the probabilistical theory. The algorithms from the second group use simple heuristic rules or score functions. The neural approach for RDP belongs to the second group of algorithms. Its main advantages are the naturally embedded principle of parallel processing, an independence of the input data's statistics and the mathematical model of the observed dynamic system.

The kinematic model of non-maneuvering aircraft is second-ordered (with nearly constant velocity), and is defined by the equations [1,2]

$$\mathbf{X}(k+1) = \mathbf{\Phi}\mathbf{X}(k) + \mathbf{\Gamma}\mathbf{w}(k) \quad (1)$$

$$\mathbf{Z}(k) = \mathbf{H}\mathbf{X}(k) + , \quad (2)$$

where $\mathbf{X}(k) = \left[\mathbf{X}_1^T(k)\ \mathbf{X}_2^T(k)\ \mathbf{X}_3^T(k)\right]$ is the state vector of the dynamic system (the aircraft) with components the state vectors $\mathbf{X}_i = [\eta_i, \dot{\eta}_i]$ for coordinate i, $i = 1, 2, 3$. Each $\mathbf{X}_i$ contains position η_i and velocity $\dot{\eta}_i$. The symbol i marks the one of the target coordinates, and is used for notational simplicity. The discrete time interval is noted by k. The matrices $\mathbf{w}$ and $\mathbf{v}$ (both of dimension three) are mutually uncorrelated random-valued processes, each with zero mean, known variance and uncorrelated with $\mathbf{X}(0)$. The first, $\mathbf{w}(k)$, gives the random velocity's changes, and the second, $\mathbf{v}(k)$, models the radar measurement errors. The noise covariance matrices and the system matrices are, as follow [1]

$$\mathbf{Q} = [q_{ij}] = E\left\{\mathbf{w}(k)\mathbf{w}^T(k)\right\};$$
$$\mathbf{R} = [r_{ij}] = E\left\{\mathbf{v}(k)\mathbf{v}^T(k)\right\};$$
$$\mathbf{\Phi} = diag[\mathbf{\Phi}_2\ \mathbf{\Phi}_2\ \mathbf{\Phi}_2]; \mathbf{\Gamma} = diag[\mathbf{\Gamma}_2\ \mathbf{\Gamma}_2\ \mathbf{\Gamma}_2];$$
$$\mathbf{H} = diag[\mathbf{H}_2\ \mathbf{H}_2\ \mathbf{H}_2],$$

where the superscript T denotes the transpose operator and

$$\mathbf{\Phi}_2 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix};\ \mathbf{\Gamma}_2 = \begin{bmatrix} T^2/2 \\ T \end{bmatrix};\ \mathbf{H}_2 = [1, 0],$$

where T is the radar sampling interval.

In this paper a comparative analysis of the accuracy of several tracking filters for non-maneuvering aircrafts using back-propagation neural network (BPNN) with different architecture and different model (prediction) order for one-step-ahead prediction is presented. The performances of these filters are evaluated and compared with standard recursive KF for 50 Monte Carlo (MC) runs with equal input data in polar coordinates for all the filters at each run. An illustrative example using real radar data is shown. The experimental results show that the tracking error and the quality of BPNN training depend on the order of the filter and the best results are obtained when more available information about the tracking history is used.

II. BPNN for Time Series Prediction

A BPNN with static structure can be used as nonlinear predictor of a stationary time series [3,4]. In this case the elements of the input vector $\mathbf{X}_{inp}$ are the past input data samples, i.e.

$$\mathbf{X}_{inp} = [x_{inp}(k-1), x_{inp}(k-2), ..., x_{inp}(k-p)]^T , \quad (3)$$

where p is the prediction order, and the BPNN's output vector $\mathbf{Y}_p(k)$ produces the estimate $\hat{\mathbf{X}}_{inp}$ of the input vector for one-step-ahead prediction as

$$\mathbf{Y}_p(k) = \hat{\mathbf{X}}_{inp} \quad (4)$$

[1]Mimi D. Daneva is with the Faculty of Communications and Communicational Technologies, 8 Kl. Ohridski str., 1000 Sofia, Bulgaria, e-mail: mimidan@tu-sofia.acad.bg

The actual values of the elements of the input vector are used as elements of the desired target vector $\mathbf{T}_{inp}$. The input vector is applied to the input neurons of the BPNN. The hidden neurons produce the weighted sum of their output signals, transformed by the sigmoid functions. By summing the weighted output signals of all the hidden units, the output neurons form the estimate of the time series. The prediction error [3,4]

$$e(k) = x_{inp}(k) - \hat{x}_{inp}(k) \tag{5}$$

for iteration k is propagates in backward manner in the neural structure using the error back-propagation algorithm.

III. Design and Training of BPNN Tracking Filters Using Different Prediction Order

Two BPNN architectures with one and two hidden layers and different number of neurons in them are designed, trained, and compared for this investigation. The input and output layers contain equal number of neurons, $n_{inp} = n_{out} = 3$. The numbers of neurons in the first and the second hidden layer are denoted by n_{hid_1} and n_{hid_2}, respectively. They are chosen in a heuristic way to achieve the optimal network architecture for the concrete case [3,5]. The input and the hidden units have bipolar sigmoid functions with biases. The output units are linear.

The input data for the BPNN tracking filter are the polar coordinates range ρ, azimuth θ, and altitude h of the multiple targets, which form the measurement vector $\mathbf{Z}(k)$ at each radar scan k. They are preprocessed by a normalization procedure, so that they have zero mean and unity variance to ensure better convergence of the training algorithm. Thus, the input vector is formulated as [3]

$$\mathbf{X}_{inp} = [\mathbf{Z}^{\bullet}(k-1),\ \mathbf{Z}^{\bullet}(k-2), \ldots \mathbf{Z}^{\bullet}(k-p)] \tag{6}$$

where $\mathbf{Z}^{\bullet}(k-1)$ is the vector of the normalized measurements for all targets. They formed the training set with length Q (Q targets). After the data processing with BPNN, the original variables are recovered by inverse normalization procedure The maximum prediction order $p_{\max} = 3$ is chosen in view of track deletion criteria [1], the absence of measurements for a given track for 3 consecutive radar scans.

The Nguyen-Widrow hidden weights initialization procedure [5] was used for the BPNN filter. It prevents them from premature saturation during the first few iterations of the training algorithm. The BPNN training is performed by the back-propagation algorithm (BPA). The standard BPA is based on the method of steepest descent, but a number of modifications and variations of the standard BPA, such as Newton's and quasi-Newton methods, conjugate gradient methods, etc. may also be used [2-4]. The equations of the standard batch mode BPA for BPNN with two hidden layers are, as follow [3,4].

Forward computations: Computing the net internal activity levels, the output signals of the neurons in the layers and the error signals by the equations

$$v_j^{(l)}(k) = \sum_{i=1}^{i_{\max}} w_{ji}^{(l)}(k)\, y_i^{(l-1)}(k) \tag{7}$$

$$y_j^{(l)}(k) = \Psi\left(v_j^{(l)}(k)\right),\ l = 1, 2, 3 \tag{8}$$

$$e_j(k) = d_j(k) - o_j(k) \tag{9}$$

Backward computations: Computing the synaptic weight corrections and the local error gradients of the neurons in the layers as

$$\begin{aligned}\triangle w_{ji}^{(3)}(k) &= \mu\delta_j^{(3)}(k)\,\tilde{x}_i^{(3)}(k) = \\ &= \mu\delta_j^{(3)}(k)\, y_j^{(2)}(k) = \mu\delta_j^{(3)}(k)\, o_j(k)\end{aligned} \tag{10}$$

$$\delta_i^{(3)}(k) = e_i(k)\,\dot{\Psi}_i^{(3)}\left(v_i^{(3)}(k)\right) \tag{11}$$

for the output layer

$$\Delta w_{ji}^{(2)}(k) = \mu\delta_j^{(2)}(k)\,\tilde{x}_i^{(2)}(k) = \mu\delta_j^{(2)}(k)\, y_j^{(1)}(k) \tag{12}$$

$$\delta_i^{(2)}(k) = \dot{\Psi}_i^{(2)}\left(v_i^{(2)}(k)\right)\sum_{n_{out}}\delta_i^{(3)}(k)\, w_{ji}^{(3)}(k) \tag{13}$$

for the second hidden layer, and

$$\begin{aligned}\Delta w_{ji}^{(1)}(k) &= \mu\delta_j^{(1)}(k)\,\tilde{x}_i^{(1)}(k) = \\ &= \mu\delta_j^{(1)}(k)\, y_j^{(0)}(k) = \mu\delta_j^{(1)}(k)\,\tilde{x}_i^{(0)}\end{aligned} \tag{14}$$

$$\delta_i^{(1)}(k) = \dot{\Psi}_i^{(1)}\left(v_i^{(1)}(k)\right)\sum_{n_{hid_2}}\delta_i^{(2)}(k)\, w_{ij}^{(2)}(k) \tag{15}$$

for the first hidden layer, where μ is the learning rate, and $\tilde{x}_i^{(l)}$ is the input signal for neuron i in layer l. The synaptic weights are computed by iterations as

$$w_{ji}^{(l)}(k+1) = w_{ji}^{(l)}(k) + \mu\delta_j^{(l)}(k)\, y_i^{(l-1)}(k),\quad l = 1, 2, 3 \tag{16}$$

In the case of BPNN with single hidden layer ($l = 1, 2$) the corresponding values of the adjustable network parameters are computed by analogy [3,4]. The net performance function

$$E = \frac{1}{Q}\sum_{q=1}^{Q}\xi_q = \frac{1}{2}\sum_{q=1}^{Q}\sum_{j=1}^{n_{out}}(d_{jq} - y_{jq})^2 \tag{17}$$

for batch mode training is minimized as the weight changes are accumulated over all training examples before the weights are actually changed [3-5].

By local approximation of the error surface in the neighborhood of the operating point $w_{ji}^{(l)}$ using Taylor series, Eq. (16) can be written in matrix form as [4]

$$\begin{aligned}\mathbf{W}(k+1) &= \mathbf{W}(k) + \mathbf{\Delta W}(k+1) = \\ &= \mathbf{W}(k) - \overline{\mathbf{H}}^{-1}(k)\,\mathbf{g}(k)\end{aligned} \tag{18}$$

where $\mathbf{g} = \nabla_w\xi_q$ is the vector of local error gradient, and $\overline{\mathbf{H}} = \nabla_w^2\xi_q$ is the Hessian matrix.

The BPNN is trained by the algorithm of Marquardt-Levenberg, which has the fastest and guaranteed convergence compared with the other BP algorithms [4]. It approximates the Hessian matrix by

$$\overline{\mathbf{H}} \approx \left[\nabla_w^2\xi_q(\mathbf{W}(k)) + \bar{\nu}\mathbf{I}\right] \tag{19}$$

where $\bar{\nu}$ is the parameter of Levenberg, and $\mathbf{I}$ is identity matrix. The training stop when the global minimum of the net

performance function is found or the maximum number of epochs or the maximum learning rate is reached. The track filtering and prediction continues till the moment when the track deletion criteria, the absence of measurements about this track for three consecutive radar scans, is satisfied.

IV. Experimental Results

Simulated and real life data for 6 tracks of non-manoeuvering targets, chosen in random way, are used for the computer modelling. The real life data are recorded from Monopulse Secondary Surveillance Radar CMSSR-401 [6] with sampling time T=10 s. The real (live) tracks with length of 140 consecutive radar scans for the experiments are shown in Fig. 1. They are used as prototypes to model the tracks for the MC simulations, as follow. After polar to Cartesian transform to equalize the dimensions of the coordinates, the measurements from the first and the last radar scan, are connected with straight line and spaced with scan time T to form the trajectory of the non-manoeuvering target. The track T_1 is modelled using the first and the last point of the section with constant altitude and the first and the last point of the section with varying altitude. Next each track is corrupted with Gaussian noises, added directly to each coordinate to model the measurement errors according to the target kinematic model with Eqs. (1) and (2). The cumulative distribution function of the noises is verificated by chi-square test with significant level $\alpha = 0.05$. Next the tracks are transformed back in polar coordinates and then filtered by the algorithm under the test. The standard deviations of the radar measurement errors are 0.05 nautical miles (NM); $0.07°$ and 100 feet for ρ, θ, and h, respectively [6]. The acceleration's standard deviation for KF

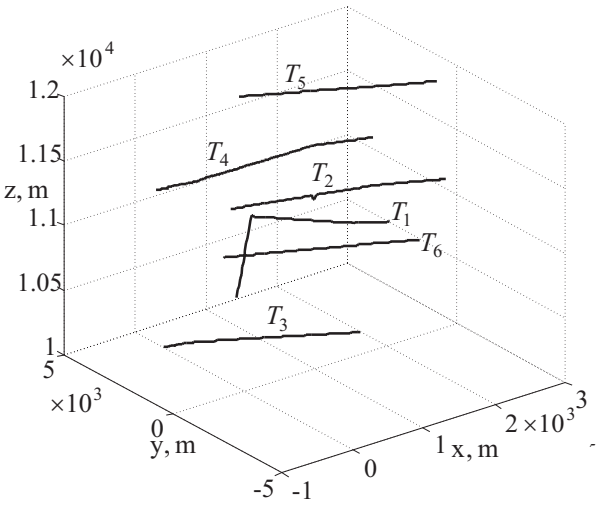


Fig. 1. Live track used for the experiments

Table 1. MC simulations results: BPNN training performance and computational costs

p	1		2		3	
$n_{hid\,1}$	12	12	12	12	12	12
$n_{hid\,2}$	-	6	-	6	-	6
$\bar{k}^T$	1	1	1	1	1	1
$\bar{E}^T_{final} \times 10^{-06}$	0,545	0,796	0,093	0,636	0,080	0,578
$\bar{t}_{CPU}$, s	0,49	0,66	0,55	0,92	0,67	1,14
$\bar{n}_{Mflops}$	1,101	4,640	2,486	8,422	4,755	1,196

Table 2. MC simulations results: BPNN and recursive Kalman filter's averaged tracking errors

Case	BPNN						KF
p	1		2		3		
$n_{hid\,1}$	12	12	12	12	12	12	p=1
$n_{hid\,2}$	-	6	-	6	-	6	
$\bar{\varepsilon}^{T_1}_{\rho}$, NM	0,25	0,24	0,19	0,20	0,24	0,23	1,10
$\bar{\varepsilon}^{T_2}_{\rho}$, NM	0,21	0,18	0,19	0,20	0,24	0,20	1,14
$\bar{\varepsilon}^{T_3}_{\rho}$, NM	0,32	0,22	0,29	0,22	0,32	0,24	1,50
$\bar{\varepsilon}^{T_4}_{\rho}$, NM	0,37	0,19	0,30	0,23	0,32	0,24	1,52
$\bar{\varepsilon}^{T_5}_{\rho}$, NM	0,27	0,18	0,23	0,15	0,25	0,18	1,23
$\bar{\varepsilon}^{T_6}_{\rho}$, NM	0,27	0,18	0,23	0,15	0,26	0,17	1,24
$\bar{\varepsilon}^{T_1}_{\theta}$, rad	0,012	0,011	0,011	0,009	0,013	0,011	0,017
$\bar{\varepsilon}^{T_2}_{\theta}$, rad	0,013	0,008	0,010	0,007	0,011	0,008	0,012
$\bar{\varepsilon}^{T_3}_{\theta}$, rad	0,014	0,010	0,013	0,011	0,014	0,011	0,008
$\bar{\varepsilon}^{T_4}_{\theta}$, rad	0,012	0,008	0,009	0,008	0,011	0,007	0,008
$\bar{\varepsilon}^{T_5}_{\theta}$, rad	0,013	0,010	0,012	0,008	0,012	0,010	0,009
$\bar{\varepsilon}^{T_6}_{\theta}$, rad	0,012	0,011	0,013	0,010	0,012	0,010	0,017
$\bar{\varepsilon}^{T_1}_{h}$, ft	20,35	12,94	18,28	13,20	16,78	12,80	39,52
$\bar{\varepsilon}^{T_2}_{h}$, ft	22,02	15,30	20,01	14,69	17,16	14,04	39,52

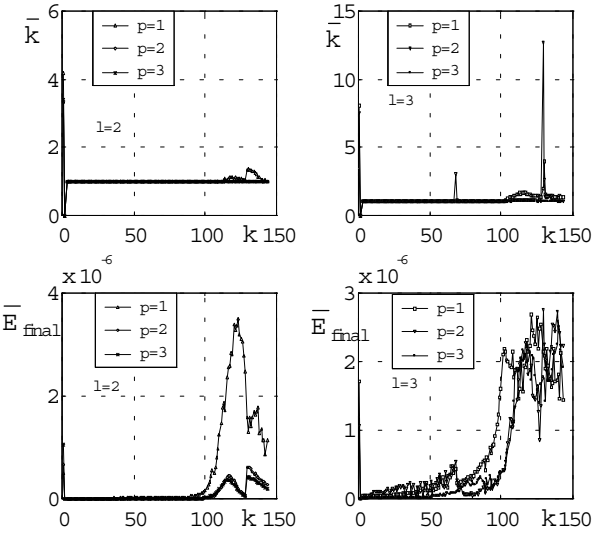


Fig. 2. MC results: BPNN training performance comparison

is 2 gm/s^2 [1]. The performances of BPNN tracking filters with prediction order p are compared with those of standard recursive KF. The tracking error $\zeta$of each filter is defined by the difference between the measurement vector and the estimated state vector [1]. The root mean square (*rms*) values

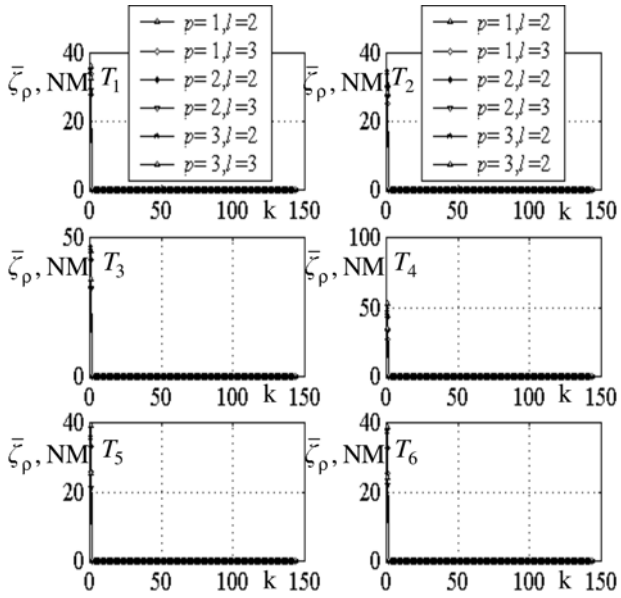


Fig. 3. MC results: BPNN tracking errors for range

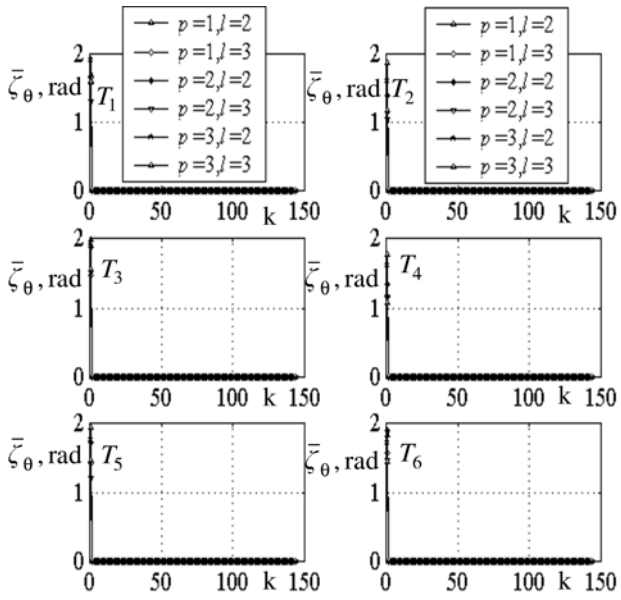


Fig. 4. MC results: BPNN tracking errors for azimuth

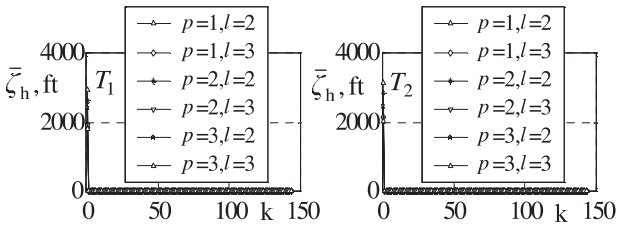


Fig. 5. MC results: BPNN tracking errors for altitude

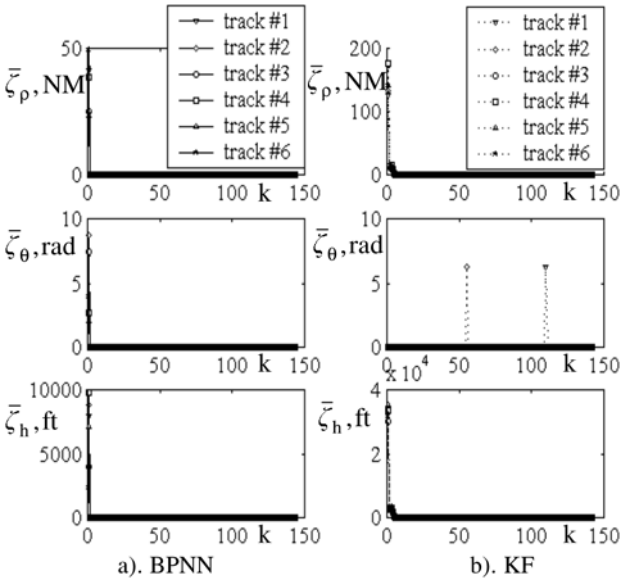


Fig. 6. BPNN and KF's tracking errors for 6 real tracks

of ζ are averaged over all MC runs to form the *rms* error at each time point. All results are obtained by Intel Celeron 500 PPGA with SDRAM 128 MB.

The results from 50 MC runs of the BPNN tracking filter with prediction order p, compared with those of KF for tracks $T_1 - T_6$are presented in Table 1 and Table 2. The averaged BPNN's parameters for the MC runs are averaged number of epochs $\bar{k}$, the averaged net performance function at the end of training $\bar{E}_{final}$, and the *rms* value of $\bar{\zeta}$ for each coordinate, averaged for the MC runs and for the whole track. The results for h coordinate only for the track with variable altitude and for one of the rest tracks are presented here, because the results for the tracks with $h = const$ are very similar. The computational complexity of each filter is estimated by the averaged parameters CPU time and number of Mflops, used for data processing of all training examples (all measurements for all the targets for scan k). The BPNN training performances are plotted in Fig. 2; the tracking errors during MC runs for ρ, θ, and h for all tracks are plotted in Figs. 3 to 5. It is seen that the relationships among the BPNN training parameters and the tracking error with respect to the prediction order are very similar. When more information about the tracking history of the targets is used, the net performance function at the end of training is more close to the global minimum. The network training time and flops increase dramatically as the number of hidden nodes in the layers increases as well as p increases. The best performance of the BPNN filter is with p=3 and BPNN with 12 hidden units in a single hidden layer, the training is stabile and requires almost constant number of epochs. An illustrative example of BPNN tracking performance for this case, compared with the KF for the real tracks is shown in Fig. 6. The BPNN tracking error has lower values than that of KF in all the cases. It is clear that the highest tracking error appears when the BPNN is initialized, and next this error decreases rapidly and keeps almost constant in time. The BPNN initial errors are different for each track, due to the different initial weights. The error

due to recovering of the original variables after the inverse normalization does not affect to the accuracy of the BPNN algorithms.

V. Conclusion

This paper has presented an investigation of several BPNN filters with respect to the used prediction order and tracking accuracy. They perform state estimation using non-linear optimization, and do not require prior statistical information on the noise and the target kinematic model. The optimal BPNN architecture and prediction order is chosen by experimental way. It ensures the highest accuracy than the other considered cases, and consistently produces smaller tracking errors than the standard recursive Kalman filter for non-manoeuvering aircrafts. It will improve the performance of the measurement-to-track association logic algorithm in future research.

References

[1] S. Blackman, *Multiple Target Tracking with Radar Applications*, Norwood, Artech House, 1986.

[2] Y. Bar-Shalom (editor), *Multiple-Target Tracking*, vol. I, 1990, and vol. II, 1992, Dedham, Artech House.

[3] S. Haykin, *Neural Networks*, New York, Macmillan College Publishing Company, 1994.

[4] A. Cichocki, R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, Stuttgart, John Wiley & Sons & B. G. Teubner, 1993.

[5] C. G. Looney, *Pattern Recognition Using Neural Networks*, New York, Oxford University Press, 1997.

[6] *Monopulse Secondary Surveillance Radar System Description*, Technical Report, Cardion Inc., *Report no. 131-162A.*

A Comparative Analysis of Some Data Preprocessing Techniques for BPNN Tracking Filter

Mimi D. Daneva[1]

Abstract – **In this paper the performances of a BPNN tracking filter depending on the method used for pattern formulation are evaluated and compared. Several input data preprocessing techniques are used for the investigation. The tracking error is compared with this of standard recursive Kalman filter using simulated and real radar database.**

Keywords – **Radar data processing, neural networks.**

I. Introduction

The algorithms for filtering and prediction of the target trajectories, also called tracking filters provide the estimate of the current and future kinematic parameters of the aircraft, used next in the data association process. The accuracy of the estimate, provided by the filter, effects to a great degree to the correct decision for data association, which by it reflects to the safety level in air traffic control [1]. As a part of the total set of algorithms for multiple target tracking, the algorithms for track filtering and prediction can be divided in two groups: probabilistic and heuristic [2]. The tracking filters based on artificial neural networks and the artificial intelligence approach belong to the second group. They differ to the first group filters in two main features [3]. First, they do not need to a priori knowledge about the statistics of the input data, and second, these target state estimators are "model-free" – contrary to the statistical estimators they can estimate input-output functions without a mathematical model of how the output of the dynamic system depends on its input. Using neural approach for data processing the choice of appropriate method of input data representation is an important factor for successful convergence of the training algorithm. In this paper, three methods for data preprocessing and possible combinations of them, used for BPNN tracking filter, are compared by Monte Carlo (MC) experiment with respect to their training performance, the tracking error, the error due to recovering of the original variables after data postprocessing, and the computational costs. The performances of the BPNN filter using different data preprocessing techniques are compared with standard recursive Kalman filter (KF) by 50 MC runs. Simulated input data according to kinematic model of non-maneuvering aircraft in polar coordinates [1], [2] and real radar data records from the plot extractor's output of Monopulse Secondary Surveillance Radar CMSSR-401 [4] are used for the investigation. All experiments are implemented in *MATLAB* environment. Interesting results are obtained.

[1]Mimi D. Daneva is with the Faculty of Communications and Communicational Technologies, 8 Kl. Ohridski str., 1000 Sofia, Bulgaria, e-mail: mimidan@tu-sofia.acad.bg

II. Input Data Representation

The input data representation, pattern formulation or input coding is a process of mapping the input feature space onto a set of input units of the neural network (NN). The choice of the most appropriate input coding effects onto the adequate NN's training performance and the error. For input data representation, normalization coding (NC) is widely used for pattern recognition, system identification or estimation purposes. It represents the input feature accurately, and is distinguished with naturalness and computational simplicity [5]. The simplest form of normalization ranges the input data in $[0; 1]$or $[-1; 1]$ (NC1). In [6] it is recommendable to distribute the input data in the interval $[0; 0.9]$ or $[-0{,}9; 0.9]$ (NC2) to avoid the high nonlinearly zones of the neurons' activation functions. The equations of NC1 and NC2 are as follow [6]

$$\mathbf{P}_n = 2\,(\mathbf{P} - P_{\min})\,/\,(P_{\max} - P_{\min}) - 1\,, \quad (1)$$

$$\mathbf{P}_n = 1{,}8\,(\mathbf{P} - P_{\min})\,/\,(P_{\max} - P_{\min}) - 0.9\,, \quad (2)$$

where $\mathbf{P}$and $\mathbf{P}_n$ are the original and the normalized pattern vectors, respectively; $P_{\min}$ and $P_{\max}$ are the minimum and maximum values of $\mathbf{P}$.

The input data can also be normalized so that they will have zero mean and unity standard deviation (NC3) as

$$\mathbf{P}_n = (\mathbf{P} - m_P)\,/\sigma_P\,, \quad (3)$$

where m_P and σ_P are the mean and standard deviation of $\mathbf{P}$, respectively (one-class normalization). This technique is effective for pattern classification purposes in the case of more one class [5].

Usually, the pattern formulation process is considered as an input space transformation, which transforms the dimension of the input space in more effective features [5]. If this transform is linear, then the function that maps the input space into the output variables for data preprocessing is well defined, and the preprocessing task reduces to determine the coefficients of this linear function with respect to minimize or maximize some optimization criterion. One approach to do it is a discrete cosine transform (DCT), which is widely used for image coding. It originally was developed as an approximation of the optimal Karhunen-Loève transform, but a number of fast algorithms have been developed for its computation [7]. In *MATLAB* environment, one way to compute the DCT is through the form DCT-IV as [8]

$$\mathbf{P}_{\mathrm{DCT}}(k) = w(k)\sum_{n=1}^{N}\mathbf{P}(n)\cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right)\,, \quad (4)$$

where $w(k) = \begin{cases} 1/\sqrt{N}, & k = 1 \\ \sqrt{2/N}, & 2 \leqslant \mathrm{k} \leqslant N \end{cases}$, N is the length of $\mathbf{P}$, and k is the discrete time. The DCT is a purely real orthogonal transform, which can be used as an additional transformation of the input space before normalization and next processing by the neural network.

The input data for the proposed BPNN tracking filter for non-maneuvering targets are the polar coordinates range ρ in nautical miles (NM), azimuth angle θ in rad, and the altitude h in feet for 6 non-manoeuvering targets. They are preprocessed independently using NC1, NC2 and NC3, and by DCT followed by NC1, NC2 and NC3. The goal is to choose the most appropriate input data preprocessing technique for the BPNN tracking filter, accordingly to the specific high accuracy requirements and acceptable computational costs [2].

The block diagram of data processing with BPNN tracking filter is shown in Fig. 1, where **W** and **B** denote the weight matrix and the bias vector of the BPNN, respectively. The prediction error of BPNN is denoted by $\mathbf{e}_{\mathrm{BPNN}}(k,q)$. It propagates in backward manner in the neural structure by the training algorithm. The relative error due to reconstruction of the original variables after the data postprocessing is defined by [8]

$$\xi_{rec} = \frac{\| \mathbf{P} - \mathbf{P}_{rec} \|}{\| \mathbf{P} \|} . \tag{5}$$

For comparison of the results the same input data as for the BPNN filter are processed by standard recursive Kalman filter (KF) [1,2]. The tracking error for a single target is defined by

$$\zeta(k+1) = \mathbf{Z}(k+1) - \mathbf{H}\hat{\mathbf{X}}(k+1) , \tag{6}$$

where **H** is the measurement matrix.

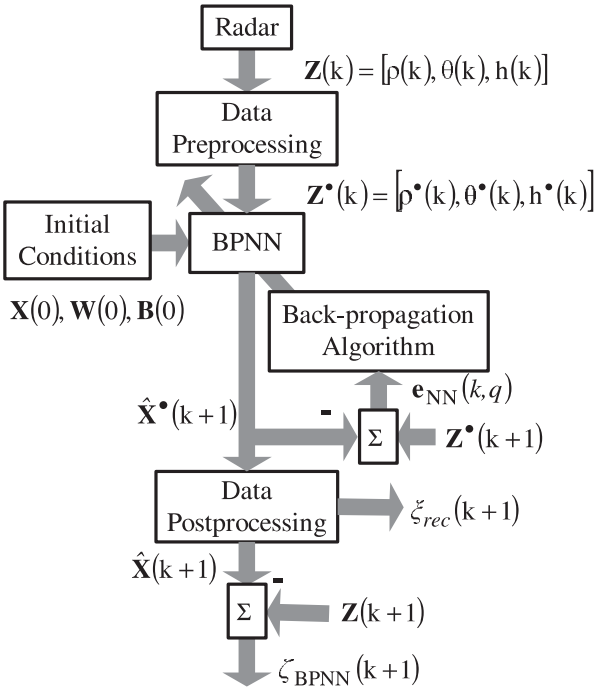


Fig. 1. Block diagram for track filtering and one-stepahead prediction with BPNN

III. Architecture and Training of BPNN Tracking Filter

The BPNN employed is fully connected and has 3 input nodes, a single hidden layer with 12 units, and 3 output units. The input and hidden neurons have bipolar sigmoid functions, the output neurons are linear. This architecture was chosen as optimal in heuristical way [6] by separated experiment. The BPNN can be used as nonlinear predictor of a stationary time series [9]. In our case the elements of the input vector are the past samples of the preprocessed radar measurements, i.e.

$$\mathbf{X}_{inp} = [\mathbf{Z}^{\bullet}(k-1),\ \mathbf{Z}^{\bullet}(k-2),\ \ldots,\ \mathbf{Z}^{\bullet}(k-p)]^{T} \tag{7}$$

where $\mathbf{Z}^{\bullet}(k-1)$is the measurement vector after preprocessing procedure, and p=3 is the model (prediction) order. The actual values of the elements of the input vector are used as elements of the desired target vector $\mathbf{T}_{inp}(k)$. The BPNN's output vector $\mathbf{Y}_P(k)$ produces the estimate $\hat{\mathbf{X}}_{inp}$ of the input vector for one-step-ahead prediction as

$$\mathbf{Y}_P(k) = \hat{\mathbf{X}}_{inp} . \tag{8}$$

The Nguyen-Widrow hidden weights initialization procedure [6] was used for network initialization. The Marquardt-Levenberg algorithm, which has the fastest and guaranteed convergence compared with the other BP algorithms [3], [9], was used for BPNN training. The training stop when the global minimum of the network performance function [9]

$$E = \frac{1}{Q}\sum_{q=1}^{Q}\xi_q = \frac{1}{2}\sum_{q=1}^{Q}\sum_{j=1}^{n_{out}}(d_{jq} - y_{jq})^2 \tag{9}$$

is reached, where d_{jq} and y_{jq} are the desired target and the actual output signal of the j-th output neuron for the $q-$th training example is found, or the maximum number of epochs or the maximum learning rate. The track filtering and prediction continues till the moment when the track deletion criteria [1], the absence of measurements about this track for three consecutive radar scans, is satisfied.

IV. Experimental Results

Simulated and live data for 6 tracks of non-manoeuvering targets, chosen in random way are used to form the training set for the computer modelling. The real data are recorded from the plot extractor's output of CMSSR-401. The measurement errors are 0.05 NM; 0.07°, and 100 feet for ρ, θ, and h, respectively; the radar sampling time is $T = 10$ s [4]. Different random input sequences are generated for each MC run to simulate the measurement errors. Their normal cumulative distribution function is verificated by chi-square test with significant level $\alpha = 0.05$. The real tracks with length of 140 consecutive scans for the experiments are shown in Fig. 2. They are used as prototypes to model the tracks for the MC simulations, as follow. After polar to Cartesian transform to equalize the dimensions of the coordinates, the measurements from the first and the last radar scan, are connected

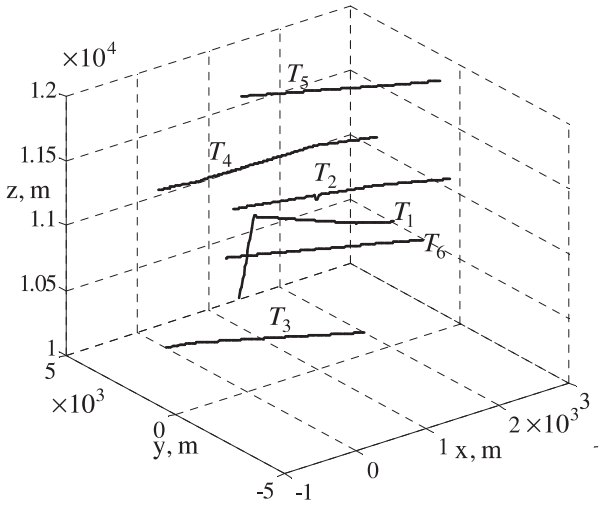


Fig. 2. Live tracks used for the experiments

Table 1. MC simulations results: BPNN and recursive Kalman filter's averaged tracking errors

FILTER / ERRORS	BPNN						KF
	NC1	NC2	NC3	DCT &NC1	DCT &NC2	DCT &NC3	
$\bar{\zeta}_{\rho}^{T_1}$, NM	0,29	0,31	0,24	3,85	4,74	3,40	1,10
$\bar{\zeta}_{\rho}^{T_2}$, NM	0,29	0,32	0,24	3,17	2,96	1,73	1,14
$\bar{\zeta}_{\rho}^{T_3}$, NM	0,36	0,39	0,32	2,09	2,28	1,87	1,50
$\bar{\zeta}_{\rho}^{T_4}$, NM	0,35	0,38	0,32	2,20	2,12	1,86	1,52
$\bar{\zeta}_{\rho}^{T_5}$, NM	0,29	0,32	0,25	1,67	2,00	1,69	1,24
$\bar{\zeta}_{\rho}^{T_6}$, NM	0,30	0,34	0,26	1,98	2,19	1,58	1,23
$\bar{\zeta}_{\theta}^{T_1}$, rad	0,014	0,015	0,013	0,021	0,032	0,021	0,017
$\bar{\zeta}_{\theta}^{T_2}$, rad	0,012	0,013	0,011	0,020	0,015	0,008	0,012
$\bar{\zeta}_{\theta}^{T_3}$, rad	0,014	0,015	0,013	0,011	0,014	0,012	0,008
$\bar{\zeta}_{\theta}^{T_4}$, rad	0,012	0,011	0,007	0,008	0,013	0,010	0,008
$\bar{\zeta}_{\theta}^{T_5}$, rad	0,013	0,014	0,010	0,014	0,015	0,011	0,009
$\bar{\zeta}_{\theta}^{T_6}$, rad	0,012	0,013	0,010	0,015	0,014	0,011	0,017
$\bar{\zeta}_{h}^{T_1}$, ft	24,67	27,27	16,78	783,8	896,8	545,3	39,52
$\bar{\zeta}_{h}^{T_6}$, ft	32,10	34,75	21,98	443,3	414,7	338,2	37,38

with straight line and spaced with scan time T to form the trajectory of the non-manoeuvering target. The track T_1 is modelled using the first and the last point of the section with constant altitude and the first and the last point of the section with varying altitude. Next each track is corrupted with Gaussian noises, added directly to each coordinate to model the measurement errors. Next the tracks are transformed back in polar coordinates and then filtered by the algorithm under the test. The tracking error $\zeta$of each filter is defined by the difference between the measurement vector and the estimated state vector [1]. The experiments include 50 MC runs and track example using real radar data. The BPNN filters with

Table 2. MC simulations results: BPNN training performance and computational costs

CASE	NC1	NC2	NC3	DCT& NC1	DCT& NC2	DCT& NC3
$\bar{k}^T$	1	1	1	1	1	1
$\bar{E}_{final}^T \times 10^{-06}$	0,036	0,032	0,080	0,480	0,229	0,196
$\bar{t}_{CPU}$, s	0,69	0,68	0,67	0,50	0,67	0,67
$\bar{n}_{Mflops}$	4,745	4,742	4,755	1,073	1,073	4,694

cases of preprocessing NCs 1 to 3, with and without DCT as an additional data preprocessing are compared with KF. The root mean square (*rms*) values of ζ are averaged over all MC runs to form the *rms* error at each time point. The acceleration's standard deviation for KF is $2g$ m/s^2 [1]. The averaged BPNN's parameters are the averaged number of epochs $\bar{k}$, the averaged net error function at the end of training $\overline{E}_{final}$, and

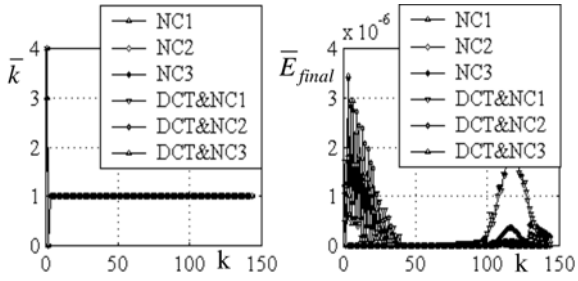


Fig. 3. MC results: BPNN training performance comparison

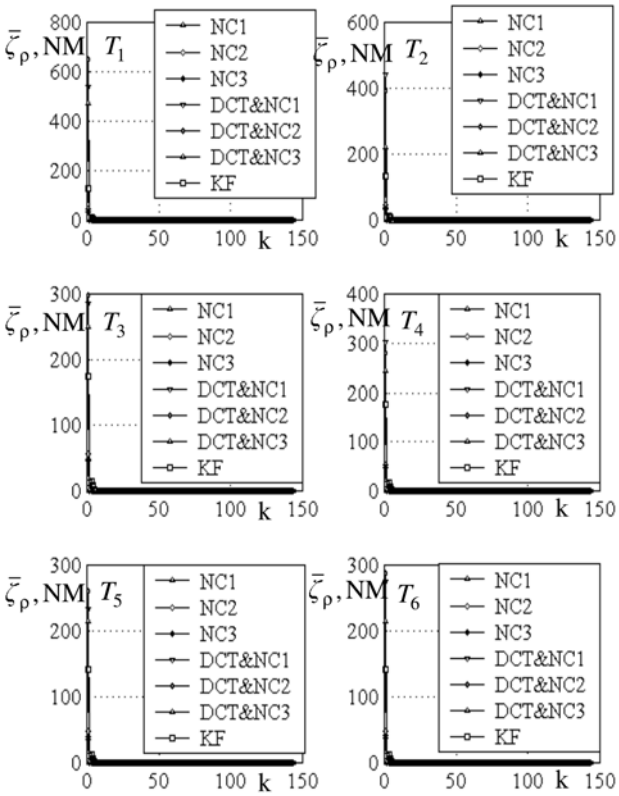


Fig. 4. MC results: BPNN and KF tracking errors for range

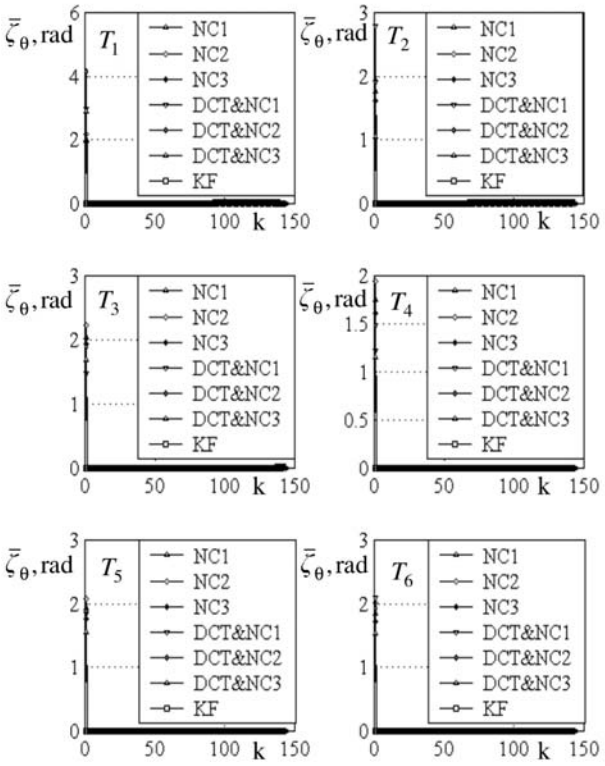


Fig. 5. MC results: BPNN and KF tracking errors for azimuth

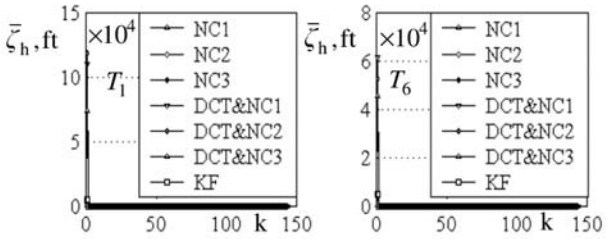


Fig. 6. MC results: BPNN and KF tracking errors for altitude

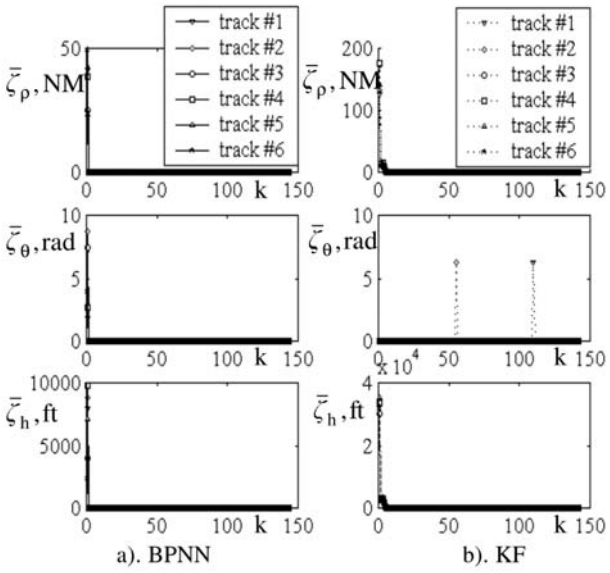


Fig. 7. BPNN and KF's tracking errors for 6 real tracks

Table 3. MC simulations results: averaged relative error due recovering

Error \ Case	NC1	NC2	NC3	DCT& NC1	DCT& NC2	DCT& NC3
$\bar{\zeta}_{\text{rec}} \times 10^{-18}$	0,056	0,015	0,021	37,7	40,4	39,1

the *rms* value of $\bar{\zeta}$ for each coordinate, averaged for the MC runs and for the whole track. The computational complexity of each BPNN filter is estimated by the averaged parameters CPU time and number of Mflops, used for data processing of all training examples (all measurements for all the targets for scan k). All results are obtained by Intel Celeron 500 PPGA with SDRAM 128 MB.

The results from the MC runs are presented in Tables 1 and 2. The BPNN training performances are plotted in Fig. 3, the tracking errors during MC runs for all the coordinate for all tracks are plotted in Figs. 4 to 6. The BPNN filters with NCs 1 to 3 produce smaller tracking errors than the KF. The best BPNN performance is achieved when NC3 is used. That is due to the specific quantities of data preprocessing in this case, which promotes the speed and learning convergence of the BPNN filter. In the cases of DCT&NCs 1 to 3 the BPNN filter needs less training time, epochs, and Mflops, but the learning convergence and the tracking performance degrade. Till this moment all BPNN filters are investigated with one and the same value of the parameter goal $g_{\min} = 0.00001$ during the training. Obviously this value is insufficiently small for the cases of all DCT&NCs. An appropriate correction of the goal will improve the training, but in Figs. 4 and 6 we can see that the BPNN initial errors for these cases are quite high (they are greater than these of KF). That is due to the specific quantities of the DCT, which increases the distances between the classes (i.e. the tracks) in the transformed space. This is suitable for data classification purposes, but in the case of time-series prediction make the initial accuracy worse. In the cases of NCs 1 to 3 the BPNN tracking errors are smaller than the KF. The best results of the BPNN filter are obtained using NC3. This conclusion is confirmed with the example of multiple-target tracking using real radar data record, shown in Fig. 7. The averaged relative error due to re-construction of the original variables after data postprocessing is presented in Table 3. It is due to rounding during the mathematical operations and does not affect to the tracking accuracy of the filters, because the recovering of the data is 100 %.

V. Conclusion

In this paper the results of an experimental comparison among three methods for input data preprocessing and some combinations between them before target track filtering and prediction then has been presented and analyzed by Monte Carlo experiment. The results show that these preprocessing techniques influence onto the BPNN training and it's tracking performance. The smaller tracking errors of the algorithm are obtained when a normalization coding with zero mean and unity standard deviation is used. It improves the accuracy of the tracking algorithm and reduces the computational costs.

References

[1] S. Blackman, *Multiple Target Tracking with Radar Applications*, Norwood, Artech House, 1986.

[2] Y. Bar-Shalom (editor), *Multiple-Target Tracking*, vol. I, 1990, and vol. II, 1992, Dedham, Artech House.

[3] B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice Hall International, 1992.

[4] *Monopulse Secondary Surveillance Radar System Description*, Technical Report, Cardion Inc., *Report No. 131-162A.*

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, New York and London, Academic Press, 1972.

[6] C. G. Looney, *Pattern Recognition Using Neural Networks*, New York, Oxford University Press, 1997.

[7] J. H. McClellan, C. Sidney Burrus, A. V. Oppenheim, T. W. Parks, R. W. Schafer, H. W. Schuessler, *Computer-Based Exer-cises for Signal Processing Using Matlab® 5*, New Jersey, Prentice Hall, 1998.

[8] *Signal Processing Toolbox User's Guide Version 4.2*, The Math Works, Inc., 1998-1999.

[9] A. Cichocki, R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, Stuttgart, John Wiley & Sons & B. G. Teubner, 1993.

IM/DD Optical System Performance in the Presence Timing Jitter and Gaussian Noise

Mihajlo Stefanovic[1], Dragan Draca[2], Petar Spalevic[3] and Aleksandra Panajotovic[4]

Abstract – **In this paper the IM/DD optical system is considered by treating the optical fiber as a linear and nonlinear medium, respectively. First, the conditions under which the dispersion effects dominate over the nonlinear effects are defined. The case of Gaussian pulse propagation is assumed. The error probability is calculated for unchirped incident pulses in the presence of timing jitter and Gaussian noise. The previous analysis repeats for a nonlinear and dispersive fiber. Also the detailed analysis of obtained results is performed.**

Keywords – **Fiber Optic Communications, IM/DD Optical System, Timing Jitter, Gaussian Noise, Bit Error Probability**

I. Introduction

The wavelength band and the optical communication system performance are determined by optical fiber characteristics. In this paper the attention is focused on the fiber dispersion and combined dispersion and nonlinear characteristics.

First, the optical system with Intensity Modulation and Direct Detection (IM/DD) is considered when the incident power and the fiber length are such that nonlinear effects can be neglected.

Dispersion is a consequence of the refractive index frequency dependence. Namely, different pulse spectral components travel with different speeds and have different time delays that lead to the pulse broadening with propagation through fiber. This generates the bit interference and the greater possibility of wrong detection. Such performance degradation is considered for Gaussian pulses.

Also, in this paper we mostly consider equal influence of both effects on pulse shape and also the case when exist self influence of dispersive effects. The main effect of dispersive influence of optical fiber is broadening an optical pulse as it propagates through the fiber. Size of these effects depends from values of GVD (group-velocity dispersion) parameter β_2.

GVD parameter can be positive or negative disregarding the light wavelength, which is below or above the zero-dispersion wavelength λ_D of fiber. Intensity dependence of the refractive index in nonlinear media occurs through SPM (self-phase modulation), a phenomenon that leads to spectral broadening of optical pulses. Size of nonlinear effect depends from values of parameter γ [1].

Direct detection is realized by sampling the photodiode current and by comparison the obtained sample with threshold. Because of timing jitter the sampling moment varieties. The error probability in the presence of timing jitter is also determined when the fiber is represented as a linear, i.e. nonlinear medium.

II. Theoretical Basis of the Pulse Propagation

Propagation short pulse, which width is between 10 fs and 50 ps, along the nonlinear-dispersive optical fiber can be described by Schrödinger equation. It is [1-3]:

$$\frac{\partial A}{\partial z} = -\frac{1}{2}\alpha A - \frac{i}{2}\beta_2 \frac{\partial^2 A}{\partial T^2} + i\gamma |A|^2 A \qquad (1)$$

where A is slowly varying amplitude of pulse envelope and $T = t - z/v_g$, v_g is group velocity, $\gamma = n_2\omega_0/(cA_{eff})$ is coefficient nonlinearity, A_{eff} is effective core area, GVD parameter is $\beta_2 = \partial^2\beta/\partial\omega^2\Big|_{\omega=\omega_0}$, n_2 is nonlinear-index refractive coefficient.

It is useful to observe eq. (1) in normalized form and then we can use following normalized parameters:

$$\tau = \frac{t - \beta_1 z}{T_0}, \quad \xi = \frac{z}{L_D}, \quad U = \frac{A}{\sqrt{P_0}} \qquad (2)$$

where T_0 is the half width (at $1/e$-intensity point) of pulse, P_0 is the peak power of the incident pulse and L_D is the dispersion length, i.e.

$$L_D = \frac{T_0^2}{|\beta_2|}. \qquad (3)$$

Than for $\alpha = 0$, eq. (1) takes normalized form:

$$i\frac{\partial U}{\partial \xi} = \mathrm{sqn}(\beta_2)\frac{1}{2}\frac{\partial^2 U}{\partial \tau^2} - N^2|U|^2 U \qquad (4)$$

where $\mathrm{sgn}(\beta_2)$ takes values +1 or -1 in dependence of dispersive regime ($\beta_2 > 0$ – normal and $\beta_2 < 0$ – anomalous dispersion regime). The eq. (5) is known as Nonlinear Shrödinger equation (NSE).

The parameter N is defined as

$$N^2 = \frac{\gamma P_0 T_0^2}{|\beta_2|} = \frac{L_D}{L_{NL}} \qquad (5)$$

[1]Mihajlo Stefanovic is with the Faculty of Electronic Engineering, University of Nis, Beogradska 14, 18000 Nis, Serbia and Montenegro, E-mail: trmka@ptt.yu or misa@elfak.ni.ac.yu

[2]Dragan Draca with the Faculty of Electronic Engineering, University of Nis, Beogradska 14, 18000 Nis, Serbia and Montenegro, E-mail: dracaa@elfak.ni.ac.yu

[3]Petar Spalevic is with the Faculty of Technical Science, University of Pristina, Kneza Milosa 7, 28000 Kosovska Mitrovica, Serbia and Montenegro, petarspalevic@yahoo.com.

[4]Aleksandra Panajotovic is with the Faculty of Electronic Engineering, University of Nis, Beogradska 14, 18000 Nis , Serbia and Montenegro, E-mail: pandrica@elfak.ni.ac.yu

and it represents nondimensional combination of the pulse and fiber parameters. Dispersion dominates for $N \ll 1$, while SPM dominates for $N \gg 1$. In eq. (5) parameter L_{NL} is the nonlinear length and it is defined as

$$L_{NL} = \frac{1}{\gamma P_0}. \tag{6}$$

Also, in this paper we shall consider the effect of GVD on pulse propagation in linear dispersive medium by setting $N = 0$ (i.e. $\gamma = 0$) in eq. (4):

$$i\frac{\partial U}{\partial \xi} = \text{sgn}(\beta_2)\frac{1}{2}\frac{\partial^2 U}{\partial \tau^2}. \tag{7}$$

This paper considers the pulse-propagation problem in IM/DD optical system with timing jitter by treating the fiber as a linear and nonlinear and dispersive medium.

III. The System Performance

For pulse widths greater than 0.1 ps, in linear dispersive medium by setting $\gamma = 0$ where the fiber loss is periodically compensated, $U(z,T)$ – a length normalized complex amplitude satisfies a linear partial differential equation given by [1].

$$i\frac{\partial U}{\partial z} = \frac{1}{2}\frac{\partial^2 U}{\partial T^2}. \tag{8}$$

The case of a Gaussian pulse is considered. The incident field is given by

$$U(0,T) = \exp\Big(-\frac{T^2}{T_0^2}\Big). \tag{9}$$

The linear partial differential equation is solved by using the Fourier method [1] and the solution, i.e. the normalized complex amplitude at any point z along the fiber is

$$\begin{aligned} U(z,T) &= \frac{1}{2\pi}X \\ X &= \int_{-\infty}^{\infty}\tilde{U}(0,\omega)\exp\Big(\frac{i}{2}\beta_2\omega^2 z - i\omega T\Big)d\omega \end{aligned} \tag{10}$$

Then, we consider incident pulses which propagated along the nonlinear and dispersive propagation regime. The power normalized eq. (1) represents nonlinear partial differential equation and we can use a great number of numerical methods for its solving. One of those methods is *"Split-step" Fourier method* [1,3], which represents pseudospectral methods. In this paper, this method is used for solving nonlinear Schrödinger equation, when incident pulse has considered Gaussian form.

The change of the pulse width due to propagation through the fiber can lead to appearance of the interference. So, the equivalent complex amplitude for the fiber length $z = L$ is

$$U_{eq}(L,T) = U(L,T) + \sum_{n=1}^{m/2} b_{\pm n}U(L, T \mp n2T_0) \tag{11}$$

if "1" is sent, or

$$U_{eq}(L,T) = \sum\nolimits_{-n=1}^{m/2} b_{\pm n}U(L, T \mp n2T_0) \tag{12}$$

if "0" is sent, where m is the number of the neighborhood pulses which interfere and $b_{\pm n}$ is the coefficient from set $\{0,1\}$ depending if 0 or 1 is transmitted.

In IM/DD optical system [3] the photodiode current sample compares with the threshold level determined for ideal case. Since the photodiode is the quadrate detector, the complex photodiode current I is

$$I = R|U_{eq}(L,T)|^2 + n(T) \tag{13}$$

where R is a conversion coefficient and $n(T)$ is a Gaussian noise. The sampling moment due to timing jitter [4] is not at $T = 0$, i.e. $T = T_J$. Then the complex amplitude I at the sampling point T_J for a given combination of bits $b_{\pm n}$, $n = 1, m/2$ is a Gaussian variable with a mean value $R|U_{eq}(L,T_J)|^2$ and variance σ^2

$$\begin{aligned} &p_1(I/b_{-m/2},...,b_{-1},b_1,...,b_{m/2}) = \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{\left[I - R\left|U(L,T_J) + \sum\limits_{n=1}^{m/2} b_{\pm n}U(L,T_J \mp n2T_0)\right|^2\right]^2}{2\sigma^2}\right) \\ &p_0(I/b_{-m/2},...,b_{-1},b_1,...,b_{m/2}) = \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{\left[I - R\left|\sum\limits_{n=1}^{m/2} b_{\pm n}U(L,T_J \mp n2T_0)\right|^2\right]^2}{2\sigma^2}\right) \end{aligned} \tag{14}$$

As bits are independent, each having a 0.5 probability of being 1 and 0, the conditional probability density functions are averaged with $1/2^m$. The error is made if 0 is sent and the current sample is greater than threshold

$$I_T = \frac{R|U(L,0)|^2}{2} \tag{15}$$

and vice versa. Then the error probability is [4,5]

$$\begin{aligned} BER &= P(D_1,H_0) + P(D_0,H_1) = \\ &= \frac{1}{2}P(D_1/H_0) + \frac{1}{2}P(D_0/H_1) = \\ &= \frac{1}{2}\sum_{b_{-m/2},...,b_{m/2}=0...0}^{1...1}\frac{1}{2^m}\frac{1}{2}\text{erfc}\left(\frac{I_T - R\left|\sum\limits_{n=1}^{m/2} b_{\pm n}U(L,T_J \mp n2T_0)\right|^2}{\sqrt{2}\sigma}\right) + \\ &+ \frac{1}{2}\sum_{b_{-m/2},...,b_{m/2}=0...0}^{1...1}\frac{1}{2^m}\frac{1}{2}\text{erfc}\left(\frac{R\left|U(L,T_J) + \sum\limits_{n=1}^{m/2} b_{\pm n}U(L,T_J \mp n2T_0)\right|^2 - I_T}{\sqrt{2}\sigma}\right) \end{aligned} \tag{16}$$

IV. Numerical Results

The obtained results for bit error probability as a function of the normalized fiber length L/L_D for signal to noise ratio $SNR = \{20, 25\}$ dB and the system without timing jitter are shown in Fig. 1. Fiber length is always normalized in accordance to L_D when dispersion in fiber is not higher than order two. For initially unchirped Gaussian pulse and linear dispersion propagation medium, the error probability monotonically increases with the increase of the fiber length. Also, we can see that BER increases when SNR decreases.

Fig. 2. represents BER as a function of L for $SNR = \{20, 25\}$ dB and the system without timing jitter. The

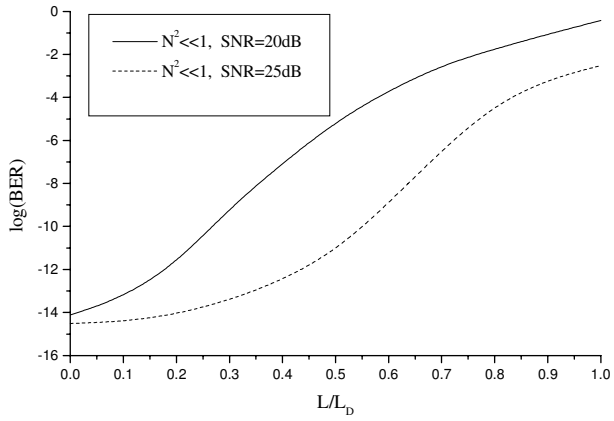


Fig. 1. BER as a function of the fiber length L for different SNR for system without timing jitter when the dispersion is dominant along fiber; $\lambda = 1.55$ μm, $\beta_2 = -20$ ps^2/km, $\gamma = 20$ (Wkm)$^{-1}$, $B = 20$ Gbit/s, $P_0 = 0.08$ mW

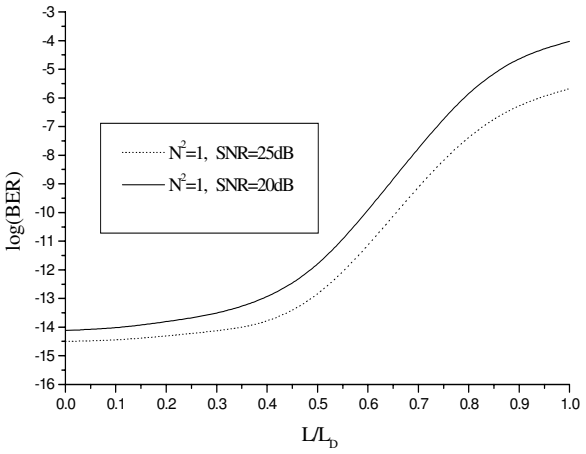


Fig. 2. BER as a function of the fiber length L for different SNR for system without timing jitter when is considered equal influence of dispersion and nonlinearity along fiber; $\lambda = 1550$ nm, $\beta_2 = -20$ ps^2/km, $\gamma = 20$ (Wkm)$^{-1}$, $B = 20$ Gbit/s, $P_0 = 1.6$ mW

unchirped Gaussian pulses are propagated along the nonlinear and dispersive medium. BER monotonically increases with the increase of the fiber length and the decrease of SNR, too. System performance is better in this propagation regime, looking BER as a function of fiber length in both cases.

Fig. 3 shows the bit error probability as a function of the signal to noise ratio SNR for fiber length $L = 0.5L_D$ and $T_J = \{T_0, T_0/10, T_0/5\}$. For initially unchirped Gaussian pulses and linear dispersion propagation medium, the bit error probability monotonically decreases with the increase of SNR and the decrease of parameter T_J.

In Fig. 4. we can see the bit error probability as a function of the signal to noise ratio SNR for fiber length $L = 0.5L_D$ and $T_J = \{T_0, T_0/10, T_0/5\}$, when signal propagated along the nonlinear and dispersive fiber. The same conclusions, regarding the bit error probability, being valid for Fig. 3 are

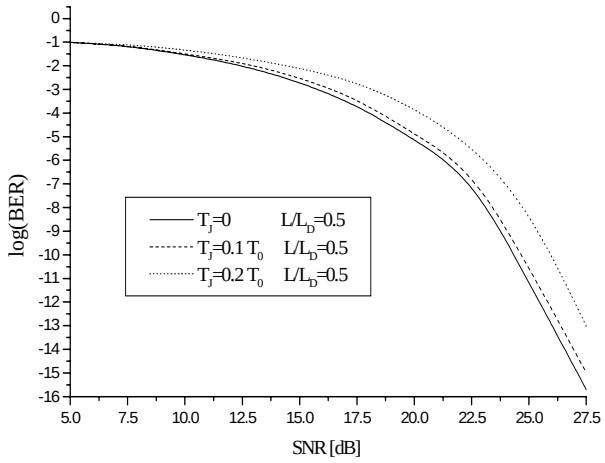


Fig. 3. BER as a function of SNR for different SNR for different values of T_J when the dispersion is dominant along fiber; $\lambda = 1.55$ μm, $\beta_2 = -20$ ps^2/km, $\gamma = 20$ (Wkm)$^{-1}$, $B = 20$ Gbit/s, $P_0 = 0.08$ mW

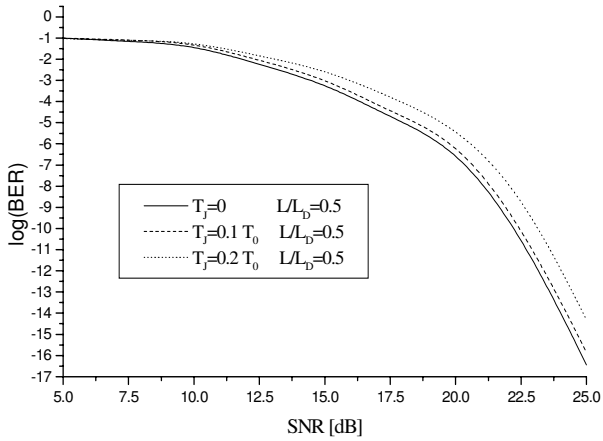


Fig. 4. BER as a function of SNR for different SNR for different values of T_J when a signal propagated along a nonlinear and dispersive fiber; $\lambda = 1.55$ μm, $\beta_2 = -20$ ps^2/km, $\gamma = 20$ (Wkm)$^{-1}$, $N^2 = 1$, $B = 20$ Gbit/s, $P_0 = 1.6$ mW

valid for Fig. 4, too. The system performance is better, when we have equal influence of dispersion and nonlinearity, looking BER as a function of SNR.

V. Conclusion

Dispersion-induced broadening of the unchirped pulse is undesirable since it interferes with the detection process leading to errors in decision. Dispersion limits the bit rate $B = 1/(2T_0)$ and the transmission distance L of a fiber-optic communication system. Timing jitter and Gaussian noise have negative influence of system detection performance in both propagation regimes, but that influence is significantly for linear dispersion propagation regime, considering equal product BL in the both cases.

References

[1] G. P. Agrawal, *Nonlinear fiber optics*, The Institute of Optics, Academic Press, University of Rochester, Rochester, New York, 1997.

[2] R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, H. C. Morris, *Solitons and Nonlinear Wave Equations*, Academic Press, London, 1982.

[3] Marincic, *Optical telecommunications* (in Serbian), University of Belgrade, Belgrade, 1997.

[4] K. Iwatsuki, S. Kawai, S. Nishi, M. Saruwatari, "Timing Jitter Due to Carrier Linewidth of Laser-Diode Pulse Sources in Ultra-High Speed Soliton Transmission", *Journal of Lightwave Technology*, vol. 13, no. 4, pp. 639-649, 1995.

[5] M. Stefanovic, A. Vidovic, "*IM/DD Optical System Performance in the Presence of Dispersion, Timing Jitter and Gaussian Noise*", vol 2., pp. 456-458, TELSIKS 2001, Nis

Suggestions for Availability Improvement of Optical Cables

Ivan Rados[1], Pero Turalija[1], Zoran Bakula[1]

Abstract – **This article analyzes suggestions for availability improvement of optical cables. There are specially analyze two different suggestions: decrease numbers of failures respectively increase mean time to failure and decrease mean time of repair optical cable. The date, which used in this article, are results of attended of failure rate during lasting several years exploitation of optical cables in SDH network HPT d.o.o. Mostar. Based on the data about failure rates, unavailability and mean down times of optical cables are made the suggestions for availability improvement.**

Keywords – **failure rate, availability, mean down time**

I. Introduction

The introduction of new services and the need of high quantity of data transmission require the high capacity transmission systems (SDH, WDM). One of the important elements of the transmission system is transmission media – in this article it is optical cable. The failures – interrupts the communication between great number of users – are making great losses for network operators [1]. Therefore, the availability performances of optical network greatly depend on availability of optical cables.

The HPT Mostar began whit installation and using of the optical cable as a transmission media in spring 1994. Ever since the optical cables have become the main transmission media at all network levels. The date about failure rate, which is analyzed, refereed to the two periods: from spring 1994. to the May 01, 1999 – we did not application suggestions for availability improvement of optical cable – and from May 01, 1999 to May 01, 2001 – we application one of the suggestions for availability improvement of optical cable. We will explain after in the article how we decided which suggestions we application on our network. Based on collected data we are analyzed the availability performances on which based the availability improvement is achieved.

II. On Availability in General

Availability A of some system in the time frame is defined as a ratio of time during which the system is functional in relation to the total operational time [2], i.e. it's probable that the system is functional in some time frame.

$$A = \frac{MTTF}{MTTF + MTTR} \tag{1}$$

where, $MTTF$ (Mean Time To Failure) is mean time till the failure occurs and $MTTR$ (Mean Time To Repair) mean time of repair.

$$MTTF = 1/\lambda \tag{2}$$

where λ is the failure rate defined as the number of failures per time unit.

$$1\ FIT\ (\text{Failure in Time}) = 1 \text{ failure per } 10^9 \text{ hours}$$

$$\lambda = \frac{m}{MT} \tag{3}$$

where n is the number of failures over monitoring time, M the length of installed cables in km and T monitoring period in hours.

For the entire optical network, comprising L kilometers of cables, mean time to failure ($MTTF$) is obtained as follows:

$$MTTF_{\text{network}} = \frac{1}{\frac{\lambda_{\text{cable}}}{\text{km}} L[\text{km}]}. \tag{4}$$

Unavailability U is probability complementary to availability [3], i.e.

$$U = 1 - A = \frac{MTTR}{MTTF + MTTR} = \lambda\, MTTR. \tag{5}$$

In reporting about system/network performances, unavailability U is often expressed as MDT (Mean Down Time) in minutes per year [3], i.e.

$$MDT = 3656024U \text{ [min/year]}. \tag{6}$$

III. Failure Analysis and Availability Calculation

In order to calculate the availability of optical cables, data on failures and time to repair of optical cables are used. The collected data referred to the period from spring 1994 to May 01, 1999 and period from May 01, 1999 to May 01, 2001. All cables are installed sub-surface, in polyethylene pipes and above them a warning tape was installed as a supplementary way of protection. Fibers of all cables are standard, single-mode with a diameter of 9/125 μm for the use of 1310 nm and 1550 nm wavelengths.

According to the collected data the main cause of most failures is outside interference (86.48%), where digging participates in 72.97% of cases [3]. The vehicle owing to the improper depth of installed cable causes two failures (5.40%) and the fire causes three failures (8.10%). Four failures (10.80%) are the consequence of the planned works by the HPT d.o.o. Mostar. These failures lasted relatively a short time because of previously well done preparations.

From the point of view of the optical failures availability we distinguish [4]:

[1]Ivan Rados, Pero Turalija, Zoran Bakula, HT d.o.o. Mostar, Kneza Branimira bb, 88000 Mostar, Bosnia and Herzegovina, e-mail: ivan.rados@tel.net.ba

- failures which break individual fibers in the cable, if there is no automatic protection, the operator has manually to direct the traffic to the correct fibers or to repair the faulty once,
- failures which simultaneously break all fibers in the cable, unless the network is ring, the operator has to repair all fibers in the cable with no regard to the existence of some protection mechanisms (optical modules). According to the collected data in 100% of causes happens the break of all fibers, either caused by digging or by fire or by vehicle.

Generally, two measures to repair are being used [2]:

- temporary repair time,
- permanent repair time.

Temporary repair time is the time needed for service restoration after the failure. This time to repair includes:

- time needed to report the failure to the maintenance team and their arrival to the telecommunication center,
- time needed for the preparation of splicing material (cable) and vehicles, - way to the failure location,
- laying of the new piece of cable (if needed) and its splicing,
- final measurements.

Table 1. Monitoring period, length and number of failures of optical cables

Monitoring period	Length (km)	Number of failures
spring 1994 – May 01, 1999	795	23
May 01, 1999 – May 01, 2001	456	14
spring 1994 – May 01, 2001	1251	37

Table 2. Failure causes of optical cables until May 01, 2001

Failure causes	No. of repor. failures	Relativ. % failures
Digging	27	72.97
Installat./workman errors	1	2.70
Defective connector	0	0.00
Fire	3	8.10
Vehicle	2	5.40
Fibers arrangement	0	0.00
Cable displacement	4	10.80
TOTAL:	37	100.00

According to the experience, the time needed to report to the maintenance team and their arrival to the telecommunication center is less than an hour. As there is no data on exact distance from the maintenance centers to the failure location, the time needed to arrive to the failure location is no special analyzes. But when the distance from the maintenance team center to the failure location is short, the influence of the arrival time to the failure location in relation to the total time to repair is insignificant.

The greatest influence on the time to repair has the type of failure, for example: difficult access to damaged cable, necessity for digging and installing the new piece of cable, cable capacity and splicing of fibers of different manufacturers, unfavorable weather conditions.

If only two fibers on one cable are actively used, regard to the availability there are exist two cases:

- repair of active fibers wherewith the system becomes available,
- repair of all fibers in the cable.

On the area covered by telecommunication network of HPT d.o.o. Mostar actively exist more transmission systems via the single mode cable, so the time needed to repair all fibers in the cable (or in more cables) is taken for the calculation of the time to repair.

Permanent repair time includes, in addition, final storage of new splicing closures, final construction works and final protection of a new cable segment.

In this article the temporary repair time is used as mean time to repair for the availability calculation owing to its influence on availability.

Until May 01, 1999 HPT d.o.o. Mostar had only one team with three members for maintains of optical cables. Two members out of three do the splicing and the one do finally measurements. The maintenance team had only one splicer and one OTDR, what practically mean that they be able do splicing only on one side of optical cable (if is necessity for installing the new piece of cable). It was 60% of the exact documentation of installing optical cables. Average time for repair during this period was 15.70 hours.

From May 01, 1999 HPT d.o.o. Mostar took precautions with aim to decrease average time to repair of optical cable (detail explanation in chapter 4). The results of this precaution were decrease average time to repair on 13.43 hours.

That we on the best way see influence to decrease average time to repair on availability of optical cables in this chapter we are presume that no decrease its that mean we are use for calculation average time to repair 15.70 hours.

Unavailability of optical cables per km is obtained as a product of failure rate per km of cable and the mean time to repair as shown in table 3.

Table 3. Failure rate (λ), unavailability (U) and mean down time (MDT) calculated for optical cables

λ (FIT/km)	$U \times 10^{-5}$	MDT (min/year)
460.61	0.72	3.80

For the unavailability calculation of the optical cable besides the mean time to repair and failures rate per km it is necessary to know the failure rates of the splices on the fiber and failure rate of connectors on the optical distribution frame. Data on failure rates of splices (30 FIT) and connectors (100 FIT) are taken from the [5] and [6]. The total length of the cable stage consists of delivered cable from factory with an avarage length of 4 km. According to this length the number

Table 4. Failure rate (λ-total), unavailability (U) and mean down time (MDT) calculated for different optical link lengths

Length (km)	No. of splices	No. of connect.	λ-total (FIT)	$MTTR$ (h)	$U \times 10^{-5}$	MDT (min/year)
20	5	2	9562.20	15.70	15.01	78.89
40	10	4	19124.40	15.70	30.02	157.78

of splices is calculated as:

$$\text{Length of cable}/4.$$

The number of connectors on the optical distribution frame is 2 for the average stage length which is used in this calculation.

The result analysis in table 4 shows that unavailability increase almost linearly to the cable length and depends on failure rate and mean time to repair of optical cables. For SDH network HPT d.o.o. Mostar, mean time to failure ($MTTF$) is obtained as follows:

$$MTTF_{\text{network}} = \frac{1}{460.61\,1251\,\text{failures}}\,\frac{h}{} = 1735h \approx 73\ \text{days}$$

IV. Suggestions for Availability Improvement

From the availability expression (1) can be seen that the availability depends on the mean time between failures and the mean time to repair of optical cables. Availability improvement can be obtained by increasing the mean time to failures and decreasing the mean time to repair of optical cables [4].

A. Increasing of mean time to failure

The increase of mean time to failures, relatively, the decrease of the number of failures can be achieved by preventive protection of optical cables against digging and by using the surveillance system for preventive maintenance. As most failures on the optical cables are caused by digging it is necessary to attract special attention to it. Although most countries have laws for preventive protection of underground cables there are still unsatisfactorily defined punishments (fees) for their infringements (digging without previous consent). The law must have to most rigid punishments (invitation prior to digging). While digging belong to the category of "instantaneous" breaks, the others belong to the category "preventive" because they are caused by complete loss of cable characteristics owing to the outside interference.

In our country still no have law about preventive protection of underground cables, what is a possible conclude on the base of number of failure during both monitoring periods, most of the failures were caused digging without previous consent and transgressor has no adequate punishment. That we are show influence mean time to failure on availability we will suppose that we already have the legal regulations regarding the protection of underground cables during the failure monitoring period and that, through change of that law, the mean time to failures increased from 73 to 84 days, which means that the number of failures decreased from 37 to 32, representing a decrease of about 13%, and that the mean time to repair remained same, i.e. 15.70 hours. As the failure rate is just proportional to the number of failures, so, decreasing the number of failures also decreases the failure rate by 13.5% or to 398.37 FIT.

As seen in the Table 5, the decrease of the number of failures resulted in the availability improvement, relatively, the decrease of MDT, for example, for $d = 20$ km – from 78.89 to 68.64 min/year or by 12.99%. Surveillance system for preventive maintenance can foresee the possible failure location using the metal protective layer on the cable as a sensor. Surveillance system alarms when the entirety of the outer sheath or the splicing point is being breaked, indicating that potential failure should be removed. As optical cables installed within the HPT d.o.o. Mostar have been exploited a short time (the first one about seven years), there were no deterioration as yet of the cable characteristics caused by the outside interference. For the preventive failure protection against outside interference (long-term exploitation) it will be necessary to install the surveillance system in order to foresee a failure. Costs for installation of such a system would be slight compared to with failure losses on the cable.

Table 5. Failure rate (λ-total), unavailability (U) and mean down time (MDT) calculated for different optical link lengths ($n = 32$ failures)

Length (km)	No. of splices	No. of connect.	λ-total (FIT)	$MTTR$ (h)	$U \times 10^{-5}$	MDT (min/year)
20	5	2	8317.40	15.70	13.06	68.64
40	10	4	16634.80	15.70	26.12	137.28

B. Decreasing of mean time to repair

From the unavailability expression can be seen that availability depends on the mean time to repair. In order to decrease the $MTTR$ it is essential to have a maintenance plan which should contain the following components [4]:

- exact documentation,
- maintenance team,
- training,
- equipment,
- plan of action,
- practice,
- continued process of improvement.

Exact documentation on optical cables is one of the most significant components for diminishing the $MTTR$. It includes: cable traces, number of failures in cable, fibers attenuation, splicing points, cable lengths, trace marking and outer-metal shield condition. Additional 34% of documentation made during period from May 01, 1999 to May 01, 2001, that mean than till now we made documentation for 94% of installing optical cable. Also it procured one mobile computer for frequent modification and bring up to date. Only two persons have access to that base and they are responsible

for its processing. Besides the exact data base for diminishing the $MTTR$ it is necessary to exactly know where are the tools and material needed to repair, as well as the key to the entrance of the building. Plan of action contains instructions on who is calling whom and when, as well as the numbers of fixed and mobile telephones. Now, maintenance team has seven members: five on the failure location and two at terminals (one in each). Four members out of five do the splicing (two teams of two members) and the fifth have the radio connections with the members at terminals.

Maintenance staff has to know to use the splicers and measuring equipment. Owing to the ever-improving measuring and connecting equipment for different types of cables, regular training of the maintenance team is very important because each improvement which leads to diminishing the time to repair increases the availability of optical cables. In HPT d.o.o. Mostar they have training two times a year at least in order to acquire new knowledge. Training in the field is more purposeful measure than the classroom teaching. Well planned and sudden exercise is the best way of the emergency staff training (one per year). The aim of each exercise is to achieve better results each time.

Quantity and kind of equipment depend on geographical spreading, network size, and the number of skilled staff. If network is too large and geographically spread there must be more maintenance teams. As network of HPT d.o.o. Mostar no geographically spread, for the now is sufficient one maintenance team of optical cable. Our maintenance team has one reflectometer (OTDR) for measuring at 1310 and 1550 nm, two splicers with tools (cutter, air, screwdrivers...), optical power meter, voltmeter and the car.

Using above mentioned suggestions, the $MTTR$ of the cable is obtained to 13.43 hours. The availability would also be improved considerably, as shown in table 6.

Table 6. Failure rate (λ-total), unavailability (U) and mean down time (MDT) calculated for different optical link lengths ($MTTR = 13.43$ h)

Length (km)	No. of splices	No. of connect.	λ-total (FIT)	$MTTR$ (h)	$U \times 10^{-5}$	MDT (min/year)
20	5	2	9562.20	13.43	12.84	67.48
40	10	4	19124.40	13.43	25.68	134.97

In the concrete, the $MTTR$ decrease of 14.45% results in the availability improvement of 14.46% or to, the decrease MDT from 78.89 to 67.48 min/year.

Every greatest availability improvement would be achieved by the simultaneous decrease of the number of failures (32) and the decrease of the mean time to repair of the cables (13.43 hours), as shown in following table 7. In the concrete, mean down time of failure is decrease for 25.57%, or to decrease from 78.89 to 58.71 min/year.

Table 7. Failure rate (λ-total), unavailability (U) and mean down time (MDT) calculated for different optical link lengths ($n = 32$ failures, $MTTR = 13.43$ h)

Length (km)	No. of splices	No. of connect.	λ-total (FIT)	$MTTR$ (h)	$U \times 10^{-5}$	MDT (min/year)
20	5	2	8317.40	13.43	11.17	58.71
40	10	4	16634.80	13.43	22.34	117.42

V. Conclusion

Data on failures and time to repair, which are analyzed in this article, refer to the 7 years exploitation of optical cables within the HPT d.o.o. Mostar transmission network. The analysis show that the most frequent cause of the optical cables break is digging (72.97%) and regardless to the break cause there has been breaks of all fibers in the cable. The analysis of temporary repair time shows that it mostly depends on the type of failure and cable capacity. Availability improvement of optical cables can be achieved by increasing of mean time to failure, relatively, decreasing the number of failures as the most frequent case of break, and decreasing the mean time to repair. The law on underground cables protection and monitoring system for preventive maintenance would be the cause of decreasing the number of failures. The mean time to repair cable is decreased considerably by using the plan of maintenance. An approximate unavailability can be used to evaluate availability of different structures.

References

[1] C. Coltro, "Evolution of Transport Network Architectures", *Alcatel Telecommunication Review*, 1st Quarter 1997, pp.10-18.

[2] I. Jurdana and B. Mikac, "An Availability Analysis of Optical Cables", *Proceedings WAON'98*, pp. 153-160.

[3] T. H. Victor, "Update on Interim Results of Fiber Optic System Field Failure Analysis", *Bellcore*, New Jersey.

[4] I. Rados, "Availability analysis of Synchronous Digital Hierarchy Network", Master thesis, University of Zagreb, 2000 (in Croatian).

[5] D. Gardan, "Availability analysis of the fibre optic local loop", *Optical Access Networks*, EFOC & N, pp. 28-32, 1994.

[6] P. N. Woolnough and N. E. Andersen, FITL System Recommendations – Final Report, Deliverable FIRST 1.0.11, pp. 1-42, 1 May 1996.

Viable Model for Diversified Path Restoration in WDM Networks

Jovan Stosic[1] and Boris Spasenovski[2]

Abstract – **In this paper we deal with planning and optimization of survivable WDM networks. We investigate routing and planning of working and spare capacity for Wavelength Path and Virtual Wavelength Path networks. Models for recommendation of k diversified paths and diversified path restoration for WDM networks are presented. The results show the dependency of spare capacity cost from number of k recommended paths and used restoration strategy.**

Keywords – **WDM networks, optimization, network planning and modeling, diversified path restoration**

I. Introduction

Most appropriate way for providing multiplexing in current optical transport networks is by WDM (Wavelength Division Multiplexing) technologies. The topic of this paper is planning and optimization of survivable WDM networks. Therefore, routing and planning of working and spare capacity are investigated. In order to determine the influence of different network parameters to the total cost of the survivable network, we've developed a tool called IOE (Integrated Object Environment) for object network modeling. Using this environment one can design the network graphical model, introduce all necessary network parameters, control the optimization process, and analyze the obtained results. Mathematical models for routing and wavelength assignment (RWA) of working and spare capacity are developed.

s	d	a	b	deltaM	p	Fpm
1	6	2	3	1	2	12
1	6	1	2	1	2	12
1	6	3	6	1	2	12
1	6	5	6	1	3	8
1	6	4	5	1	3	8
1	6	1	4	1	3	8
1	8	1	7	1	2	10
1	8	7	8	1	2	10

Fig. 1. Clip from the table representing links in working paths

In Section II, we present models for recommendation of k diversified paths and diversified path restoration in Section III. The models are written in MPL (Mathematical Programming Language) and solved by CPLEX 7.1 MIP (Mixed Integer Programming) solver.

[1]Jovan Stosic is with Macedonian Telecommunications, MTcom, Orce Nikolov BB, 1000, Skopje, Macedonia, E-mail: jstosic@mt.net.mk

[2]Boris Spasenovski is with Faculty of Electrical Engineering, University "Ss. Cyril and Methodious", Karpos 2, 1000, Skopje, Macedonia, E-mail: boriss@etf.ukim.edu.mk

Two types of WDM networks are considered: WP (Wavelength Path) and VWP (Virtual Wavelength Path) with diversified path restoration (PRd) strategy in case of single link failure. This strategy is compared with link restoration (LR) and path restoration (PR). Results of this comparison, dependency of spare capacity cost from number of recommended k-MSP (Most Suitable Paths) and used restoration strategy are presented in Section IV.

II. Model for Recommendation of K Diversified Paths for Rerouting

This model is combination of k shortest paths and min-max flow network problems, a kind of approach that gives two basic benefits. Firstly, it is solved with classical ILP techniques, and secondly, it is aware of links' capacities and residual network capacity that could be used for rerouting. In this model it is not important to observe the interrupted link and search for k-MSP for restoration. Actually in the model the interrupted link is not mentioned at all. The crucial information is the disrupted path, regardless the fact we investigate the restoration in case of one link failure. This is due to the demand for diversification. The model should recommend k-MSP not having link in common with working path, regardless of which link of the working path is in failure. This is realized with diversity constraint (Eq. 7). Part from an important table for VWP model is given on Fig. 1, where routing of working paths is shown. It is obvious that traffic demand of 1-6 (sd) pair amounting 20 units is transported over two routes (1-2-3-6 and 1-4-5-6). Index p is identifying the working paths of sd pair. For example, for 1-6 traffic pair $p = 2$ represents path 1-2-3-6, and $p = 3$ represents working path 1-4-5-6. For every sd pair and for each path between this nodes the model search for k-MSP for restoration that don't have link in common with working path. In the model we use the following decision variables that have to be adapted by means of optimization technique:

$UL_{p,k}^{m,j}$: binary variable with value 1 if restoration route k which carries part from the traffic flow of m-th sd pair which has been passing through p-th working route in failure, is passing through link j, otherwise 0;

$PATHS_{p,k}^{m}$: binary variable with value 1 if route k is used for rerouting of p-th working route in failure, otherwise 0;

$FMAX$: integer variable representing the flow of most loaded link in the network;

$HD_k^{m,p}$: integer variable that represents the number of hops in k-th restoration route for p-th working route from m-th s,d traffic pair. It is used in the sorting algorithm.

following parameters given from the graphical model or derived from output of previous mathematical model (i.e. model for routing and assignment of working capacities):

d_m : traffic demand of m-th s,d pair;

Λ_j : number of wavelengths per optical fiber;

$IR_p^{m,j}$: 1 if p-th working route of m-th s,d pair passes through link j, otherwise 0[3];

MF_j : maximal number of optical fibers per direction (total number of fibers in the cable is twice as large);

F_j : flow of the working traffic through link j[3];

F_p^m : flow of m-th s,d pair that have been passing through p-th working route in failure[3];

$MaxP$: number of different working routes per s,d pair (we take the value of traffic pair that has largest number of working routes);

and following indices:

$j=\{a,b\}$: index that represents the link j $(j:1...L)$ with adjacent nodes: a,b $(a:1...N, b:1...N)$;

$m=\{s,d\}$: index denoting s,d pair of nodes with traffic demand equal to d_m $(m:1..M)$;

n : an index that denotes the node in observed network topology $(n:1...N)$;

k : index denoting the different restoration routes;

p : index denoting the different routes used for routing of working traffic.

Objective is to minimize the flow of maximally loaded link:

$$MIN \qquad FMAX \tag{1}$$

Decision variables are bounded by number of constraints which define the dependence between these variables and the given input parameters.

1) With this constraint we bound a variable that represents the flow of most loaded link.

$$FMAX \geq \sum_{m=1}^{M}\sum_{k=1}^{F_p^m}\sum_{p=1}^{MaxP} UL_{p,k}^{m,j}, \qquad \forall j. \tag{2}$$

2) The constraint given with (3) is for flow conservation. The amount of flow that enters a given node has to leave it. It should be noticed that upper bound for the number of k-MSP is given by the flow of p-th working path.

$$\sum_{\substack{b=1\\a\neq d}}^{N} UL_{p,k}^{s,d,a,b} - \sum_{\substack{b=1\\a\neq s}}^{N} UL_{p,k}^{s,d,a,b} = \begin{cases} PATH_p^{m,k} & \text{if} \quad s=a \\ -PATH_p^{m,k} & \text{if} \quad d=a \\ 0 & \text{otherwise} \end{cases}$$
$$\forall s,d,a; \qquad \forall k=1,2,...,F_p^m. \tag{3}$$

[3]It is obtained as output from previously solved model for RWA of working capacity.

3) This constraint is concerning the capacity of the links and compared to the models for RWA of working traffic in the RHS (Right Hand Side) of the statement the flow allocated to the working traffic is taken in account. The number of restoration routes that pass through a given link j should not be greater than the number of free wavelengths.

$$\sum_{m=1}^{M}\sum_{k=1}^{F_p^m}\sum_{p=1}^{MaxP} UL_{p,k}^{m,j} \leq MF_j\Lambda_j - F_j, \qquad \forall j. \tag{4}$$

4) For the p-th working path in failure the model would recommend as many k-MSP for rerouting as is the value of the flaw[4] for this path (F_p^m).

$$\sum_{k=1}^{F_p^m} PATH_p^{m,k} = F_p^m \qquad \forall m, \forall p. \tag{5}$$

5) The constraint for symmetry in rerouting is imposed for each p working path. In other words, for p-th working path of sd traffic pair and for p-th working path for ds traffic pair, the k restoration routes should be symmetrical i.e. pass the same links but in opposite direction.

$$\begin{aligned} &UL_{p,k}^{s,d,a,b} = UL_{p,k}^{d,s,b,a} \\ &\forall s,d,a,b,p, \quad \forall k=1,2,...,F_p^m. \end{aligned} \tag{6}$$

6) This constraint is essential for the PRd models. It states that any used link in k-th restoration route should not be the same with any link passed by p-th working path.

$$\begin{aligned} &UL_{p,k}^{m,j} + IR_p^{m,j} \leq 1 \\ &\forall m,\ \forall j;\ \forall k=1,2,...,F_p^m;\ \forall p=1,2,...,MaxP. \end{aligned} \tag{7}$$

7) The following two constraints prevents occurrence of cycles (paths that return to the previously traversed nodes) in the k recommended routes. The constraint (8) excludes the possibility for an originating node of one link to become a destination node of the restoration route, as well the possibility a termination node of one link to become a source node of the restoration route. Second constraint (9) precludes the route from using the links with same adjacent nodes but opposite directions.

$$\sum_{b=1}^{N}\sum_{k=1}^{F_p^m} UL_{p,k}^{s,d,a=d,b} + \sum_{a=1}^{N}\sum_{k=1}^{F_p^m} UL_{p,k}^{s,d,a,b=s} = 0;\ \forall s,d,p \tag{8}$$

$$\begin{aligned} &UL_{p,k}^{s,d,b,a} + UL_{p,k}^{s,d,a,b} \leq 1 \\ &\forall s,d,a,b;\ \forall k=1,2,...,D_x^m;\ \forall p=1,2,...,MaxP. \end{aligned} \tag{9}$$

8) The constraint (10) actually is not a constraint but only an auxiliary statement for definition of the HD (Hop Distance) variable, which represent the number of hops traversed by k-th restoration route. It is used in an algorithm for sorting of proposed k-MSP by ascending number of hops.

$$\begin{aligned} &HD_k^{m,p} = \sum_{j=1}^{L} UL_{p,k}^{m,j} \\ &\forall m,\ \forall k=1,2,...,F_p^m;\ \forall p=1,2,...,MaxP. \end{aligned} \tag{10}$$

[4]Unit for measurement of the flow is a wavelength.

9) Last but not least constraint is to impose integer values for some variables.

$$UL_{p,k}^{m,j},\ PATHS_{p}^{m,k} \in \{0,1\},$$
$$\forall j = 1,2,...,L;\ \forall m = 1,2,...,M;\ \forall k \qquad (11)$$
$$FMAX,\ HD_{k}^{m,p} \in Z^{+}$$

From $UL_{p,k}^{m,j}$ variable by means of database programming we generate table containing links in failure. The link in failure is represented by two dimensional index $x = \{e, f\}$. Namely, in the PR and PRd models the paths to be restored are identified by the link in failure (x) and by the order number (m) they have occupied in that link. In Fig. 2 parts of two tables are shown: a) recommendation of k-MSP for PRd model and b) recommendation of k-MSP for PR model. Considering that working path 3-4 is routed through 3-8-4 the recommended diversified k-MSP should not pass through links 3-8 and 8-4. This condition (Fig. 2a) is fulfilled in the case of model for recommendation of diversified routes and not fulfilled for model with non-diversified routes (Fig. 2b). In second case this is confirmed by the fact that link 8-4 is used in the $k = 2$ and $k = 4$ recommended restoration routes.

a)

s	d	a	b	k	dm	p	m	e	f	HD
3	4	2	1	1	6	1	1	3	8	3
3	4	1	4	1	6	1	1	3	8	3
3	4	3	2	1	6	1	1	3	8	3
3	4	2	7	2	6	1	1	3	8	3
3	4	7	4	2	6	1	1	3	8	3
3	4	3	2	2	6	1	1	3	8	3
3	4	3	6	3	3	1	1	3	8	3
3	4	5	4	3	3	1	1	3	8	3
3	4	6	5	3	3	1	1	3	8	3
3	4	8	5	4	3	1	1	3	8	4
3	4	6	8	4	3	1	1	3	8	4
3	4	3	6	4	3	1	1	3	8	4
3	4	5	4	4	3	1	1	3	8	4
3	4	3	2	5	1	1	1	3	8	4
3	4	8	7	5	1	1	1	3	8	4
3	4	2	8	5	1	1	1	3	8	4
3	4	7	4	5	1	1	1	3	8	4
3	4	3	6	6	1	1	1	3	8	4
3	4	8	7	6	1	1	1	3	8	4
3	4	6	8	6	1	1	1	3	8	4
3	4	7	4	6	1	1	1	3	8	4

b)

s	d	a	b	k	dm	m	e	f
3	4	6	5	1	9	1	3	8
3	4	3	6	1	9	1	3	8
3	4	5	4	1	9	1	3	8
3	4	3	6	2	2	1	3	8
3	4	6	8	2	2	1	3	8
3	4	8	4	2	2	1	3	8
3	4	3	2	3	4	1	3	8
3	4	2	7	3	4	1	3	8
3	4	7	4	3	4	1	3	8
3	4	8	4	4	1	1	3	8
3	4	2	8	4	1	1	3	8
3	4	3	2	4	1	1	3	8
3	4	2	1	5	4	1	3	8
3	4	1	4	5	4	1	3	8
3	4	3	2	5	4	1	3	8

Fig. 2. Recommendation of k-MSP for PRd and PR models: a) diversified routing; b) non-diversified routing

III. Model for Diversified Path Restoration

Chosen paths by the models for recommendation of k-MSP for rerouting are feasible but not optimal. Optimization task is done by the model for diversified path restoration (PRd) which uses recommended paths as input parameters. Actually we use well known "non-diversified" PR model as presented in [2], i.e. [1]. The objective of optimization is to minimize the total cost of network resources used for spare capacity assignment, with requirement of 100% survivability in case of single link failure. For VWP models it is presented by following objective function:

$$MIN \sum_{j=1}^{L} (\beta_j SF_j + \gamma_j SC_j) + \sum_{n=1}^{N} \sum_{i=1}^{I} C_i \delta_i^n \qquad (12)$$

while for WP network:

$$MIN \sum_{j=1}^{L} \left(\beta_j SF_j + \gamma_j \sum_{\lambda=1}^{\Lambda} SC_{\lambda}^{j}\right) + \sum_{n=1}^{N} \sum_{i=1}^{I} C_i \delta_i^n \qquad (13)$$

where: SF_j is variable that represents the number of spare optical fibers in link j; SC_j is variable that represents the number of spare optical channels in link j; SC_{λ}^{j} is variable that represents the number of spare optical channels on wavelength λ in link j; δ_i^n is variable denoting the type i of the node n; β_j, γ_j are parameters denoting the cost of the link (β is cost related to a fiber and γ to a channel); C_i cost of using the node of type i.

The constraints bounding the variables are based on the constraints presented in [2] and are thoroughly elaborated in [1]. One crucial constraint that is important for understanding the model for VWP is:

$$SC_j \geq \sum_{m} \sum_{p} \delta_{m,p}^{j,x} F_p^{x,m} \quad \forall j,\ \forall x, j \neq x \qquad (14)$$

and for WP network:

$$SC_{j,\lambda} \geq \sum_{m} \sum_{p} \delta_{m,p}^{j,x} F_{p,\lambda}^{x,m} \quad \forall j,\ \forall x,\ \forall \lambda, j \neq x \qquad (15)$$

where: $F_p^{x,m}$ is variable representing the flow of p-th restoration route for m-th working path which has been passing through x-th interrupted link; $F_{p,\lambda}^{x,m}$ is variable denoting the flow of p-th restoration route for m-th path on wavelength λ which has been passing through x-th interrupted link; $\delta_{m,p}^{j,x}$ is binary parameter with value 1 if the p-th restoration route for m-th path passing through x-th interrupted link is using link j, otherwise 0; m is index denoting the working path with traffic demand d_m^x which has been passing through x-th interrupted link. It doesn't contain information for s, d traffic pair but enumerates paths that have been passing through x-th interrupted link; x is index denoting the interrupted link.

In PR and PRd models it is possible to reuse the capacity that is released by working path in failure. In this case the LHS of the constraints (14) and (15) should be augmented by this capacity.

s	d	a	b	Fxmp	m	p	e	f
3	4	2	1	9	1	1	3	8
3	4	1	4	9	1	1	3	8
3	4	3	2	9	1	1	3	8
3	4	7	4	5	1	2	3	8
3	4	2	7	5	1	2	3	8
3	4	3	2	5	1	2	3	8
3	4	3	6	6	1	3	3	8
3	4	5	4	6	1	3	3	8
3	4	6	5	6	1	3	3	8

Fig. 3. Part of the path rerouting table

On Fig. 3 it is shown that traffic demand of path 3-4 in case of failure of link 3-8 would be rerouted over three restoration routes (3-2-1-4, 3-2-7-4, 3-6-5-4). These three routes are chosen from the six recommended routes (Fig. 2) in order to minimize the total cost of spare capacity.

IV. Results

Fig. 4 depicts the studied network. We have chosen an arbitrary physical topology and arbitrary traffic matrix. By means

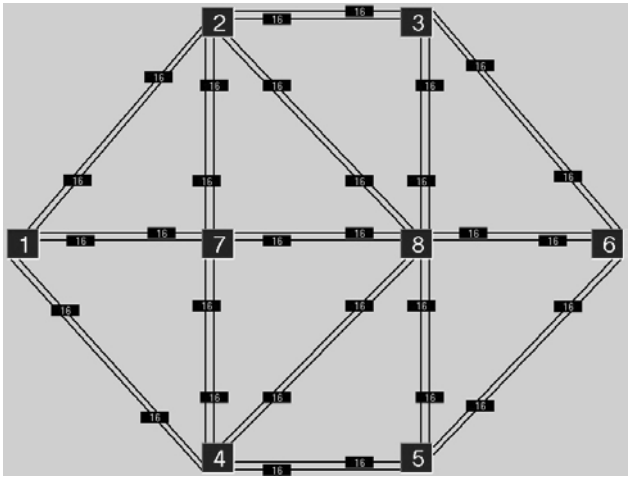


Fig. 4. Topology of the observed network

of IOE other physical topologies and traffic matrices could be easily constructed.

Dependency of the cost of spare capacity from the number of recommended k-MSP for rerouting is shown on Fig. 6. The costs are in arbitrary currency. On same figure, beside PRd and PR models results for LR are given. For WP PRd and PR models two cases are considered: case where optical transmitters and receivers are tunable, and the restoration route on another wavelength can be used (WPa), and case where transmitter wavelength is fixed and the restoration route must be found on the same wavelength (WPb). It is found that by increasing of the considered k-MSP for rerouting, the cost of spare capacities could be up to 20% decreased. Increasing of k above 3 gives very small reduction of costs (less then 10%); however more computational effort is required. By imposing the integer constraints (Eq. 11) CPLEX solver automatically uses MIP strategies for solving the problems. Used algorithm is branch and cut (B&C). Sometimes, for WPa models getting an optimal integer solution takes more time, therefore the B&C algorithm has to be stopped on n-th feasible integer solution. However for considered network B&C algorithm had found an optimal solu-

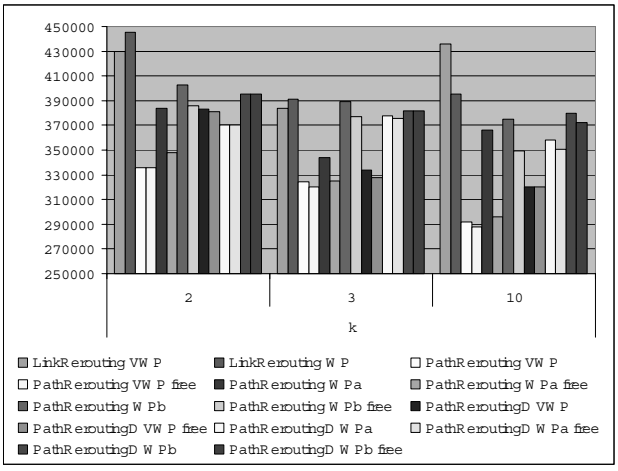


Fig. 5. Network cost dependency from k-MSP

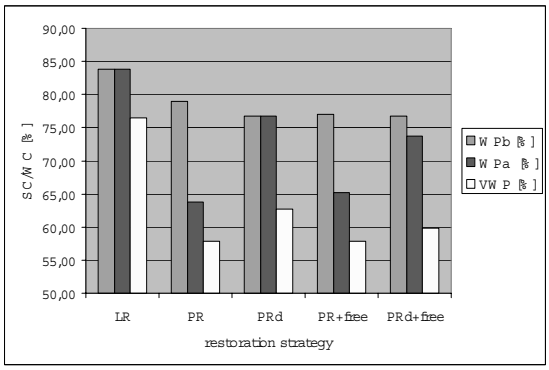


Fig. 6. Network cost dependency from restoration strategy

tions in reasonable time period (not larger then 30 min on Pentium III with 128 MB RAM). We obtained these results with default CPLEX settings except we applied a rounding heuristic on every 5th node.

We stress again that obtained results strongly depend from network meshing degree and traffic matrix. Greater benefit could be gained from using higher value for k-MSP in networks with larger meshing degree and larger traffic demands. From Fig. 5 it is obvious that the difference in cost of spare capacity between PR and PRd is negligible (also see Fig. 6). The difference is smaller for lower number of k-MSP. The fact that diversified path restoration starts the restoration in case of any link failure affecting the specific path, in authors' opinion is great advantage for implementation in the network management by using the current transport networks' control plane standards. Therefore we think investing in a bit more expensive spare capacity for diversified path restoration is worth of having easier implementation.

On Fig. 6 the dependency of the cost of the network from the used restoration strategy is shown. The results are expressed as ratio between spare and working optical channels. From aspect of largest requirement of spare capacity link restoration is most expensive. PR and PRd strategies are cheaper. In PR and PRd strategies by using of released capacity of the disrupted paths extra saving of capacity could be gained (PR+free, PRd+free). Moreover large difference between PRd VWP and WPa models is determined (much larger compared to PR) i.e. using of wavelength converters in PRd models could be justified. One more reason for using of PRd VWP models instead PRd WPa models is latter's long computation time.

V. Conclusions

In this paper routing and planning of working and spare capacity in WDM networks are investigated with specific approach for diversified path restoration. This strategy was compared with link restoration and path restoration. The model for diversified path restoration is presented as model that gives results similar to path restoration technique, and is most practical for implementation in present transport networks. Although it might give even greater benefit in networks with larger line system, the use of wavelength converters in diversified path restoration networks is justified.

References

[1] Jovan Stosic, “Modeling, planning and optimization of survivable WDM networks” – Masters thesis, Faculty of Electrical Engineering, University “Ss. Cyril and Methodious”, Skopje, Macedonia, November 2002.

[2] Bart Van Caenegem, Wim Van Parys, Filip De Turck, and Piet M. Demeester, “Dimensioning of Survivable WDM Networks”, *IEEE Journal on Selected Areas In Communications*, vol. 16, No.7, September 1998.

[3] Biswanath Mukherjee, “Optical Communication Network,” McGraw-Hill, 1997.

Functional Analysis of Basic Security Models

Nikoleta H. Hristova[1], Vencislav G. Trifonov[2], Ivailo I. Atanasov[3]

Abstract – **This paper investigates some of the basic security models. These models are examined in terms of their security properties and divided into three groups with respect to their definitions of security. It has been made a description of the areas where they could be used and some proposals of model's applicability in practise.**

Keywords – **security models, access control, security classes, information flow**

I. Introduction

Information contained in an automated system must be protected from three kinds of threats: (1) the *unauthorized disclosure* of information, (2) the *unauthorized modiftcation* of information, and (3) the *unauthortzed withholding* of information (usually called *denial of service*). To achieve protection, that fully covers these kind of threats it is very important clearly to understand and implement the system's security requirements. The purpose of a security model is to express those requirements precisely.

The term security model is used to describe any formal statements of a system's confidentiality, availability, or integrity requirements. A security model does not deal with all variables and functions of the system, but it is concerned only with security relevant ones.

In this article we try to make a brief explanation of some of the well-known models of security and to compare them with respect to motivation, approach, view of security and use in practice.

II. Basic Security Models

A *finite-state machine model* describes a system as an abstract mathematical state machine; in such a model, *state variables* represent the state of the machine, and *transition functions* or *rules of operation* describe how the variables change.

The *lattice model* [2] uses a lattice as a building base.A lattice is a finite set together with a partial ordering on its elements such that for every pair of elements there is a least upper bound and a greatest lower bound.

[1]Nikoleta H. Hristova is with the Faculty of Communications and Communications Technologies,Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: nhh@suntu.vmei.acad.bg

[2]Vencislav G. Trifinov is with the Faculty of Communications and Communications Technologies,Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: vgt@vmei.acad.bg

[2]Ivailo I. Atanasov is with the Faculty of Communications and Communications Technologies,Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: iia@vmei.acad.bg

The *Access Matrix model* [2] is a state machine model which represents the security state of the system as a large rectangular array containing one row per subject and one column for subject and objects. Each entry specifies the modes of access the object has to a subject or to other subject. A variant of the access model is the *information flow model*, which – rather than checking a subject's access to an object – attempts to control the transfer of information from one object into another object, constrained according to the two objects' security attributes.

The *Bell and LaPadula model* [2], may be summarized in two axioms: (1) No user may read information classified above his clearance level ("No read up");(2) No user may lower the classification of information ("No write down"). The full statement of the model includes several more axioms and is quite complex.

The *high-water mark model* [1] takes its name from the "History functions" which record the highest authority assigned to the object and the union of all categories assigned to the object since its creation. The model works with four types of objects, and each object is described by an ordered triple of properties, called Authority (A), Category (C), and Franchise (F). The model also defines an ordering on these triplets that corresponds to the lattice model.

UCLA Data Secure Unix (DSU) model [1] is a finite state machine model, with the state defined by the following four components: (a) process objects; (b) protection-data objects, with values being sets of capabilities; (c) general objects (comprising both pages and devices); and (d) a current-process-name object, whose value is the name of the currently running process.

Take-grant models [1] use graphs to model access control. Although couched in the terms of graph theory, these models are fundamentally access matrix models. The protection state of a system is described by a directed graph that represents the same information found in an access matrix. Nodes in the graph are of two types, one corresponding to subjects and the other to objects. It implies a new property called "Take", which means that grants may be taken from another subject to receive rights for a given object.

Filter models imply security policies as a filter of input functions on system inputs.

Strong dependency [1] is a model build on an approach which is based on the notion, fundamental to information theory, that information is the transmission of *variety* from a sender to a receiver.

A *constraint* specifies a sequence of states that cannot occur.It may be a part of other model, or may be used a base to a separate group of models.

III. Comparison of Models

It is very hard to find rates suitable for assessing the security of a particular security model, that are either precise and general enough to imply for different types of models, so that a quantitative analysis could be made. That is a result of that all of the models define security as absolute: an operation is either secure or not secure. Therefore we can make only a qualitative comparison of models.

Tables 1 shows a comparison of models, described above with respect of motivation, approach, view if security and use.

Table 1. Comparison of properties of models in terms of motivation

Properties	Models									
	A. M	UCLA	T-G	HWM	B + L 1	B + L 2	Flow	Filt	S. D	Cons
Motiwation										
Developed primarily to represent existing systems	×	×[b]		×[b]						
Developed to guide construction of future systems			×		×	×	×	×	×	×
View of security										
Models access to objects without regard to contents	×	×	×	×	×					
Models flow of information among objects						×	×			
Models inferences that can be made about project data								×	×	×
Approach										
Model focuses on system structures (files, processes)	×	×	×	×	×	×				×
Model focuses on language structures (variables, statements)							×	×	×	
Model focuses on operations on capabilities		×	×							
Model separates protection mechanism and security policy	×	×	×					×		
Systems based on or represented by this model have been implemented	×	×		×		×				

With respect to their definitions of security, the models can be divided roughly into three groups:

- those that are concerned only with controlling direct access to particular objects (access matrix model, UCLA DSU model, take-grant model);
- those that are concerned with information flows among objects assigned to security classes (information-flow model, revised Bell and LaPadula model);
- and those that are concerned with an observer's ability to deduce any information at all about particular variables (filters, strong dependency, constraints). (The high-water-mark model falls between the first and second groups, since it is concerned with the security levels of the objects a process touches over time, but it only controls direct accesses to objects.)

Models in the first category may be used in systems that require high degree if security, with addition of security policy for the concrete implementation. Those in the second group are probably the closest in structure to the requirements for telecommunication applications, but applications often require more flexibility than these models permit. The models in the third category are the least tested and would probably be the most difficult to use. Security properties of the UCLA DSU model were proved to hold for substantial portions of that system, but only the Bell and LaPadula model has been applied in more than one formally specified system. The properties specified by the high-water-mark, access matrix, and take grant models could probably be stated in a form suitable for automated verification techniques. The properties required by the constraint, strong dependency, and filter models could be expressed similarly, but actually developing a system specification in the terms required by those models requires unduly amount of work. Most of the secure system developments using the (revised) Bell and LaPadula model have been based on the concept of a security kernel, and there have been problems in extending its use beyond the operating system to application systems. The lattice model will probably fit many of the requirements for security and privacy in the private sector. An alternative to adding special models for trusted processes on top of the Bell and LaPadula model for the operating system is to develop integrated models tailored to particular applications. A security model designed for a particular application could be used as a basis for the development of an application-specific security kernel. A key problem in this approach is to ensure that the model incorporates the desired notion of security while permitting the operations required in the application. Further off, if capability-based systems are successfully developed, models more appropriate to their structures may be used. The take and grant model is a possible candidate in this area, though it would require tailoring for specific applications.

IV. Conclusions

Some of the described models are good candidates for secure system building, but even if such a system truly simulates any of the models it is impossible to say, that it is unconditionally secure, because each model defines its own concept of security, and a that system will be secure only in the sense defined by that model.

In conclusion it is clear that none of the basic models of security is suitable for the new generation complex systems, where many security requirements need to be taken into consideration. These models may be used only as a reference point when designing the security of the system.

References

[1] Shimeall T., Williams P., "Models of Information Security Trend Analysis " – CERT® Analysis Center, Software Engineering Institute Carnegie Mellon University

[2] Amoroso E., "Fundamentals of Computer security Technology". Prentice Hall, 1994

Model of Fail-Safe Self-Modifying Finite Automata

Vencislav G. Trifonov[1] and Ivailo I. Atanasov[2]

Abstract – **The paper presents a model of fail-safe finite state machine with self-modifying functions. Self-modifying functions are fail-safe and change automata states, ordered by a safety set of conditions.**

Keywords – **Fail-safe finite machines, Self-modifying automata, Discrete Event Systems**

I. Introduction

Finite automata (FA, FM) are one of methods for mathematical description of discrete event systems. Fail-safe finite machines (FSFM) are a subclass of main automata class. [1-3] This class includes two types of function:

— normal (conventional, work) functions;

— fail-safe functions.

Fail-safe functions used to achieve a high-level of safety and security it case of hazardous fault in conventional operations.

Definition 1: Finite machine (FM) is a 6-tulpe

$$FM = \langle X, Y, Q, Int(\bullet), Out(\bullet) \rangle \qquad (1)$$

where $X : \{x_1, ..., x_n\}$ is input alphabet of FM, $Y : \{y_1, ..., y_n\}$ is output alphabet of FM. Q"$\{q_1, ..., q_n\}$ is internal set of automata states. $Int(\bullet)$ is automata states function and $Out(\bullet)$ is automata output function.

Formal description of automata behavior is given:

$$\begin{cases} Q(t+1) = Int(Q(t), X(t)) \\ Y(t+1) = Out(Q(t), X(t)) \end{cases} \qquad (2)$$

Fig. 1 presents an abstract FM.

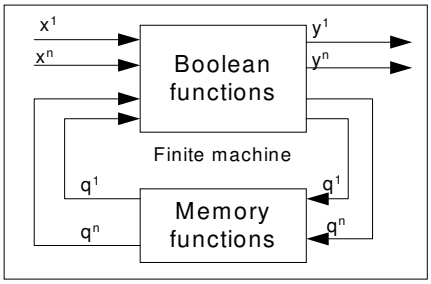


Fig. 1. Finite state machine

Definition 2: **Fail-safe behavior** is each automata transaction that changes current state to safe state after hazardous fault detection [1].

[1]Vencislav G. Trifonov is with Telecom Department at Technical University of Sofia, "Kl. Ohridsky" Blvd. 8, 1756 Sofia, Bulgaria, e-mail: vgt@vmei.acad.bg

[2]Ivailo I. Atanasov is with Telecom Department at Technical University of Sofia, "Kl. Ohridsky" Blvd. 8, 1756 Sofia, Bulgaria, e-mail: iia@vmei.acad.bg

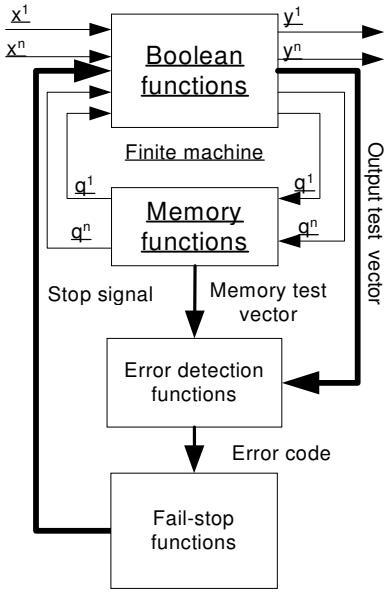


Fig. 2. Fail-safe finite machine

Definition 3: **Hazardous failure** is a failure that changes current state to a dangerous state [2].

Definition 4: **Dangerous input sequence** is each sequence of input alphabet that puts automata to a dangerous state [2].

For successfully recognize a kind of automata states should by defined criteria for fail-safe behavior and criteria for safety fault classification, described in detail in [1-3].

To achieve fail-safe behavior FM abstract structure on Fig. 1 extends to FSFM structure Fig. 2

II. States Ordering by Fail-Safe Degrees

The paper presents another point of view about the automata states. For most of fail-safe finite machines have possibility to define a linear ordered relation between their states [4,5].

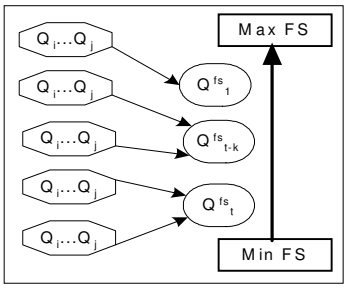
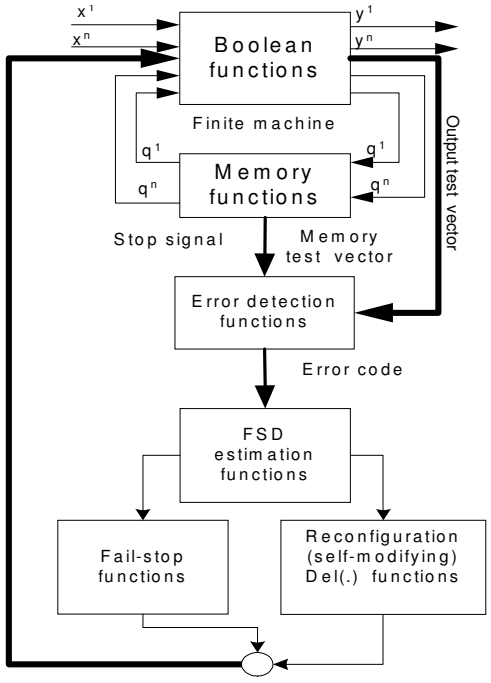


Fig. 3. Fail-safe state order

Each set of automata states should by presents as:

$$Class(Q) = \sum_{1}^{j} Subc; ass_j(Q_i) \tag{3}$$

where: $Class(Q)$ is a set of all FM states and $Sublass_j(Q_i)$ is a set of FM states with equivalent fail-safe degree, $j : \{1, ...n\}$ where $n = Gard\{Subclass_i\}$ and $i : \{1, ..., m\}$ is number of all internal states with same fail-safe degree [4,5]. The automata states should by present as follow (Fig. 3):



Let FM A is a reverse counter with input alphabet $X : \{a, b, c\}$; output alphabet $Y : \{0, 1\}$; and set of internal states: $Q\{1, 2, ..., N\}$. FM model is presents in Table 1.

Table 1. Reverse counter

	a	**b**	**c**
1	*2*	*1*	*1*
2	*3*	*2*	*1*
N	*N*	*N*	*N-1*

FM recognizes each valid input sequence with a set of final functional states

$$Final\_set(Q_i) = \{2, 4, 6, 8, ..., n/2\}.$$

A sample distribution to subclasses is:

$$\begin{array}{l} Subclass\ (1) : \{2, 4, 6\}; \\ Subclass\ (2) : \{8, 14, 18, 34, 98\}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Subclass\ (n) : \{q_i, ..., q_k\}. \end{array} \tag{4}$$

For each subclass defines fail-safe degree (FSD). A fail-safe relation orders FSD as follow [4]:

$$FSD_{\min} < FSD_1 < \cdots < FSD_n < FSD_{\max}. \tag{5}$$

For Subclass set defines a map function to FSD set:

$$\begin{array}{l} Subclass\ (1) \in FSD_1; \\ Subclass\ (2) \in FSD_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ Subclass\ (n) \in FSD_k. \end{array} \tag{6}$$

III. Fail-Safe Reconfiguration Functions

Structure of a Fail-Safe Automata with State Restriction is present on Fig. 4. To realize new self-modifying function **Del()** should by defines it as follows:

$$\begin{array}{l} D(g) : Class(Q_{old}) \rightarrow Class(Q_{new}); \\ Gard(Class(Q_{new})) > Gard((Class(Q_{old})). \end{array} \tag{7}$$

Definition 5. Del(.) function is fail-safe function if:

1. $Class(Q_{new})$ is FSD ordered set;
2. $D(t) : Class(Q_{old}(t-1)) \rightarrow Class(Q_{new}(t))$;
3. Each possible input sequence FSFM maps as follow:
   – if $X(i)$ is valid input word then
   $FSFM : \{x_1, ..., x_n\} \rightarrow Final\_set(Class_{new}(Q))$;
   – else:
   $FSFM : \{x_1, ..., x_n\} \rightarrow Final\_stop\_set(Class_{new}(Q))$
4. After **Del(.)** contraction, subtraction subclasses are total inaccessible.

IV. Conclusion

The paper presents a model for self-modifying fail-safe finite state machine and definition for fail-safe self-modifying function.

Self-modifying property based on automata states contraction, witch subtract dangerous states for each detected hazardous fault.

References

[1] H. Hristov, V. Trifonov , “New problems in fail-safe automata”, II Международная научно-техническа конференция “Актуалные проблеми развития железодорожного транспорта” Россия, Москва, 24-25 сентября 1996 г. стр. 125.

[2] Trifonov V, “Algoritmus fur Syntese des ungefarlichen Endautomat bei geanderte Verteilung des Eingangraums nach der Absage” TRANSCOM’97, University of Zilina, Slowak Republik, June 25-26, 1997 Volume 2, стр. 73.

[3] Hristo Hristov, Trifonov Vencislav, “A method for fail-safe degradation of final state machines”, The 4 th INTERNATIONAL SCIENTIFIC CONFERENCE OF RAILWAYS EXPERT, Yugoslawia, Vrnjacka Banja, Oktober 2-4 1997 г., стр. 328.

[4] V. Trifonov, “Method for synthesis a fail-safe automata with restricted degradation”, The 5 th INTERNATIONAL SCIENTIFIC CONFERENCE OF RAILWAY EXPERTS - JUЖEL, Yugoslvia, Vrnjcka Banja, October 28-30, 1998 г. стр. 91.

[5] Joze Eyzell, Jose Cury, Exploiting Symmetry in the Synthesis of Supervisor for Discrete Event Systems, American Control Conference 1998.

An Overview of Presence Service Architecture and Functionalities

Milan Jankovic , Borislav Odadzic[1]

Abstract – **This paper in the first half describes Presence service Architecture and Functional description, and Presence service specific functionalities of existing network elements and possible new network element. The Session Intitiation Protocol (SIP) extension for Presence and future work based on studying the available literature and documents is presented too.**

Keywords – **Presence service, Architecture, SIP extension for Presence**

I. Introduction

Presence and availability technologies provide the ability to determine the event in which a mobile user is present in a certain location and available for certain events. Presence gives the user the ability to show their availability to communicate, as well as their current location, to the ones they chose. The user will be able to move between different networks and devices, while maintaining their Presence status. This gives Service Providers the opportunity to offer new and existing services to their users.

The ability to determine the presence of mobile user may be accomplished via both ”push” and ”pull” technologies. Push technologies include the ability to set static triggers in network nodes such as Visitor Location Registers (VLR-s) to report the presence of mobile user at a serving switch location based on GSM MAP (Mobile Application Part) messaging. Pull technologies include the ability for the network to proactively poll for presence information about selected mobile users.

The ability to determine availability is based on the possibility of the system to state (on/off) of the mobile device; it’s capabilities and is the mobile user willing to engage in certain activities. All of these network abilities depend on the existence of programs residing on applications that provide this information once the presence status is determine. There are too many applications to adequately introduce them. The application include location based services, mobile commerce, mobile advertising, and many more.

II. Presence Service Reference Architecture

This section describes the reference architecture, the reference points and interfaces used for Presence Service. The generic reference architecture for providing presence service is depicted in Figure 1 [1].

[1]M. Jankovic and B. Odadzic are with Community of Yugoslav PTT, Palmoticeva 2, 11000 Belgrade, Serbia and Montenegro, e-mail: miljankovic@zj.ptt.yu

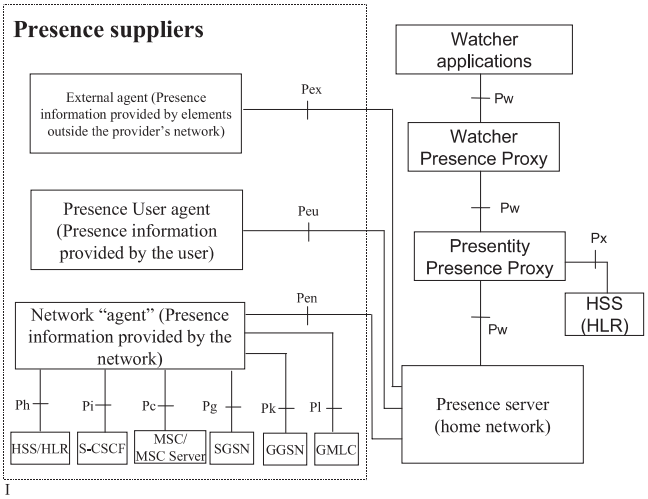


Fig. 1. Reference architecture to support a presence service

A. *Functional description of network elements*

This section describes the Presence Service specific functionalities of existing network elements and possible new network elements.

Presence Server – The Presence Server resides in the presentity’s home network. The Presence Server shall manage presence information that is uploaded by the Presence User/Network/External agents, and is responsible for combining the presence-related information for a certain presentity from the information it receives from multiple sources into a single presence document.

The mechanisms of combining the presence related information will be defined based on presence attributes, and according to certain policy defined in the Presence Server. The Presence Server is not required to interpret all information, the information that the Presence Server is not able to interpret shall be handled in a transparent manner. The Presence Server shall also allow users to fetch and subscribe for receiving presence information.

The Presence Server shall support internetwork operability mechanisms to allow for an interoperable Presence Service across multiple operators’ networks and domains (e.g. external Internet). Mechanisms for locating the Presence Server shall be developed, especially with respect to these internetwork operability aspects.

The Presence Server shall support SIP (Session Initiation Protocol)-based communications with the Presentity Presence Proxy. In the IP Multimedia Sub-system (IMS) the Presence Server is seen as a SIP Application Server, and is lo-

cated using SIP URLs, standard SIP and existing IMS mechanisms – (SIP routing, HSS query, ISC filtering, etc.).

The Presence Server shall support authorization and security mechanisms, at least the following levels of authorization are foreseen: providing presence information to any Watcher application that requests it and provide presence information to only those Watcher applications in an "allowed" list.

The Presence Server may also support authorization and security mechanisms that is based on asking permission from the Presence User agent on a case-by-case basis.

The Presence Server may support rate-limiting or filtering of the presence notifications based on local policy in order to minimize network load.

The Presence Server could be extended to a generic State Agent, supporting subscriptions and notifications regarding other types of events than presence as well. An example for such event is the combined presence of a whole buddy list.

Watcher and Presentity Presence Proxy – When a Watcher application intends to access some presence information of a presentity, it first needs to find the Presence Server containing this information. The Watcher Presence Proxy provides the following functionality: address resolution and identification of target networks associated with a presentity, authentication of watchers, interworking between presence protocols for watcher requests and generation of accounting information for watcher requests.

The Presentity Presence Proxy provides the following functionality: determination of the identity of the presence server associated with a particular presentity and generation of accounting information for updates to presence information.

The Presentity and or the Watcher Presence Proxies may also be responsible for providing network configuration hiding.

Communications between the Presentity Presence Proxy and the Watcher Presence Proxy shall be based on SIP as shown in Figure 2. Other IP-based mechanisms may also be needed to support the delivery of large amount of presence information. Support for non-SIP based Watchers may be provided by the use of an interworking functions located at the Watcher Presence Proxy.

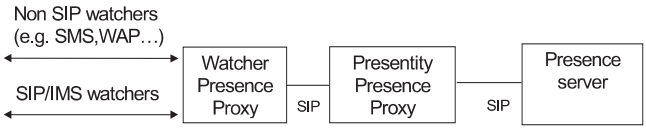


Fig. 2. Communications between the Presentity Presence Proxy and the Watcher Presence Proxy for Watchers

External Agent – The Agent elements in the Presence Architecture are functionally distinct from the Presence Server functional element. The generic function of the Agent elements is to make presence information available to the Presence Server element in standardized formats across standardized interfaces.

The External Agent element provides the following functionality: the External Agent supplies Presence information from external networks, the External Agent sends the Presence information across the Pex interface according to the format standardized for the Pex interface and the External Agent handles the interworking and security issues involved in interfacing to external networks.

Examples of Presence Information that the External Agent may supply, include: third party services (e.g. calendar applications, corporate systems), Internet Presence Services and other Presence Services.

Presence User Agent – The Agent elements in the Presence Architecture are functionally distinct from the Presence Server functional element. The generic function of the Agent elements is to make presence information available to the Presence Server element in standardized formats across standardized interfaces.

The Presence User Agent (PUA) element provides the following functionality: the Presence User Agent collects Presence information associated with a Presentity representing a Principal, the Presence User Agent assembles the Presence information in the format defined for the Peu interface, the Presence User Agent sends the Presence information to the Presence Server element over the Peu interface, the Presence User Agent shall be capable of managing the Access Rules and the Presence User Agent shall handle any necessary interworking required to support terminals that do not support the Peu reference point.

From a conceptual view, the PUA element resides between the presence server and the user's equipment as illustrated in the reference architecture in Figure 1. In reality, a Presence User Agent may be located in the user's terminal or within a network entity.

Where the PUA is located in a terminal, the terminal shall support the Peu interface to the presence server as illustrated in Figure 3.

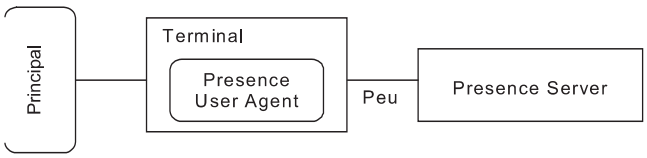


Fig. 3. Terminal based Presence User Agent

Where the PUA is located within the network, the particular network entity shall support the Peu interface to the presence server as illustrated in Figure 4. In such a case an additional functionality may be required to resolve the location of the presence server associated with the presentity. In this case, the interface between the terminal and the Presence User agent is outside of the scope of standardisation of the presence service.

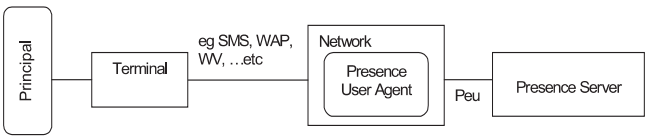


Fig. 4. Network based Presence User Agent

Network Agent – The Agent elements in the Presence Architecture are functionally distinct from the Presence Server functional element. The generic function of the Agent elements is to make presence information available to the Presence Server element in standardized formats across standardized interfaces.

The Network Agent element provides the following functionality:

- The Network Agent receives Presence information from network elements within the Operator's network.
- The Network Agent associates Presence information with the appropriate Subscriber/Presentity combination.
- The Network Agent converts the Presence information into the format standardized for the Pen interface.
- The Network Agent sends the Presence information across the Pen interface.
- The Network Agent may Push Presence information to the Presence Server alternatively some network elements may be queried, signaled, or provisioned to deliver Presence information. For those elements that require querying or signaling, the Presence Server makes a request to the Network Agent directing it to acquire the Presence information. The Network Agent then issues the appropriate commands to the element.

III. Presence and Availability Management (PAM)

This section of the document specifies the Presence and Availability Management Service Capability Feature (SCF) aspects of the interface. All aspects of the Presence and Availability Management SCF are defined here, these being: Sequence Diagrams, Class Diagrams, Interface specification plus detailed method descriptions, State Transition diagrams, Data Definitions and IDL Description of the interfaces.

The process by which this task is accomplished is through the use of object modelling techniques described by the Unified Modelling Language (UML).

Consider the following simple but desirable scenario for a communication service: An end-user wishes to receive instant messages from her management at any time on her mobile phone, from co-workers only on her desktop computer, and in certain cases for the messages to be forwarded to e-mail or even a fax machine/printer. The senders may know her availability for various forms of communication in the way she chooses to reveal it or alternatively the senders may never know how she will be receiving their messages. This scenario spans over multiple services and protocols and can only be solved currently by a proprietary solution that maintains the required information in an ad-hoc fashion within the application.

PAM is not a replacement for the protocols being standardized for various communication and network services. PAM attempts to standardize the management and sharing of presence and availability information across multiple services and networks.

The PAM specification is motivated by the observations that

- The notions of Identity, Presence and Availability are common to but independent of the various communication technologies, protocols and applications that provide services using these technologies.
- Presence does not necessarily imply availability. End-users or organizations require greater control over making
- Themselves available through various communication devices.
- Presence based services need to address privacy concerns on who can access presence information and under what conditions.

Management of availability will span over multiple communication services and service providers

IV. Session Initiation Protocol (SIP)

The Session Initiation Protocol (SIP) is an application-layer control (signaling) protocol for creating, modifying and terminating sessions with one or more participants. These sessions include Internet multimedia conferences, Internet telephone calls and multimedia distribution. Members in a session can communicate via multicast or via a mesh of unicast relations, or a combination of these. SIP invitations used to create sessions carry session descriptors which allow participants to agree on a set of compatible media types. SIP supports user mobility by proxying and redirecting requests to the user's current location. Users can register their current location. SIP is not tied to any particular conference control protocol. SIP is designed to be independent of the lower-layer transport protocol and can be extended with additional capabilities.

SIP is a request-response protocol in that clients send SIP requests to servers which send back responses to these requests. A typical application is generally made up of both client and server functionality. A User Agent is an intelligent end point. A User Agent initiates a session by creating and sending an INVITE request. This request can either be sent directly to another User Agent or one or more proxies can be traversed. Proxies forward requests based on local policyand information contained in the SIP request. A typical SIP call / session set up and tear down between two User Agents, traversing a SIP proxy can be seen in Figure 5.

In the Figure 5, a User Agent acting as a User Agent Client, initiates an INVITE request and sends it to a second User Agent, acting as a User Agent Server, via a proxy. The server returns an OK response. When the client receives the OK response, an ACK request is sent to acknowledge the receipt of the final response and the communication between the two user agents is set up; a session is now in progress. The ACK may not traverse the proxy as the network path has been established. The BYE request indicates that one side wants to terminate the session.

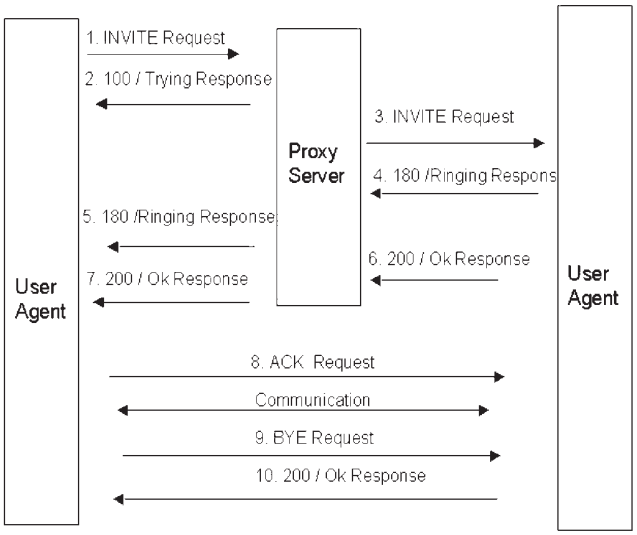


Fig. 5. A typical call set-up and tear down

V. SIP extension for Presence

Presence is defined, as subscription to and notification of changes in the communications state of a user. This communications state consists of the set of communications means, communications address, and status of that user. A presence protocol is a protocol for providing such a service over the Internet or any IP network.

This section proposes the usage of the Session Initiation Protocol for presence [9]. This is accomplished through a concrete instantiation of the general event notification framework defined for SIP, and as such, makes use of the SUBSCRIBE and NOTIFY methods defined there. User presence is particularly well suited for SIP. SIP registrars and location services already hold aspects of user presence information; it is uploaded to these devices through REGISTER messages, and used to route calls to those users. Furthermore,SIP networks already route INVITE messages from any user on the network to the proxy that holds the registration state for a user. As this state is user presence, those SIP networks can also allow SUBSCRIBE requests to be routed to the same proxy. This means that SIP networks can be reused to establish global connectivity for presence subscriptions and notifications.This event package is based on the concept of a presence agent, which is a new logical entity that is capable of accepting subscriptions, are changes in user presence. The entity is defined as a logical one, since it is generally co-resident with another entity. This event package is also compliant with the Common Presence and Instant Messaging (CPIM) framework that has been defined in [7]. This allows SIP for presence to easily interwork with other presence systems compliant to CPIM.

When an entity, the subscriber, wishes to learn about presence information from some user, it creates a SUBSCRIBE request. This request identifies the desired presentity in the request URI, using a SIP URI, SIPS URI or a presence URL [9,7]. The subscription is carried along SIP proxies as any other request would be. It eventually arrives at a presence server, which can either terminate the subscription (in which case it acts as the presence agent for the presentity), or proxy it on to a presence client. If the presence client handles the subscription, it is effectively acting as the presence agent for the presentity. The decision at a presence server about whether to proxy or terminate the SUBSCRIBE is a local matter; however, we describe one way to effect such a configuration, using REGISTER.

VI. Conclusion

The SIP and PAM study looked at a relatively new aspect of the interaction between PAM and SIP. This work is related to the problem that Parlay's generic nature makes the relationship with order technologies such as Session Initation Protocol (SIP) difficult to define. The issue arise when analysing how the functionality of new protocols, SIP in particular, can be mapped onto the current Parlay interfaces.

At first glance the proposed SIP Presence Extensions seem to map well onto the Parlay PAM Event Management Interfaces. More network is required to see how the order Parlay PAM Interfaces fit into a SIP enabled network. In particular one could look at how Parlay PAM could be used to manage a presentity's presence information, which would require looking at an application that actually registers its information with a presence server, rather than subscribing to another presentity's presence information.

References

[1] 3GPP: Technical Specification Group Services and System Aspects; Presence service, TR 23.841, v1.1.0 (2002-03), Release 6.

[2] 3GPP: TS 32.200, Charging Principles.

[3] 3GPP: TS 32.225, Charging Data Description for the IMS domain.

[4] 3GPP: TS 33.210, Network Domain Security.

[5] 3GPP: TS 23.228, IM Subsystem Stage-2.

[6] 3GPP: TS 23.218, IP Multimedia (IM) Session Handling; IP Multimedia (IM) call model.

[7] IETF: CPIM Presence Information Data Format, Internet Draft in IMPP WG http://www.ietf.org/internet-drafts/draft-ietf-impp-cpim-pidf-01.txt, October 2001

[8] http://www.pamforum

[9] http://www.sipforum

[10] Rosenberg et al.: SIP Extensions for Presence, Internet-Draft in SIMPLE WG, http://www.ietf.org/internet-drafts/draft-ietf-simple-presence-06.txt, April 2002.

[11] http://www.3gpp.org

Evaluation of Optimum Orthogonal Code Allocation in CDMA Downlinks

Stiliyan Y. Paunov[1] and Seferin T. Mirtchev[2]

Abstract – **This paper presents an optimum orthogonal code allocation in case of the multiple-scrambling-code approach. The multiple-scrambling-code approach that was used in the aforementioned capacity results for the code limitation cases fully utilizes the soft capacity feature of CDMA. If the orthogonal code is short in a cell, then a new orthogonal code set is made appending a new scrambling code. However, we can no longer expect to avoid intracell interference between signals assigned different scrambling codes, and therefore the assignment of the orthogonal codes to the respective signals is an important study item if we continue to consider the multiple-scrambling-code approach.**

I. Introduction

The orthogonal code limitation affects CDMA (Code-division multiple access) performance: The fewer the orthogonal codes, the smaller the capacity. This code limitation problem is also discussed extensively in [1], which takes a theoretical approach.

Two general approaches are used to cope with the code limitation: the call-blocking approach and the multiple-scrambling-code approach. In call blocking, the arrival of a new call is blocked if the number of MSs (mobile stations) connected exceeds the number of orthogonal codes available. This approach sacrifices some of the soft capacity feature of CDMA because a call will be rejected in case of orthogonal code shortage even though the interference has not reached the specified limit. However, the multiple-scrambling-code approach that was used in the aforementioned capacity results for the code limitation cases fully utilizes the soft capacity feature of CDMA; if the orthogonal code is short in a cell, then a new orthogonal code set is made appending a new scrambling code. However, we can no longer expect to avoid intracell interference between signals assigned different scrambling codes, and therefore the assignment of the orthogonal codes to the respective signals is an important study item if we continue to consider the multiple-scrambling-code approach.

In cellular CDMA systems, downlink signals are spread by orthogonal spreading codes in order to minimize the interference between the signals [2]. However, because the number of orthogonal codes is limited, the downlink capacity is also limited once the number of MSs connected exceeds the number of orthogonal codes available. The impact of code limitation depends on such factors as the data transmission rate and the forward error control coding rate and is especially significant when soft handoff is used. This is because soft handoff accelerates the consumption of the orthogonal codes in accordance with the number of base stations connected to an MS. The number of downlink channels can be increased by enabling multiple scrambling codes to be allocated to a single BS. But a downlink signal is then subject to strong interference from the other signals assigned different scrambling codes. In this paper we discuss, with the help of a general genetic algorithm toolkit implemented in Java , an optimum code allocation maximizing the average SIR(Signal-to-interference power ratio) measured at MSs within a cell. A genetic algorithm is a search/optimization technique based on natural selection. Successive generations evolve more fit individuals based on Darwinian survival of the fittest. The genetic algorithm is a computer simulation of such evolution where the user provides the environment (function) in which the population must evolve [3].

II. Average SIR

The term *code allocation* throughout this section means an allocation of codes given as products of the multiplication of a basic orthogonal code set and multiple scrambling codes. We denote as S the number of basic orthogonal codes and denote as N the number of orthogonal code sets provided for a cell. The transmission power for downlink signals is fixed, and the orthogonal code occupancy ratio k_i is the ratio of the number of MSs assigned the i-th orthogonal code set to the number n of all MSs within a cell, where the sum of $k_1, k_2, ...,$ and k_N is 1. Defining the code occupancy ratio vector of $k = (k_1, k_2, ..., k_N)$, we can write the following equation for the average SIR measured at MSs within a cell:

$$\bar{\Gamma}(k) = \sum_{i=1}^{N} k_i \frac{P_g}{\varepsilon k_i n + (1 - k_i)n} = \frac{P_g}{n} \sum_{i=1}^{N} \frac{1}{\varepsilon + 1/k_i - 1} \quad (1)$$

where P_g and ε denote the processing gain and the interference figure. The value of ε ranges from 0 to 1, and $\varepsilon = 0$ denotes no multipath distortion in the downlink channel. In (1), $\varepsilon = 1$ is assumed between downlink signals assigned different scrambling codes, hence the signals with different scrambling codes interfere completely with each other. We also assume in (1) a single-cell environment, that is, we exclude the intercell interference in measurement of the average SIR.

[1]Stiliyan Paunov is with the Faculty of Communication Technology, bul. Kl. Ohridsky 8, 1000 Sofia, Bulgaria, stp@vmei.acad.bg

[2]Seferin T. Mirtchev is with the Faculty of Communication Technology, bul. Kl. Ohridsky 8, 1000 Sofia, Bulgaria, stm@tu-sofia.acad.bg

III. Extreme Value of Average SIR

When $k^* = (k_1^*, k_2^*, ..., k_N^*)$ is the code occupancy ratio vector giving an extreme value of $\bar{\Gamma}(k)$ and when $d = (d_1, d_2, ..., d_N)$ is the variation vector, the equation

$$\frac{d\bar{\Gamma}(k^* + \alpha d)}{d\alpha}\Big|_{\alpha=0} = 0 \quad (2)$$

is satisfied for an arbitrary variation vector d because $\bar{\Gamma}(k)$ takes an extreme value at $k = k^*$. From (1) and (2), we can derive

$$\frac{P_g}{n}\sum_{i=2}^{N} d_i \left[\frac{-1}{\left((\varepsilon-1)\left(1-\sum_{j=2}^{N} k_j^*\right)+1\right)^2} + \frac{1}{\left((\varepsilon-1)k_i^*+1\right)^2}\right] = 0 \quad (3)$$

which is an identical equation with respect to d_i because the equation must be satisfied for an arbitrary variation vector d, and hence all coefficients of d_i must be 0. Eventually, simultaneous equations with respect to k_i are given as follows:

$$\frac{-1}{\left((\varepsilon-1)\left(1-\sum_{j=2}^{N} k_j^*\right)+1\right)^2} + \frac{1}{\left((\varepsilon-1)k_i^*+1\right)^2} = 0 \quad (4)$$

Solving simultaneous equations led by (4), we have

$$k_i^* = 1 - \sum_{j=2}^{N} k_j^* = k_1^* \quad \because \quad i = 2 \sim N \quad (5)$$

which leads to a unique solution to the simultaneous equations of interest, $k^* = (1/N, 1/N, ..., l/N)$. To determine which $k^* = (1/N, 1/N, ..., l/N)$ gives a minimum or maximum SIR, we use Figure 1, which is a plot of the average SIR against k_1 and k_2 in the case for which $N = 3$ is assumed. The assumed case gives $k^* = (k_1, k_2, k_3) = (1/3, 1/3, 1/3)$. From Figure 1, we can see that the point at $(k_2, k_3) = (1/3, 1/3)$ shows the lowest SIR. The sequence developed so far tells us that the average SIR is smallest when each orthogonal code set is equally used.

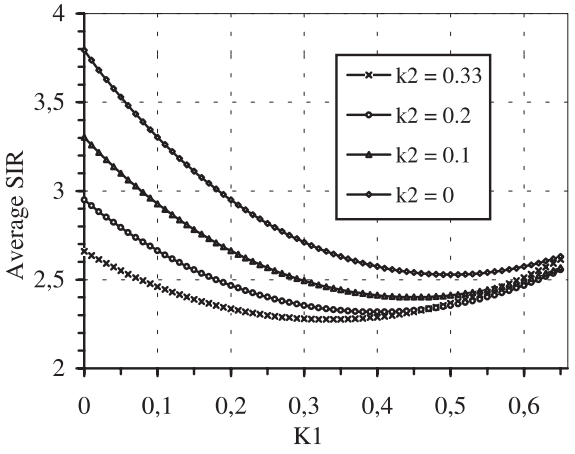


Fig. 1. The average SIR for $N = 3$

IV. Optimum Code Allocation

Figure 1 and Table 1 also shows that the average SIR $\bar{\Gamma}(k^* + \alpha d)$ becomes larger as the absolute value of α increases. There is, however, a limit to the value of α because all elements of the vector $k^* + \alpha d$ have to be between 0 and S/n $(n \geq S)$. Defining the limit of α as α_1, we can write the code occupancy vector $k^* + \alpha_1 d$ at α_1 as

$$k^* + \alpha_1 d = \left(\frac{S}{n}, \frac{S}{n}, ..., \frac{S}{n}, r_1, r_2, ..., r_R, 0, 0, ..., 0\right) = k \quad (6)$$

where $0 < r_1 < r_2 < \cdots < r_R < S/n$. We call k_r a code occupancy bound vector, and the code occupancy vector maximizing $\bar{\Gamma}$ is one of these code occupancy bound vectors.

From Figure 1 and Table 1 we can see that as the number of fully occupied orthogonal code sets increases, the average SIR can be larger and also as the number of empty orthogonal code sets increases, and hence as the use of scrambling codes is reduced, the average SIR can be larger.

Finally, an optimum code allocation can be described as follows: Optimum code allocation is an allocation scheme that maximizes the number of code sets fully occupied and minimizes the use of scrambling codes.

Table 1. Average SIR against N for $S = 128$, $n = 270$, $P_g = 512$ and $\varepsilon = 0.5$

N	2	3	4	5	8	10
Average SIR, dB	4.03	3.57	3.36	3.24	3.06	3.00

V. Evaluation of Optimum Code Allocation

The average SIRs for the optimum code allocation were compared with those for the worst code allocation in which every orthogonal code set enabled by multiple scrambling codes is equally occupied. When $N = 5$, $S = 128$, $n = 270$, $P_g = 512$, and $\varepsilon = 0.5$, the average SIR was found to be 3.24 dB for the worst allocation and 3.90 dB for the optimum allocation. That is, the optimum code allocation improved the average SIR by about 0.7 dB. And when $N = 2$, $S = 128$, $n = 150$, $P_g = 512$, and $\varepsilon = 0.5$, the average SIRs given by the optimum and worst code allocations, respectively, were 6.58 and 7.50 dB; the average SIR was about 1 dB better with the optimum allocation.

References

[1] Furocawa, H., " Theoretical Capacity Evaluation of Power Controlled CDMA Downlink", *Proc. VTC 2000-Spring*, Vol. 2, May 2000, pp. 997-1001.

[2] Adachi, F., "Effects of Orthogonal Spreading and Rake Combining on DS-CDMA Forward Link in Mobile Radio", *IEICE Trans. On Communication*, Vol. E80-B, No. 11, Nov. 1997, pp. 1703-1712.

[3] Goldberg, David, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, 1989.

[4] Furocawa, H., "An Optimum Code Allocation Scheme for CDMA Downlink", *Proc. IEICE Nat. Conf.*, B-5-93, Oct. 2000.

[5] Wang, J., Ng, TS., "Advances in 3G Enhanced Technologies for Wireless Communications", Artech House, 2002.

Planning Wireless Code Division Multiple Access Network Considering Customer Traffic

Pavlina Hr. Koleva[1]

Abstract – **In this paper the behavior of the CDMA cell is investigated - using clustered Poisson process. The density function of the CDMA cell radius is modeled using the traffic intensity and the capacity of the cell. The investigations are done with singe cell.**

Keywords – **CDMA, Number of calls, Traffic intensity**

I. Introduction

Several system access methods are known from the references: FDMA, TDMA and CDMA. In an FDMA system, the time-frequency plane is divided into m discrete frequency channels. During any particular time, the user transmits signal energy in one of these m frequency channels with 100% duty cycle. In a TDMA system, the time-frequency plane is divided into m discrete timeslots. During any particular time, the user transmits signal energy in one of these timeslots with low duty cycle. In a CDMA system, the signal energy is continuously distributed throughout the entire time-frequency plane. The time-frequency plane is not divided among subscribers, as done in the FDMA and TDMA systems. Each end user employs a wideband coded signaling waveform. The technology that supports broadband wireless access to the end-users is WCDMA. It applies CDMA technique with broadened spectrum. This technology is applied in UMTS system.

II. Network Parameter Modeling

A cell in a CDMA network with Base Transceiver Station (BTS), witch supports a number of calls is considered -= Fig. 1. At the observation instant there are m calls to be supported and power-controlled in the cell.

Let us observe a subscriber i in conversation phase. This is the period of time, when the user transmits activity burst during his call. These bursts are separated by idle phases. The probability that the customer gets a link with acceptable Quality of Service (QoS) is a function of the distance to the base station and the current interference. The interference is not only depending on distance to base station x, but is also a function of the distribution of the calls currently supported in the cell.

This paper is an extension of the works presented in [2] and [3]. In [2] is investigated the blocking probability of a

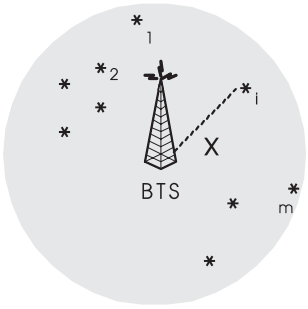


Fig. 1. CDMA cell with m supported calls

CDMA cellular system and the relation between dimension of the expanded frequency band and the number of users in the current cell (Erlang capacity). Also were made investigations when the frequency band is expanded/reduced twice and how this reflects on the number of served users. In [3] is investigated the blocking probability of the CDMA system while the data rate of mobile users are different. In bout works are taken into account the interference between users and interference between neighboring cells (multiple cells). But in bout cases is not taken into account the distance to the base station.

By modeling the location of the subscribers with spatial process, an analytic description of the user distribution within the cell could be obtained. A point process to characterize the relationship between number and location of the subscribers must be used. This paper deals with the homogeneous Poisson process.

III. Customer Traffic And Basic Relations

The distribution of the random variable M_A of calls on a surface with area A is Poisson distributed as:

$$P(M_A = m) = \frac{(\lambda A)^m}{m!} \exp(-\lambda A) \qquad (1)$$

where λ (in calls per km^2) denotes the spatial traffic intensity. The distribution of M_A given above is valid at any arbitrary observation instant.

Based on this Poisson process assumptions now consider a cell modeled by a circle with radius R_C. One active call is assumed to be on the circle and $m-1$ connections are inside the circle Fig. 1. The corresponding coverage area is $A = \pi R_C^2$, where both A and R_C are random variables. To give a precise analytic description the random variable A as the surface of the smallest circle containing m points must be defined [1,4]. Due to the property of the spatial Poisson

[1]Pavlina Hr. Koleva is with Telecom Department, Faculty of Communications and Communications Technologies at Technical University of Sofia, "Kliment Ochridsky" Blvd. 8, 1756 Sofia, Bulgaria, e-mail: p_koleva@vmei.acad.bg

process, the size of the surface A is distributed according to an Erlang-distribution of order m. It is more useful to consider the radius of the cell rather than its surface, as this can translate directly to the distance between subscriber and base station. The distribution of the radius R_C can be derived as:

$$R_C(x) = 1 - \sum_{i=0}^{m-1} \frac{(\lambda\pi x^2)^i}{i!} \exp(-\lambda\pi x^2). \qquad (2)$$

The probability density function is given by following equation:

$$r_C = \frac{\lambda(\lambda\pi x^2)^{m-1}}{(m-1)!} \exp(-\lambda\pi x^2)(2\pi x). \qquad (3)$$

With Eq. (3) the probability that we have a cell radius of x for a cell currently supporting m calls at an intensity of λ, could be calculated.

IV. Results and Analysis

Fig. 2 depicts the density function of the cell radius for different values on number of calls. The graphics are obtained using Eq. (3) with traffic intensity of $\lambda = 10$ [calls/km^2]. Plots for different values of number of users, an exactly $m = 10$, 20 and 30, are shown in the figure.

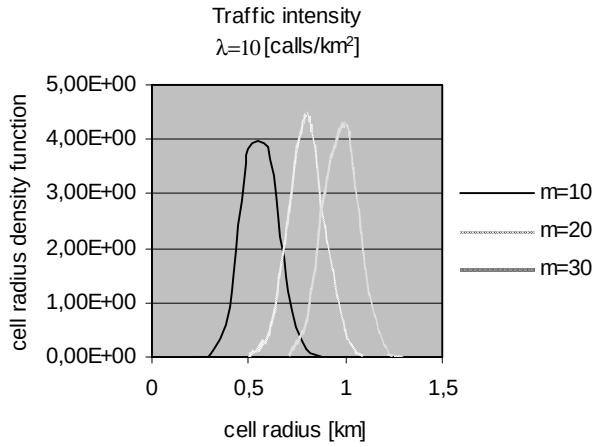


Fig. 2. Density function of the cell radius for different number of calls and $\lambda = 10$ [calls/km^2]

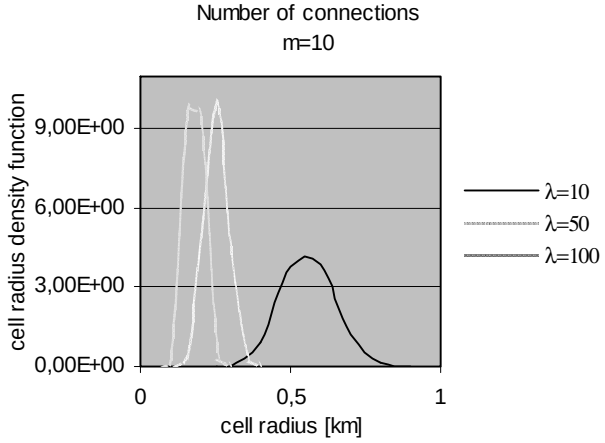


Fig. 3. Density function of the cell radius for different traffic intensity and $m = 10$

It shows that to support fewer calls, the mean cell radius is in general smaller than for larger values of number of calls, for fixed traffic intensity. The shape and variance of the cell radius density function for different m stay the same.

In the next graphic Fig. 3 are presented the curves for density function of the cell radius for different spatial traffic intensity. The value of calls number are fixed – $m = 10$. It indicates that for areas with high values of traffic intensity, e.g. urban or dense urban regions, the cell radius is more clearly defined than for areas with lower intensity, like the curve for $\lambda = 10$. The range of the radius is here more than double the size compared to traffic intensity $\lambda = 100$ [calls/km^2].

In Fig. 4 are shown the plots for density function of the cell radius for different values of calls number and for different values of traffic intensity. These values for number of calls are: $m = 10$, 20 and 30. The values for traffic intensity are: $\lambda = 10$, 50 and 100. From the curves on this figure it can be seen clearly what is the influence on traffic intensity and number of calls over cell radius.

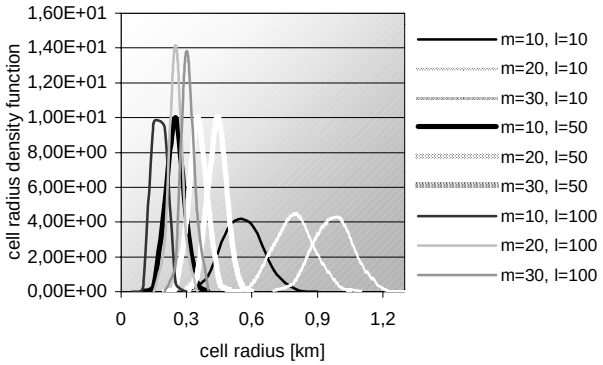


Fig. 4. Density function of the cell radius for different number of calls and different traffic intensity

V. Conclusion

An approach to obtain some relations about the influence of the number of calls and traffic intensity, over the cell radius is proposed in this work.

From the experiments and results that have been obtained it can be concluded that the augmentation of a number of calls lead to increase of the cell radius. An exact relationship is observed and when traffic intensity is increased. It is necessary to note that the investigations are made only for a singe cell.

References

[1] J. Sam Lee, L. Miller, "CDMA Systems Engineering Handbook", 1998

[2] P. Koleva, G. Balabanov, "Teletraffic aspects of mobile ATM", *ICEST'2002*, October 2002

[3] P. Koleva, Sl. Pavlova, "Teletraffic aspects in mobile ATM", *TTONY'2002*, September 2002

[4] P. Tran-Gia, N. Jain, K. Leibnitz, "CDMA wireless network planning considering clustered spatial customer traffic", *ITNPS*, Italy, October 1998

Analysis of Access Network in ATM Networks

Kamelia S. Ivanova[1]

Abstract – **This paper is a part of ATM access network analysis implemented with an ATM multiplexer. The ATM multiplexer on ATM layer is a buffer. The aim is to define: buffer overflow probability for different incoming streams and cell loss ratio with variable queue length, multiplexing gain and number of sources.**

Keywords – **ATM, multiplexing, overflow, queue**

I. Introduction

Planning and designing of Asynchronous Transfer Mode - based (ATM) networks is a challenge in the view of heterogeneous services, statistical multiplexing and additional logical layer that manages virtual paths. Describing ATM as multiplexing and switching modes defines both ATM multiplexer and ATM switch as underlying components in Broadband Integrated Services Digital Network (B-ISDN). They consist of a physical ATM link with a buffer. Since packet switching networks are based on queuing model, their efficiency is bound up with the serving queue ones.

II. Principles of Multiplexing

The main feature of ATM multiplexer is multiplexing into one and the same outgoing ATM line. If there is multiplexing of cell streams generated by separate sources the multiplexer is called primary. On the other hand the multiplexer is called secondary if the flows come from great number of incoming ATM lines [3].

The above mentioned division of ATM multiplexers is strongly related to the nature of incoming heterogeneous traffic. In that sense the stream has specific quality of services requirements. On ATM layer, from traffic point of view the multiplexer is a buffer [4]. The multiplexer may implement different queue disciplines or scheduling policies. In this paper we assume first in first out (FIFO) discipline. The cell transfer rate depends on the outgoing line speed and transfer time of each cell is constant due to its fixed length (fig. 1).

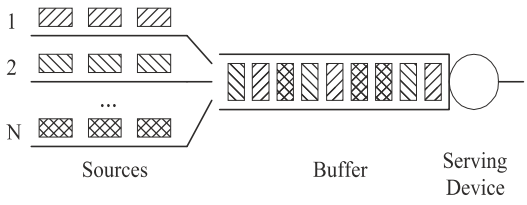


Fig. 1. ATM multiplexer

[1]Kamelia S. Ivanova is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohgidski 8, 1000 Sofia, Bulgaria, E-mail: ksi@vmai.acad.bg

The queue length and the incoming traffic intensity could lead to the loss of cells originating from different sources. Cell losses are due to cell delay or lack of queuing positions.

III. Buffer Overflow Probability

A. Constant bit rate (CBR) streams multiplexing

Buffer occupancy state for any fixed set of streams is a periodic process of period T. To render the probability of this occurrence less than a target level (say, 10^{-9}), the multiplexer buffer may be dimensioned so that, for a random incoming traffic, the probability of buffer overflow $Q(B)$ (where B is buffer capacity in cells) in an arbitrary instant is less than the target value [1].

For the small probabilities generally considered, $Q(B)$ can be likened to the saturation probability of a buffer of capacity B. An approximate formula which gives good order of magnitude estimates at load (N/D) greater than 0.8 is [2]:

$$Q(B) = \exp\Big\{-2B\Big(\frac{B}{N} + \frac{D-N}{N}\Big)\Big\}. \qquad (1)$$

For the very small probabilities of interest, $Q(B)$ constitutes a tight upper bound on the cell loss ratio. In general, the streams do not have the same rate. In this case, the calculation of the buffer overflow probability $Q(B)$ proves complicated.

The $M/D/1/m$ system may be used as a tool for worst-case dimensioning, without restriction of number of multiplexed connections. The results of this traffic model provide conservative estimate of buffer requirements and constitute a good approximation when the number of sources is high and the multiplexer load is not too close to 1. An accurate approximation for the queue length is [2]:

$$Q(B) \approx Ce^{-rB} \qquad (2)$$

where $C = (1-\rho)/(\rho e^r - 1)$ and r is the solution of the equation and ρ is multiplexer load.

$$\rho(e^r - 1) - r = 0. \qquad (3)$$

The assumption of Poisson arrivals corresponds to a worst-case traffic model for any superposition of periodic streams (homogeneous or heterogeneous) having the same overall average arrival rate in the sense that all quantiles of the delay distribution are greater [1].

In particular, the Cell Loss Ratio (CLR) estimated by Q(B) is greatest for Poisson arrival (Fig. 2).

The Eq. (2) can be used to estimate the buffer saturation probability for corresponding batch arrival systems on replacing B by B/k. For example, the buffer saturation probability when cells arrive in batches of k according to a Poisson process (the $M^{(k)}/D/1$ queue) may be estimated by

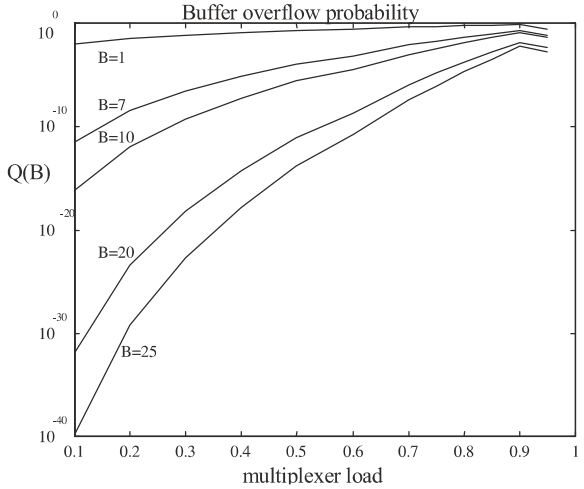


Fig. 2. $Q(B)$ for Poisson arrival process

$Q(B/k)$ where $Q()$ is given by Eq. (2). An example is given in Fig. 3.

B. Variable bit rate (VBR) streams multiplexing

Variable bit rate steam called on/off stream allow statistical multiplexing, it can be performed by assuring that the combined instantaneous input rate is not greater than the multiplexer service rate.

The purpose being to keep the cell arrival rate within the limit defined by the service rate is referred to as Rate Envelope Multiplexing (REM) [1].

Cell loss ratio (CLR) is decomposed at into "burst-scale" and "cell-scale" components. The first, CLRbs corresponds to losses due to rates greater than multiplexer capacity and second CLRsc corresponds to a correction term.

When the rate of multiplexed traffic streams is defined with negligible cell delay variation with respect to a k-batch Poisson process the "burst-scale" component may be sufficiently approximate with buffer overflow probability given

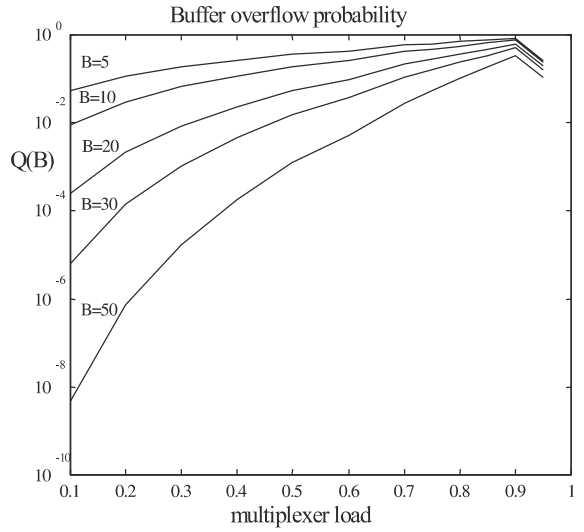


Fig. 3. $Q(B)$ for $M^{(k)}/D/1$ queue

with Eq. (2):

$$CLR_{CS} + Q(B/k). \tag{4}$$

When N identical on/off sources of peak rate p and mean rate m are multiplexed on a link of capacity c. CLR_{bs} is then estimated by:

$$CLR_{bs} = \sum_{ip>c'} (ip-c) \binom{N}{i} \left(\frac{m}{p}\right)^i \left(1-\frac{m}{p}\right)^{N-i} \frac{1}{Nm}. \tag{5}$$

IV. Results and Analysis

The buffer overflow probability as a function of buffer length is displayed in Fig. 4 (Eq. 1). In this case the multiplexer load is fixed at 0.9 and the number of CBR sources is variable. It is shown that the buffer overflow probability becomes smaller when the queue length increase. When queue length is greater than 20 $Q(B)$ is negligible.

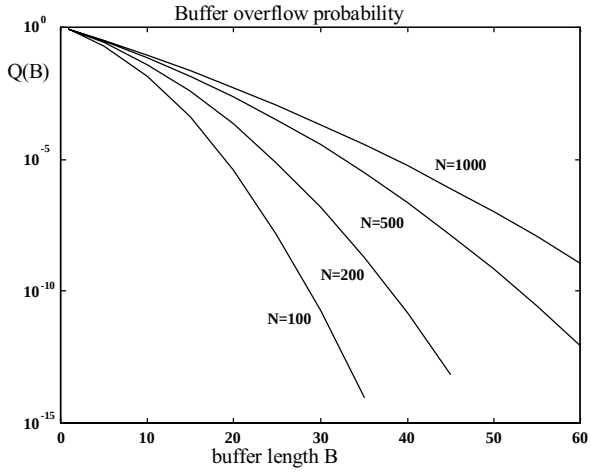


Fig. 4. $Q(B)$ at multiplexer load 0.9

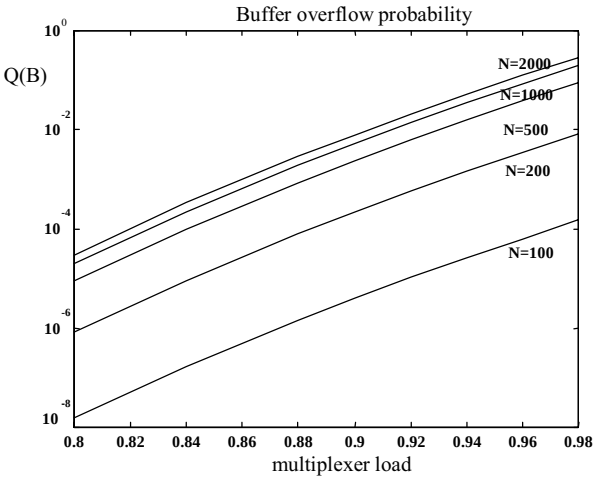


Fig. 5. $Q(B)$ at buffer length equal to 20

The buffer overflow probability as a function of multiplexer load is shown in Fig. 5. It shown that, as much the number of streams increases as the difference between $Q(B)$ decrease.

V. Conclusion

The results of this investigation may be used for the multiplexer dimensioning. The CLR or $Q(B)$ requirements (around 10^{-6} or 10^{-9}) are achieved by a queue length with capacity over 20-50 cells.

References

[1] E 736 Recommendation – Methods for cell level traffic control in B-ISDN, 2000.

[2] J. Roberts, U. Mocci, J. Virtamo, (Eds): "Broadband Network Teletraffic", *Lecture Notes in Computer Science*, Vol. 1155, Springer-Verlag 1996.

[3] B. Tsankov, L. Hadjiivanov, ATM – principles and performance, ISRT, Sofia, 1996.

[4] S. Mirchev, ATM – communications, Novi Znanya, Sofia, 2001.

Application of Genetic Algorithms in Access Network Planning

Alexander Tsenov[1]

Abstract – **In recent work a possible presentation of location-allocation problems in access network planning was introduced [1] – the set-based presentation. The new presentation was tested over many topological problems and has shown acceptable results as well.**

However, one important problem has arisen - the application of standard genetic operators, such a crossover and mutation has leaded to generation of many invalid individuals (unusable decisions of the topological problem).

Therefore, in this paper new genetic operators are introduced – set-based crossover, based on Random Transmitting Recombination, and set-based Mutation. Both are oriented to the set-based representation of location - allocation problems. The paper presents the functionality of these types of operators and some practical results as well.

Keywords – **Network design, Optimization, Genetic Algorithms, Crossover, Mutation**

I. Introduction

The problem of optimally designing a network in order to meet a given set of specifications (such as prescribed traffic requirements, achieving a desired level of reliability, respecting a given maximum transit time), while minimizing total cost, arises in a wide variety of contexts: computer networks, telecommunications networks, transportation networks, distribution systems.

Network design algorithms draw an increasing amount of attention nowadays. Considering the complexity, high cost factor and fast deployment times of today's communications systems (such as IP and ATM backbones, optical networks, numerous types of access structures etc.), network operators can benefit a lot from the use of network design tools. These tools can help speeding up and 'automating' the design process, ensuring superior quality (i.e. lower cost and/or better Quality of Service) and more justifiable solutions. Network design tools typically incorporate a wide range of functionality, such a geographical database handling, traffic estimation, link dimensioning, cost calculation, equipment configuration databases etc. The real benefit of using these tools, however, comes from the possibility of using the algorithmic network optimization approaches. In this way, there arises a possibility for finding solutions of better quality in much shorter time, as compared to the manual network design.

This paper introduces new genetic operators that are appropriate to new type of presentation of location - allocation problems for access network planning and applies them to some simple network optimization problems.

The first section describes a new set-based crossover operator based on Random Transmitting Recombination [2] and the second section – the set-based mutation operator, both used for application in genetic set-based representations in the context of a general location-allocation problem.

II. Set-Based Crossover

The set-based representation described in [1] is such that traditional crossover and mutation operators will not perform well. The representation is not orthogonal as for example

$$\xi_{a\,b} \cap \xi_{bc} \cap \xi_{\overline{ac}} = \emptyset$$

The non-orthogonality displays itself as dependencies between forma ("forma" is every simple of the representation): that means that a traditional operator would generate many invalid individuals. This is illustrated using an example:

Let consider a simple case where there are four customers to be connected. If the following two parents are selected for crossover witch represent the sets $\{\{a\},\{b,c,d\}\}$ and $\{\{a,b\},\{c,d\}\}$, Fig. 1 shows how an illegal child can be generated when uniform crossover is used. In fact, of the eight possible children under uniform crossover, five would be infeasible. Infeasible means that a customer is specified as sharing a set with a customer *and* not to sharing. E.g. a shares with b, b shares with c and a does not share with c.

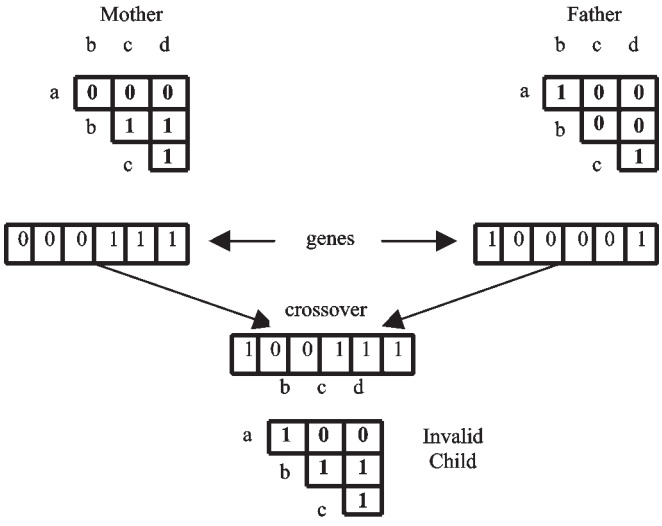


Fig. 1. Example uniform crossover, generating illegal children with set-based representation

[1]Alexander Tsenov is with Telecom Department at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: akz@vmei.acad.bg

As a consequence of this non-orthogonality and the fact the forma are separable (as assortment and respect are compatible) an operator based on Random Transmitting Recombination (RTR) was developed. The operator is a version of uniform crossover adapted for problems where the alleles are non orthogonal.

The operator functions in a number of stages:

1. Step 1: Gene values that are common to both parents are transmitted to the children;
2. Step 2: For each remaining uninitialized gene in the child: randomly select a parent and take the allele at the same location, set the child's gene to this allele. Update all dependant values in the array;
3. Repeat Step 2 until all values are specified.

Figs. 2-4 show a working example of the described operator. "1" indicates that two customers are sharing a set, (e.g. physically they are connected to the same node of the network) and "0" indicates not sharing.

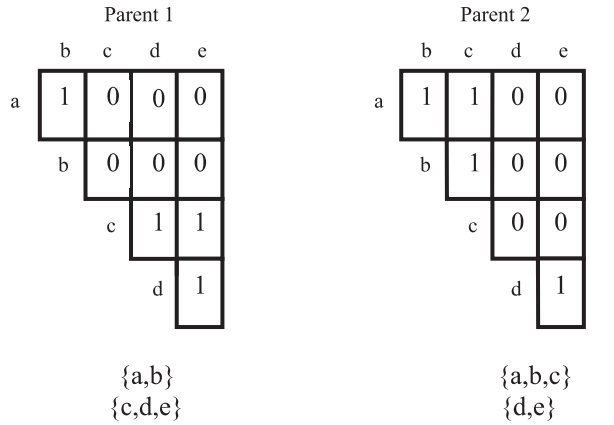


Fig. 2. Two example parents and the sets that they represent

So, in the first stage of crossover a child is built that contains all the allele values that are common to both parents. These values are shown on Figure 3, where for example the gene for customer a and b have an allele value of "1" in both parents. This applies to all common allele values.

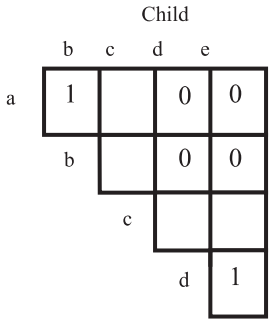


Fig. 3. The new child after the first stage of crossover

In the next stage of crossover a random uninitialized gene is selected and the corresponding value from a random parent is chosen to place there. After this value has been set the array must be updated to show any consequential changes. Figure 4 illustrates all the possible children that can be generated by this process given the two parents shown in Figure 2.

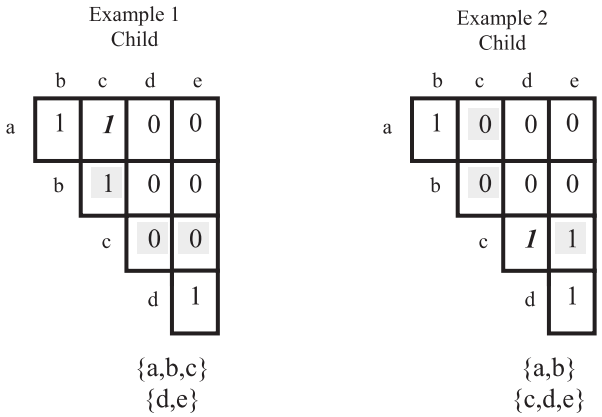
In example 1 in the figure for customers a and c is chosen and randomly assigned the value "1" from parent 2 (shown with italic in the Figure). As a consequence of this the following happens (shown with gray background in the Figure):

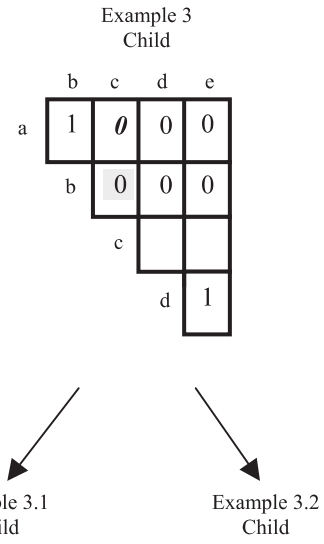
- because $(a, c) = 1$ and $(a, b) = 1$; that means a shares with c and a shares with b, therefore, b must share with c. So, $(b, c) = 1$.
- because $(b, c) = 1$ and $(b, d) = 0$; that means b shares with c and b doesn't share with d, therefore c cannot share with d. So $(c, d) = 0$.
- because $(b, c) = 1$ and $(b, e) = 0$; that means b shares with c and b doesn't share with e, therefore c cannot share with e. So $(c, e) = 0$.

In this case, there are no longer any uninitialized genes, therefore crossover has finished. Same reasonings may be made in example 2.

In examples 3 and 4 this is not the case and Step 2 is repeated to initialize all the remaining genes.

The two parents in the example are very similar, and there are not many customers to allocate to sets, therefore the children are similar, in fact there are children same as the parents.





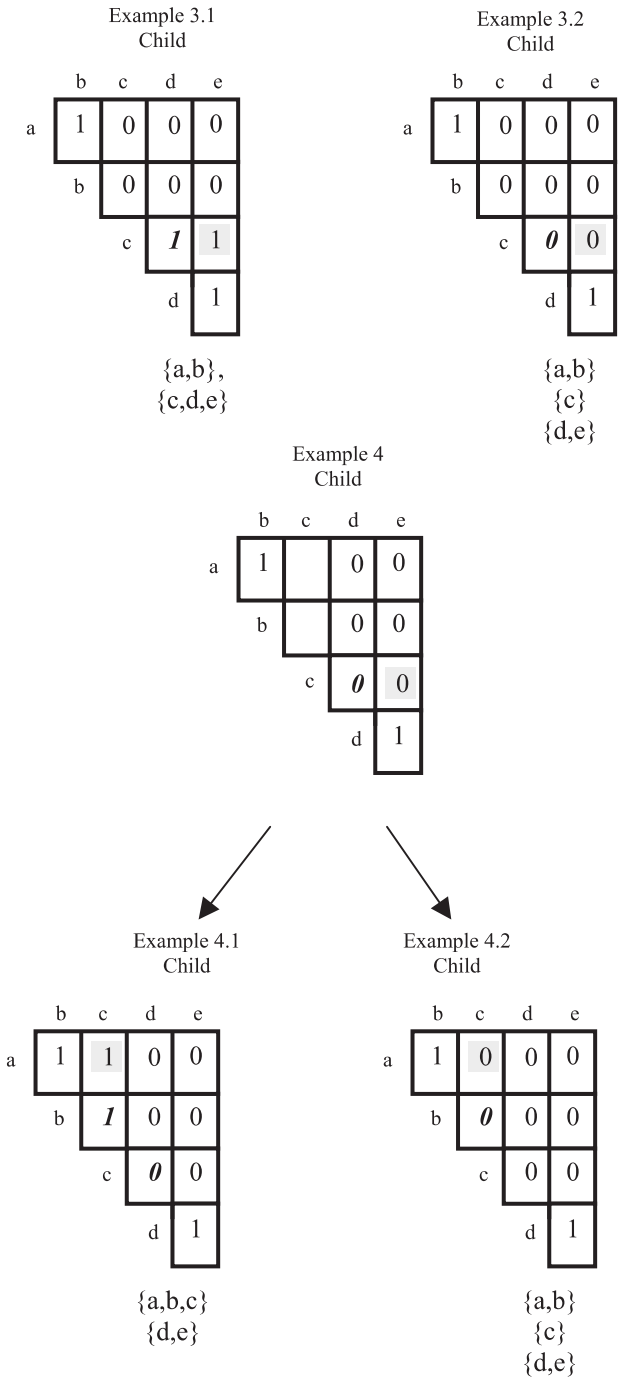


Fig. 4. All the possible children that can be generated for the parents shown in Figure 1. Those values shown in *italic* indicate the randomly chosen allele for a randomly chosen gene. Those values shown with gray background are consequential changes

Clearly, for larger examples much more variety is present in the children.

III. Set-Based Mutation

The aim of a mutation operator is to help the algorithm explore new areas of the search space by generating new genetic material. Given the representation described in [1] it is necessary to devise a compatible mutation operator. Typically genetic algorithms mutate an individual by randomly flipping the value of a gene in the individual; this corresponds to changing the value of a bit at a single point in the array. The nature of the representation is such that a change to the value of a single point in the array will lead to a cascade of changes throughout the array.

The effect of a change in the value of a bit can be understood in terms of the effect that it will have in the problem domain. If a value is changed from zero to one, this means that a customer that previously did not belong so a set of customers now does. That is, the customer changes its connection from one site to another – Fig. 5.

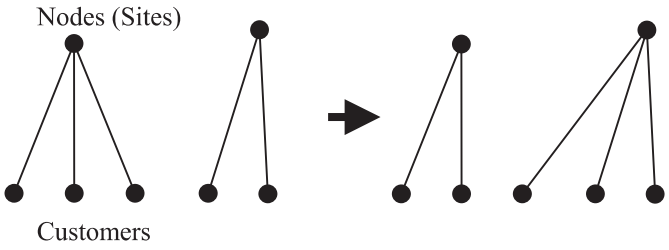


Fig. 5. The effect of mutation – changing an array value from “0” to “1” causes a customer to swap sites

Changing the value from one to zero means that s customer is removed from a set but not added to another – Fig. 6 The effect of this is to create a new set of customers with a single member. This means a new site must be initialized too.

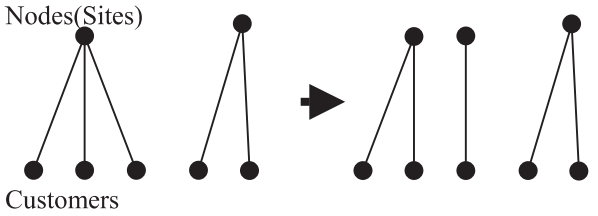


Fig. 6. The effect of mutation – changing an array value from “1” to “0” causes a customer to be removed from a set

The next section provides a worked example of the operation of the simple set-based mutation operator. The operator is implemented so that it has one of two effects: either the customer is moved from one set to another, or a new set is formed with a single customer.

Fig. 7 demonstrates the latter case where a gene is chosen randomly – in this case gene (a, b). This gene has the value “1”, which is mutated to “0”. This change leads no consequential changes, and customers a and b are split into two separate sets. If more then two customers where in the initial set, e.g. $\{a, b, c\}$, and the same gene had been chosen, then there would be two possible children: $\{\{a, b\}, \{c\}\}$ and $\{\{a\}, \{b, c\}\}$. The mutation operator chooses randomly between the two possible children.

Figure 8 demonstrates the mutation operator when the gene chosen has a value of “0” which is mutated to “1”. This is implemented so that one of the customers associated with the gene is moved to a new set. So, in the figure gene (a, c) is chosen and changing its value leads to two possible children

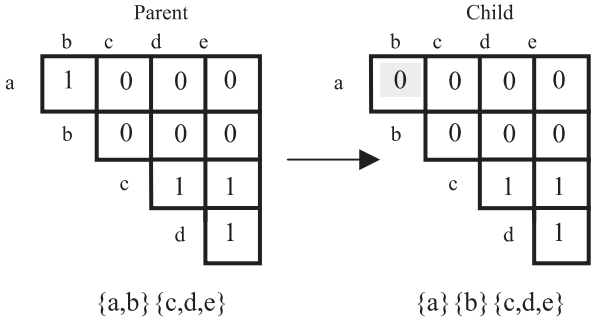


Fig. 7. The effect of mutation – changing an array value from "1" to "0" causes a customer to be removed from a set and new set is build.

one where customer a is moved to another set and the other when customer *c* is moved to another set. The mutation operator chooses randomly between the two possible children. It can be seen from the diagram that there are many consequential changes (shown in *italic*) produced by this change.

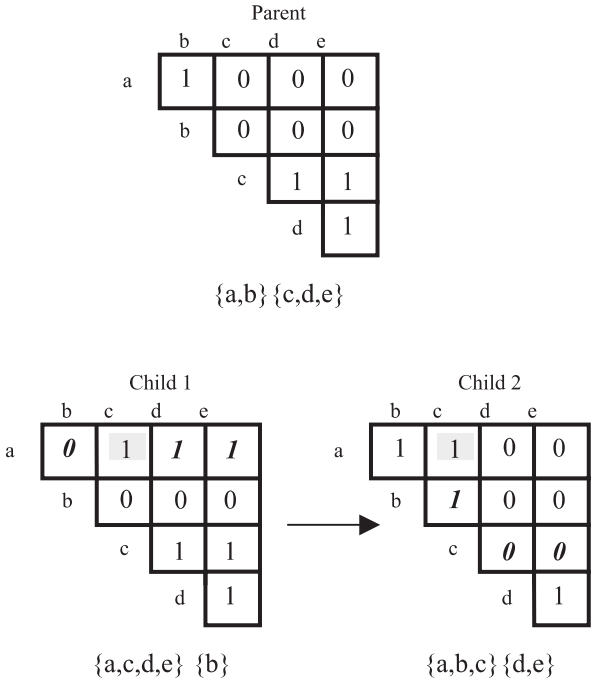


Fig. 8. The effect of mutation – changing an array value from "0" to "1" causes consequential changes (shown with *italic*)

IV. Conclusion

In this work two new genetic operators where introduced – the set-based crossover and the set-based mutation. Both are oriented to a new type of presentation of location-allocation problems in access network planing. It was proved that the implementation of these operators overcomes the disadvantages of implementing the standard crossover and mutation operators by using the set-based representation. This work is a part of the dissertation work of the author, where the implementation of both new operators was developed further and has led to good results.

References

[1] Tsenov A., "Presentation of location problems", *ICEST 2002*, pp. 299-302, Nish, Yugoslavia, 2-4 October 2002

[2] Radcliffe N.J, The Algebra of Genetic Algorithms, *Annals of Mathematics and Artificial Intelligence*, 10, pp. 339-384, 1994

[3] Routen T., Genetic Algorithms and Neural Network Approaches to Local Access Network Design, *Proceedings of the 2nd International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems*, 1994

[4] Kratica J., V. Filipovitch, V. Ljeljum, D. Toljijg, Solving of the Uncapacitated Warehouse Location Problem Using a Simple Genetic Algorithm, *Proceedings of the XIV International Conference on Material handling and warehousing*, 3.33-3.37, Belgrade, 1996

[5] Kratica J., Improvement of Simple Genetic Algorithm for Solving the Uncapacitated Warehouse Location Problem, *Proceedings of the 3rd On-line World Conference on Soft Computing (WSC3)*, 1998

[6] P. Chardaire, A. Kapsalis, J.W. Mann, V.J. Rayward-Smith and G.D. Smith, Applications of Genetic Algorithms in Telecommunications. In J. Alspector, R. Goodman, T.X. Brown (Eds.), *Proceedings of the 2nd International Workshop on Applications of Neural Networks to Telecommunications*, pp. 290-299, 1995

[7] N.J. Radcliffe The Algebra of Genetic Algorithms, *Annals of Mathematics and Artificial Intelligence*, 10, pp. 339-384, 1994.

[8] Hoang, H.H.: A Computational Approach to the selection of an Optimal Network, *Management Science*, vol. 19, pp. 488-498,1973

Priority Traffic Shaping Analyses in ATM Networks

Rossitza Goleva[1], Pavel Merdjanov[2], Tashko Nikolov[3], Kostadin Golev[4], Maria Goleva[5]

Abstract – **In this paper the traffic shaping analyses of voice, data and video traffic sources with priorities are given. ATM traffic sources are specified by their quality class. Shaping techniques with and without priorities are implemented and compared. Results show the delay distribution and its influence on end-to-end cell delay.**

Keywords – **ATM, traffic source, traffic shaping, ATM multiplex, QoS, priority**

I. Introduction

ATM technology is the unique technology in the world that may fully support so called traffic contract with end-users. This specific and useful feature is important when the user would like to commit and be sure on the Quality of the Service (QoS) received. ATM technology is going to be blown away from the market by IP network, but the work necessary to support QoS in IP-based network is still far away from the really comparative status.

The network establishes a separate traffic contract with the user of each ATM Virtual Path Connection (VPC) or Virtual Channel Connection (VCC). It is regarding a set of traffic parameters of the cell flow including Quality of Service (QoS) parameters. One of the major problems is the high bandwidth that usually the end-users ask for. There are many different approaches already known from the literature on decreasing the bandwidth. The most prominent but still not well implemented is traffic shaping. Because of lack enough collected experience traffic shaping devices are almost not available on the market.

It is considered that new generation ATM (Asynchronous Transfer Mode) terminals will obtain shaping functionality. Traffic shaping may influence significantly on the Quality of Service. It may play with the balance between cell delay and cell loss in real-time and non-real-time services and optimize the load to the transmission lines and ATM switches. Therefore, it is expected that the overall price will decrease due to the offered optimization.

[1]Rossitza Goleva is with the Department of Telecommunications at Technical University of Sofia, Kl. Ohridski blvd 8, 1756, Sofia, Bulgaria, Email: rig@vmei.acad.bg, Member of IEEE

[2]Pavel Merdjanov is with the Department of Telecommunications at Technical University of Sofia, Kl. Ohridski blvd 8, 1756, Sofia, Bulgaria, Email: pim@vmei.acad.bg.

[3]Tashko Nikolov is with the Department of Telecommunications at Technical University of Sofia, Kl. Ohridski blvd 8, 1756, Sofia, Bulgaria, Email: tan@vmei.acad.bg.

[4]Kostadin Golev is with the Department of Telecommunications at Technical University of Sofia, Kl. Ohridski blvd 8, 1756, Sofia, Bulgaria.

[5]Maria Goleva is with the Department of Telecommunications at Technical University of Sofia, Kl. Ohridski blvd 8, 1756, Sofia, Bulgaria.

Network Performance (NP) parameters have significance only at network side. The mapping between Network Performance and Quality of Service parameters is performed at Service Access Reference Point (SAP).

II. Traffic Sources

When the end-user signs traffic contract many QoS parameters might be identified as: Cell Delay (CD) – the delay of the cell at given node of the network; end-to-end Cell Delay – the delay of the cell between two end-terminals; Cell Delay Variation (CDV) – the variation of the cell delay arrival times at given network node. Cell delay and cell delay variation have an great impact on the traffic shaping. The ATM forum has defined QoS classes and has specified certain values for the QoS parameters for each class. The classification is as follows:

- Service class A – connection-oriented service, circuit emulation, Constant Bit Rate (CBR) applications like voice, video, audio, etc.;
- Service class B – connection-oriented service, Variable Bit Rate (VBR) sources like audio and video;
- Service class C – connection-oriented data transfer, VBR sources like Frame Relay data source, LAN emulation source, etc.;
- Service class D – connectionless data transfer with VBR sources for WWW, email, etc.

QoS classes identified also by the ATM Forum are five:

- Class 0 – unspecified quality for protocols like “Best effort”. It is implemented now in Internet;
- Class 1 – specified quality for service class A;
- Class 2 – specified quality for service class B;
- Class 3 – specified quality for service class C;
- Class 4 – specified quality for service class D.

All the other QoS classes without class 0 directly map to the service classes. QoS class 0 is a special case of lack of quality negotiation. Furthermore, a simulation of voice and data traffic sources is presented. Voice traffic is delay and delay variation sensitive. Data traffic is loss sensitive. The main problem with data traffic is its self-similarity. That means that the burstiness of the data traffic remains almost the same even after few ATM multiplexes. It is seen that the data traffic congested the transmission line and suppressed the voice traffic.

III. Traffic Shaping

In order the user to derive maximum benefit from the quaranteed QoS from ATM, the device connecting to the network should ensure that the cells sent to the network conform with the parameters in traffic contract. Traffic shaping is used accordingly to reorder and shift service cells in the way to satisfy all the contracts simultaneously. A network may also employ shaping, when transferring a cell flow to another network in order to meet the conditions of a network-to-network traffic contract, or in order to ensure that the receiving user application operates in an acceptable way. Traffic shaping might be also applied by different network nodes in order to avoid congestion or cell loss.

There are many algorithms for traffic shaping and some of them are: buffering; Leaky Bucket buffering; spacing; scheduling; burst length limiting; source rate limitation; framing.

IV. Delay and Loss Priority Control

Priority control helps achieving the full range of QoS loss and delay parameters required by the applications. Delay priority involves the use of separate queues served in a prioritized manner. Loss priority involves thresholds within the buffers for different traffic types and Cell Priority (CLP) bit in the cell header.

Priority queuing implements multiple queues in the switch, such that traffic on certain Virtual Path Connection/Virtual Channel Connection that are not tolerant on delay can go ahead of those that are more tolerant on delay. ATM switches employ priority queuing between VPCs and VCCs in different QoS classes or service categories to meet different delay and loss priorities simultaneously (Fig. 1).

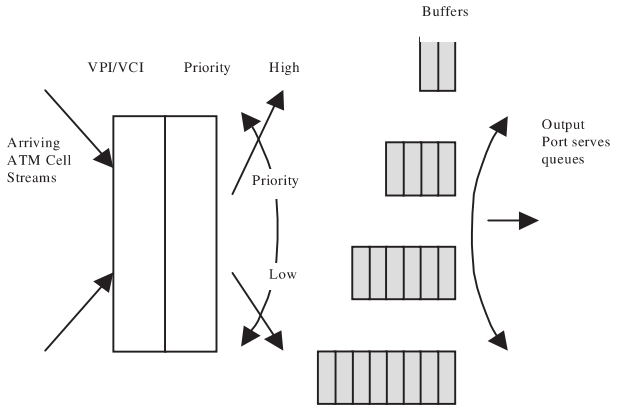


Fig. 1. ATM switch port

In the example shown on figure 1 the priority queuing function operates on the output port of an ATM switch. The switch takes arriving cell streams from multiple input ports, looks up an internal priority value. It directs the cell to one of several corresponding queues on the output port. The output side of the ATM port serves each of the queues according to a particular scheduling algorithm.

In some algorithms the cells from the highest priority nonempty queue are served first. Then the cells in the lower priority queue are served. The process continues for each successively lower priority queue. Thus it is ensured that the highest priority buffer has the least loss, delay, and delay variation. As a result lower priority queues may experience significant delay variation.

Other algorithms spread out the variation in delay across the multiple queues. For example, a schedular could send out cells just before reaching the maximum delay variation value for cells in the higher priority queues. Thus delay variation in the lower priority queues is decreased.

V. Priority Discard

The priority order of cells can be determined by discard thresholds. On the switch port each delay priority queue may be segmented into several regions via the thresholds. Arriving cell streams have the CLP bit. The switch port consults a lookup table to determine the Queue Loss Priority with a value of either high (H) or low (L). For example, when we use four thresholds they determine the priority order for cell discard based upon buffer occupacy as follows. Starting from right to left the first cells to be discarded are with Queue Loss Priority – QLP=L, CLP=1. Next come for QLP=H, CLP=1 cells, following by QLP=L, CLP=0 cells and finally cells with QLP=H, CLP=0 are discarded only if the entire buffer is full. Other choices of discard thresholds are also possible.

VI. Results and Conclusion

Traffic shaping phenomena is simulated on computer by Visual Basic program. Different priority schema is applied for different services as well as its influence on the probability density function (pdf) of cell delay. The main results from this simulation are an investigation of the self'similarity of the traffic, shaping effect on voice, data, and mixed voice and data traffic. On Fig. 2 the probability density function of the cell delay for 25 data sources at the entrance of the ATM multiplex is shown.

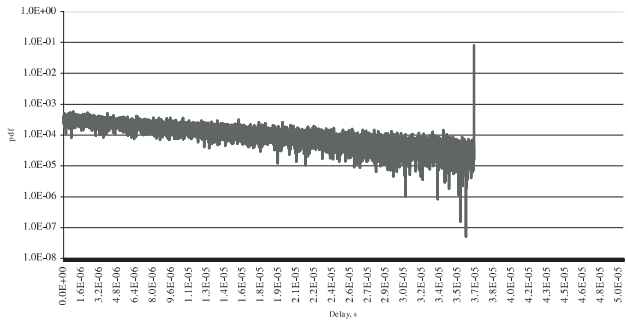


Fig. 2. Pdf function of the delay distribution of 25 data sources before ATM multiplex

Figure 3 shows the pdf function of the same traffic mixture after pure FIFO queue. The shaping is very limited. Figure 4 represents the pdf of the delay of the source with the highest

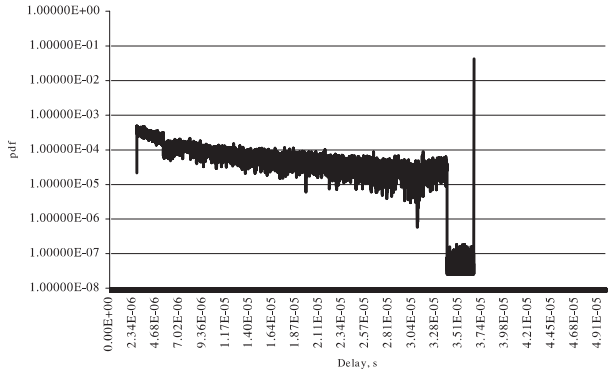


Fig. 3. Pdf function of the delay distribution of 25 data sources after ATM multiplex

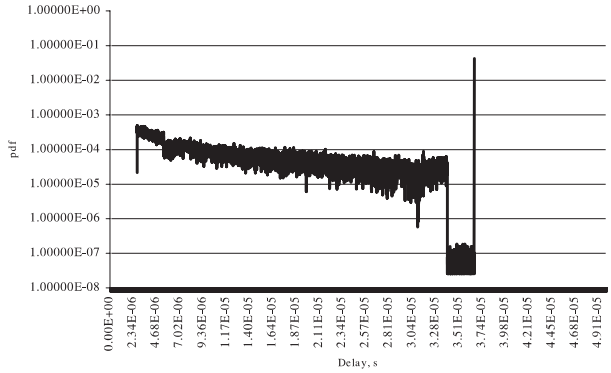


Fig. 4. Pdf function of the delay distribution of the first data source after ATM multiplex

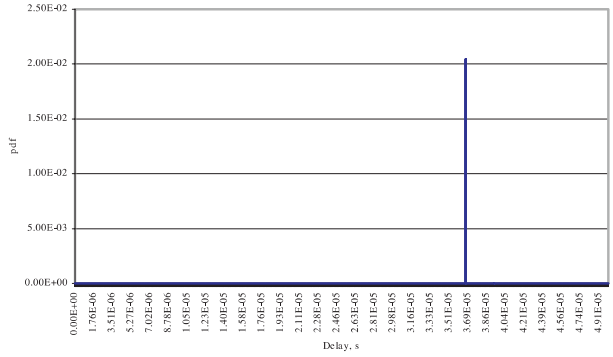


Fig. 5. Pdf function of the delay distribution of the last data source after ATM multiplex

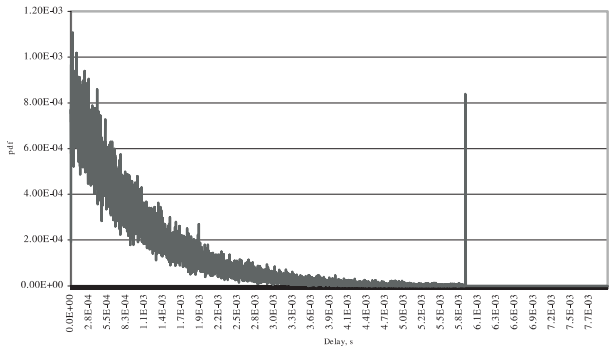


Fig. 6. Pdf function of the delay distribution of the 25 voice sources before ATM multiplex

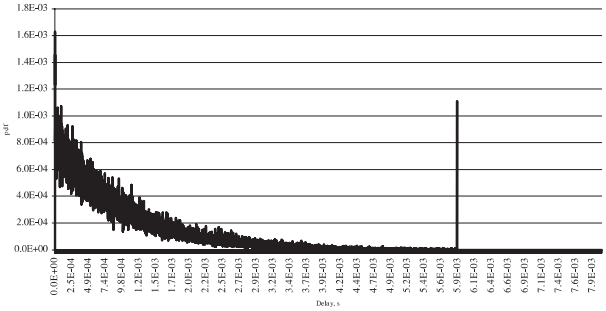


Fig. 7. Pdf function of the delay distribution of the 25 voice sources after ATM multiplex

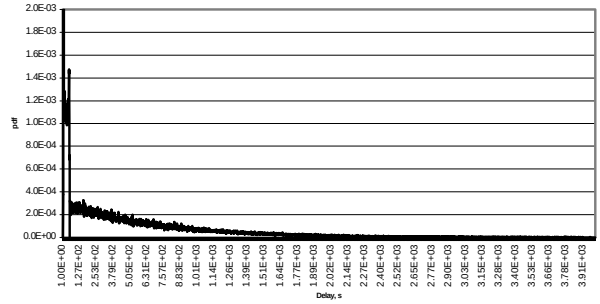


Fig. 8. Pdf function of the delay distribution of the 1 data and 50 voice sources after ATM multiplex

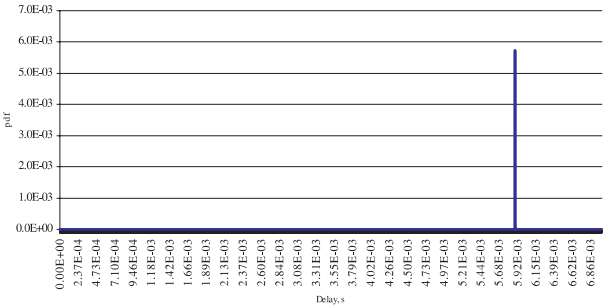


Fig. 9. Pdf function of the delay distribution of the first voice source among 1 data and 50 voice sources after ATM multiplex

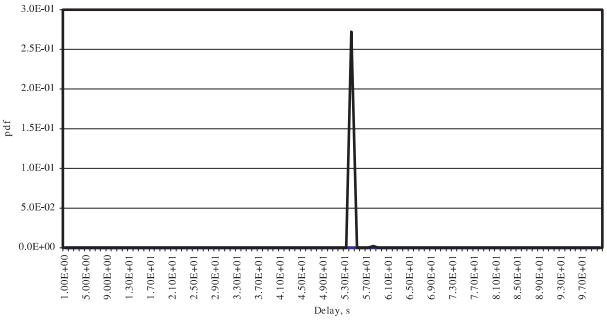


Fig. 10. Pdf function of the delay distribution of the data source among 1 data and 50 voice sources after ATM multiplex

priority as well as Figure 5 shows the pdf of the source with the lowest priority of the same traffic sample.

Pdf of the delay before and after the ATM multiplex of the sample of 25 voice sources is represented on figures 6 and

7. The peak in the figures is due to the packetization delay. Figure 8 shows the mixture of the typical office traffic of 50 voice and 1 LAN emulation data sources. Voice traffic has the highest priority.

Figure 9 and 10 represents the pdf of the delay of the first (with highest priority) voice source and the data (with the lowest priority) source.

The authors continue to work on the problem trying to estimate cell delay variation and its derivatives. The problem is also referred to pure packet switching networks like IP based networks.

References

[1] Collivignarelli, M., A. Daniele, G. Gallassi, F. Rossi, G. Valsecchi, L. Verri, System and Performance Design of the ATM Node UT-XC, IEEE Selected Areas On Communication, 1994.

[2] Kadrev, V., R.Goleva, Analyses of ATM Multiplex Behavior Influenced by Different Types of Traffic Sources, Proceedings from Telecom'98, Varna, Bulgaria, 1998.

[3] Goleva, R., Analyses of Traffic Shaping and its Influence on Quality of Service in BISDN, Proceedings from Telecom'2001, Varna, Bulgaria, 2001 (on publishing)

[4] Kadrev, V., R.Goleva, Analyses of Cell Servicing in ATM Multiplex with Deterministic Traffic Sources, Proceedings of ITC Sponsored St. Petersburg Regional International Teletraffic Seminar "Teletraffic Theory as a Base for QoS: Monitoring, Evaluation, Decisions", LONIIS, Saint- Petersburg, June-1 -7, 1998, pp. 377-388.

[5] Kadrev, V., R.Goleva, Cell Delay and Cell Loss Ratio with Different Type of Traffic Sources in ATM Multiplex, Proceedings from Telecom'98, Varna, Bulgaria, 1998.

[6] Kim, Y. C., Dong Eun Lee, Bong Ju Lee, Dynamic Channel Reservation Based on Mobility in Wireless ATM Networks, IEEE Communications Magazine, Nov. 1999.

[7] McDysan, D., ATM Theory and Application, McGraw-Hill, 1999.

[8] Information Technologies and Sciences, COST 224, Performance Evaluation and Design of Multiservice Networks, Edited by J. W. Roberts, Oct. 1991.

Traffic and Performance Analysis of Personal Wireless Communication Networks

Toni Janevski[1], Tomislav Kartalov[2], Boris Spasenovski[3]

Abstract – **Analysis of wireless networks are performed in this paper. Using simulation techniques, we obtain the dependence of performance parameters, such as call blocking and call dropping probability, from traffic and mobility parameters, such as cell capacity, traffic intensity, number and distribution of subscribers, mobility pattern and guard channels.**

Keywords – **Call blocking probability, call dropping probability, handovers per call**

I. Introduction

Cellular mobile networks are characterized with different traffic characteristics than wired networks. This is mainly due to cellular structure of the network, in which the coverage area is divided in smaller areas called cells. In such case, during a single ongoing call a subscriber is allowed to handover from its current cell to another neighboring cell. In the handover process the user is releasing the channel in the old cell and occupies another channel in the target cell. However, if all channels in all target cells are busy then the call is dropped. On the other side, new calls can also be blocked if there are no available channels in the serving cell at the call initiation.

In our analysis we consider channel-based allocation policy in the mobile network. Our aim is to analyze the Grade of Service (GoS) of the mobile network under various mobility and traffic scenarios. Mobility pattern of the users is very important in such analysis, because it directly influences handover intensity and hence the GoS. For modeling the mobility we use two-state mobility model, with a stationary state and mobile state. Mobility modeling in personal communication networks is investigated in details in [1]. Furthermore, overwhelming traffic analysis of cellular mobile networks with single-state mobility model can be found in [2]. Application of well-known loss formulas (e.g., Erlang loss formula) to cellular networks is performed in [3,4].

However, different users show different mobility behavior. So, with two-state mobility model we are targeting two main types of users: users at office or at home (stationary-state), and users in cars or trains (mobile-state).

[1]Toni Janevski is with the Faculty of Electrical Engineering, University "Sv. Kiril i Metodij", Karpos 2 bb, 1000 Skopje, Macedonia, E-mail: tonij@cerera.etf.ukim.edu.mk.

[2]Tomislav Kartalov is with the Faculty of Electrical Engineering, University "Sv. Kiril i Metodij", Karpos 2 bb, 1000 Skopje, Macedonia.

[3]Boris Spasenovski is with the Faculty of Electrical Engineering, University "Sv. Kiril i Metodij", Karpos 2 bb, 1000 Skopje, Macedonia, E-mail: boriss@cerera.etf.ukim.edu.mk.

In this paper, we analyze the dependence of call dropping probability, call-blocking probability, and carried traffic, from variations in network, traffic, or subscriber parameters. In the analysis we consider channel allocation scheme with prioritized handover.

The paper is organized as follows. In next section we define mobility and traffic models used in the analysis. Simulation results are presented in Section 3. Finally, Section 4 concludes the paper.

II. Mobility and Traffic Models

We created a simulation environment in Matlab. Here, we briefly report on our simulation model.

The coverage area of the mobile network is divided into hexagonal cells. Each cell is further divided into elements to be able to track and locate subscribers in the network. Every cell is assigned a certain number of channels. The subscribers can be served only by one cell at a moment that is the current cell.

We define a users (i.e. subscribers) matrix. That is a look up table for all subscribers in the wireless network, which contains the most important data for each subscriber, such as: his current position (location in the area), his motion status (whether he is a stationary one or a moving subscriber), his speed, current cell, direction of his movement, and number of handovers that he has performed. Speed of each subscriber in mobile state is modeled with normal (Gaussian) distribution truncated at 0, given by:

$$P_{\text{speed}}(x) = k\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-v)^2}{2\sigma^2}}, \quad x \geq 0 \qquad (1)$$

where v is the average speed, and σ^2 is the variance.

Position of each subscriber in the users matrix is represented with 2 coordinates (x and y). Direction of subscriber movement is defined with 2 coefficients, one for the movement in x-direction and the other for y-direction. Coefficients for subscriber elementary movements are randomly obtained, but combination of both zeros is prevented, because that subscriber would be stationary subscriber. So, subscriber movement is defined in the following manner: in every step of the simulation, each subscriber is allowed to move in either x- and y-directions or in both. The length of the trajectory is dependent upon the user speed. However, each user may be in one of two possible states: moving or stationary.

In next step, one should define traffic matrix, which consists of traffic data for all subscribers during the simulation. The traffic matrix has number of rows same as number of

subscribers in the area, and number of columns same as number time slots (steps) in the simulation. For each subscriber in each step corresponds exactly one element of the traffic matrix. For example, if that element is "1", the subscriber is using the network (e.g., he is talking on the phone) and he occupies one channel in his cell; if that element is "0", the subscriber is inactive.

We consider primarily voice traffic in a mobile network. Call arrival process for voice traffic is usually modeled with Poisson distribution [4]. Furthermore, call duration time for voice connections are traditionally modeled with exponential distribution.

The last preparation step is defining the cell matrix, which contains information for each cell in the area. For each cell we assign the total number of channels in that cell (its hard capacity), as well as number of channels reserved for handovers only (i.e. guard channels). The aim of guard channels (which number is always small amount of total channels) is to obtain higher reliability of ongoing calls compared to the new ones, because blocking of a new call is less sensitive to the user than dropping of an ongoing call. The call is dropped when there are no free channels in the target cell at handover. However, number of guard channels should be small compared to the cell capacity. In our model, we use channel policy "blocked calls cleared", that is, an already blocked new call or handover is cleared from the system.

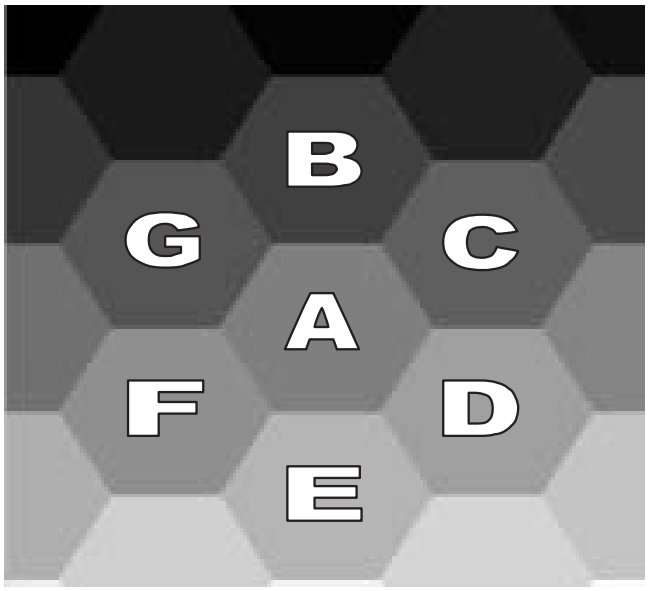


Fig. 1. Hexagonal cell structure

Finally, we may state the main assumptions used in the simulation model:

- Each subscriber in every time slot can be in one of two mobility states: in stationary state with probability P_S, or in mobile state with probability $1 - P_S$.
- Direction and speed for each subscriber are initially set at the start of the simulation, and latter they change randomly in each simulation step.
- We assume equilibrium of the subscribers in the whole coverage area of the mobile network.

III. Simulation Results

We performed several simulation experiments to obtain performances of mobile networks with guard channels and two-state mobility model. With the given input data we have the network structure shown in Figure 1. We consider centered cell A, and six neighboring cells (i.e. cells B to G).

However, one should remember that the traffic parameters in the simulations refer to the "busy hour". So, if we obtain certain call blocking probability in the "busy hour", it is straightforward to conclude that blocking probability will be smaller in the rest of the time.

For the first simulation scenario we use the following mobility and traffic parameters: average number of subscribers per cell $= 80$; probability of stationary subscribers $= 0.5$; average number of calls per subscriber per one hour $= 3$; mean call duration $= 120$ seconds (120 time slots); total number of channels in one cell $= 12$; number of "guard channels" $= 2$.

The results of this simulation are presented in the Figure 2, which shows the dependence of handovers per call from the user velocity and standard deviation σ.

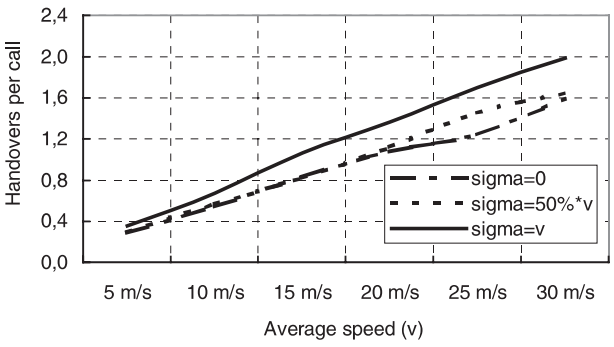


Fig. 2. Dependence of number of handovers per call from average subscriber speed, and standard deviation sigma

The average number of handovers per call increases with the speed of subscribers, showing a quite good linearity. Also, average number of handovers per call slightly increases with the standard deviation of the speed. This can be explained with higher standard deviation of the speed, when the probability of appearance of subscribers with high speeds is higher, which are increasing the average number of handovers. On the other side, we also have higher probability of appearance of subscribers with low speeds, but they cannot generate less than zero handovers in their calls, so their influence is insufficient to compensate that of the users with higher speeds.

In the second simulation experiment we investigate the GoS as a function of the average number of subscribers per cell. In this case, we use the following parameters: average subscriber speed $= 10$ m/s; standard deviation of average subscriber speed $= 2$ m/s; average number of calls per subscriber per one hour $= 3$; mean call duration $= 120$ seconds (120 time slots); total number of channels in one cell $= 12$; number of "guard channels" $= 2$.

The results of this simulation are presented in Figures 3-5; in which we show dependence of call dropping, call blocking

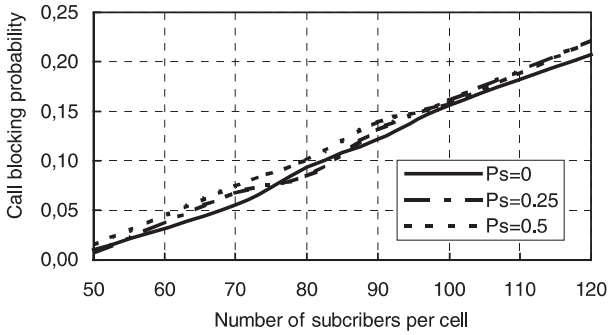


Fig. 3. Dependence of call blocking probability from number of subscribers per cell, and stationary probability P_S

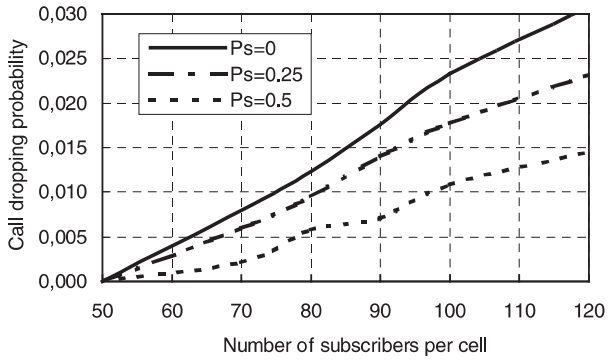


Fig. 4. Dependence of call dropping probability from number of subscribers per cell, and stationary probability P_S

probability, and carried traffic, from the average number of users per cell and probability of stationary user P_S.

As it is shown in Figures 3 and 4, both call blocking and call dropping probabilities increases with the number of subscribers per cell, but only call dropping probability shows significant dependence from stationary probability P_S. This result can be explained by the nature of call dropping, that is, it occurs at handovers. So, lower P_S results in more users in mobile state, leading to higher handover intensity per user and hence higher dropping probability. In the area in which all subscribers are stationary, there are no handovers and call-dropping probability equals zero.

According to our previous discussion on user sensitivity to call blocking and call dropping events, we have deliberately chosen system parameters (i.e. cell capacity, number of guard channels etc.) to have dropping probability for one order of magnitude smaller than call blocking probability. Figure 5 shows the carried traffic, which is higher for larger number of subscribers. In this case carried traffic does not depend upon the mobility parameters, because call-dropping probability is many times smaller then new call blocking probability.

We also examined the behaviour of the GoS at different traffic parameters for the cellular access network. Here, the following parameters are common in this simulation: average subscriber speed $= 10$ m/s; standard deviation of the subscriber speed $= 2$ m/s; average number of subscribers per cell $= 80$; probability of stationary subscribers $= 0.5$; total number of channels in one cell $= 12$; number of "guard channels" $= 2$.

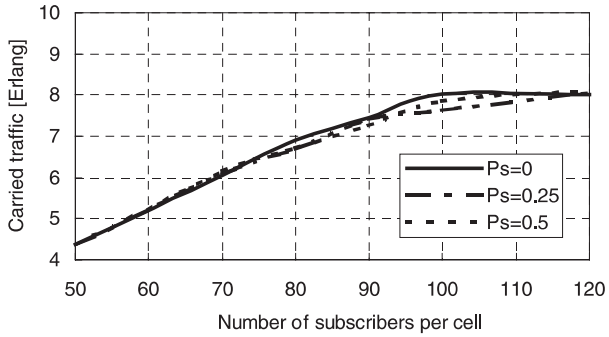


Fig. 5. Dependence of carried traffic from number of subscribers per cell, and stationary probability P_S

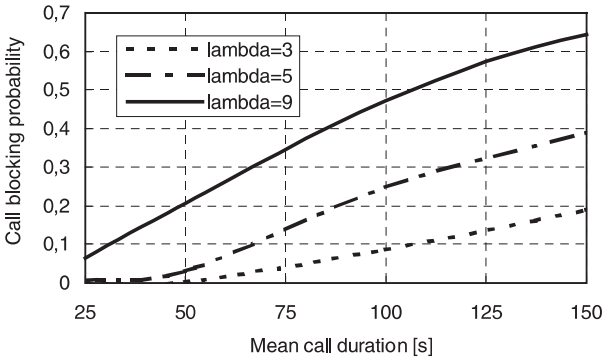


Fig. 6. Call blocking probability vs. mean call duration and average number of calls per subscriber per hour λ

The results of this simulation are presented in Figures 6-8, where are shown call blocking probability, call dropping probability, and carried traffic, versus mean call duration (t), and average number of calls per hour (lambda) respectively.

In this simulation call blocking probability (Figure 6) and call dropping probability (Figure 7) show similar dependence upon both input parameters: mean call duration and average number of calls per user per hour. This is a consequence of the fact that increasing each of the input parameters causes increasing in the offered traffic to the network. However, traffic increase results in higher losses as well.

If we analyze the results shown in Figure 8, we can observe some kind of saturation in the carried traffic that is a result of exploiting the whole cell capacity. When all channels are busy in a given cell, arrival calls are rejected either new calls or handovers. As expected, higher call arrival intensity leads to faster saturation of the carried traffic.

Finally, we analyze the cell capacity in a cellular network with prioritized handovers. In these simulations the following set of parameters is used: average number of subscribers per cell 80; average subscriber speed 10 m/s; standard deviation of subscriber speed 2 m/s; probability of stationary subscribers 0.5; average number of calls per subscriber per one hour 3; and 150 seconds mean call duration (150 time slots). The results of these simulations are presented in Figures 9-10, which show the dependences of call dropping probability, and call blocking probability, upon cell capacity (c) and number of guard channels (th).

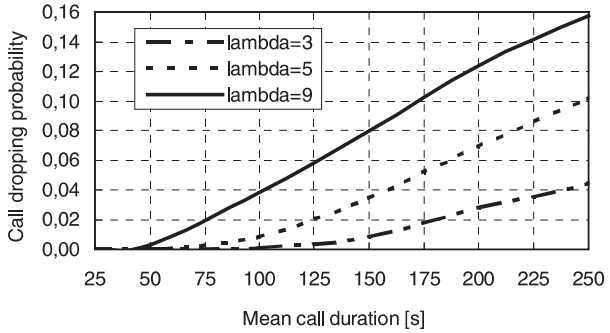


Fig. 7. Call dropping probability vs. mean call duration and average number of calls per subscriber per hour λ

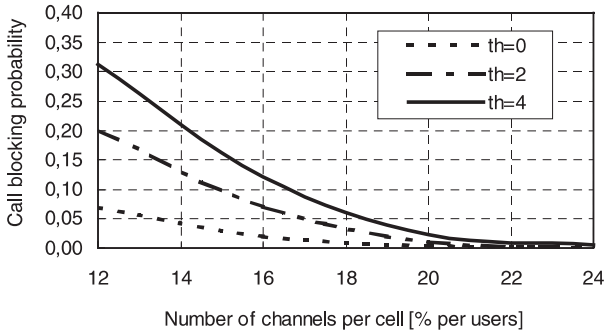


Fig. 9. Call blocking probability vs. number of channels per cell, for different numbers of guard channels "th"

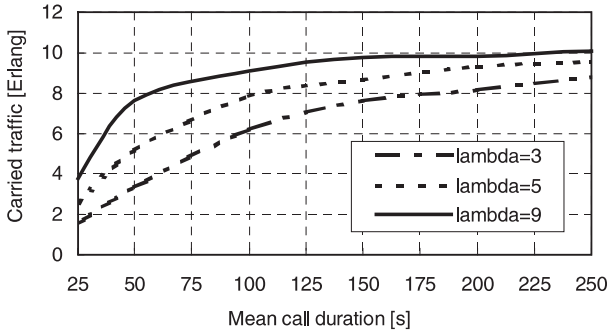


Fig. 8. Carried traffic vs. mean call duration and average number of calls per subscriber per hour λ

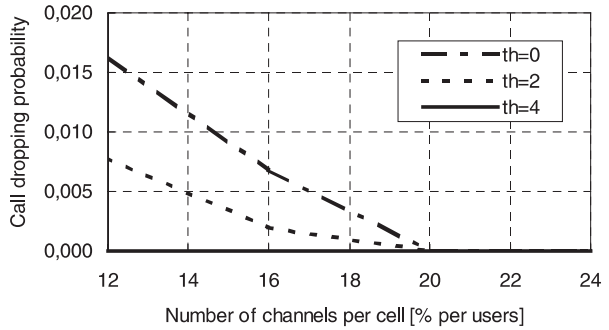


Fig. 10. Call dropping probability vs. number of channels per cell, and number of guard channels "th"

From Figures 9 and 10, it is easy to notice the influence of guard channels on blocking probabilities. However, guard channels can be allocated only to handovers calls. So, higher number of guard channels causes decrease of the call dropping probability because less new calls are accepted by the network, but at the same time it results in increase of the new call blocking probability, and vice versa. Considering the dependence of these two GoS parameters from the cell capacity, it is straightforward to conclude that higher cell capacity results in lower blocking and dropping probabilities.

Finally, it is a subject of an optimization process to derive the optimum number of guard channels under given constraints on blocking and dropping probabilities in the mobile network.

IV. Conclusion

In this paper we presented performance analyses of personal mobile communication networks with prioritized handovers. The performance of the networks was considered through its Grade of Service (GoS), defined by new call blocking and call dropping probabilities. For the purpose of the analyses we performed simulations with various mobility and traffic parameters in cellular access network. The influence of the guard channels of the network performance was investigated. Simulation analyses showed the dependence between the GoS, network capacity and mobility pattern of the users. Using the two-state Markov model for mobility, we showed that network performance is strongly dependent upon the mobility of the users (i.e. percentage of stationary and mobile users).

Higher mobility causes higher handover intensity, leading to higher dropping probability. On the other hand, it was shown that we can control the dropping probability by using prioritized handovers, i.e. dedicating small number of guard channels to handovers calls only. However, prioritized handovers cause lower carried traffic and higher new call blocking probability. Practically, we should use the offered traffic and given constraints on GoS at the design phase of the mobile network, to obtain the optimal cell capacity and number of channels dedicated to handovers.

Finally, the performance analysis in this paper is done on a call-level basis and for hard capacity in the mobile network. Future work should include packet-level analysis as well as soft capacity in the cellular access network.

References

[1] D. Lam, D.C. Cox, J. Widom, "Teletraffic Modeling for Personal Communication Services", IEEE Communication Magazine, February 1997.

[2] Liljana Gavrilovska, Toni Janevski, "Modeling Techniques for Mobile Communication Systems", Globecom 98, Sydney, Australia, November 1998.

[3] G. Haring, R. Marie, R. Puigjaner, K. Trivedi, "Loss formulas and their application to optimization for cellular networks", IEEE Transactions on Vehicular Technology, Vol. 50, No. 3, May 2001.

[4] Toni Janevski, "Traffic Analysis and Design of Wireless IP Networks", Artech House Inc., 2003.

Statistical Analysis and Modeling of the Internet Traffic

Toni Janevski[1], Dusko Temkov[2], Aleksandar Tudjarov[3]

Abstract – **In this paper we show empirical data from Internet traffic measurements. Collected measurements are analyzed for different protocols, such as TCP and UDP. We perform statistical analysis through the correlation coefficients, covariance, and self-similarity degree i.e. Hurst parameter. Our experimental studies captured traffic with Hurst parameter around 0.7-0.75, which is near half way between values of 0.5 (it is not a self-similar) and 1 (strong self-similar properties). We use Maximum Likelihood approach to fit the obtained time series to existing distributions, such as Pareto and exponential distribution, where the first one is a self-similar process and the second is not. The analysis pointed out that Internet traffic with such values for the Hurst parameter could be modeled with similar accuracy using either distribution, Pareto and exponential.**

Keywords – **Internet, traffic, analysis, Pareto distribution, exponential distribution**

I. Introduction

Internet (IP) traffic is shown to be self-similar and bursty by nature [1,2]. Aggregate IP traffic is consisted of different traffic types, which are based on different protocols or applications. One may perform classification of the Internet traffic by analyzing different traffic types [1].

IP packets have varying length and they are generated with varying data rates, which is dependent upon the transport and application protocols, as well as link capacity. Experimental studies have shown that Internet traffic can have different characteristics than traditional voice traffic [3]. Further, some authors have shown that IP traffic properties are dependent upon the buffer size, because small buffers cannot capture self-similar behavior [4].

In this paper we present statistical analysis of measured traces for main traffic types in today's IP networks: TCP traffic, UDP traffic. Our main goal is targeted to statistical analysis and modeling of the IP traffic intensity. We capture traffic traces from a live network and then examine the self-similarity of IP traffic with so-called Hurst parameter and autocorrelation function. Also, we obtain probability distribution function (PDF) of the captured traffic for each of the Internet traffic types. Furthermore, we compare Pareto and exponential distributions for modeling the Internet traffic by fitting traffic types to each of them.

[1]Toni Janevski is with the Faculty of Electrical Engineering, University "Sv. Kiril i Metodij", Karpos 2 bb, 1000 Skopje, Macedonia, E-mail: tonij@cerera.etf.ukim.edu.mk.

[2]Dusko Temkov is with Macedonian Telecommunications, MT-net department, Orce Nikolov bb, 1000 Skopje, Macedonia, E-mail: duletem@mt.net.mk.

[3]Aleksandar Tudzarov is with Komercijalna Banka, KBnet department, 1000 Skopje, Macedonia, E-mail: tucko@kbnet.com.mk.

The paper is organized as follows. Section 2 gives the background. We present IP traffic measurements and modeling in Section 3. Finally, Section 4 concludes the paper.

II. Background

In this section we provide mathematical basis for the Internet traffic statistical analysis and modeling.

IP packets arrivals at a given network node, are described mathematically as point processes, consisting of arrivals at instants in time $T_0, T_1, ..., T_n$. A mathematically equivalent description is the interarrival time process $\{A_n\}_{n=0}^{\infty}$, where the continuous function $A_n = T_n - T_{n-1}$ is the time separating the n-th arrival from the previous one. If all An are identically and independently distributed, one gets a renewal process.

We will compare the measured traffic traces with two type of distributions: exponential distribution, which is widely used in teletraffic modeling of voice traffic; and Pareto distribution, which is almost standardized for modeling the self-similar packet traffic. In the following, we describe both distributions, as well as a notion of self-similarity.

Self-similar properties of a random process are slow-decay variance and long-tailed autocorrelation. By a definition, long-tailed distribution is a distribution which complementary cumulative distribution function has the following asymptotic behavior (regardless of its shape for small values of the random variable):

$$P(T \geq t) \sim t^{-\alpha} \quad \text{for} \quad t \to \infty, \text{ and } 0 < \alpha < 2. \qquad (1)$$

The Pareto CDF is a power curve. It is given by:

$$F(t) = \begin{cases} 1 - \left(\frac{t}{k}\right)^{-\alpha-1} & \text{if } t > k \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

The probability for $t < k$ is zero; k is the minimum value which can occur in a sample set. The probability density function (PDF) is given by:

$$f(t) = \begin{cases} \alpha k \left(\frac{t}{k}\right)^{-\alpha-1} & \text{for } t > k \\ 0 & \text{otherwise.} \end{cases} \qquad (3)$$

A stationary time series $X = (Xt, t = 1, 2, 3, ...)$ is statistically exact second-order self-similar if it has the same autocorrelation function $r(k) = E[(X_t - \mu)(X_{t+k} - \mu)]$ as the series X^m for all m, where $X^m = (X_k^{(m)} : k = 1, 2, 3, ...)$ is the m-aggregated series obtained by summing the original series X over non-overlapping blocks of size m:

$$x_k^{(m)} = \frac{1}{m}(X_{km-m+1} + X_{km-m+2} + \cdots + X_{km}). \qquad (4)$$

A stationary time series X is statistically asymptotically second order self-similar if autocorrelation $r^{(m)}(k)$ of X^m, for large k, agrees asymptotically with the autocorrelation $r(k)$ of X. Self-similar traffic patterns can be detected by visual observation of traffic plots on different time scales. It looks similar over many time scales whereas short-range dependent time series look like noise after aggregating them. The degree of self-similarity can be defined using the so-called Hurst parameter H, which expresses the speed of decay of the autocorrelation function. The range of values is $0.5 < H < 1$. For $H \to 0.5$ the time series is short range dependent, while for $H \to 1$ the process becomes more and more self-similar. Since slow decaying variance and long range dependence (i.e. slow decaying autocorrelation functions) are both related to self-similarity, it is possible to determine the degree of self-similarity using either of these properties. In this work, we obtain the Hurst parameter of traffic traces by using the so-called variance-time plot, which relies on the slow decaying variance of every self-similar process. For a self-similar time series, the variance of an aggregated process decreases linearly (for large m) in log-log plots over m. The slope of the curve β can be estimated using a linear regression. Then, the Hurst parameter is determined by the following equation:

$$H = 1 - \frac{\beta}{2}. \tag{5}$$

On the other side, exponential distribution is defined with a single parameter λ, and its CDF is given by:

$$F(t) = 1 - e^{-\lambda t}. \tag{6}$$

III. IP Traffic Measurements and Modeling

For the purpose of the analysis we created traces with application CommView, scanning the Internet traffic on user Network interface toward Internet Service Provider (ISP) on client side. Actually the traces are made on Ethernet interface on local PC attached on network which has direct connection to Internet by using a leased line.

CommView text file is parsed and stored in Microsoft access base and after that it is converted in an SQL base for analysis purposes. All mathematical statistics and processing are made in Matlab. Figs. 1 to 3 illustrate the measured traffic intensity for aggregated traffic, TCP and UDP traffic. It can be noticed that TCP traffic has highest volume. This was expected due to the popularity of TCP-based applications today such as WWW. On the other side, UDP traffic intensity is respectively smaller then TCP, because in most cases it is due to DNS traffic. Each session is identified by unique source/destination address and source/destination port.

In practice we characterize statistical processes by their first two moments. Hence, we target our statistical analysis of IP traffic traces to autocorrelation function and variance. We obtain correlation coefficients from the traffic trace in the following manner: for a given measurement with N samples, $y_1, y_2, ..., y_N$, at time moments $x_1, x_2, ..., x_N$, the lag k correlation coefficient is defined as:

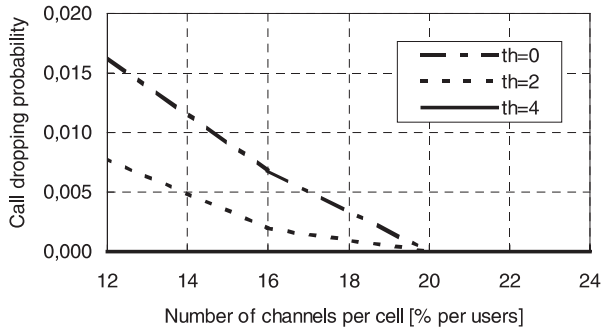


Fig. 1. Aggregated traffic trace

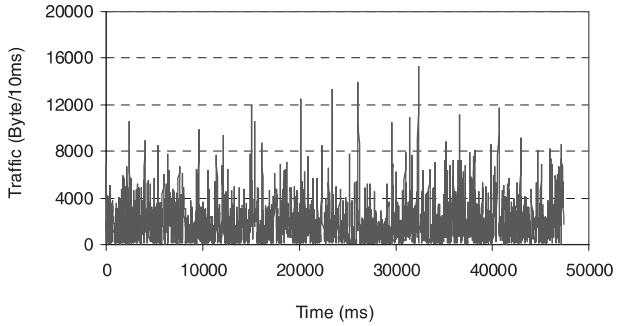


Fig. 2. TCP traffic trace

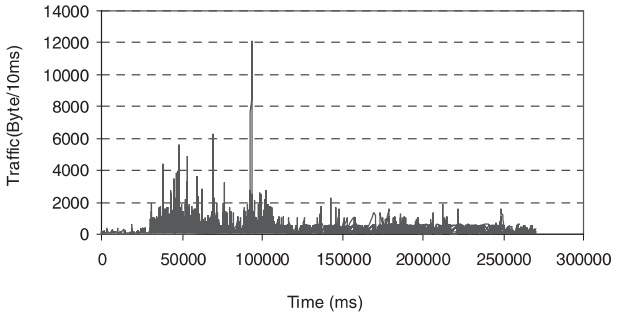


Fig. 3. UDP traffic trace

$$r_k = \frac{\sum_{i=1}^{N-k}(yi - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^{N}(y_i - \bar{y})^2}. \tag{7}$$

Autocorrelation functions of the captured traffic types are shown on Figs. 4 to 6. One can observe that TCP and aggregated traffic traces have autocorrelation function which oscillates around zero, while UDP traffic experiences long-tailed autocorrelation. It leads to stronger self-similar properties of UDP traffic compared to the TCP traffic from the measurements.

We obtain the Hurst parameter from traffic traces by us-

Table 1. Hurst parameter and slope β

Traffic	β	H
Aggregate	0.632055	0.6840
TCP	0.533801	0.7331
UDP	0.465238	0.7674

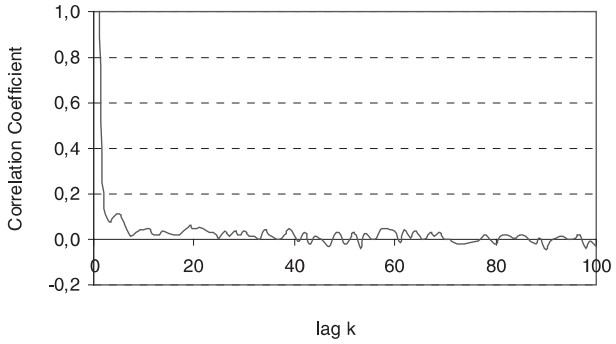


Fig. 4. Correlation coefficients for aggregated traffic

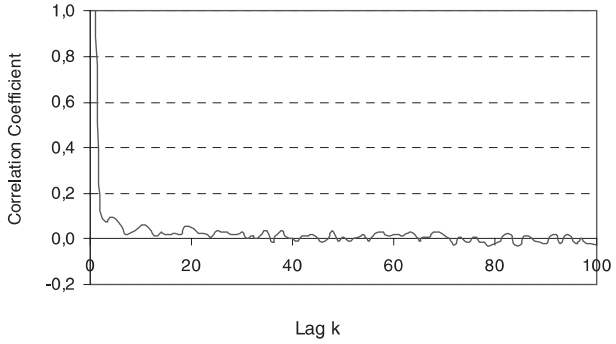


Fig. 5. Correlation coefficients for TCP traffic

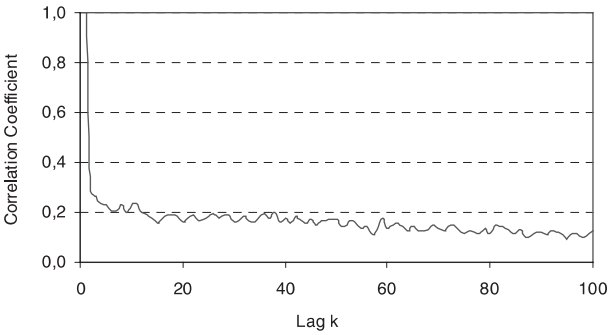


Fig. 6. Correlation coefficients for UDP traffic

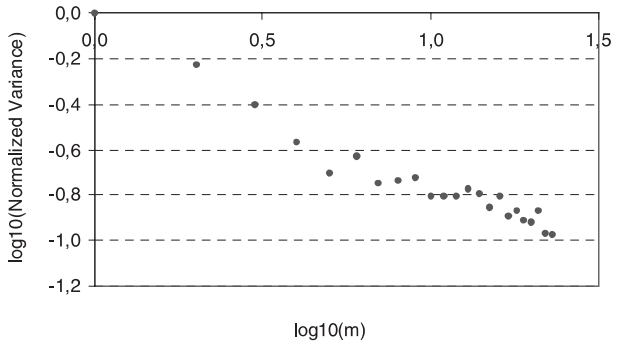


Fig. 7. Variance-Time Plot for aggregate traffic

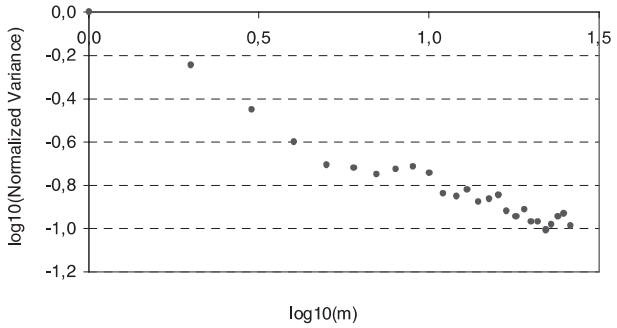


Fig. 8. Variance Time Plot for TCP traffic

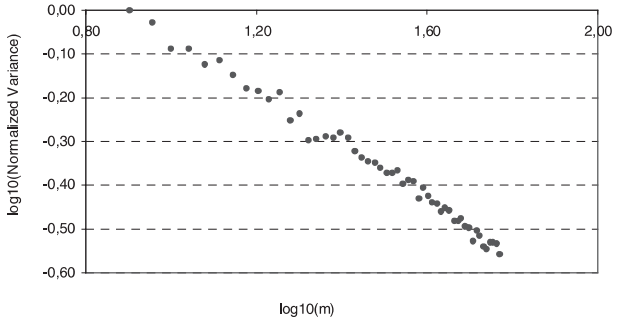


Fig. 9. Variance Time Plot for UDP traffic

ing variance time plot method. For a self-similar process, the variance of an aggregated process decreases linearly (for large m) in log-log plots over m. The slope β can be estimated using linear regression, leading to the Hurst parameter defined as above. Hurst parameter can be calculated from the slope β using Eq. (5).

Figs. 7 to 9 show variance-time plots for traffic traces, and the obtained parameters, β and H, are given in Table 1.

Further, our aim is to fit the first two moments of the measured Internet traffic traces with the two distributions. In other words, we need to obtain the optimal parameters of the Pareto distribution as well as exponential distribution that best model the measured traffic data.

We start with fitting the measured traces to Pareto distribution. With analysis of traffic traces (shown in Figs. 1-3) we obtain the value for α parameter of the Pareto distribution, using its relation with the Hurst parameter, Eq. (5). It results in $\alpha = 1.16$. Then, for $k = 60$ to 5000 we perform minimization of mean square error between the measured samples and Pareto distribution.

Further, we continue with fitting the measurements data to the exponential distribution. In this case, our aim is to find the optimal parameter λ of the exponential distribution that provides the best match between the exponential PDF and the measured traffic. For this purpose we search for $\lambda_{\min}$ that minimizes mean square error between the measured traffic PDF and the exponential PDF. These results are shown in Fig. 10-12 for aggregated, TCP and UDP traffic, respectively.

The comparison of fitting the measured data to exponential and Pareto distributions is given in Table 2. One can notice that mean square error is slightly slower for the exponential distribution than for Pareto. However, such result may be expected due to lower values of the Hurst parameter for

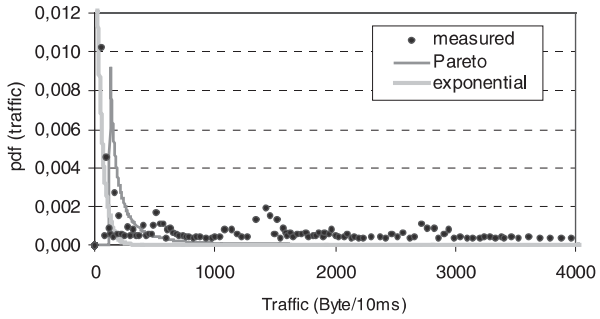


Fig. 10. Measured, Pareto and exponential PDF for aggregated traffic

Table 2. Mmean square error for EXP and PARETO

Traffic	$\lambda_{\min}$	MSE for EXP	$k_{\min}H$	MSE for Pareto
Aggregate	0.021	1.038E-05	1500	1.0998E-05
TCP	0.016	1.585E-05	1510	1.6753E-05
UDP	0.016	1.429E-04	550	1.4729E-04

the measured traffic ($H \approx 0.70$-0.75). So, when $H \rightarrow 1$ we should use Pareto for modeling bursty data traffic, while when $H \rightarrow 0.5$ then exponential distribution should be the best model. However, if we are in the middle of the range for the Hurst parameter, then according to our results presented in this paper we may equally choose either, exponential or Pareto distribution, for modeling the Internet traffic.

IV. Conclusions and Future Work

We performed statistical analysis of the captured Internet traffic from a real network. We analyzed the traffic per protocol, i.e. TCP and UDP, as well as aggregated traffic.

Then, we used obtained statistics to fit the measured traffic data to the exponential and to the Pareto distribution. The results showed that for traffic with Hurst parameter in the range 0.7-0.75 we can use each of the distributions for modeling the Internet traffic with equal accuracy.

Possible future extension to this work is to use a linear combination of the two distributions for modeling the Internet traffic.

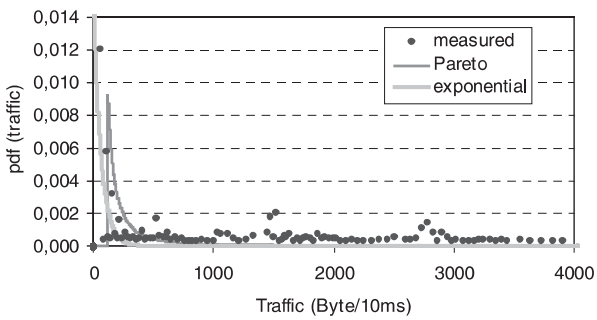


Fig. 11. Measured, Pareto and exponential PDF for TCP traffic

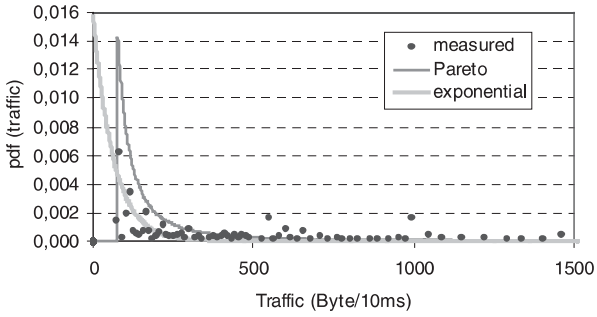


Fig. 12. Measured, Pareto and exponential PDF for UDP traffic

References

[1] Toni Janevski, *Traffic Analysis and Design of Wireless IP Networks*, Artech House Inc., 2003.

[2] Vern Paxson and Sally Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, June 1995, pp.226-244.

[3] Walter Willinger and Vern Paxson, "Where Mathematics meets the Internet", *Notes of the American Mathematical Society*, Vol.45, No.8, August 1998, pp.961-970.

[4] F. Huebner, D. Liu and J.M. Fernandez, "Queuing Performance Comparison of Traffic Models for Internet Traffic", *GLOBECOM'98*, Sydney, Australia, November 8-12, 1998, pp.1931-1936.

Williamson-Hadamard Transforms: Design and Fast Algorithms

Sos Agaian[1], Hakob Sarukhanyan[2], Karen Egiazarian[3] and Jaakko Astola[3]

Abstract – **- In this paper, algorithms for fast computation of a special type of Hadamard transforms, namely 4tpoint (t is an order of so-called Williamson type matrices) Williamson-Hadamard transforms, are presented. These transforms are based on Williamson's construction of Hadamard matrices. Comparisons revealing the efficiency of the proposed algorithms with respect to the known ones are given. Also, of numerical examples are presented.**

Keywords – **Hadamard matrices, Hadamard transforms, fast algorithms**

I. Introduction

In the past decade fast orthogonal transforms have been widely used in many areas, such as data compression, pattern recognition and image reconstruction, interpolation, linear filtering, spectral analysis, watermarking, cryptography and communication. The computation of unitary transforms is a complicated and time consuming task. However, it would not be possible to use the orthogonal transforms in signal and image processing applications without effective algorithms is calculate them. An important question in many applications is how to achieve the highest computation efficiency of the Then discrete orthogonal transforms (DOTs). Among DOTs a spe cial role plays a class of Hadamard transforms based on the Hadamard matrices [1,-13,19-21] and they do not require any multiplication operation in their computation.

In this paper we have utilized Williamson's construction of Hadamard matrices in order to develop efficient computational algorithms of a special type of Hadamard transforms, called Williamson-Hadamard transforms.

The paper is organized as follows. Section 2 described the Hadamard matrix construction from Williamson matrices. Sections 3 and 4 present the block representation of Williamson-Hadamard matrices and fast Williamson-Hadamard block transform algorithm. In Section 5 Williamson-Hadamard transform algorithm on add/shift architecture is developed. Section 6 and 7 present a fast Williamson-Hadamard transform algorithms based on the multiplicative theorems. Section 8 gives the complexities of developed algorithms, and also a comparative estimate, revealing the efficiency of the proposed algorithms with respect to the known once.

[1]Sos Agaian is with the University of Texas at San Antonio, San Antonio, USA, sagaian@utsa.edu

[2]Hakob Sarukhanyan is with the Institute for Informatics and Automation Problems of national academy of sciences of Armenia, Yerevan, Armenia, hakop@ipia.sci.am

[3]Karen Egiazarian and Jaakko Astola are with the Institute of Signal Processing, Tampere University of Technology, Tampere, Finland, karen@cs.tut.fi, jta@cs.tut.fi

II. How to Build Hadamard Matrices from Williamson Matrices

In this section we first give a definition of Hadamard matrices and then describe an algorithm for generation of Williamson-Hadamard matrices.

Definition 2.1: A *Hadamard matrix* H_n of order n is an orthogonal matrix consisting of elements ± 1:

$$H_n H_n^T = H_n^T H_n = nI_n,$$

where T is a transposition sign, I_n is an identity matrix of order n.

Let as briefly describe the Williamson's approach to the Hadamard matrices construction.

Theorem 2.1: (Williamson [15]) Suppose there exist four (± 1)-matrices A, B, C, D of order n satisfying

$$\begin{aligned} &PQ^T = QP^T, \quad P,Q \in \{A,B,C,D\}, \\ &AA^T + BB^T + CC^T + DD^T = 4nI_n. \end{aligned} \tag{1}$$

Then

$$W_{4n} = \begin{pmatrix} A & B & C & D \\ -B & A & -D & C \\ -C & D & A & -B \\ -D & -C & B & A \end{pmatrix} \tag{2}$$

is Hadamard matrix of order $4n$.

The matrices A, B, C, D with properties (1) are called *Williamson matrices*. The matrix (2), is called the *Williamson-Hadamard matrix*.

If A, B, C, D are cyclic symmetric (± 1)-matrices of order n, then the first relation of (1) is satisfied automatically and the second condition becomes

$$A^2 + B^2 + C^2 + D^2 = 4nI_n.$$

The first rows of the Williamson type cyclic and symmetric matrices A, B, C, D of order n, $n = 3, 5, ..., 25$ can be found in [2,11,16]. For the more complete list of Williamson matrices see in J. Seberry's web page [17].

Note that any cyclic symmetric matrix A can be represented as $A = \sum_{i=0}^{n-1} a_i U^i$, where U is a cyclic matrix of order n with the first row $(0, 1, ..., 0)$, and $U^{n+i} = U^i$, $a_i = a_{n-i}$, +, for $i = 1, 2, ..., n-1$.

Below we give an algorithm of the WilliamsonHadamard matrices generation.

Algorithm 2.1: **Hadamard matrix generation via cyclic symmetric Williamson matrices.**

Input: the vectors $(a_0, a_1, ..., a_{n-1})$, $(b_0, b_1, ..., b_{n-1})$, $(c_0, c_1, ..., c_{n-1})$ and $(d_0, d_1, ..., d_{n-1})$.

Step 1. Construct matrices A, B, C, D by

$$A = \sum_{i=0}^{n-1} a_i U^i, \qquad B = \sum_{i=0}^{n-1} b_i U^i,$$
$$C = \sum_{i=0}^{n-1} c_i U^i, \qquad D = \sum_{i=0}^{n-1} d_i U^i.$$

Step 2. Substitute matrices A, B, C, D into the array (2).

Output: Williamson-Hadamard matrix W_{4n}.

Example 2.1: The following matrices are the Williamson type matrices of order 3.

$$A = \begin{pmatrix} + & + & + \\ + & + & + \\ + & + & + \end{pmatrix}, \quad B = C = D = \begin{pmatrix} + & - & - \\ - & + & - \\ - & - & + \end{pmatrix}.$$

Substituting these matrices in (2) we obtain a Williamson-Hadamard matrix W_{12} of order 12

III. Block Representation of Williamson-Hadamard Matrices

In this section we present an approach of block Hadamard matrices construction equivalent to the Williamson-Hadamard matrices. This approach is useful for designing fast transforms algorithms.

We begin with an example. Let (a_0, a_1, a_1), (b_0, b_1, b_1), (c_0, c_1, c_1) and (d_0, d_1, d_1) be the first rows of Williamson type cyclic symmetric matrices of order 3. Using Algorithm 2.1, we can construct the following matrix of order 12:

$$\begin{pmatrix} \begin{pmatrix} a_0 & a_1 & a_1 \\ a_1 & a_0 & a_1 \\ a_1 & a_1 & a_0 \end{pmatrix} & \begin{pmatrix} b_0 & b_1 & b_1 \\ b_1 & b_0 & b_1 \\ b_1 & b_1 & b_0 \end{pmatrix} & \begin{pmatrix} c_0 & c_1 & c_1 \\ c_1 & c_0 & c_1 \\ c_1 & c_1 & c_0 \end{pmatrix} & \begin{pmatrix} d_0 & d_1 & d_1 \\ d_1 & d_0 & d_1 \\ d_1 & d_1 & d_0 \end{pmatrix} \\ -\begin{pmatrix} b_0 & b_1 & b_1 \\ b_1 & b_0 & b_1 \\ b_1 & b_1 & b_0 \end{pmatrix} & \begin{pmatrix} a_0 & a_1 & a_1 \\ a_1 & a_0 & a_1 \\ a_1 & a_1 & a_0 \end{pmatrix} & -\begin{pmatrix} d_0 & d_1 & d_1 \\ d_1 & d_0 & d_1 \\ d_1 & d_1 & d_0 \end{pmatrix} & \begin{pmatrix} c_0 & c_1 & c_1 \\ c_1 & c_0 & c_1 \\ c_1 & c_1 & c_0 \end{pmatrix} \\ -\begin{pmatrix} c_0 & c_1 & c_1 \\ c_1 & c_0 & c_1 \\ c_1 & c_1 & c_0 \end{pmatrix} & \begin{pmatrix} d_0 & d_1 & d_1 \\ d_1 & d_0 & d_1 \\ d_1 & d_1 & d_0 \end{pmatrix} & \begin{pmatrix} a_0 & a_1 & a_1 \\ a_1 & a_0 & a_1 \\ a_1 & a_1 & a_0 \end{pmatrix} & -\begin{pmatrix} b_0 & b_1 & b_1 \\ b_1 & b_0 & b_1 \\ b_1 & b_1 & b_0 \end{pmatrix} \\ -\begin{pmatrix} d_0 & d_1 & d_1 \\ d_1 & d_0 & d_1 \\ d_1 & d_1 & d_0 \end{pmatrix} & -\begin{pmatrix} c_0 & c_1 & c_1 \\ c_1 & c_0 & c_1 \\ c_1 & c_1 & c_0 \end{pmatrix} & \begin{pmatrix} b_0 & b_1 & b_1 \\ b_1 & b_0 & b_1 \\ b_1 & b_1 & b_0 \end{pmatrix} & \begin{pmatrix} a_0 & a_1 & a_1 \\ a_1 & a_0 & a_1 \\ a_1 & a_1 & a_0 \end{pmatrix} \end{pmatrix} \quad (3)$$

Now we want to use this matrix in order to make an equivalent block cyclic matrix. The first block P_0 we form as follows: a) from the first row of above matrix the first, 4-th, 7-th, and 10-th elements (a_0, b_0, c_0, d_0) and make the first row of block P_0, b) from the 4-th row of above matrix the first, 4-th, 7-th, and 10-th elements $(-b_0, a_0, -d_0, c_0)$ we make the second row of block P_0, and so on. Hence, we obtain

$$P_0 = \begin{pmatrix} a_0 & b_0 & c_0 & d_0 \\ -b_0 & a_0 & -d_0 & c_0 \\ -c_0 & d_0 & a_0 & -b_0 \\ -d_0 & -c_0 & b_0 & a_0 \end{pmatrix}. \quad (4)$$

The second (and third) block P_1 we form as follows: a) from the first row of above matrix the second, 5-th, 8-th and 11-th elements we make the first row (a_1, b_1, c_1, d_1) of block P_1, b) from the 4-th row of above matrix the second, 5-th, 8-th, and 11-th elements $(-b_1, a_1, -d_1, c_1)$ we make the second row of block P_1, and so on. Hence, we obtain

$$P_0 = \begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ -b_1 & a_1 & -d_1 & c_1 \\ -c_1 & d_1 & a_1 & -b_1 \\ -d_1 & -c_1 & b_1 & a_1 \end{pmatrix}. \quad (5)$$

From (3), (4) and (5) we obtain

$$[BW]_{12} = \begin{pmatrix} P_0 & P_1 & P_1 \\ P_1 & P_0 & P_1 \\ P_1 & P_1 & P_0 \end{pmatrix} \quad (6)$$

which is a block cyclic block symmetric Hadamard matrix.

Using the properties of the Kronecker product, we may rewrite (6) as $[BW]_{12} = P_0 \otimes I_3 + P_1 \otimes U + P_1 \otimes U^2$.

In general, any Williamson-Hadamard matrix of order $4n$ can be presented as

$$W_{4n} = \sum_{i=0}^{n-1} Q_i \otimes U^i,$$
$$Q_i(a_i, b_i, c_i, d_i) = \begin{pmatrix} a_i & b_i & c_i & d_i \\ -b_i & a_i & -d_i & c_i \\ -c_i & d_i & a_i & -b_i \\ -d_i & -c_i & b_i & a_i \end{pmatrix}, \quad (7)$$

where $Q_i = Q_{n-i}$, $a_i, b_i, c_i, d_i = \pm 1$ and $\otimes$ is a sign of the Kronecker product [11].

The Hadamard matrices of the form (7) are called *block-cyclic block-symmetric Hadamard matrices* [2]. The Williamson-Hadamard matrix W_{12}(see previous section) can be represented as a block-cyclic block-symmetric matrix:

$$[BW]_{12} = \begin{pmatrix} ++++ & +--- & +--- \\ -+-+ & +++- & +++- \\ -++- & +-++ & +-++ \\ --++ & ++-+ & ++-+ \\ \\ +--- & ++++ & +--- \\ +++- & -+-+ & +++- \\ +-++ & -++- & +-++ \\ ++-+ & --++ & ++-+ \\ \\ +--- & +--- & ++++ \\ +++- & +++- & -+-+ \\ +-++ & +-++ & -++- \\ ++-+ & ++-+ & --++ \end{pmatrix}$$
$$= \begin{pmatrix} Q_0(+,+,+,+) & Q_4(+,-,-,-) & Q_4(+,-,-,-) \\ Q_4(+,-,-,-) & Q_0(+,+,+,+) & Q_4(+,-,-,-) \\ Q_4(+,-,-,-) & Q_4(+,-,-,-) & Q_0(+,+,+,+) \end{pmatrix}$$

From (7) we may see that all the blocks are Hadamard matrix of Williamson type of order 4. In [18] it was proved that cyclic symmetric Williamson-Hadamard block matrices can be constructed using only 5 different blocks such as

$$Q_0 = \begin{pmatrix} + & + & + & + \\ - & + & - & + \\ - & + & + & - \\ - & - & + & + \end{pmatrix}, \quad Q_1 = \begin{pmatrix} + & + & + & - \\ - & + & + & + \\ - & - & + & - \\ + & - & + & + \end{pmatrix},$$

$$Q_2 = \begin{pmatrix} + & + & - & + \\ - & + & - & - \\ + & + & + & - \\ - & + & + & + \end{pmatrix}, \quad Q_3 = \begin{pmatrix} + & - & + & + \\ + & + & - & + \\ - & + & + & + \\ - & - & - & + \end{pmatrix}, \quad (8)$$

$$Q_4 = \begin{pmatrix} + & - & - & - \\ + & + & + & - \\ + & - & + & + \\ + & + & - & + \end{pmatrix}$$

For example, Williamson-Hadamard block matrix $[BW]_{12}$ was constructed using matrices Q_0 and Q_4 only.

Note that when we fix first block then one needs maximum 4 blocks to design any Williamson-Hadamard block matrix and these 4 blocks is defined uniquely up to a sign. Thus, if the first row of the first block consists of an even +1, then the first rows of the others 4 blocks consist of an odd +1. And if the first row of the first block consists of an odd +1, then the first rows of the others 4 blocks consist of an even +1. The set of blocks with fixed first block with odd +1 is following $Q'_0 = Q_0(+,+,+,-)$, $Q'_1 = Q_1(+,-,-,+)$, $Q'_2 = Q_2(-,+,-,+)$, $Q'_3 = Q_3(-,-,+,+)$, $Q'_4 = Q_4(+,+,+,+)$.

IV. Fast Block Williamson-Hadamard Transforms

In this section we describe two algorithms for forward block Williamson-Hadamard transform calculation:

$$F = [BW]_{4n} f.$$

Let us split the vector-column f into n 4-dimensional vectors as $f = \sum_{i=0}^{n-1} X_i \otimes P_i$, where P_i are column-vectors of dimension n, whose i-th element is equal to 1, and the remaining elements are equal to 0, and $X_i = (f_{4i}, f_{4i+1}, f_{4i+2}, f_{4i+3})$, $i = 0, 1, ..., n-1$.

Now, using (7), we have

$$[BW]_{4n} f = \Big(\sum_{i=0}^{n-1} Q_i \otimes U^i\Big)\Big(\sum_{j=0}^{n-1} X_j \otimes P_j\Big) = \sum_{i,j=0}^{n-1} Q_i X_j \otimes U^i P_j. \quad (9)$$

We may cheek $U^i P_j = P_{n-i+j}$, $j = 0, 1, ..., n-1$, $i = j+1, ..., n-1$.

Hence, the equation (9) can be presented as

$$[BW]_{4n} f = \sum_{i,j}^{n-1} Q_i X_i \otimes U^i P_j = \sum_{j=0}^{n-1} B_j, \quad (10)$$

where $B_j = \sum_{i=0}^{n-1} Q_j X_j \otimes U^i P_j$.

From (10) we have that for performing the fast Williamson-Hadamard transform we need to calculate the spectral coefficients of the block transforms, such as $Y_i = Q_i X$. Here Q_i, $i = 0, 1, 2, 3, 4$ have the form (8), and $X = (x_0, x_1, x_2, x_3)^T$, and $Y_i = (y_i^0, y_i^1, y_i^2, y_i^3)^T$ are the input and output column-vectors, respectively.

The flow graph of the algorithm for joint computation of five 4-point Williamson-Hadamard transforms $Q_i X$, $i = 0, 1, ..., 4$ is given in Figure 1.

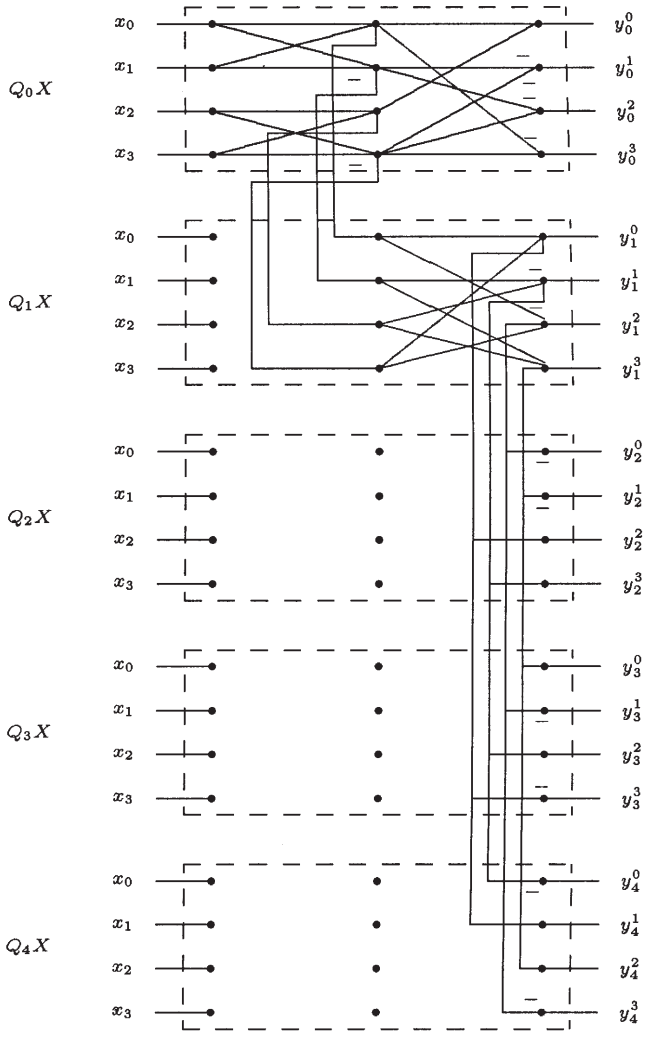


Fig. 1. Flow graph for the joint $Q_i X$ transforms, $i = 0, 1, ..., 4$

The joint computation of 4-point transforms $Q_i X$, $i = 0, 1, ..., 4$ requires only 12 addition/subtraction operations. Note that the separate calculation of Q_j, $j = 0, 1, ..., 4$ requires 40 addition/subtraction operations.

Really, from Figure 1 we can check that the transform $Q_0 X$ requires 8 addition/subtraction operations, and the transform $Q_1 X$ requires 4 addition/subtraction operations. We can see also that the joint computation of all 4-point transforms $Q_i X$, $i = 0, 1, ..., 4$ requires only 12 addition/subtraction operations.

Now we give a detailed description of 36-point block Williamson-Hadamard fast transform algorithm.

Example 4.1: **36-point fast Williamson-Hadamard transform using 396 operations.**

Input : vector-column $F_{36} = (f_i)_{i=0}^{35}$ and blocks Q_0, Q_1 and Q_2

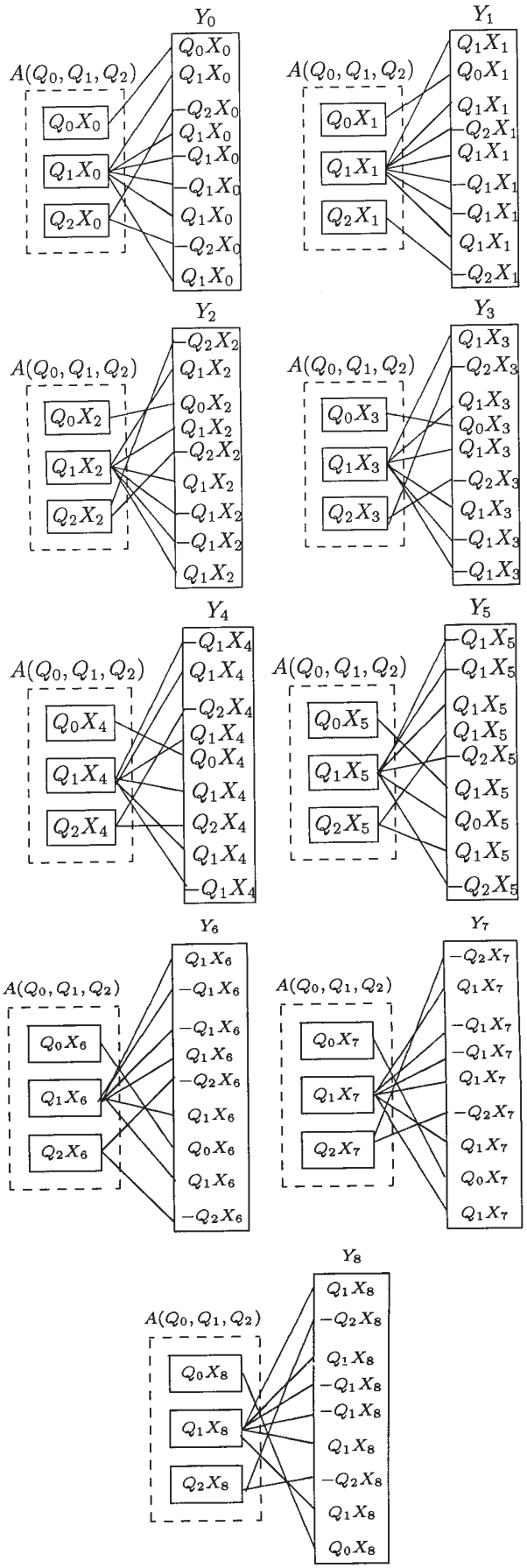


Fig. 2. Flow graphs of 36-dimensional v ectors Y_i, $i = 0, 1, ..., 8$ computation

Step 1. Split vector F_{36} into 9 parts as $F_{36}^T = (X_0^T, X_1^T, ..., X_8^T)$, where $X_i^T = (f_{4i}, f_{4i+1}, f_{4i+2}, f_{4i+3})$, $i = 0, 1, ..., 8$.

Step 2. Compute the vectors Y_i, $i = 0, 1, ..., 8$, as shown in Figure 2.

Note that the subblocks $A(Q_0, Q_1, Q_2)$in Figure 2 can be computed using the scheme in Figure 1.

Step 3. Evaluate the vector $Y = Y_0 + Y_1 + \cdots + Y_8$.

Output: transform coefficients, i.e. vector Y.

As we have seen from the fast algorithm of 4-point Williamson-Hadamard transforms (see Figure 1), the joint computation of the transforms Q_0X_i, Q_1X_i and Q_2X_i requires only 12 addition/subtraction operations. From (8) we can see that only these transforms are there in each vector Y_i. Hence, for all these vectors it is necessary to perform 108 operations. Finally, the 36-point Hadamard transform requires only 396 addition/subtraction operations, but in the case of direct computation it would require 1260 addition/subtraction operations.

Note that we have developed a fast Walsh-Hadamard transform algorithm without knowing the existence of Williamson-Hadamard matrices. We can speed up this algorithm if we would know a construction of these matrices. The first block-rows of the block-cyclic block-symmetric (BCBS) Hadamard matrices of Williamson type of order $4n$, $n = 3, 5, ..., 25$ can be found in [2,16].

V. Williamson-Hadamard Transform on Add/Shift Architectures

In this section we describe add/shift for Williamson-Hadamard transform.

Denoting by $z_1 = x_1 + x_2 + x_3$, $z_2 = z_1 - x_0$, and using (8), we can calculate $Y_i = (y_i^0, y_i^1, y_i^2, y_i^3) = Q_iX$ as:

$$
\begin{aligned}
&y_0^0 = z_1 + x_0, && y_0^1 = z_2 - 2x_2,\\
&y_0^2 = z_2 - 2x_3, && y_0^3 = z_2 - 2x_1;\\
&y_1^0 = y_0^0 - 2x_3, && y_1^1 = z_2,\\
&y_1^2 = y_0^2 - 2x_1, && y_1^3 = y_0^0 - 2x_1;\\
&y_2^0 = -y_1^2, && y_2^1 = -y_1^3,\\
&y_2^2 = y_1^0, && y_2^3 = y_1^1;\\
&y_3^0 = y_1^3, && y_3^1 = -y_1^2,\\
&y_3^2 = z_2 = -y_1^1, && y_3^3 = -y_1^0;\\
&y_4^0 = -z_2 = -y_1^1, && y_4^1 = y_1^0,\\
&y_4^2 = y_1^3, && y_4^3 = -y_1^2.
\end{aligned}
$$

It is easy to check that the joint 4-point transforms computation requires less operations than their separate computations. The separate computations of transforms Q_0X and Q_1X require 14 addition/subtraction operations and 6 one-bit shifts, but for their joint computation it is necessary only 10 addition/subtraction operations and 3 one-bit shifts.

So, using this fact, the complexity of the fast Williamson-Hadamard transform will be discussed next.

VI. Multiplicative Theorem Based Williamson-Hadamard Matrices

In this section we describe of Williamson-Hadamard matrices constructions based on the following multiplicative theorem.

Theorem 6.1: (Multiplicative Theorem [14]) Let there exist Williamson-Hadamard matrices of orders $4m$ and $4n$. Then there exist Williamson-Hadamard matrices of order $4(2m)^i n$, $i = 1, 2, ...$

Theorem 6.2: Let there exist Williamson matrices of order n and Hadamard matrix of order $4m$. Then there exists a Hadamard matrix of order $8mn$.

The prove of Theorems 6.1 and 6.2 is presented in Appendix.

Algorithm 6.1: **Generation of Williamson-Hadamard matrix of order $8mn$ from Williamson-Hadamard matrices of orders $4m$ and $4n$.**

Input : Williamson matrices A, B, C, D and A_0, B_0, C_0, D_0 of orders m and n, respectively.

Step 1 . Construct matrices X and Y as

$$X = \frac{1}{2}\begin{pmatrix} A+B & C+D \\ C+D & -A-B \end{pmatrix}, \quad Y = \frac{1}{2}\begin{pmatrix} A-B & C-D \\ -C+D & A-B \end{pmatrix}. \tag{11}$$

Step 2 . For $i = 1, 2, ..., k$ construct recursively the following matrices

$$\begin{aligned} A_i &= A_{i-1} \otimes X + B_{i-1} \otimes Y, \\ B_i &= B_{i-1} \otimes X - A_{i-1} \otimes Y, \\ C_i &= C_{i-1} \otimes X + D_{i-1} \otimes Y, \\ D_i &= D_{i-1} \otimes X - C_{i-1} \otimes Y. \end{aligned} \tag{12}$$

Step 3 . For $i = 1, 2, ..., k$ construct the Williamson-Hadamard matrix as

$$[WH]_i = \begin{pmatrix} A_i & B_i & C_i & D_i \\ -B_i & A_i & -D_i & C_i \\ -C_i & D_i & A_i & -B_i \\ -D_i & -C_i & B_i & A_i \end{pmatrix}$$

Output : Williamson-Hadamard matrices $[WH]_i$, $i = 1, 2, ..., k$.

Example 6.1: **Construction of Williamson matrices**.

Using Williamson matrices of order 3 and 5 from the Example 2.1 and (11), we obtain

a) for $n = 3$:

$$X = \begin{pmatrix} + & 0 & 0 & + & - & - \\ 0 & + & 0 & - & + & - \\ 0 & 0 & + & - & - & + \\ + & - & - & - & 0 & 0 \\ - & + & - & 0 & - & 0 \\ - & - & + & 0 & 0 & - \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & + & + & 0 & 0 & 0 \\ + & 0 & + & 0 & 0 & 0 \\ + & + & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & + & 0 & + \\ 0 & 0 & 0 & + & + & 0 \end{pmatrix}.$$

b) for n=5:

$$X = \begin{pmatrix} + & - & - & - & - & + & 0 & 0 & 0 & 0 \\ - & + & - & - & - & 0 & + & 0 & 0 & 0 \\ - & - & + & - & - & 0 & 0 & + & 0 & 0 \\ - & - & - & + & - & 0 & 0 & 0 & + & 0 \\ - & - & - & - & + & 0 & 0 & 0 & 0 & + \\ + & 0 & 0 & 0 & 0 & - & + & + & + & + \\ 0 & + & 0 & 0 & 0 & + & - & + & + & + \\ 0 & 0 & + & 0 & 0 & + & + & - & + & + \\ 0 & 0 & 0 & + & 0 & + & + & + & - & + \\ 0 & 0 & 0 & 0 & + & + & + & + & + & - \end{pmatrix}$$

$$Y = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & + & - & - & + \\ 0 & 0 & 0 & 0 & 0 & + & 0 & + & - & - \\ 0 & 0 & 0 & 0 & 0 & - & + & 0 & + & - \\ 0 & 0 & 0 & 0 & 0 & - & - & + & 0 & + \\ 0 & 0 & 0 & 0 & 0 & + & - & - & + & 0 \\ 0 & - & + & + & - & 0 & 0 & 0 & 0 & 0 \\ - & 0 & - & + & + & 0 & 0 & 0 & 0 & 0 \\ + & - & 0 & - & + & 0 & 0 & 0 & 0 & 0 \\ + & + & - & 0 & - & 0 & 0 & 0 & 0 & 0 \\ - & + & + & - & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now let $A_0 = (1)$, $B_0 = (1)$, $C_0 = (1)$, $D_0 = (1)$ and $A = (+ + +)$, $B = C = D = (+ - -)$ be cyclic symmetric matrices of order 1 and 3, respectively. Then, from the equation (12), matrix we obtain Williamson matrices of order 6, i.e.

$$A_1 = A_3 = \begin{pmatrix} A & C \\ D & -B \end{pmatrix}, \quad A_2 = A_4 = \begin{pmatrix} B & D \\ C & -A \end{pmatrix},$$

$$A_1 = A_3 = \begin{pmatrix} + & + & + & + & - & - \\ + & + & + & - & + & - \\ + & + & + & + & - & + \\ + & - & - & - & + & + \\ - & + & - & + & - & + \\ - & - & + & + & + & - \end{pmatrix},$$

$$A_2 = A_4 = \begin{pmatrix} + & - & - & + & - & - \\ - & + & - & - & + & - \\ - & - & + & - & - & + \\ + & - & - & - & - & - \\ - & + & - & - & - & - \\ - & - & + & - & - & - \end{pmatrix}.$$

Let $A_0 = (1)$, $B_0 = (1)$, $C_0 = (1)$, $D_0 = (1)$ and $A = B = (+ - - - -)$, $C = (+ + - - +)$, $D = (+ - + + -)$ be cyclic symmetric matrices of order 1 and 5, respectively. Then, from the equation (12), we obtain Williamson matrices of order 10, i.e.

$$A_1 = A_3 = \begin{pmatrix} + & - & - & - & - & + & + & - & - & + \\ - & + & - & - & - & + & + & + & - & - \\ - & - & + & - & - & - & + & + & + & - \\ - & - & - & + & - & - & - & + & + & + \\ - & - & - & - & + & + & - & - & + & + \\ + & - & + & + & - & - & + & + & + & + \\ - & + & - & + & + & + & - & + & + & + \\ + & - & + & - & + & + & + & - & + & + \\ + & + & - & + & - & + & + & + & - & + \\ - & + & + & - & + & + & + & + & + & - \end{pmatrix},$$

$$A_2 = A_4 = \begin{pmatrix} + & - & - & - & - & + & - & + & + & - \\ - & + & - & - & - & - & + & - & + & + \\ - & - & + & - & - & + & - & + & - & + \\ + & - & - & + & - & + & + & - & + & - \\ - & + & - & - & + & - & + & + & - & + \\ + & + & - & - & + & - & + & + & + & + \\ + & + & + & - & - & + & - & + & + & + \\ - & + & + & + & - & + & + & - & + & + \\ - & - & + & + & + & + & + & + & - & + \\ + & - & - & + & + & + & + & + & + & - \end{pmatrix}.$$

VII. Multiplicative Theorem Based Fast Williamson-Hadamard Transforms

In this section we present a fast transform algorithm based in Theorems 6.1 and 6.2. First we give an algorithm of generation of Hadamard matrix based on Theorem 6.2.

Algorithm 7.1: **Generation of Hadamard matrix via Theorem 6.2.**

Input : Williamson matrices A, B, C and D of order n and Hadamard matrix H_1 of order $4m$.

Step 1 . Construct the matrices X and Y according to (11).

Step 2 . Construct Hadamard matrix as

$$P = X \otimes H_1 + Y \otimes S_{4m} H_1, \qquad (13)$$

where S_{4m} is a monomial matrix with conditions: $S_{4m}^T = -S_{4m}$, $S_{4m} S_{4m}^T = I_{4m}$.

Output Hadamard matrix P.

Example of a monomial matrix of order 8 is given below

$$\begin{pmatrix} 0 & + & 0 & 0 & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & + & 0 & 0 & 0 & 0 \\ 0 & 0 & - & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & + & 0 & 0 \\ 0 & 0 & 0 & 0 & - & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + \\ 0 & 0 & 0 & 0 & 0 & 0 & - & 0 \end{pmatrix}.$$

Algorithm 7.2: **Fast transform with matrix (13).**

Input vector-column $F^T = (f_1, f_2, ..., f_{8mn})$, and Hadamard matrix P from (13).

Step 1 . Perform P as

$$P = (X \otimes I_{4m} + Y \otimes S_{4m})(I_{2m} \otimes H_1). \qquad (14)$$

Step 2 . Split vector F as $F = (F_1, F_2, ..., F_{2n})$, where $F_j = (f_{4m(j-1)+1}, f_{4m(j-1)+2}, ..., f_{4m(j-1)+4m})$.

Step 3 . Compute the transform $Q_i = H_1 F_i$, $i = 1, 2, ..., 2n$.

Step 4 . Split vector $Q = (Q_1, Q_2, ..., Q_{2n})$ into $4m$ $2n$-dimensional vectors as $Q = (P_1, P_2, ..., P_{4m})$, where

$$P_j = (f_{2n(j-1)+1}, f_{2n(j-1)+2}, ..., f_{2n(j-1)+2n}).$$

Step 5 . Compute the transforms XP_j and YP_j.

Output : transform coefficients.

Let us give an example of computation of transforms XF and YF $(F = (f_1, f_2, ..., f_6))$, where A, B, C, D are Williamson matrices of order 3, and X and Y from the Example 2.1. First we compute

$$XF = \begin{pmatrix} f_1 + f_4 - (f_5 + f_6) \\ f_2 + f_5 - (f_4 + f_6) \\ f_3 + f_6 - (f_4 + f_5) \\ f_1 - f_4 - (f_2 + f_3) \\ f_2 - f_5 - (f_1 + f_3) \\ f_3 - f_6 - (f_1 + f_2) \end{pmatrix}, \; YF = \begin{pmatrix} f_2 + f_3 \\ f_1 + f_3 \\ f_1 + f_2 \\ f_5 + f_6 \\ f_4 + f_6 \\ f_4 + f_5 \end{pmatrix} \qquad (15)$$

From (15) it follows that the joint computation of XF and YF requires only 18 addition/subtruction (see Figure 3).

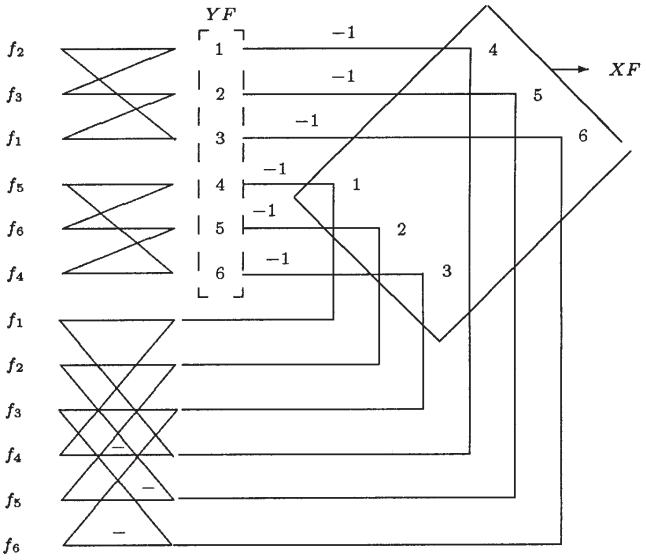


Fig. 3. Flow graph of joint computation of XF and YF

Then, from (14), we can conclude that the complexity of PF transform algorithm can be obtained by

$$C(24m) = 48m(2m + 1).$$

Note, that if X, Y are matrices of order k defined by (11), H_m is an Hadamard matrix of order m, and S_m is a monomial matrix of order m, then for any integer n

$$H_{mk^n} = X \otimes H_{mk^{n-1}} + Y \otimes S_{mk^{n-1}} H_{mk^{n-1}} \qquad (16)$$

is an Hadamard matrix of order mk^n.

Remark 7.1: For $A = B = C = D = (1)$ from (11) we have $X = \begin{pmatrix} + & + \\ + & - \end{pmatrix}$, $Y = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. And if $H_2 = \begin{pmatrix} + & + \\ + & - \end{pmatrix}$, then the matrix (16) is the Walsh-Hadamard matrix of order $2n + 1$ [1].

Algorithm 7.3: **Construction of Hadamard matrices of order** $m(2n)^k$.

Input : Williamson matrices A, B, C, D of order n and Hadamard matrix of order m.

Step 1 . Construct matrices X and Y according to (11).

Step 2 . Construct the matrix $H_{2nm} = X \otimes H_m + Y \otimes S_m H_m$.

Step 3 . If $i < k$, then $i \leftarrow i+1$, $H_{m(2n)^i} \leftarrow H_{m(2n)^{i+1}}$, $S_{m(2n)^i} \leftarrow S_{m(2n)^{i+1}}$, and go to the **Step 2**.

Output : Hadamard matrix $H_{m(2n)^k}$.

Let us represent a matrix H_{mk^n} as a product of sparse matrices.

$$\begin{aligned} H_{mk^n} &= (X \otimes I_{mk^{n-1}} + Y \otimes S_{mk^{n-1}})(I_k \otimes H_{mk^{n-1}}) \\ &= A_1(I_k \otimes H_{mk^{n-1}}). \end{aligned}$$

Continue this factorization process for all matrices $H_{mk^{n-i}}$, $i = 1, 2, ..., n$, we obtain

$$H_{mk^n} = A_1 A_2 \cdots A_n (I_{k^n} \otimes H_m), \tag{17}$$

where

$$A_i = I_{k^{i-1}} \otimes (X \otimes I_{mk^{n-i}} + Y \otimes S_{mk^{n-i}}), \quad i = 1, 2, ..., n.$$

Example 7.1: Let H_m be a Hadamard matrix of order m, and X and Y have the form as in Example 2.1, and let $F = (f_i)_{i=1}^{6m}$ be an input vector. Then we have a Hadamard matrix of order $6m$ of the form

$$H_{6m} = X \otimes H_m + Y \otimes S_m H_m.$$

Like in (17), we have $H_{6m} = A_1(I_6 \otimes H_m)$, where $A_1 = X \otimes I_m + Y \otimes S_m$, and

$$X \times I_m = \begin{pmatrix} I_m & 0 & 0 & I_m & -I_m & -I_m \\ 0 & I_m & 0 & -I_m & I_m & -I_m \\ 0 & 0 & I_m & -I_m & -I_m & I_m \\ I_m & -I_m & -I_m & -I_m & 0 & 0 \\ -I_m & I_m & -I_m & 0 & -I_m & 0 \\ -I_m & -I_m & I_m & 0 & 0 & -I_m \end{pmatrix}$$

$$Y \otimes S_m = \begin{pmatrix} 0 & S_m & S_m & 0 & 0 & 0 \\ S_m & 0 & S_m & 0 & 0 & 0 \\ S_m & S_m & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & S_m & S_m \\ 0 & 0 & 0 & S_m & 0 & S_m \\ 0 & 0 & 0 & S_m & S_m & 0 \end{pmatrix} \tag{18}$$

The input column-vector is represented as $F = (F_1, F_2, ..., F_6)$, where F_i is an m-dimensional vector.

Now we estimate the complexity of transforms

$$\begin{aligned} H_{6m}F &= A_1(I_6 \otimes H_m)F \\ &= A_1 \cdot \mathrm{diag}\{H_m F_1, H_m F_2, ..., H_m F_6\}. \end{aligned} \tag{19}$$

Denote $T = (I_6 \otimes H_m)F$. Computing $A_1 T$, where $T = (T_1, ..., T_6)$, from (18) we obtain

$$(X \otimes I_m)T = \begin{pmatrix} T_1 + T_4 - (T_5 + T_6) \\ T_2 + T_5 - (T_4 + T_6) \\ T_3 + T_6 - (T_4 + T_5) \\ T_1 - T_4 - (T_2 + T_3) \\ T_2 - T_5 - (T_1 + T_3) \\ T_3 - T_6 - (T_1 + T_2) \end{pmatrix},$$

$$(Y \otimes S_m)T = \begin{pmatrix} S_m(T_2 + T_3) \\ S_m(T_1 + T_3) \\ S_m(T_1 + T_2) \\ S_m(T_5 + T_6) \\ S_m(T_4 + T_6) \\ S_m(T_4 + T_5) \end{pmatrix} \tag{20}$$

From (19) and (20) it follows that the computational complexity of transform H_{6mF} is

$$C(H_{6m}) = 24m + 6C(H_m),$$

where $C(H_m)$is a complexity of an m-point Hadamard transform.

VIII. Complexity and Comparison

A. *Complexity of block-cyclic block-symmetric Williamson-Hadamard Transform*

Since every block-row of block-cyclic block-symmetric Hadamard matrix contains block Q_0 and other blocks are from a set $\{Q_1, Q_2, Q_3, Q_4\}$ (see Appendix), it is not difficult to find that the complexity of the block Williamson-Hadamard transform of order $4n$ can be obtained from the following formula

$$C(H_{4n}) = 4n(n+2).$$

From representation of a block Williamson-Hadamard matrix we can see that some of block pairs are repeated.

Two block sequences of length k $(k < n)$ in the first block row of block Williamson-Hadamard matrix of order $4n$ $(-1)^{p_1}Q_i, (-1)^{p_2}Q_i, ..., (-1)^{p_k}Q_i$ and $(-1)^{q_1}Q_j, (-1)^{q_2}Q_j, ..., (-1)^{q_k}Q_j$ $(i, j = 0, 1, 2, 3, 4)$, where $p_t, q_t \in \{0, 1\}$ and for all $t = 1, 2, ..., k$ $q_t = p_t$, or $q_t = \bar{p}_t$ $(\bar{1} = 0, \bar{0} = 1)$, we call *cyclic congruent circuits* if $dist[(-1)^{p_t}Q_i, (-1)^{p_{t+1}}Q_i] = dist[(-1)^{q_t}Q_j, (-1)^{q_{t+1}}Q_j]$ for all $t = 1, 2, ..., k-1$, where $dist[A_i, A_j] = j - i$, for $A = (A_i)_{i=1}^m$.

For example, in first block row of the block-cyclic block-symmetric Hadamard matrix of order 36 there are 3 cyclic congruent circuits of length 2. These circuits are underlined as

$$Q_0, \underline{Q_1}, \underline{\underline{-Q_2}}, \underline{\underline{Q_1}}, -Q_1; -Q_1, \underline{Q_1}, \underline{\underline{-Q_2}}, \underline{\underline{Q_1}}.$$

With this observation, one can reduce some operations in summing up the vectors Y_i (see Step 3 above example and corresponding ?ow graphs).

Let m be a length of the cyclic congruent circuits into block-row of the block-cyclic block-symmetric Hadamard matrix of order $4n$, t_m be a number of various cyclic congruent circuits of length m, $N_{m,j}$ be a number of cyclic congruent circuits of type j and length m. Then the complexity of the Hadamard transform of order $4n$ takes the form

$$C_r(H_{4n}) = 4n\Big(n + 2 - 2\sum_{i=2}^{m}\sum_{j=1}^{t_m}(N_{m,j} - 1)(i-1)\Big). \tag{21}$$

The values of parameters n, m, t_m, $N_{m,j}$ and the complexity of the Williamson type Hadamard transform of order $4n$ are given in the following table 1.

Thus, the complexity of the block Williamson-Hadamard transform can be calculated from the formula

$$C^{\pm} = 2n(2n+3), \qquad C^{sh} = 3n,$$

where $C^{\pm}$ is a number of addition/subtractions, and C^{sh} is a number of shifts.

Table 1.

No.	4n	m	t_m	$N_{m,j}$	$C_r(H_{4n})$	direct comp.
1	12				60	132
2	20				140	380
3	28	2	1	2	224	756
4	36	2	1	3	324	1260
5	44	2	1	2	528	1892
6	52	3	1	2	676	2652
7	60	2	3	3, 2, 2	780	3540
8	68	2	2	2, 3	1088	4558
9	76	2	3	2, 4, 3	1140	5700
10	84	2	3	2, 2, 5	1428	6972
11	92	2	3	4, 2, 2	1840	8372
12	100	2	3	2, 7, 2	1850	9900

Now, using repetitions of additions of vectors Y_i and the same notations as in the previous subsection (see equation (21)), the complexity of Williamson-Hadamard transform can be presented as

$$C_r^{\pm} = 2n\Big(2n+3-2\sum_{i=2}^{m}\sum_{j=1}^{t_m}(N_{m,j}-1)(i-1)\Big),$$
$$C^{sh} = 3n.$$

Formulas of complexities of the fast Williamson-Hadamard transforms without repetitions of blocks and with repetitions and shifts, and their numerical results are given in formula (22) and in Table 2, respectively,

$$\begin{aligned} C &= 4n(n+2),\\ C_r &= 4n\Big(n+2-2\sum_{i=2}^{m}\sum_{j=1}^{t_m}(N_{m,j}-1)(i-1)\Big),\\ C^{\pm} &= 2n(2n+3),\\ C^{sh} &= 3n,\\ C_r^{\pm} &= 2n\Big(2n+3-2\sum_{i=2}^{m}\sum_{j=1}^{t_m}(N_{m,j}-1)(i-1)\Big),\\ C^{sh} &= 3n. \end{aligned} \tag{22}$$

Table 2.

No.	4n	C	$C^{\pm}$	C^{sh}	C_r	$C_r^{\pm}$	direct comput.
1	12	60	54	9	60	54	132
2	20	140	130	15	140	130	380
3	28	252	238	21	224	210	756
4	36	396	378	27	324	306	1260
5	44	572	550	33	528	506	1892
6	52	780	754	39	676	650	2652
7	60	1020	990	45	780	750	3540
8	68	1292	1258	51	1088	1054	4558
9	68	1292	1258	51	1020	986	4558
10	76	1596	1558	57	1140	1102	5700
11	84	1932	1890	63	1428	1386	6972
12	92	2300	2254	69	1840	1794	8372
13	100	2700	2650	75	1900	1850	9900

B. Complexity of Hadamard transform from multiplicative theorem

Recall that if X, Y are matrices from (11) of order k and H_m is an Hadamard matrix of order m, then the Hadamard matrix constructed recursively

$$H_{mk^n} = X \otimes H_{mk^{n-1}} + Y \otimes S_{mk^{n-1}} H_{mk^{n-1}},$$

can be factorized as

$$H_{mk^n} = A_1 A_2 \cdots A_n (I_{k^n} \otimes H_m),$$

where

$$A_i = I_{k^{i-1}} \otimes (X \otimes I_{mk^{n-i}} + Y \otimes S_{mk^{n-i}}). \tag{23}$$

Let us now evaluate the complexity of a transform

$$H_{mk^n} F, \qquad F^T = (f_1, f_2, ..., f_{mk^n}).$$

First, we find the required operations for transform $A_i P$, $P^T = (p_1, p_2, ..., p_{mk^n})$.

Represent $P^T = (P_1, P_2, ..., P_{k^{i-1}})$, where

$$P_j = \Big((j-1)mk^{n-i+1} + t\Big)_{t=1}^{mk^{n-i+1}},$$

and $j = 1, 2, ..., k^{i-1}$.

Then, from (23), we have

$$\begin{aligned} A_i P = \operatorname{diag}\{&(X \otimes I_{mk^{n-i}} + Y \otimes S_{mk^{n-i}}) P_1, ...,\\ &..., (X \otimes I_{mk^{n-i}} + Y \otimes S_{mk^{n-i}}) P_{k^{i-1}}\}. \end{aligned} \tag{24}$$

Denote the complexities of transforms XQ and YQ by C_X and C_Y, respectively. We have

$$C_X < k(k-1), \qquad C_Y < k(k-1).$$

From (24) we obtain the complexity of transform $\prod_{i=1}^{n} A_i P$, $(C_X + C_Y + k)mnk^{n-1}$.

Table 3.

H_m	Complexity
$H_m = X = H_2$ (see Remark 7.1)	$(n+1)2^{n+1}$
Walsh-Hadamard	$(C_X + C_Y + k)m\,nk^{n-1} + k^n m \log_2 m$.
BCBS Williamson-Hadamard: a) with block repetition	$(C_X + C_Y + k)m\,nk^{n-1} + k^n m\,(\frac{m}{4} + 2)$
b) with block repetition and congruent circuits	$(C_X + C_Y + k)m\,nk^{n-1} + k^n m\,(\frac{m}{4} + 2 - 2\sum_{i=2}^{r}\sum_{j=1}^{t_r}(N_{r,j}-1)(i-1))$
c) with block repetition and shifts	$(C_X + C_Y + k)m\,nk^{n-1} + k^n \frac{m}{2}(\frac{m}{2} + 3)$; $k^n \frac{3m}{4}$
d) with block repetition and congruent circuits, and shifts	$(C_X + C_Y + k)m\,nk^{n-1} + k^n \frac{m}{2}(\frac{m}{2} + 3 - 2\sum_{i=2}^{r}\sum_{j=1}^{t_r}(N_{r,j}-1)(i-1))$; $k^n \frac{3m}{4}$

Hence, the total complexity of transform $H_{mk^n}F$ is

$$C(H_{mk^n}) < (C_X + C_Y + k)mnk^{n-1} + k^n C(H_m),$$

where $C(H_m)$is a complexity of m-point Hadamard transform.

For a given matrices X and Y we can compute the exact value of C_X, C_Y, and therefore we can obtain the exact complexity of the transform. For example, for $k = 6$, from (15) we see that $C_X + C_Y = 18$, hence $m6^n$-point Hadamard transform requires only $24mn6^{n-1} + 6nC(H_m)$ operations.

IX. Conclusion

Three new efficient algorithms of $4t$-point (t is a 'arbitrary' integer number, for which there is a construction of a Hadamard matrix) Williamson-Hadamard transforms computation are developed. The design algorithms are based on block representation of Williamson-Hadamard matrices, on multiplicative theorem, and on iterative constructions.

Using the structures of the existing Williamson matrices the computational complexity of the developed algorithm is greatly reduced. Williamson-Hadamard transform algorithms on add/shift architectures are also described. The complexity of developed algorithms are demonstrated. Comparative estimates revealing the efficiency of the proposed algorithms with respect to the known ones are given. The results of numerical examples were presented.

Appendix

Proof of Theorem 6.1 . Let A, B, C, D be Williamson matrices of order m. Introduce $(0, \pm 1)$-matrices X, Y of order $2n$

$$\begin{aligned} X &= \tfrac{1}{2}\begin{pmatrix} A+B & C+D \\ C+D & -A-B \end{pmatrix}, \\ Y &= \tfrac{1}{2}\begin{pmatrix} A-B & C-D \\ -C+D & A-B \end{pmatrix}. \end{aligned} \tag{25}$$

One can check that matrices X, Y satisfy the conditions

$$\begin{aligned} &X \odot Y = 0, \quad \odot \text{ is a Hadamard product}, \\ &XY^T = YX^T, \\ &X \pm Y \quad \text{is a } (\pm 1)\text{-matrix}, \\ &XX^T + YY^T = 2mI_{2m}. \end{aligned} \tag{26}$$

Let A_0, B_0, C_0, D_0 be Williamson type matrices of order n. Introduce the matrices

$$\begin{aligned} A_i &= A_{i-1} \otimes X + B_{i-1} \otimes Y, \\ B_i &= B_{i-1} \otimes X - A_{i-1} \otimes Y, \\ C_i &= C_{i-1} \otimes X + D_{i-1} \otimes Y, \\ D_i &= D_{i-1} \otimes X - C_{i-1} \otimes Y. \end{aligned} \tag{27}$$

Prove that for any natural number i A_i, B_i, C_i, D_i are Williamson type matrices of order $(2m)^i n$. From the formulas (27) for $i = 1$ we obtain

$$\begin{aligned} A_1A_1^T =& A_0A_0^T \otimes XX^T + B_0B_0^T \otimes YY^T \\ &+ A_0B_0^T \otimes XY^T + B_0A_0^T \otimes YX^T, \\ B_1B_1^T =& B_0B_0^T \otimes XX^T + A_0A_0^T \otimes YY^T \\ &- B_0A_0^T \otimes XY^T - A_0B_0^T \otimes YX^T. \end{aligned}$$

Taking into account conditions (25) and (26) and summarizing last expressions, we find

$$A_1A_1^T + B_1B_1^T = (A_0A_0^T + B_0B_0^T) \otimes (XX^T + YY^T).$$

Similarly, it turns out that

$$C_1C_1^T + D_1D_1^T = (C_0C_0^T + D_0D_0^T) \otimes (XX^T + YY^T).$$

Now, summarizing last two equations and taking into account that A_0, B_0, C_0, D_0 is a Williamson type matrices of order n, and X and Y satisfy to conditions (26), we have

$$A_1A_1^T + B_1B_1^T + C_1C_1^T + D_1D_1^T = 8mnI_{2mn}.$$

Let's prove now equality $A_1B_1^T = B_1A_1^T$. Really, from (27)

$$\begin{aligned} A_1B_1^T =& A_0B_0^T \otimes XX^T - A_0A_0^T \otimes XY^T \\ &+ B_0B_0^T \otimes YX^T - B_0A_0^T \otimes YY^T, \\ B_1A_1^T =& B_0A_0^T \otimes XX^T + B_0B_0^T \otimes XY^T \\ &- A_0A_0^T \otimes YX^T - A_0B_0^T \otimes YY^T. \end{aligned}$$

Comparing both expressions, we conclude, that $A_1B_1^T = B_1A_1^T$. Similarly, it can be shown, that $PQ^T = QP^T$, where $P, Q \in \{A, B, C, D\}$. Thus, the matrices A_1, B_1, C_1, D_1 are Williamson type matrices of order $2mn$.

Now we assume that the theorem is correct for $k = i \geq 1$, i.e. A_k, B_k, C_k, D_k are Williamson matrices of order $(2m)^k n$. Let's prove, that A_{i+1}, B_{i+1}, C_{i+1}, D_{i+1} are also Williamson matrices. Let's check up only the second condition of the equation (25). Compute

$$\begin{aligned} A_{i+1}A_{i+1}^T =& A_iA_i^T \otimes XX^T + A_iB_i^T \otimes XY^T \\ &+ B_iA_i^T \otimes YX^T + B_iB_i^T \otimes YY^T, \\ B_{i+1}B_{i+1}^T =& B_iB_i^T \otimes XX^T - B_iA_i^T \otimes XY^T \\ &- A_iB_i^T \otimes YX^T + A_iA_i^T \otimes YY^T, \\ C_{i+1}C_{i+1}^T =& C_iC_i^T \otimes XX^T + C_iD_i^T \otimes XY^T \\ &+ D_iC_i^T \otimes YX^T + D_iD_i^T \otimes YY^T, \\ D_{i+1}D_{i+1}^T =& D_iD_i^T \otimes XX^T - D_iC_i^T \otimes XY^T \\ &- C_iD_i^T \otimes YX^T + C_iC_i^T \otimes YY^T. \end{aligned}$$

Summarizing the obtained equations, we find

$$\begin{aligned} &A_{i+1}A_{i+1}^T + B_{i+1}B_{i+1}^T + \cdots + D_{i+1}D_{i+1}^T \\ &\quad = (A_iA_i^T + \cdots + D_iD_i^T) \otimes (XX^T + YY^T). \end{aligned} \tag{28}$$

Since A_i, B_i, C_i, D_i are Williamson matrices of order $(2m)^i n$, and the matrices X, Y satisfy to conditions (26), then the equation (28) has a form

$$\begin{aligned} A_{i+1}A_{i+1}^T + B_{i+1}B_{i+1}^T + \cdots + D_{i+1}D_{i+1}^T = \\ = 4(2m)^{i+1}nI_{(2m)^{i+1}n}. \end{aligned}$$

S_{4m} is a monomial matrix with condition $S_{4m}^T = -S_{4m}$ and H_1 is a Hadamard matrix of order $4m$.

Thus, $P = X \otimes H_1 + Y \otimes S_{4m}H_1$ is a Hadamard matrix of order $8mn$.

References

[1] Ahmed, Rao. Orthogonal Transforms for Digital Signal Processing. Springer-Verlag, New York, 1975.

[2] Agaian S.S. Hadamard Matrices and Their Applications. Lecture Notes in Mathematics, vol. 1168, 1985.

[3] G.R. Reddy, P. Satyanarayana. Interpolation Algorithm Using WalshHadamard and Discrete Fourier/Hartley Transforms. IEEE, 1991, p. 545-547.

[4] CheungFat Chan. Efficient Implementation of a Class of Isotropic Quadratic Filters by Using WalshHadamard Transform. IEEE Int. Symposium on Circuits and Systems, June 912, Hong Kong, 1997, p. 2601-2604.

[5] Brian K. Harms, Jin Bae Park, Stephan A. Dyer. Optimal Measurement Techniques Utilizing Hadamard Transforms. IEEE Trans. on Instrumentation and Measurement, vol. 43, No. 3, June 1994, p. 397-402.

[6] Chen Anshi, Li Di, Zhou Renzhong. A Research on Fast Hadamard Transform (FHT) Digital Systems. IEEE TENCON 93/Beijing, 1993, p.541-546.

[7] Sarukhanyan H.G. Hadamard matrices: Construction methods and applications. In Proc. of The Workshop on Transforms and Filter Banks, Feb. 21-27, Tampere, Finland, 1998, 35p.

[8] Sarukhanyan H.G. Decomposition of the Hadamard matrices and fast Hadamard transform. Computer Analysis of Images and Patterns, Lecture Notes in Computer Science, vol. 1296, 1997, p.

[9] Yarlagadda Rao R.K., Hershey E.J. Hadamard Matrix Analysis and Synthesis with Applications and Signal/Image Processing, 1997.

[10] Sylvester J.J. Thoughts on Inverse Orthogonal Matrices, Simultaneous Sign Successions and Tesselated Pavements in Two or More Colours, with Applications to Newton's Rule, Ornamental Til-Work, and the Theory of Numbers. Phil. Mag., vol. 34, 1867, p. 461-475.

[11] Seberry J., Yamada M. Hadamard Matrices, Sequences and Block Designs. Surveys in Contemporary Design Theory, Wiley-Interscience Series in Discrete Mathematics. Jhon Wiley, New York, 1992.

[12] Samadi S., Suzukake Y., Iwakura H. On Automatic Derivation of Fast Hadamard Transform Using Generic Programming. Proc. 1998 IEEE Asia-Pacific Conference on Circuit and Systems, Thailand, 1998, p. 327-330.

[13] Coppersmith D., Feig E., Linzer E. Hadamard Transforms on Multiply/Add Architectures. IEEE Trans. on Signal Processing, vol. 42, No. 4, 1994, p. 969-970.

[14] Agaian S.S., Sarukhanian H.G. Reccurent Formulae of the Construction Williamson Type Matrices. Math. Notes, vol. 30, No. 4, 1981, p. 603-617.

[15] Williamson J. Hadamard Determinant Theorem and Sum of Four Squares. Duke Math. J., No. 11, 1944, p. 65-81.

[16] H. Sarukhanyan, S. Agaian, K. Egiazarian, J. Astola. On Fast Hadamard Transforms of Williamson Type. EUSIPCO, 2000.

[17] http://www.cs.uow.edu.au/people/jennie/lifework.html

[18] Agaian S.S. Construction of Planar and Spatial Block Hadamard Matrices. Proc. of Computer Centre of Armenian Academy of Sciences,(in Russian), vol. 12, 1984, p. 550.

[19] Z. Li, H.V. Sorensen, C.S. Burrus. FFT and Convolution Algotithms an DSP Microprocessors. Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing. 1986, pp. 289-294.

[20] R.K. Montoye, E. Hokenek, S.L. Runyon. Design of the IBM RISC System/6000 floating point execution unit. IBM J. Res. Develop., 1990, vol. 34, pp. 71-77 p. 5-50.

[21] http://www.top500.org/reports/1993/section2-12-7.html

Least Squares Image Resizing

Atanas Gotchev[1], Karen Egizarian[2], and Tapio Saramaki[3]

Abstract – **- This invited paper considers the problem of reducing the size of a digital image as a problem of constructing a proper polynomial spline approximation to a function $s(t) \in L_2(R)$. From this point of view, the crucial problem is to design an appropriate reconstruction (basis) function. Then, the analysis (projecting) function can be formed as biorthogonal to the reconstruction one. The basis functions are chosen among the class of the symmetrical and compactly supported modified B-splines having both good approximation properties and efficient realization structures. We review the theory of orthogonal projections both in continuous (minimizing the L2 norm) and discrete domain (minimizing the l_2 norm) and propose a constructive compromise yielding an efficient decimation structure possessing good anti-aliasing properties. It is shown, by means of examples, that with a considerably lower computational complexity the proposed structure provides practically the same quality for the restored images as the best existing structures.**

Keywords – **decimation, orthogonal projection, B-splines**

I. Introduction

Many image processing applications demand an effective resizing algorithm. These applications include, among others, digital zooming effects, equalizing the resolution for different imaging and printing devises, and building multiresolution pyramids. The problem of image interpolation (reconstruction) can be accomplished by first fitting an appropriate continuous model to the discrete data and then resampling the resulting image on a *finer* grid [1,2]. The problem of image size reduction is rather different, since the decimation is vulnerable to aliasing effects if one samples the same continuous model at a *sparser* grid. The classical digital signal processing theory dictates that a preliminary low-pass (*anti-aliasing*) filtering is needed. The ideal low-pass filtering is not a practically realizable option because it corresponds to the use of the sinc impulse response that is infinitely supported. Hence, the investigation efforts have been concentrated in searching for solutions based on decimation (pre-filtering) functions that are compactly supported and have good anti-aliasing proper ties. Some designs using polynomial or B-spline combinations have seemed to be perspective candidates [1,2]. These solutions have been originally designed to be good interpolators (with good anti-imaging properties). The research question is how to adapt them to the decimation problem. More formally, by considering those kernels as basis functions generating polynomial spline subspaces, the problem of down-sampling can be transferred to the problem of designing a new signal approximation with the minimum loss of information, thereby naturally leading to a certain form of *least squares* (LS).

For B-spline bases, algorithms for rational [3] and for arbitrary [4] scale conversion have been proposed, aimed at minimizing the continuous L_2 error norm. In order to have an input *continuous* function at hands, those algorithms advise a preliminary spline model fitting, as in the interpolation case, and a subsequent continuous-time processing.

Another standpoint insists on minimizing the discrete l_2 norm, based on the fact that the images are discrete and the quality variations are assessed by means of the signal-to noise ratio (SNR) that is a discrete l_2 measure. The drawback of using the classical discrete LS is that there is a demand to compute a *pseudo-inverse* matrix [9]. This computation is not preferable even for the banded matrices that express compactly supported bases.

In this paper, a new way is proposed for treating the L_2 and l_2 solutions for the image decimation problem and it is shown how they can be merged into a high-quality and cost-efficient hybrid scheme.

II. Signal Decimation Problem

By assuming a separable basis functions, we downgrade the image size reduction problem to the 1-D signal decimation problem.

Consider an initial discrete signal $x(k)$ for $k = 0, 1, ..., L_{in} - 1$ as taken from a continuous function $s(t)$ sampled over a uniform grid in the interval $[a, b]$: $\boldsymbol{\tau} = [\tau_0, ..., \tau_{L_{in}-1}]$, $\tau_{k+1} - \tau_k = h_{in}$, $\tau_0 = a$, $\tau_{L_{in}-1} = b$ that is, $x(k) = s(\tau_k)$. This signal is desired to be decimated into a signal $y(l)$ for $l = 0, 1, ..., L_{out} - 1$ over a new uniform grid, determined by a larger step h_{out} with $h_{out} > h_{in}$. Without loss of generality it is assumed that $h_{out} = 1$, $a = 0$, and $b = L_{out} - 1$. Hence, the output grid will be placed on the integer coordinates. The quantity $h \equiv h_{in} = (L_{out} - 1)/(L_{in} - 1) < 1$ indicates the decimation ratio. Fig. 1 illustrates the process.

The decimated signal should preserve the original signal features as well as possible, i.e., having a proper reconstruction function one should be able to get a good approximation to the original signal from the decimated one. Hence, we seek for a solution in a form of a projection onto some linear space generated by this well chosen reconstruction basis. In Section III, we briefly review the problem of construct-

[1]Atanas Gotchev, Member of IEEE is with the Institute of Signal Processing, Tampere University of Technology, P.O.Box 553, FIN-33101 Tampere, Finland E-mail: agotchev@cs.tut.fi

[2]Karen Egiazarian, Senior Member of IEEE is with the Institute of Signal Processing, Tampere University of Technology, P.O.Box 553, FIN-33101 Tampere, Finland E-mail: karen@cs.tut.fi

[3]Tapio Saramaki, Fellow of IEEE is with the Institute of Signal Processing, Tampere University of Technology, P.O.Box 553, FIN-33101 Tampere, Finland E-mail: ts@cs.tut.fi

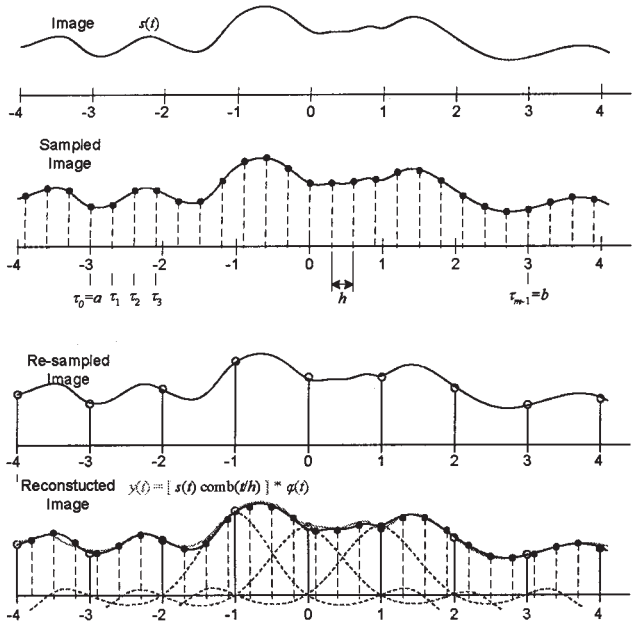


Fig. 1. Image decimation problem in 1-D

ing different polynomial spline approximations to a function $s(t) \in L_2(R)$ being parameterized by its scale. Then we return back to the problem of signal decimation and comment how it can be solved using the proposed formalism.

A. An interpolative solution

Before reviewing the LS approaches, we present a solution that has been considered as good candidate for continuous (at an arbitrary sparser grid) signal decimation [5,11]. Consider a continuous model fitting for the discrete data given at the coordinate grid $\boldsymbol{\tau}$. To preserve the values of the given discrete sequence we write the model as:

$$x_h(t) = \sum d_h(i)\varphi(t/h - i) \tag{1}$$

where the modeling coefficients are obtained by the recursive pre-filtering with the all-pole filter formed by the basis function sampled at the integers as follows:

$$d_h(k) = \sum_i x(i)(q)^{-1}(k-i) \quad \text{where} \quad q(k) = \varphi(k). \tag{2}$$

Equivalently,

$$h_h(t) = \sum_{i=1}^{m} x_h(i)\varphi_{\text{int}}(t/h - 1), \tag{3}$$

where

$$\varphi_{\text{int}}(t) = \sum_{i=1}^{m} (q)^{-1}(i)\varphi(t-i). \tag{4}$$

Eq. (3) can be interpreted as continuous domain filtering of the impulse train $\sum x(i)\delta(t/h - i)$ by the function $\varphi_{\text{int}}(t)$, rescaled to have its nodes over the grid $\boldsymbol{\tau}$. Its Fourier transform $\Phi_{\text{int}}(\omega) = \Phi(\omega)/\Phi(e^{j\omega})$ has zeros clustered around multiples of $F_h = 1/h$. If we now resample $x_h(t)$ at the integers we will encounter aliasing effects due to the fact that $F_h > F_1 = 1$. In the contrary, we can take the continuous filter as function specified at the integers (non-rescaled), with zeros around $F_1 = 1$, thereby generating good anti-aliasing properties. This is equivalent to changing the order of operation: first the continuous filtering and resampling is performed followed by the recursive digital filtering. Writing the continuous filtering and resampling as

$$x_h(k) = \sum_{i=0}^{L_{in}-1} x(i)\varphi(k - hi) \tag{5}$$

the subsequent digital filtering leads to

$$y(l) = \sum_k q^{-1}(k) \sum_{i=0}^{L_{in}-1} x(i)\varphi(l - k - hi). \tag{6}$$

While this solution has shown satisfactory performance for certain class of signals [5], it is not optimal in LS sense. This fact has some effect when resizing images.

III. Orthogonal Projection Paradigm

A. Function spaces and generating bases

Consider the following *shift-invariant* function space $V(\varphi)$ that is a closed subspace of L_2 and generated by a function φ as:

$$V(\varphi) = \left\{ q(t) = \sum_{l=-\infty}^{\infty} c(i)\varphi(t-i) \quad c \in l^2 \right\}. \tag{7}$$

Any function $s(t) \in L_2$ can be orthogonally projected into $V(\varphi)$ by finding the corresponding discrete-time sequence $c(k)$. It gives the LS approximation with respect to the L_2 norm defined by

$$||s||_{L_2}^2 = \langle s, s \rangle = \int s(\xi)s(\xi)d\xi. \tag{8}$$

The resulting approximation is given by

$$\tilde{s}(t) = \sum_{i=-\infty}^{\infty} c(i)\varphi(t-i). \tag{9}$$

If the basis $\phi(t-i)$ is not orthogonal, but only linearly independent, then the orthogonal projection of $s(t) \in L_2$ onto V is obtained by

$$c(i) = \langle s(\xi)\hat{\varphi}(\xi - i) \rangle = \int s(\xi)\hat{\varphi}(\xi - i)d\xi \tag{10}$$

Here, $\hat{\varphi}(t)$ is the dual (bi-orthogonal) basis of $\varphi(t)$ satisfying

$$\hat{\varphi}(t) \in V(\varphi) \tag{11}$$

and

$$\hat{\varphi}(t) = \sum_{i=-\infty}^{\infty} (p)^{-1}(i)\varphi(t-i) \tag{12}$$

where

$$p(i) = \int \varphi(\xi)\varphi(\xi - i)d\xi. \tag{13}$$

The sequence $p^{-1}(i)$ is the convolution inverse of the autocorrelation sequence $p(i)$ [6]. If φ is a symmetrical and compactly supported function over N, then $p(i)$ is a symmetrical sequence of length $2N+1$.

B. Spline-like basis functions

We choose our basis function φ to be a compactly supported piece-wise function constructed by B-splines:

$$\varphi(t) = \beta^{\text{mod}}(t) = \sum_{n=0}^{N} \sum_{m \in 0} \gamma_{nm} \beta^n (t-m) \qquad (14)$$

where $\beta^n(t)$ is the B-spline of degree n being symmetrical around the origin defined as

$$\beta^n(t) = \Delta^{n+1} * \frac{x_+^n}{n!} * \delta\Big(t + \frac{n+1}{2}\Big) \qquad (15)$$

where

$$t_+^n = \begin{cases} t^n & \text{if } t \geq 0 \\ 0 & \text{otherwise,} \end{cases} \qquad (16)$$

and where $\Delta = \delta(t) - \delta(t-1)$ denotes the backward *finite difference*, and $\delta(t)$ is the Dirac's mass distribution [4].

Examples of such functions, apart the B-splines themselves, are the *modified* B-splines and *moms* studied in [1,2]. They have shown their superiority as bases for image reconstruction (interpolation) both because of their good performance and low computational complexity. B-splines are compactly supported over the interval $[N/2, N/2)$ and are the most regular functions having the maximum order with the given support. While the combinations (14) are not so regular, because of the added low-degree terms, they can be optimized to have good asymptotic approximation properties [2] or good anti-imaging properties in frequency domain needed for the signal reconstruction [1]. Furthermore, the function given by Eq. (14) can be formed in such a way as to have the support of the highest degree B-spline attending the combination, thereby resulting in the same computational complexity as with classical B-splines of the same degree. Those, so-called *splines of minimal support* [7] can be presented also as consisting of polynomial pieces of degree N at every interval between integers for the interval of its support, that is $N+1$:

$$\varphi(t) = \sum_{i=0}^{N} \sum_{m=0}^{N} c_m(i) \Big(t + \frac{N+1}{2} - i\Big)^m. \qquad (17)$$

The latter form is more suitable for practical realizations. The polynomial coefficients $c_m(i)$ in the i-th interval are formed as linear combinations of the corresponding polynomial coefficients $c_m^n(i)$ of the B-spline of degree n, attending the i-th interval of the modified kernel

$$c_m(i) = \sum_{n=0}^{N} \gamma_{mn} c_m^n(i). \qquad (18)$$

IV. Least Squares Solutions

There are two main problems in making a down-scale projection (decimation). First, the continuous function $s(t)$ is not known. Hence, we cannot perform a true L_2 orthogonal projection in the form of Eqs. (9), (10). One can avoid this inconvenience by modeling this function with a continuous model. Thus, the subsequent projection would minimize the squared error between the approximation and the model. Second, even if an appropriate continuous function is provided, solving the integral given by Eq. (10) is rather problematic. This is due to the fact that, while the synthesis (reconstruction) function is compactly supported, the analysis one is not. Hence, in the continuous convolution described by Eq. (10), there are two functions being infinitely supported.

Equally well, we can try to stay entirely in the discrete domain seeking for a solution minimizing a certain discrete norm. We shall present this alternative in the next subsection.

A. Minimizing l_2 norm

This solution is very well known and described in most of the textbooks considering function approximations, bases, and signal expansions [8,9]. We present it here briefly, in order to use it later as a reference when comparing different solutions.

The discrete norm to be minimized is induced by the following discrete inner product over the grid $\boldsymbol{\tau}$:

$$\langle u, v \rangle \equiv h \sum_{i=0}^{L_{in}-1} u(\tau_i) v(\tau_i) \qquad (19)$$

being an approximation to the continuous inner product. It induces a semi-norm given by

$$||s||_{l^2}^2 = \langle s, s \rangle = h \sum_{i=0}^{L_{in}-1} s(\tau_i) s(\tau_i). \qquad (20)$$

This semi-norm serves in quantifying the distance between the approximation and the initial samples as follows:

$$||s - \tilde{s}||_{l^2}^2 = \min_{y \in V} ||s - y||_{l^2}^2. \qquad (21)$$

Here, s is assumed to be known only on the grid $\boldsymbol{\tau}$, that is, $s(hk) = x(k)$. Assume $\tilde{s}(t)$ is reconstructed by the basis $\varphi(t-i)$, as in (9), i.e. $\tilde{s}(\tau_k) = \sum_{j=0}^{L_{out}-1} c_j \varphi(\tau_k - j) = x(k)$. The solution for the unknown coefficients c_j, $j = 0, 1, ..., L_{out}-1$ is given by the following system of normal equations [8]:

$$\begin{aligned} \sum_{j=0}^{L_{out}-1} \Big(\sum_{k=0}^{L_{in}-1} \varphi(\tau_k - i) \varphi(\tau_k - j) \Big) c_j = \\ = \sum_{k=0}^{L_{in}-1} \varphi(\tau_k - i) x(k) \\ \text{for} \quad i = 0, ..., L_{out} - 1 \end{aligned} \qquad (22)$$

or in the matrix form as follows:

$$(\boldsymbol{\Phi}^T \boldsymbol{\Phi}) \boldsymbol{c} = \boldsymbol{\Phi}^T \boldsymbol{x}, \qquad (23)$$

$$\boldsymbol{c} = (\boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \boldsymbol{x}. \qquad (24)$$

The matrix $[\boldsymbol{\Phi}]_{L_{in} \times L_{out}}$ where $L_{in} \geq L_{out}$, takes the coefficients' L_{out}-dimensional vector $\boldsymbol{c}$ to the reconstructed signal version $\tilde{s}(\tau_k)$ (L_{in}-dimensional). The matrix $[\boldsymbol{P}]_{L_{out} \times L_{out}} = \boldsymbol{\Phi}^T \boldsymbol{\Phi}$ Grammian of the reconstruction matrix $[\boldsymbol{\Phi}]$. Usually, the system given by Eq. (22) is solved by efficient methods for the matrix (pseudo) inversion taking advantage of the band structure of the matrix $\boldsymbol{P}$. These methods include, among others, the Cholesky factorization, and the

Givens or Householder QR decompositions [9]. For example, when performing the Cholesky factorization, the total number of floating point operations is (cf. [9]) $0.5(L_{in}N(N+3)+L_{out}(N-1)(N+1))+N(2N-1)+O(L_{in}+L_{out})$.

B. Minimizing L_2 norm

Remember that the initial signal is a discrete sequence $x(k)$ generated by uniformly sampling a certain *unknown* continuous function $s(t)$ according to $x(k)=s(\tau_k)=s(h_k)$. This makes it impossible to minimize directly the error $||s-\tilde{s}||^2_{L_2}$. Instead, a continuous model can be used:

$$x_s(t)=\sum x(k)K_\varepsilon\Big(\frac{t}{h}-k\Big) \tag{25}$$

for minimizing the error $||x_s-\tilde{s}||^2_{L_2}$. The function K_ε, possibly depending on a certain scale parameter ε, takes the samples $x(k)$ into a continuous function $x_\varepsilon(t)\in L_2$. Having the model given by Eq. (25), one can confidently apply the theory of Subsection III.A, namely, the coefficients $c(i)$ can be determined as in Eq. (10) as follows

$$\begin{aligned}
c(i)&=\langle x_s(\xi)\hat{\varphi}(\xi-i)\rangle=\int x_s(\xi)\hat{\varphi}(\xi-i)d\xi\\
c(i)&=\sum_n p^{-1}(n)\int x_s(\xi)\varphi(i-n-\xi)d\xi \qquad (26)\\
c(i)&=\sum_n p^{-1}(n)\int\Big\{\sum_k x(k)K_\varepsilon\Big(\frac{\xi}{h}-k\Big)\Big\}\varphi(i-n-\xi)d\xi\\
c(i)&=\sum_n p^{-1}(n)\Big\{\sum_k x(k)\int K_\varepsilon\Big(\frac{\xi}{h}-k\Big)\varphi(i-n-\xi)d\xi\Big\}
\end{aligned}$$

Consider the convolution integral

$$\phi(t)=\int k_\varepsilon\Big(\frac{\xi}{h}\Big)\varphi(t-\xi)d\xi. \tag{27}$$

It involves two continuous kernels with different sizes. By properly selecting the kernel K_ε this integral can be solved yielding the following linear model:

$$d(t)=h\sum_k x(k)\phi(t-hk). \tag{28}$$

Then, the procedure continues by sampling the latter at the integers followed by digital filtering with the convolution inverse p^{-1}. For a symmetrical sequence $p(k)$ with length $2N+1$ and $P(z)=\sum_{k=-N}^{N}p(k)z^{-k}$, digital filtering with $p^{-1}(k)$ means an efficient realization of the IIR filter $1/P(z)$ [10]. Fig. 2 shows the entire algorithm.

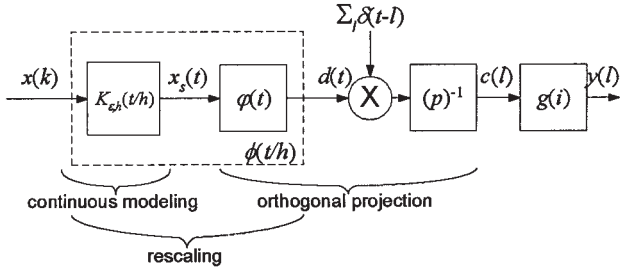


Fig. 2. Continuous LS decimation

1) *Solution of Munoz et al.* A very elegant solution entirely consistent with the continuous domain framework established in Section II has been developed in [4]. Their work concerns primarily B-splines of degree N as basis functions.

The first step in the algorithm is to match an interpolative, *continuous-domain, spline model* to the discrete data $x(k)$ by constraining the new function to have the same values at the coordinate grid τ. Thus, the modeling spline coefficients can be obtained applying Eqs. (3) and (4). This is equivalent to substituting the kernel K_ε by the interpolating function

$$K_\varepsilon(t)=\sum_{i=1}^{m}(b^N)^{-1}(i)\beta^N(t-i). \tag{29}$$

The benefit of the method proposed in [4] is that it involves in the integral given by Eq. (27) two splines. A convolution of two splines (of degrees N and M, correspondingly) is again a spline of degree $N+M+1$ [8,10]. When splines are represented by their B-spline expansion this integration can be separated into low-complexity discrete-time operations such as *finite differences* and *running sums* and the continuous interpolation with the B-spline kernel of degree $N+M+1$ [4]. After some rearrangements in Fig. 2, the final algorithm can be expressed as in Fig. 3 with the following steps [4]:

- Inverse filtering with $(b^N)^{-1}$ for obtaining the initial spline coefficients.
- $(N+1)$ running sums $\Delta^{-(N+1)}$.
- Re-sampling by a factor of $1/h$ using the spline model of degree $2N+1$.
- $(N+1)$ finite differences $\Delta^{(N+1)}$.
- Inverse filtering with $(b^{2N+1})^{-1}$ for generating the new spline coefficients.
- FIR filtering with the kernel b^N for obtaining the resized signal.

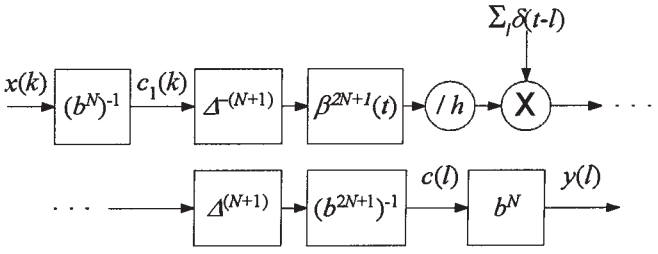


Fig. 3. Continuous LS signal decimation using finite differences

V. A Hybrid Method

To overcome the difficulties in solving the integral given by Eq. (27) we model the kernel K_ε by the Gaussian function

$$K_\varepsilon(t)=e^{-x^2/(4\varepsilon)}/(2\sqrt{\pi\varepsilon}). \tag{30}$$

While this is not an interpolating kernel, it allows us to approximate the integral given by Eq. (27) when $\varepsilon\to 0$, as

$$\lim_{\varepsilon\to 0}K_\varepsilon(t)=\lim_{\varepsilon\to 0}e^{-x^2/(4\varepsilon)}/(2\sqrt{\pi\varepsilon})=\delta(x). \tag{31}$$

It can be though as emphasizing the fact that we are interested in minimizing the error in L_2 sense especially around the given discrete points. This approximation has sense, especially for a relatively small sampling step h, which is the practical case in most of the images of interest. We will illustrate this assertion later by our experiments.

By using the above simplification, we get a particular form of Eq. (28), where $\phi(t)$ is formally replaced by $\varphi(t)$ as:

$$d(t) = h \sum_k x(k)\varphi(t - hk) \tag{32}$$

and

$$c(i) = \sum_n p^{-1}(n)\Big\{h \sum_k x(k)\varphi(i - n - hk)\Big\}. \tag{33}$$

Eq. (33) can be regarded as a form of Eq. (24) where the matrix inversion $(\mathbf{\Phi}^T\mathbf{\Phi})^{-1}$ is substituted by the digital recursive filtering $1/P(z)$.

When the compactly supported piece-wise polynomial functions in the form of Eq. (14) are used, the inner sum in Eq. (33) can be realized very efficiently by using the so-called transposed modified Farrow structure [11,5,12]. We will discuss this realization in the next subsection.

A. *Transposed Farrow structure*

Recall that our basis functions are composed from B-splines up to degree N in a combination preserving the support of the highest degree B-spline. Hence, they are formed of polynomial pieces as in Eq. (17).

By substituting Eq. (17) into Eq. (32) and sampling at the integers we get

$$d(l) = h \sum_k x(k) \sum_{i=0}^{N} \sum_{m=0}^{N} c_m(i)\Big(l + \frac{N+1}{2} - i - hk\Big)^m. \tag{34}$$

Changing the order of the summations results in the following practically realizable form

$$d(l) = h \sum_{m=0}^{N} \sum_{i=0}^{N} c_m(i) \sum_k x(k)\Big(l + \frac{N+1}{2} - i - hk\Big)^m. \tag{35}$$

The innermost sum contains input signal samples inside an interval of the *unity length* weighted by fraction values raised to various powers of m for $m = 0, 1, , N$. Denoting those fractions by

$$\mu_k = |j - \tau_k| \quad \text{for} \quad l \le \tau_k < l + 1 \tag{36}$$

results in the practical implementation scheme shown in Fig. 4, known as the transposed modified Farrow structure [11,5,12]. It contains $N+1$ filters with coefficients $c_m(i)$ determined by the B-splines involved. The sampling rate conversion [innermost summation in Eq. (35)] is made in the accumulator blocks. The outputs of these blocks are used as inputs to the fixed filters when $\mu_k < \mu_{k-1}$. Then, the accumulators are reset for the new sample summation.

This structure realizes exactly the matrix multiplication between the transposed basis sampled at the grid $\boldsymbol{\tau}$ and the signal vector: $\mathbf{\Phi}^T\boldsymbol{x}$. In $\mathbf{\Phi}$, the norm of each basis vector $\varphi(\boldsymbol{\tau} - \boldsymbol{l})$ differs from $1/h$. This would not be a problem if we would realize the l_2 error norm minimization given by Eq. (24). But, formally replacing it by Eq. (35), we have no preservation of the constant in the output. For some re-sampling ratio range this can cause visible periodic texture artifacts in the smooth areas of the reconstructed image. Therefore, we have to take care of the constant preservation in the *intermediate* stage given by Eq. (34) before the IIR filtering. It has turned out the above-mentioned problem can be solved by using the local scaling factors h_l as follows:

$$d(l) = h_l \sum_k x(k)\varphi(l - hk) \tag{37}$$

where

$$(1/h_l) = \sum_k \varphi(l - hk). \tag{38}$$

The local scaling factors h_l depend on $\boldsymbol{\tau}$. When performing uniform re-sampling on images, the same basis and the same grid are used for all column-wise (row-wise) transformations, hence the scaling factors are the same. To obtain them, what is needed is one more re-sampling operation on a vector with the column (row) length composed of unities.

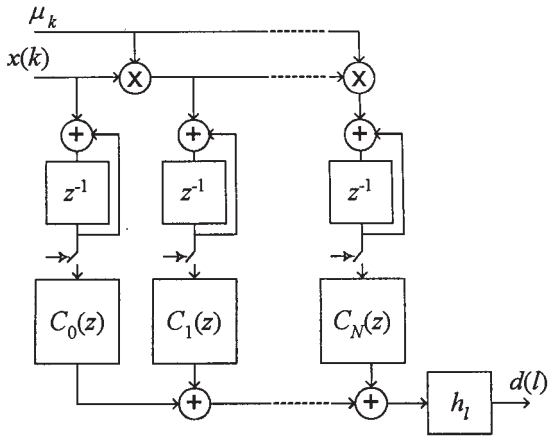


Fig. 4. Transposed modified Farrow structure for piece-wise polynomial signal decimation

B. *Relation with the l_2 solution*

The method can be represented in the following matrix form

$$\boldsymbol{c} = (\boldsymbol{H}\boldsymbol{P})^{-1}\mathbf{\Phi}^T\boldsymbol{x}. \tag{39}$$

Here, the matrix $[\boldsymbol{P}]_{L_{out}\times L_{out}}$ is a band matrix containing $p(k)$ along its rows and $[\boldsymbol{H}]_{L_{out}\times L_{out}}$ is a diagonal matrix containing the scaling factors as given by Eq. (38). To illustrate how close this solution is to the corresponding l_2 solution, the norm or the error matrix

$$\boldsymbol{E} = \mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})^{-1} - (\boldsymbol{H}\boldsymbol{P})^{-1}\mathbf{\Phi}^T \tag{40}$$

has been computed for different values of h. Fig. 5 shows the result for the case of cubic B-spline basis function.

It can be seen that the error between the true discrete LS and the hybrid method is relatively low for decimation ratios up to some value close to 0.9. After this value, the error increases considerably because the proposed model is no longer adequate.

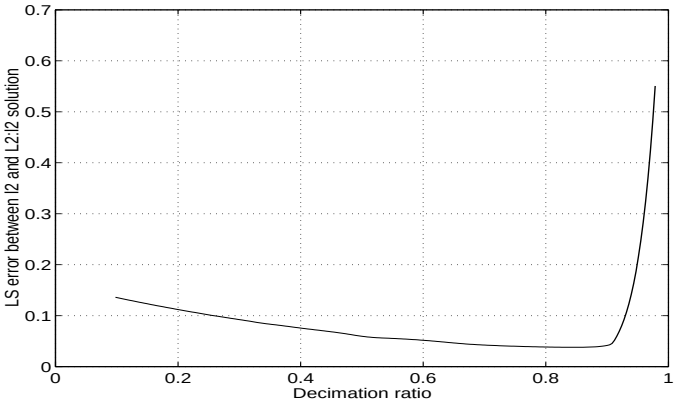


Fig. 5. Error norm $||\boldsymbol{E}||_{l_2}$ for cubic B-spline basis. $L_{in} = 512$

VI. Comparative Analysis of Different Decimation Schemes

A. *Frequency-domain analysis*

The schemes presented above can be considered as forms of continuous-domain filtering of the impulse train $x_(t) = \sum_k x(k)\delta(t/h - k)$. This allows us to compare their (s continuous frequency responses.

The interpolative scheme of Subsection II.A for the grid on the integers has the frequency response given by

$$\Phi_{\text{int}}(\omega) = \Phi(\omega)/\Phi(e^{j\omega}). \tag{41}$$

The continuous LS scheme of Subsection IV.A can be presented in the frequency domain as

$$\Phi_{L_2}(\omega) = \frac{\Phi(\omega)\Phi(h\omega)}{P(e^{j\omega})\Phi(e^{jh\omega})}, \tag{42}$$

where $P(e^{j\omega})$ is the frequency response of the autocorrelation sequence given by Eq. (13).

The new scheme takes in the frequency domain the following rather simple form:

$$\Phi_{l_2}(\omega) = \Phi(\omega)/P(e^{j\omega}). \tag{43}$$

The frequency responses are shown in Fig. 6 for the case of cubic B-spline basis, that is, $\varphi(t) = \beta^3(t)$. As can be seen,

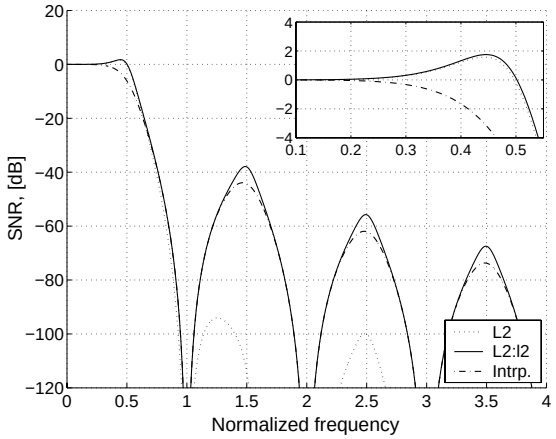


Fig. 6. Continuous frequency responses of different decimation systems

all schemes possess anti-aliasing properties around multiples of $F_{\text{out}} = 1$ and are rather similar in the transition band. The main difference appears in the pass-band where the LS schemes provide better performances especially for frequencies close to 0.5. The bump appearing in this region reflects the duality in decimation and interpolation processes [13].

Next, we compare three different kernels realizing the hybrid LS method. The first kernel is the following linear combination of the cubic and linear B-splines: $\varphi_{\text{mod}}(t) = \beta^3(t) + \sum_{i=-1}^{1} \gamma_{3i}\beta^1(t-i)$ [1]. The second kernel is the following linear combination of the cubic B-spline and its second derivative $\varphi_{moms}(t) = \beta^3(t) + \gamma\frac{d^2\beta^1(t)}{dt^2}$ [2]. The third kernel is the cubic B-spline itself. Their frequency characteristics for some well-optimized γ's are shown in Fig. 7. What is seen is that the modified kernel, initially optimized to possess good anti-imaging properties [1], has a flatter frequency response for the dual function. This, combined with the flatter region of the interpolator near to the cut-off frequency, would give better preservation of the high-frequency details in the processed image.

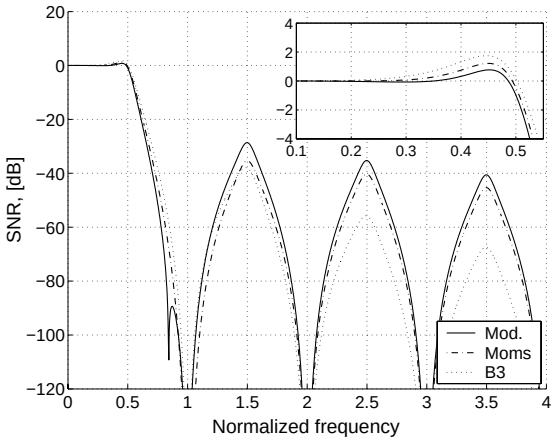


Fig. 7. Continuous frequency responses of systems realizing l_2 LS signal decimation with different kernels. For moms case $\gamma = 1/42$; for modified case $\gamma_{30} = 2\gamma_{31} = -0.0714$

B. *Quality assessment*

We compare the quality by measuring the SNR between the original image and the image resulting from a two-step processing: *decimation* followed by *reconstruction* to the initial size. In these two steps, the respective operators have been chosen to be dual each other, thus complying with the LS paradigm. We have realized also the interpolative scheme described in Subsection II.A.

The well-known *Barbara* and *Lena* test images of size 512×512 have been processed. We changed the target size from 51×51 to 511×511 and built the SNR curves for different methods, as shown in Figs. 8 and 9. We related also the SNRs ?60 of the L_2 solution and our solution (denoted as $L_2 : l_2$) to the pure discrete (l_2) solution. The proposed method performs equally well compared to the original l_2 solution up to a certain target size close to the original image size. After that, it fails. The same is true also for the

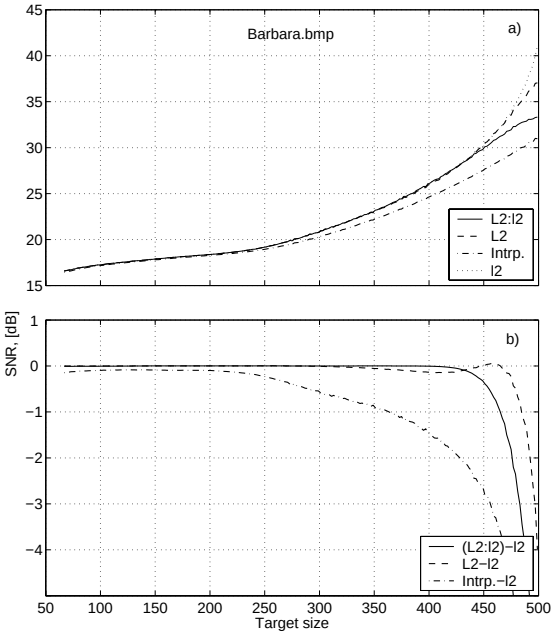


Fig. 8. Fig. 8. a) SNR for different decimation schemes; b) relative SNR to Target size l2 scheme. *Barbara* image processed

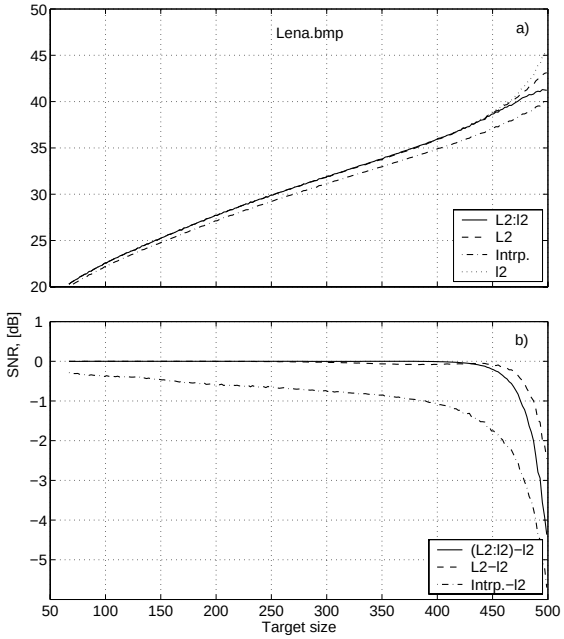


Fig. 9. Same as in Fig. 8 for *Lena* image

L_2 solution when compared to the l_2 solution, although for a higher size, because its corresponding continuous model is more adequate than ours for large and close to unity sampling steps. Although the relative differences to the l_2 solution can be quite large (up to 5 dB), the absolute values of SNR are higher than 32 dB, indicating subtle variances from the original image.

As far as the interpolative method is concerned, it performs always worse than the LS methods. This is less visible for small target sizes, where more aliasing is introduced for the frequency content close to 0.5. However, for moderate

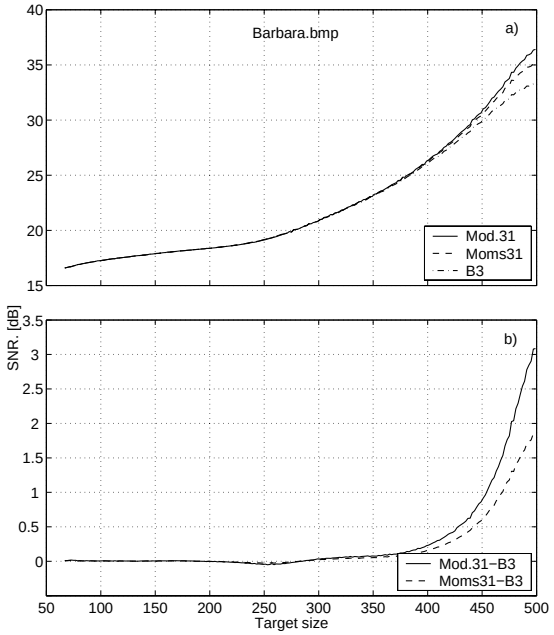


Fig. 10. (a) SNR for different decimation kernels. (b) Relative SNR to the cubic spline kernel. *Barbara* image processed

decimation ratios it is evident that providing the good anti-aliasing properties is not enough. Some better performance in the pass-band is needed and it is provided by the higher order IIR filtering assuring the LS optimality.

Three different kernels are compared in terms of their LS decimation performance in Fig. 10. The benefit from using modified B-spline basis vs. other bases is outlined by the higher SNR, especially for low decimation ratios.

C. Computational complexity assessment

The computational complexity has been measured in terms of the number of multiplications and additions required to process one image row. We refer to Figs. 2 and 3 that help to evaluate the number of operations for each step (see also [5,4] for details).

For a kernel with the highest degree being N and the input

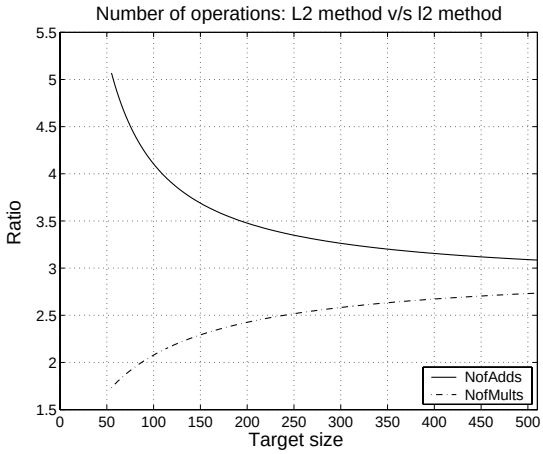


Fig. 11. 12 NOR: L_2 method versus $L_2 : l_2$ method

and output signals of lengths L_{in} and L_{out}, respectively, the numbers are summarized in Tables 1 and 2.

Table 1. Number of operations for L_2 method [4]

Algorithm step		Multiplications	Additions
1.	Pre-filtering	(N 1)/2)(2m-1)	(N 1)/2)(2m 1)
2.	Running sums	—	(N+1)m
3.	Interpolation		
	- Output grid	(2N+1)(2N+2)n	(2N+1)(2N+2)n
	- Filtering	(2N+1)n	(2N+1)n
4.	Running diff.	—	(N+1)n
5.	IIR filtering	N(2n 1)	N(2n 1)
6.	FIR filtering	(N+1)/2)n	(N+1)/2)n
	Total:	(N 1)/2)(2m 1)+ + (4N^2+10.5N+4)n-N	3mN (N 1)/2)+ + (4N^2+11.5N+5)n-N

Table 2. Number of operations for $L_2 : l_2$ method

Algorithm step		Multiplications	Additions
1.	Pre-filtering	(N 1)/2)(2m-1)	(N 1)/2)(2m 1)
2.	Running sums	—	(N+1)m
3.	Interpolation		
	- Output grid	(2N+1)(2N+2)n	(2N+1)(2N+2)n
	- Filtering	(2N+1)n	(2N+1)n
4.	Running diff.	—	(N+1)n
5.	IIR filtering	N(2n 1)	N(2n 1)
6.	FIR filtering	(N+1)/2)n	(N+1)/2)n
	Total:	(N 1)/2)(2m 1)+ + (4N^2+10.5N+4)n-N	3mN (N 1)/2)+ + (4N^2+11.5N+5)n-N

Fig. 11 shows the number-of-operations ratio (NOR) between two methods. As can be seen, the proposed method reduces the computational complexity at least by factor of two in terms of the number of multiplications and at least by a factor of three in terms of the number of additions. This reduction is mainly due to the efficient computational structure discussed in Subsection V.A and being a result of the compromise made when considering the continuous signal model given by Eq. (32).

In order to be entirely consistent with the continuous LS, the method described in [4] demands a pre-filtering to get the continuous spline model and an interpolation with higher degree spline. The proposed $L_2 : l_2$ method, in turn, helps to lower the computations by compromising between continuous norm minimization and its discrete counterpart. The L_2 method also involves running sums in its second step. While this procedure can be realized rather easily, it should be maintained carefully because it risks introducing overflow or round-off errors, especially for large size images. Munoz *et al*. [4] have successfully avoided those potential problems by introducing some small amount of additional computations (not included in the comparison above). As far as our method is concerned, it is not risky in this sense. We have already commented the need of local normalization that, in the case of images, is not so costly additional operation. The computational complexity of our method can be reduced even further by exploiting the symmetries of the sub-filters $C_m(z)$ in the transposed Farrow structure of Fig. 4 (see [5] for a detailed analysis).

VII. Conclusions

In this paper, we reviewed the recent advances in performing image resizing and presented a novel method based on a hybrid form of the LS. By taking the limit case in our continuous model, we managed to simplify the L_2 norm minimization problem and to make it resembling to its discrete counterpart. However, the new solution is not the classical discrete LS since, in fact, we replaced the matrix inversion – the most costly operation in the discrete case – by a more efficient IIR filtering. We commented some important details in realizing the transposed Farrow structure, namely the need for a local scaling (normalization). We showed that our method dramatically lowers the computational complexity when the quality remains practically the same as in the classical discrete case for a wide range of decimation ratios. Our method competes successfully in quality with the state-of-the-art method developed in [4] while being at least twice faster. We compared also the performance of a number of spline-based kernels, all of them having the same computational complexity, but different approximation or anti-aliasing properties. It was demonstrated that by properly optimizing the basis functions, the proposed solution can be made very close to the classical l_2 solution.

Acknowledgement

This work was supported by the Academy of Finland, project No. 44876 (Finnish Centre of Excellence Program).

References

[1] A. Gotchev, K. Egiazarian, and T. Saramäki, "Optimization Techniques in Designing Piece-wise Polynomial Interpolators of Minimal Support", *Int. Journal Wavelets, Multiresolutiom and Information Processing*, in press.

[2] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-Order Interpolation of Minimal Support", *IEEE Trans. Image Proc*., vol. 10, no. 7, pp. 1069-1080, July 2001.

[3] Yu-Ping Wang, "Scale-space Derived from B-spline", *IEEE Trans. PAMI*, 1998.

[4] A. Munos, T. Blu, and M. Unser, "Least-Squares Image Resizing Using Finite Differences", *IEEE Trans. Image Proc.*, vol. 10, No.9, pp. 1365-1378, September, 2001.

[5] D. Babic, J. Vesma, T. Saramäki, and M. Renfors, "Implementation of the transposed Farrow structure", in *Proc. Int. Symposium Circuits and Systems, ISCAS'2002*, Phoenix, Arizona, USA, May 2002, vol. IV, pp. 5-8.

[6] A. Aldroubi and M. Unser, "Sampling procedures in function spaces and asymptotic equivalence with Shannon's sampling theory", *Numer. Funct. Anal. Optim*. vol. 15, No.1/2, pp. 1-21, 1994.

[7] A. Ron, "Factorization theorems for univariate spline on regular grids", *Israel Journal of Mathematics*, vol. 70, No.1, pp. 48-68, 1990.

[8] C. de Boor, *A Practical Guide to Splines*, Springer-Verlag, Second Edition, 2001.

[9] A. Björk, *Numerical methods for least squares problems*, SIAM, 1996.

[10] M. Unser, A. Aldroubi, M.Eden, “B-spline signal processing: Part I – Theory and Part II – Efficient design and applications”, *IEEE Trans. Signal Processing*, vol. 41, pp. 821-848, 1993.

[11] A. Gotchev, J. Vesma, T. Saramaki, and K. Egiazarian “Multiscale image representations based on modified B-splines”, in *Proc. Int. Workshop on Spectral Techniques and Logic Design*, SPEGLOG’2000, Tampere, Finland, June 2000, pp. 431-452.

[12] T. Hentschel, and G. Fettweis, “Continuous-time digital filters for sam-ple-rate conversion in reconfigurable radio terminals”, *Proc. European Wireless*, Dresden, Germany, Sept. 2000, pp. 55-59.

[13] M. Unser, A. Aldroubi and M. Eden, “Polynomial spline signal approximation: Filter design and asymptotic equivalence with Shannon’s sampling theorem”, *IEEE Trans. Inform. Theory*, vol. 38, pp. 95-103, January 1992.

Direct Design of Transitional Butterworth-Chebyshev IIR Filters

Nikolić V. Saša[1] and Vidosav S. Stojanović[1]

Abstract – **In this paper the procedure for direct design of selective IIR digital filters has been described. These filters can not be obtained using transformations from analogous domain. The complete procedure for determining of coefficients of the magnitude squared characteristics has been presented. Slope of the amplitude characteristics has been also investigated. These filters provide much sharper cutoff slope characteristics, especially in the design of narrow band lowpass filters.**

Keywords – **IIR filters, Transitional filters, Minimax amplitude characteristics, Cutoff slope**

I. Introduction

Digital filters are playing an important and increasing role in design of modern telecommunication systems. Recursive digital filters are mostly used in design of selective amplitude characteristics. Transfer function of these filters can be obtained by indirect and direct methods. Indirect methods are well known, but it is not possible to design all kind of filters. In this paper the direct design of transitional Butterworth-Chebyshev digital filters with selective amplitude characteristics will be described.

The direct design of transitional Butterworth-Chebyshev digital filters with all zeroes at the origin has been described in [1]. The procedure for introducing single or multiple zero pairs on the unit circle directly in z domain has been described in [2]. By this way it is possible to obtain more selective amplitude characteristic. The introducing of zeros on the unit circle results in loosing of mini-max characteristic of the attenuation in the passband.

In order to improve again selectivity, the procedure for direct design of transitional filters with equal pass-band ripples, where characteristic function is obtained using simple iterative procedure, has been proposed in this paper. Obtained filters have increased selectivity if we compare them with those described in [2]. These filters can not be obtained using transformations from analogous domain.

II. Approximation

Magnitude squared functions of a transitional Butterworth-Chebyshev filter with m pairs of identical zeros at $e^{\pm j\omega_z T}$ is given by

$$|H_{n,k,m}(e^{j\omega T})|^2 = \frac{1}{1+\varepsilon^2 K_n^2(x)}, \tag{1}$$

where ε is a real constant, x is a frequency variable in ωT suitable for low-pass filtering given by

$$x = \frac{\sin\frac{\omega T}{2}}{\sin\frac{\omega_c T}{2}} \tag{2}$$

and x_z is multiple zero on the unit circle

$$x_z = \frac{\sin\frac{\omega_z T}{2}}{\sin\frac{\omega_c T}{2}}, \tag{3}$$

where $\omega_c T$ is cutoff frequency and $\omega_z T > \omega_c T$ and $n \geq 2m$ where parameter n is $n = k+l$ and m is a number of multiple zero pairs on the unit circle at $w_z T$.

Characteristic function $K_n(x)$ is a rational function and can be written as

$$K_n(x) = P_n(x)\Big(\frac{x_z^2-1}{x^2-x_z^2}\Big)^m, \tag{4}$$

where $P_n(x)$ is a real polynomial equal to

$$P_n(x) = x^l(a_0 + a_2x^2 + \cdots + a_kx^k), \tag{5}$$

where parameter k is always odd, while parameter l is natural number and for even l one can get even order of filter and for odd l odd order of filter.

The introduction of the zeros on polynomial transfer function will improve the amplitude characteristic and the cutoff slope but Chebyshev filter will loose minimax amplitude characteristic in the passband. In order to keep minimax characteristic in the passband it is necessary to find new coefficients of the polynomial $P_n(x)$.

Coefficients a_i of real polynomial $K_n(x)$ are obtained so that the characteristic function $K_n(x)$ has minimax characteristic in the passband. These coefficients can be calculated using an iterative procedure. More information about this calculation of coefficients of the polynomial $P_n(x)$ can be found in [5].

In table 1 coefficients of the polynomial $P_n(x)$ for two pairs of zeros on the unit circle in $\omega_z = 0.45\pi$, for different values of parameters k and l, are displayed.

When the polynomial $P_n(x)$ is determined, than the magnitude squared function can be written in the more convenient form

$$|H_{n,k,m}(e^{j\omega T})|^2 = \\ = \frac{1}{1+\varepsilon^2 x^{2l}(b_0+\cdots+b_{2k}x^{2k})\Big(\frac{x_z^2-1}{x^2-x_z^2}\Big)^{2m}}, \tag{6}$$

[1]The authors are with University of Niš, Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: caci@elfak.ni.ac.yu

Table 1. Coefficients of polynomial $P_n(x)$, $n = 8$, $k = 0$, $m = 2$ and $\omega_z T = 0.45\pi$

l	0	2	4	6	8
k	8	6	4	2	0
a_0	3.825	-56.495	62.314	-17.508	1
a_2	-93.736	294.355	-155.343	18.508	—
a_4	399.622	-457.631	94.030	—	—
a_6	-573.442	220.771	—	—	—
a_8	264.732	—	—	—	—

where coefficients b_i can be obtained using the next simple relation

$$b_{2i} = \sum_{j=0}^{i} a_{2j} a_{2i-2j} \quad \text{for} \quad i = 0, ..., k. \tag{7}$$

After some rewriting the magnitude squared function can be written as

$$|H_{n,k,m}(e^{j\omega T})|^2 = \frac{(x^2 - x_z^2)^{2m}}{(x^2 - x_z^2)^{2m} + c_{2n}x^{2n} + \cdots + c_{2l}x^{2l}}, \tag{8}$$

where coefficients c_i can be obtained from the following relation

$$c_{2i} = (x_z^2 - 1)^{2m} \varepsilon^2 b_{2(i-1)}, \quad \text{for} \quad i = l, ..., n. \tag{9}$$

Substituting $x^2 = (zl)^2/(-4\alpha z)$ in (8) we obtain the next function

$$G(z) = \Big(\frac{(z-1)^2}{-4\alpha z} - x_z^2\Big)^{2m} \Bigg[\Big(\frac{(z-1)^2}{-4\alpha z} - x_z^2\Big)^{2m} + \\ + c_{2n}\Big(\frac{(z-1)^2}{-4\alpha z}\Big)^{n} + \cdots + c_{2l}\Big(\frac{(z-1)^2}{-4\alpha z}\Big)^{l} \Bigg]^{-1}, \tag{10}$$

which is equal to $|H_{n,k,m}(e^{j\omega T})|^2$ when evaluated along the unit circle and $\alpha = \sin^2(\omega T/2)$.

After some calculations in equation (10) the function $G(z)$ obtains form

$$G(z) = \frac{z^{n-2m}[z^2 + 2(2\beta - 1)z + 1]^{2m}}{z^{n-2m}[z^2 + 2(2\beta - 1)z + 1]^{2m} + F(z)}, \tag{11}$$

where parameter β is $\beta = \sin^2(\omega_z T/2)$ and

$$F(z) = f_0 z^{2n} + \cdots + f_n z^n + \cdots + f_0. \tag{12}$$

The function $F(z)$ is mirror-image polynomial because a sum of a mirror-image polynomial of order s and a mirror-image polynomial of order t multiplied by a mirror-image polynomial of order st is also a mirror-image polynomial of order s.

Coefficients f_i of the mirror-image polynomial $F(z)$ are given by

$$f_{2n-1} = (-4\alpha)^{2m} \sum_{j=0}^{2n-1} \frac{(-1)^j c_{2(i+j-n)} \binom{2(i+j-n)}{j}}{(-4\alpha)^{i+j-n}}, \\ \text{for} \quad i = n, n+1, ..., 2n. \tag{13}$$

The first part in the nominator of the transfer function (11) is also a mirror-image polynomial with multiple zeros on the unit circle and can be written in the extended form

$$[z^2 + 2(2\beta - 1)z + 1]^{2m} = \\ = z^{4m} + \cdots + e_{2m}z^{2m} + \cdots + 1, \tag{14}$$

where coefficients e_i can be obtained from the next relation

$$e_i = \sum_{j=0}^{i} \binom{2m}{j} \binom{4m-2j}{i-j} (-1)^j (4\beta)^j, \\ \text{for} \quad i = 0, ..., 2m. \tag{15}$$

Collecting the same order coefficients of the mirror-image polynomials (12) and (14) we finally have

$$G(z) = \frac{z^{n-2m}[z^2 + (4\beta - 2)z + 1]^{2m}}{g_0 + \cdots + g_n z^n + \cdots + g_1 z^{2n-1} + g_0 z^{2n}}, \tag{16}$$

where coefficients g_i are given by

$$g_i = \begin{cases} f_i & \text{for} \quad i = 0, ..., n - 2k - 1 \\ f_i + e_{i-n+2k} & \text{for} \quad i = n - 2k, ..., n. \end{cases} \tag{17}$$

Equating the denominator of (16) with zero, the roots occur in reciprocal pairs. Now, it is easy to find poles because the poles of the transitional filters are merely roots inside the unit circle.

III. Results of Approximation

Normalized frequency response of the new filter of eighth order and one pair of zeros at $\omega_z T = 0.45\pi$ is displayed in Figure 1. Normalized cutoff frequency is $\omega_c T = 0.3\pi$, and maximal passband attenuation is $A_{\max} = 1$ dB. Passband attenuation ripples are equalized. As parameter l is lower, the selectivity of filter is improved. Minimal stopband attenuation significantly differs and depends from the value of the parameter l, so for $l = 8$ minimal stopband attenuation is

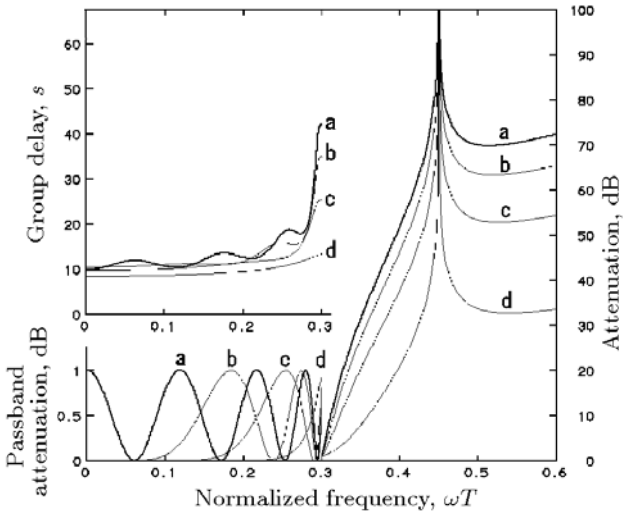


Fig. 1. Characteristics of transitional filters for different values of parameters l; $m = 1$, $\omega_c T = 0.3\pi$, $A_{\max} = 1$ dB; a) $l = 0$, b) $l = 4$, c) $l = 6$ and d) $l = 8$

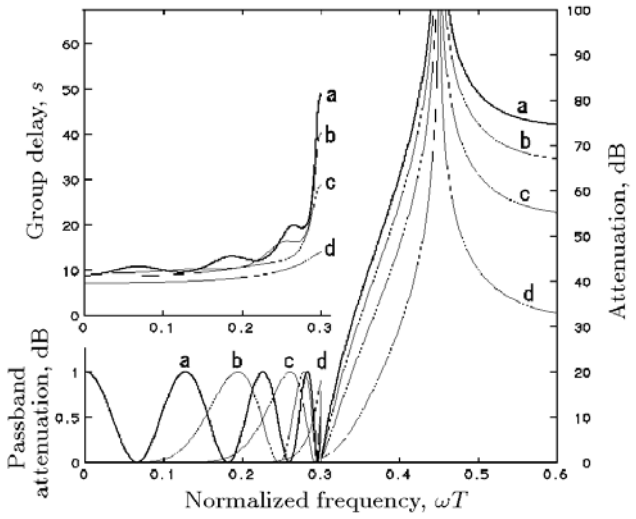


Fig. 2. Characteristics of transitional filters for different values of parameters l; $m = 2$, $\omega_c T = 0.3\pi$, $A_{\max} = 1$ dB; a) $l = 0$, b) $l = 4$, c) $l = 6$ and d) $l = 8$

near 32 dB, while for $l = 0$ it is slightly less than 70 dB. Minimums of stopband attenuation are all at the same value of normalized frequency $\omega_p T$. Group delay characteristics has been also displayed and it is obviously that the characteristic becomes flatter increasing parameter l. For $l = 8$ the group delay is in the range between 9 s and 11 s and for $l = 0$ it is between 10 s and 42 s.

Influence of the multiplicity order of zeroes m at the unit circle to the frequency characteristic of filter is shown in Figure 2. The frequency characteristic of the filter with two pairs of zeros at the unit circle in $\omega_z = 0.45\pi$ is displayed. In order to compare frequency characteristics, normalized cutoff frequency and maximal passband attenuation are the same as in the previous example. Increasing multiplicity order m, selectivity also increases and minimal stopband attenuation is bigger. For $l = 0$ minimal stopband attenuation is more than 75 dB, while for $l = 8$ it is about 34 dB. Group delay characteristic has more deviation comparing with previous example. For $l = 8$ group delay is now in the range from 8 s until 12 s and for $l = 0$ it is between 9 s and 48 s.

Differentiating (1) in ωT, it is obviously that the slope of the amplitude characteristics in the frequency $\omega T = \omega_c T$ is given by

$$S_{n,k} = \frac{-\varepsilon^2 \sum_{i=0}^{k} a_i \Big\{ \sum_{i=0}^{k} (i+l) a_i + \dfrac{2m}{x_z^2 - 1} \sum_{i=0}^{k} a_i \Big\}}{2 \tan\left(\dfrac{\omega_c T}{2}\right) \sqrt{(1+\varepsilon^2)^3}} \qquad (18)$$

Sum of coefficients of the polynomial $P_n(x)$ depends from multiplicity order of introduced zeroes and has value $(1)^m$. It means that for even m it is equal to 1, and for odd m it is equal to 1.

The cutoff slope of the amplitude characteristic depends from the width of the passband and it is smaller if the passband is wider. If the normalized value of the passband is π the cutoff slope is equal to zero. From above reasons this approximation is more suitable for design of narrow band low-pass digital filters. Introducing simple or multiple zero pairs on frequency $\omega_z T > \omega_c T$ it is possible to obtain satisfactory value for the slope of the amplitude characteristics on the cutoff frequency even for the low pass digital filters with the wider passband.

We can conclude from equation (18) that if the cutoff frequency $\omega_c T$ and the order of filter n are known, we can choose the cutoff slope changing four different parameters. These parameters are:

- Parameter l (i.e. parameter k),
- Zero of the transfer function $\omega_z T$,
- Multiplicity order of zero pairs m and
- Maximal passband attenuation $A_{\max}$.

The cutoff slope will be increased if parameter l is lower, $\omega_z T$ is closer to the cutoff frequency $\omega_c T$, multiplicity of the zero pairs m is bigger and if the maximal passband attenuation is bigger.

This conclusion can be confirmed from Figure 3. Dependency of the slope of the amplitude characteristics from the passband value $\omega_c T$ for transitional filters with $m = 2$ zero pairs on the unit circle at $\omega_z T = \omega_c T + 0.1\pi$ has been shown in this Figure. The order of filters is $n = 8$. The cutoff slope decreases if the passband is wider. It confirmed our conclusion that this type of filters is more suitable for design of narrow band low pass digital filters.

In order to check influence of zero pairs on the unit circle on the cutoff slope, in Figure 4 is displayed cutoff slope dependency from value of the $\omega_z T$. We can see that if the $\omega_z T$ is closer to the boundary of the passband, the cutoff slope will be increased. The cutoff slope will increase also if the parameter l becomes lower. The cutoff slopes of filters for $l = 0$ and $l = 2$ are very similar and it is a consequence of the similar frequency characteristics.

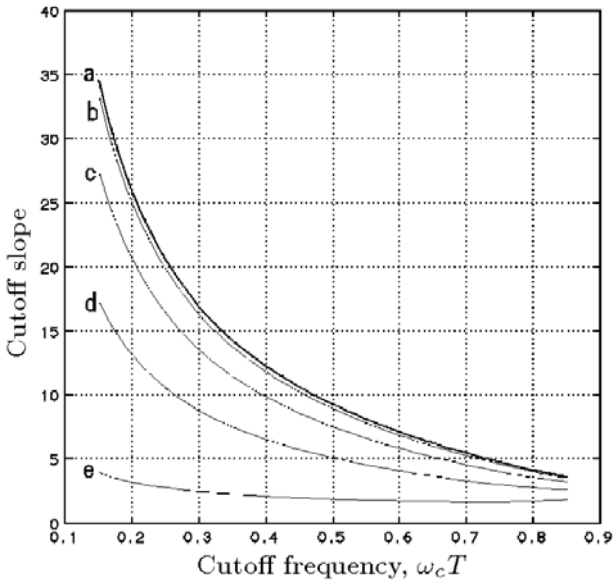


Fig. 3. Cutoff slope characteristics for different values of cutoff frequency $\omega_c T$; $A_{\max} = 1$ dB; a) $l = 0$, b) $l = 2$, c) $l = 4$, d) $l = 6$, e) $l = 8$

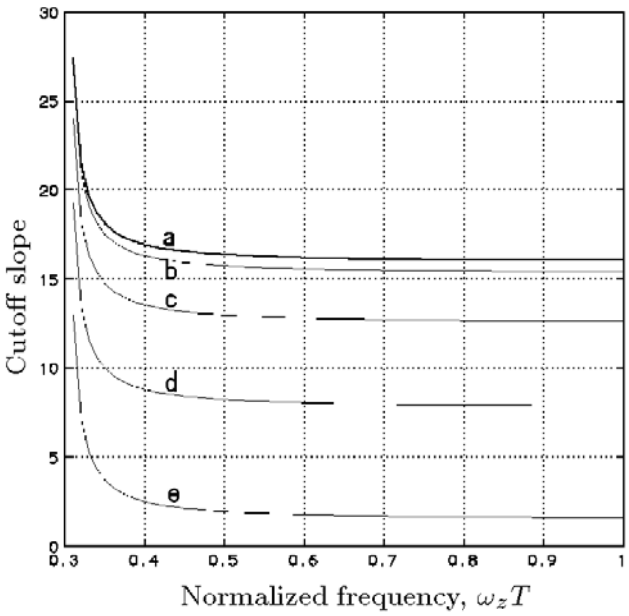


Fig. 4. Cutoff slope characteristics for different values of of parameter $\omega_c T$; $A_{\max} = 1$ dB; a) $l = 0$, b) $l = 2$, c) $l = 4$, d) $l = 6$, e) $l = 8$

IV. Conclusion

In this paper a procedure for design of transitional Butterworth-Chebyshev digital filters with a single or multiple zero-pairs on the unit circle and mini-max amplitude characteristic in the passband has been described. These filters can not be obtained using transformations from analogous domain. Proposed procedure is suitable for design of narrow band low pass digital filters.

References

[1] V.S. Stojanović, S. V. Nikolić, "Direct design of transitional Butterworth-Chebyshev recursive digital filters", *Electronics Letters*, Vol. 29, No. 3, (1993), pp. 286-287.

[2] V.S. Stojanović, S.V. Nikolić, "Direct design of sharp cut-off low pass recursive digital filters", *International Journal of Electronics*, vol.85, No.5, 1998, pp. 589-596.

[3] C.M. Rader, B. Gold, "Digital filter design techniques in the frequency domain", *Proceedings of IEEE*, vol. 55, February 1967, pp. 149-171.

[4] J.J. Soltis, M.A. Sid-Ahmed, "Direct design of Cheby-shev type recursive digital filters", *International Journal of Electronics*, 1991, vol. 70, pp. 413-419.

[5] S.V. Nikolić, "Direct design of selective recursive digital filters", *Master Thesis*, Faculty of Electronic Engineering, Niš, 1996.

A Turbo Codes Delay Reduction Based on Probability Density Evolution

Zafir Popovski[1], Tatjana Ulcar-Stavrova[2]

Abstract – **The turbo codes [1] are decoded in a iterative decoding scheme [2] performing a predefined number of iterations before the final decision comes up. Since the method is time consuming, stopping rules can be apply to prematurely quit the process, for the decoder has already done its job. This technique requires a reasonable trade-off which should results in a average decoding speed increase while not sacrificing the decoder performance.**

Keywords – **Stopping rules, number of iterations**

I. Introduction

The turbo codes (TC) are a class of FEC (forward error control) codes, known as parallel concatenation of two or more recursive systematic convolutional codes (RSCC), produced by a turbo encoder composed of two or more constituent RSC encoders (CEs) with input to each but one constituent encoder permuted by an interleaver of length N.

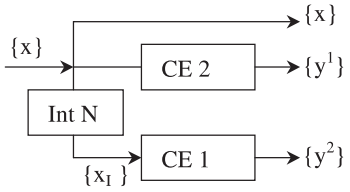


Fig. 1. Turbo encoder structure

Such a composition allows for replacement of the optimal but rather complex maximum likelihood decoding algorithm by two or more relatively simple constituent decoders to decode corresponding constituent codes one by one.

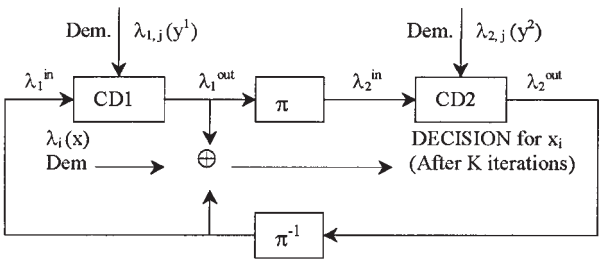


Fig. 2. Iterative TD with two MAP CDs

[1]Zafir Popovski is with the Faculty of Electrical Engineering, "St.Cyril and Methodius" University – Skopje, E-mail: zpopovski@yahoo.com

[2]Tatjana Ulcar-Stavrova is with the Faculty of Electrical Engineering, "St.Cyril and Methodius" University – Skopje, E-mail: tanjaus@etf.ukim.edu.mk

A simple SISO (Soft-In Soft-Out) maximum a posteriori (MAP) decoder [3] which minimise the probability of bit error appears to be the best solution for component decoders (CDs).

We consider only TCs with two CEs for which the coding and decoding processes are shown on Figs. 1 and 2, respectively.

The MAP algorithm provide as an output a real number which is a measure of the probability of error in decoding a particular bit. Since both CDs decode the same information bit xi, coded twice but in different order, it becomes possible this extra information, named extrinsic information, li, to be passed as input to the second CD allowing it to improve its own output extrinsic which will be passed to the first CD in the next iteration. The process is than iterated until reaching a satisfactory degree of confidence regarding the received noisy examples contained in length N sequence. This "satisfactory degree of confidence" does not depend on the pre-set number of iterations and can be reached in any of them. So, the idea is to follow the extrinsic's density evolution until it reaches a "confidential threshold" and then cease the iterative decoding process.

In the follow-up we first briefly present some published rules and a "magic genie rule" [4] used as a benchmark for other rules to be compared with. Later we adopt a model for probability density evolution of the extrinsic information suitable to impose a confidential threshold and then explore few threshold values and compare them with known results through a "C^{++}" turbo code simulation programme. The criteria for comparison are decoding speed, BER and FER, and computational complexity. At the end a conclusion is given.

II. Stopping Rules

Basically, all the stopping rules, by checking the stopping condition at the end of each or each half iteration, attempt to determine the moment when a frame can be reliably decoded. If the condition is true the iterative process is terminated. Otherwise, it continues with the next iteration and, if the stopping condition is never met, stops after a pre-set maximum number of iterations to prevent an endless loop.

Mainly, the stopping rules that cause the decoder to use a variable number of iterations are divided in two main types, hard and soft decision rules, both of which provide for a more or less easy computation based on the data available during the decoding process.

Hard decision (HD) rules evaluate the tentative decoded bits (hard bit decisions) at the end of each or half iteration

and the decision is taken after detecting a consistency in two, three or more successive full or half iterations.

Soft decision (SD) rules are based on comparing a metric on bit reliabilities (soft bit decisions) with a threshold. Suitable metrics used in [4] are the average and minimum reliabilities of the information bits.At each iteration, the turbo coder computes the relevant metric and compares it with preset threshold value.

Our rule belongs to the SD type rules but the turbo decoder has to perform less computations than in any previously published SD rule because it uses the actual density evolution of exchanged extrinsics that contribute to the information bit reliabilities. Since the extrinsics density evolves from iteration to iteration, the turbo decoder's CPU only needs to compare this evolution with a pre-set threshold. The only problem we experienced is to conduct a comparison after the first iteration, for some skewness in the density evolution appears on the first constituent decoder's noisy output which might result in falsely decoded bits at the end of the first iteration. This mainly due to the lack of a extrinsic information for the first constituent decoder at the first iteration.

The "magic genie" rule is an unrealisable (in reality) rule useful for establishing an unbeatable performance benchmark against which the other rules are measured. For this rule, the correct decoded codeword is immediately recognised due to the foreknowledge of the transmitted bit sequence and it stops the iterative process in exactly the minimum number of iterations required to produce the correct codeword.

III. Density Evolution Model

The iterative decoder can be considered as a non-linear dynamical feedback system as shown on Fig. 2.

Extrinsic information messages λ_i are passed from one to the other constituent decoder. The message λ_i measures the log-likelihood ratio for the i-th bit based on input massages λ_j from all other bits but the i-th. So, if we assume that the all-zero codeword is transmitted (with BPSK modulation it corresponds to transmission of "+1"s on the channel) then a positive value of the extrinsic information, $\lambda_i > 0$, for each i, will represent a favourable evidence toward determining the true value of the i-th bit.

When the interleaver is large and random, the extrinsics $lambda_i$ are independent and identically distributed with probability density function $f(\lambda)$. As shown in [5], this pdf is consistent ($\lambda = \log[f(\lambda)/f(-\lambda)]$), its mean, $\mu = E(\lambda)$, is discrimination between the two densities $f(\lambda)$ and $f(-\lambda)$, and the error probability, $e = \Pr\{\lambda < 0\}$, can be evaluated as $e = \mathrm{E}\{1/(1 + e^{|\lambda|})\}$. Computed histograms of the λ_{in} and λ_{out} extrinsics at the input and output of a SISO MAP decoding module for a 4 states, rate 1/3, $[1,5/7]_8$ turbo code are plotted on Fig. 3 for a number of iterations. As it can be seen, the empirical probability densities $f(\lambda_{\text{in}})$ and $f(\lambda_{\text{out}})$ evolve with successive decoder iterations from narrow densities concentrated nearby $\lambda = 0$, to broader Gaussian-shaped densities with increasing means as the iterations continue. Ignoring some irregularities at the beginning of the process,

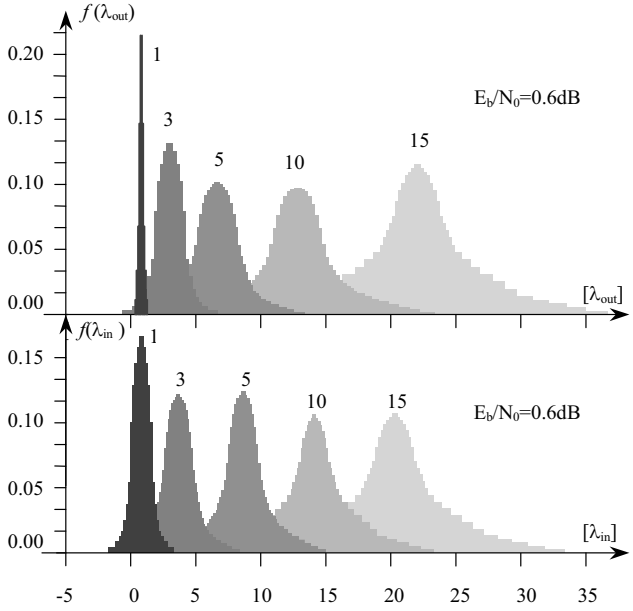


Fig. 3. Evolution of the input and output extrinsics

this probability density function can be approximated by a Gaussian density in which case its statistics depend on two parameters: its mean $\mu = \mathrm{E}(\lambda)$ and its variance $\sigma^2 = \mathrm{Var}(\lambda)$.

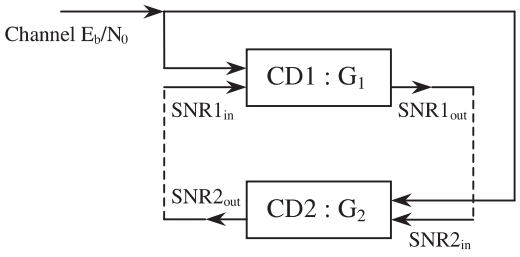


Fig. 4. Analytical density evolution model

A signal-to-noise ratio for such random variable can be defined as SNR $= \mu^2/\sigma^2$ but since it is both Gaussian and consistent, then $\sigma^2 = 2\mu$ and, consequently, SNR $= \mu/2$. This evaluation gives the best approximation of the empirically measured variance.

Now we can observe the input and output SNRs for each decoder denoted as SNR1_{in}, SNR1_{out}, SNR2_{in}, and SNR2_{out}, at each iteration as shown on Fig. 4. A non zero E_b/N_0 from the channel helps CD1 to produce a non zero SNR1_{out} for the extrinsic information despite starting with $\text{SNR1}_{\text{in}} = 0$. So, for given value of E_b/N_0 the output SNR of each CD is a non-linear function of its input, denoted as G1 for CD1 and G2 for CD2. Thus we have:

$$\text{SNR1}_{\text{out}} = \text{G1}(\text{SNR1}_{\text{in}}, E_b/N_0) \quad (1)$$

$$\text{SNR2}_{\text{out}} = \text{G2}(\text{SNR2}_{\text{in}}, E_b/N_0) \quad (2)$$

From the Fig. 5 it follows that $\text{SNR2}_{\text{in}} = \text{SNR1}_{\text{out}}$, so we have

$$\text{SNR2}_{\text{out}} = \text{G2}(\text{G1}(\text{SNR1}_{\text{in}}, E_b/N_0), E_b/N_0) \quad (3)$$

The G1 and G2 functions can be evaluated either directly from the histogram of output λ's from the previous decoder

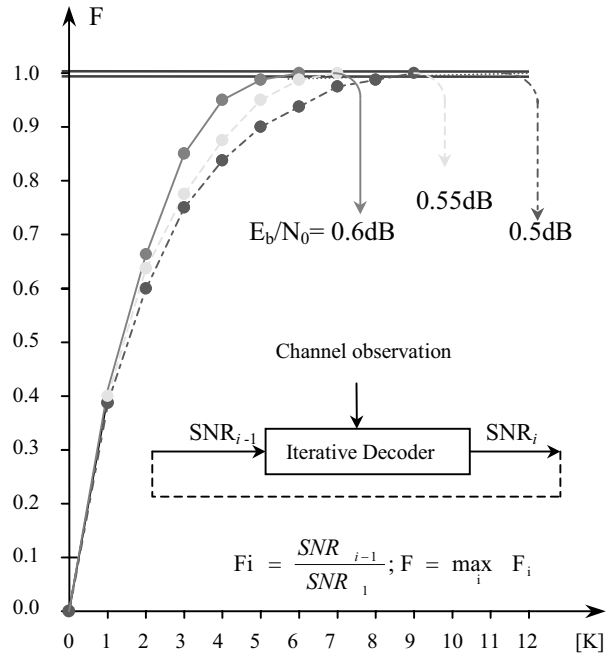


Fig. 5. Noise figure versa the number of iterations

or to generate input λ's from a consistent Gaussian density with mean μ and variance 2μ. In our simulations we use the former model to suit our intention and SNRs are computed from the actual λ-histograms as $\mathrm{E}\{\lambda\}/2$.

The decoder convergence is assessed by measuring the change in the extrinsic information's SNRs from one to the next iteration which allows for defining a noise figure F as a ratio of the input SNR of CD1 at the beginning of the iteration to the output SNR of CD2 at the end of the iteration,

$$F_i = \mathrm{SNR}_i 1_{\mathrm{in}} / \mathrm{SNR}_i 2_{\mathrm{out}} = \mathrm{SNR}_{i-1} / \mathrm{SNR}_i \qquad (4)$$

So, the noise figure is bounded by 0 dB and, if its value at a given iteration is less than 1, this indicates an improvement in the SNR of the extrinsic information from the beginning to the end of a iteration. If that is the case for the entire range of CD1's input SNRs, then, according to [6], the turbo decoder will converge to the correct codeword. The increase of F is plotted on Fig. 5 for the rate 1/3, $N = 128$, [1,(5/7,5/7)] turbo code, for some values of E_b/N_0.

Let now observe a typical variation of codeword error rate, or frame error rate (FER), with the number of iterations, for the same code at $E_b/N_0 = 0.6$ dB, given on Fig. 6.

A careful observation allows for drawing an analogy between Figs. 5 and 6. We can see on Fig. 6 that a FER of about 3×10^{-3} is achieved with 10 iterations and a further increase of the number of iterations do not significantly reduce this error rate, reaching a value of about 1×10^{-3} at 20 iterations. However, below 10 iterations, the difference is great – one order of magnitude between 6 and 4 iterations. The conclusion is that after the 10th iteration, the decoder is wasting effort (and time) by continuing to iterate on about 95% of frames that are already decoded by then.

On Fig. 5, there is a dramatic increase of the F value in approximately first $6 \div 8$ iterations, reaching $F > 0.9$. The following next iterations allow only for a negligible increase. So, this $F \geq 0.9$ might be appropriate values for a threshold to cut off further iterations because the decoder has already gained enough reliability to come to a correct decision.

IV. Simulation Results

The performance results for this new stopping rule are presented in this section. We simulated rate 1/3 turbo codes with very short block size, $N = 128$ bits, with $K = 10$ iterations, for some threshold values chosen to cover a reasonable range of undetected frame error rates (FER). The performance for each threshold is then compared with the performance of the magic genie rule and the performance of the turbo decoder operating with fixed 10 iterations.

Fig. 7 shows the results concerning the average number of iterations per decoded frame as a function of the bit signal to noise ratio, E_b/N_0, for the chosen threshold values.

From these plots we can conclude that the average number of iterations is roughly between 4 and 6 for E_b/N_0 near the so called "waterfall" region where the bit error rate (BER) changes most abruptly. The number of iterations for each threshold matches the corresponding number of iterations on Fig. 5. Es expected, the average decoding speed increases

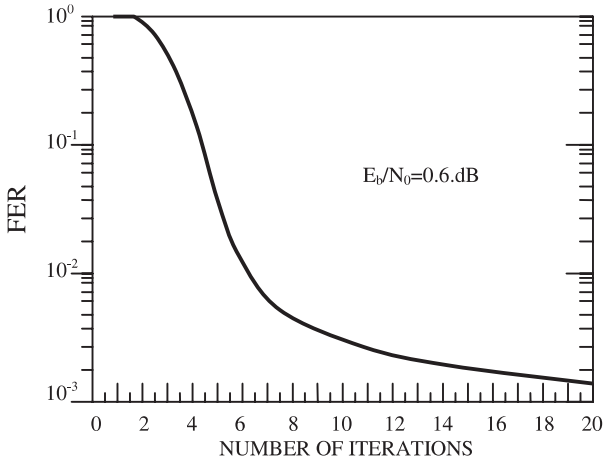


Fig. 6. Variation of FER with the number of iterations

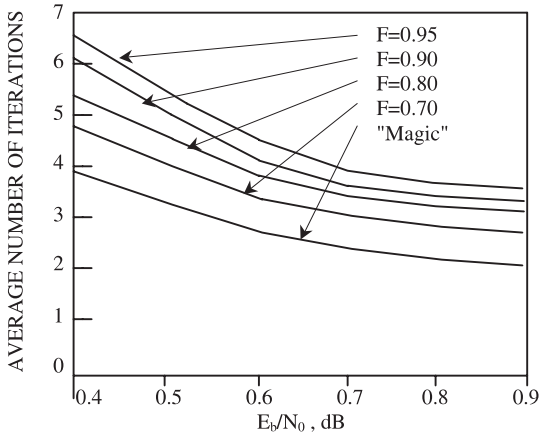


Fig. 7. Average number of iterations

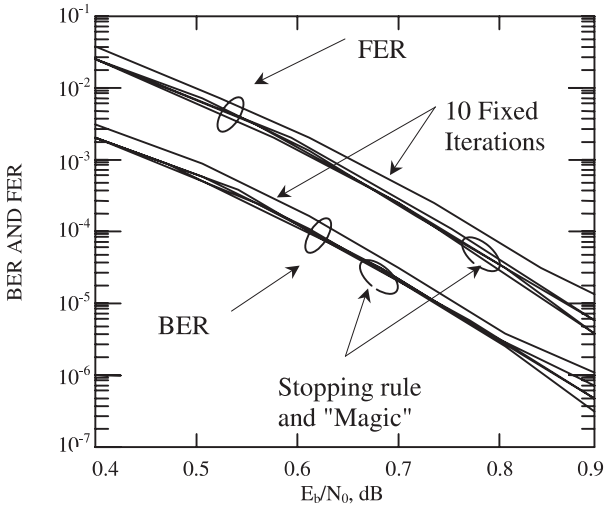


Fig. 8. FER and BER for the stopping rule

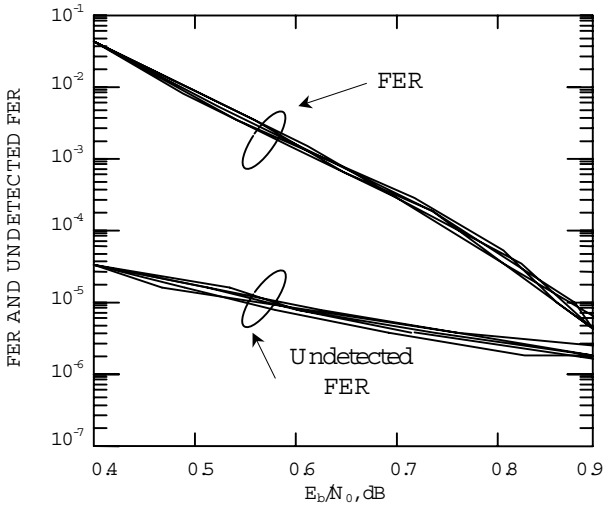


Fig. 9. FER and undetected FER for the stopping rule

with E_b/N_0, for the decoded sequences converge to the correct codeword in fewer iterations. Also, as it is common for all soft rules, there is a consistent offset from the average number of iteration required by the magic rule.

Both, the FER and BER performance of this rule are compared with the reference FER and BER performance curves for $K = 10$ fixed iterations per decoded frame and for the magic stopping rule with 20 iterations. The results for the chosen thresholds are shown on Fig. 8.

Having in mind the skewness of the statistics in the first iteration we didn't allow the decoder to take decisions after the first iteration, for there is a small number of first step decisions and they usually result in falsely decoded bits. By applying this measure, we noticed that the improvement in BER and FER is much greater than loss in speed, so it has a positive influence on overall performance.

The Fig. 8 indicates that the overall error rates of the turbo code employing this stopping rule are noticeable better than those for 10 fixed iterations as in any other soft stopping rule case. In fact they are nearly equal to the error rates achieved by a decoder using 20 iterations. Furthermore, it can also be seen that this rule has no characteristic error floor of its own (a matter of the length of the simulations), but the floor constrain is set by the inherent turbo code floor.

To get a deeper insight into the FER composition one should consider four different conditions that can occur when stopping rules are used by turbo decoder, depending whether the decoded sequence is detected to be reliable or unreliable and whether it is actually correct or in error. First is the case of correct decoding when stopping rule is satisfied at some iteration, $k < K_{\max}$, and the decoded sequence is correct. Second case is when the rule is satisfied but the sequence is actually in error which produces an undetected error. Next case is when the rule fails to stop the decoder in $k = K_{\max}$ iterations and the decoded sequence is indeed incorrect which corresponds to a detected error. The last case is when the sequence is declared as unreliable but is actually correct which means a falsely detected error. A good stopping rule has a small undetected error probability and a small probability of falsely detected errors.

The Fig. 9 plots the overall FER and undetected FER for this rule. We can observe that the undetected error rates decrease slowly over the entire range of tested E_b/N_0 values. Naturally, there is a small difference in FER values due to the level of the thresholds. At low E_b/N_0 the detected errors dominate the overall FER, whereas at high E_b/N_0 values the undetected errors are more contributory.

If the performance of this rule is compared with the other mentioned soft rules, one can see neither significant advantage in favor of this rule nor significant disadvantage. However, if we consider the soft rules from computational effort and implementation complexity aspects, then this rule certainly has a comparative advantage.

V. Conclusion

In this article we propose a new type of stopping rules that can be used to reduce the average number of iterations to decode a turbo code. This type of stopping rules belongs to the class of soft stopping rules but, for difference, it is based on the evolution of the probability density function (pdf) of the extrinsic information. A noise figure is defined through the pdf parameters as a measure of quality of the extrinsic and its values are used to set up appropriate stopping threshold.

We assessed our stopping rule in accordance with the evaluation process applied to evaluate other soft rules [4].

Though the performance simulation doesn't show any significant performance improvement for the range of tested thresholds, this rule has an advantage of low computational burden and low complexity requirements which make it more suitable for implementation.

References

[1] C.Berrou, A.Glavieux and P.Thitimajshima, "Near Shannon Limit Error Correcting Coding: Turbo Codes", *Proceedings 1993 IEEE Int.Conf. on Comm.*, pp.1064-1070, Geneva, May 1993.

[2] L.Bahl, J. Cocke, F. Jelinek and J. Raviv, “Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate”, *IEEE Transactions on Information Theory*, vol.IT-20, 1974.

[3] S.Benedetto, D.Divsalar, G.Montorsi and F.Pollara, “Soft-Input Soft-Output Maximum A Posteriori (MAP) Module to Decode Parallel and Serial Concatenated Codes”, *The Telecomm. and Data Acquisition Prog. Rep.*, Jet Propulsion Laboratory, Pasadena, Sep’96

[4] A.Matache, S.Dolinar and F.Polara, “Stopping Rules for Turbo Decoders”, *TMO Progress Report* 42-142, August 15, 2000.

[5] T.Richardson, A.Shokrollahi and R.Urbanke, “Design of probably Good LDPC Codes”, *IEEE Transactions on Information Theory*.

[6] S. ten Brink, “Convergence of Iterative Decoding”, *Electronics Letters*, vol.35, May 24, 1999.

[7] D. Divsalar, S. Dolinar, and F. Pollara, “Iterative Turbo Decoder Analysis Based on Density Evolution”, *TMO Progress Report*, pp.42-144, February 15, 2001.

A Turbo Codes Research Tool Based on Probability Density Evolution

Zafir Popovski[1], Tatjana Ulcar-Stavrova[2]

Abstract – **The turbo codes [1] are decoded in a iterative decoding scheme [2] performing a predefined number of iterations before the final decision comes up. Since the method is time consuming, stopping rules can be apply to prematurely quit the process, for the decoder has already done its job. This technique requires a reasonable trade-off which should results in a average decoding speed increase while not sacrificing the decoder performance.**

Keywords – **Stopping rules, number of iterations**

I. Introduction

Richardson et al. [3,4] used similar probability density evolution to compute iterative decoding thresholds for LDPC codes over a binary input AWGN channel. Similarly, S. ten Brink [5] developed a method for analysing the convergence of the decoder based on the evolution of mutual information. D.Divsalar et al.[6] apply these analyses to gain new insights into designing new turbo like code structures. The method in this article is similar to all of these approaches. Our main contribution is a tool which is capable to easily distinguish the qualitative values of compared turbo code structures. Since the method is based on the extrinsic's evolution, the extension to all iteratively decoded binary codes is straightforward.

In the next section we shortly explore the components of a turbo code and than develop a model to track the density evolution of the extrinsic's probability density function (pdf). Finally we apply the models in a simulation process to confirm some results known from the analytical constituent encoder design process. In our simulations we use modular "C", "C^{++}" and "MATLAB" programs which are modified to support statistical evaluation of the extrinsics. Being independent of the program in use , the results prove our feelings that for any given turbo code, the speed and manner of its extrinsics probability density evolution somehow represents a unique comparable "fingerprint" for the code.

II. Turbo Codes Background

The turbo codes (TC) are a class of FEC (forward error control) codes, known as parallel concatenation of two or more recursive systematic convolutional (RSC) codes (RSCC) produced by a turbo encoder (TE) composed of two or more component RSC encoders (CEs) with input to each but one CE permuted by an interleaver of length N. For further considerations we will use TCs with only two concatenated CEs, CE1 and CE2, with overall code rate $R = 1/3$, as shown on the Fig. 1. It is proved that these codes can perform very close to the Shannon limit.

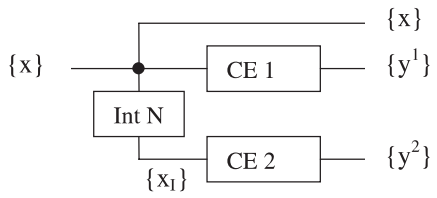


Fig. 1. Turbo encoder structure

Due to the interleaver, both the coding and decoding procedures are carried out on blocks of information sequences of length N plus some tail bits used to drive the constituent encoder trellises to all zero path, for decoder to know the beginning and the end of the codewords. Unfortunately, the length of N information bits can, and usually contains sequences, termed as self terminating (ST) sequences, which prematurely drives the CEs to the all zero path thus producing error events. Note that a codeword is also an "error event".

Basically, the performance of a turbo-code composition is determined by five factors: constituent encoders design, interleaver design, decoding algorithm, interleaver size and number of iterations. To achieve better coding gain, the latter three factors require either increases in delay or in complexity. The former two factors, however, can be thought of as a matter of "a good choice".

The most important impact of the interleaver on the overall turbo code performance is its information weight distribution from one to the other component encoder input.

As it can be seen from the Fig. 1, the parity vector $\boldsymbol{y}^1$ depends on $\boldsymbol{x}$ and CE1, while the $\boldsymbol{y}^2$ is determined by $\boldsymbol{x}$, CE2 and the interleaver. An explanation of the interleaver's role is almost obvious: If the sum of Hamming weights of $\boldsymbol{x}$ and $\boldsymbol{y}^1$, $W(\boldsymbol{x})$ and $W(\boldsymbol{y}^1)$, are "low", then, in order to brake up the ST sequence for the second encoder, the interleaver should provide such a permutation $\boldsymbol{x}_I$ which will yield a "high" weight contribution by the second parity vector $\boldsymbol{y}^2$.

A satisfactory overall weight of the TC naturally depends on length of N. Yet, there is no mathematical framework to

[1]Zafir Popovski is with the Faculty of Electrical Engineering, "St.Cyril and Methodius" University – Skopje, E-mail: zpopovski@yahoo.com

[2]Tatjana Ulcar-Stavrova is with the Faculty of Electrical Engineering, "St.Cyril and Methodius" University – Skopje, E-mail: tanjaus@etf.ukim.edu.mk

solve this problem for all ST sequences and the value of N is often heuristically determined. In other words, for a given structure of the constituent RSCCs and the size of N, the goal mapping for each input sequence can not be determined. One may find some efforts in the literature but the problem of a perfect interleaver is still away from an exact science and, for given conditions, a "good choice" should be worked out.

The design of turbo code component encoders should also help solving the problems caused by ST sequences.

A constituent RSCC is commonly denoted in matrix D form as $[1, d(D)/g(D)]$ where the first term provides for the systematic outputs when multiplied by input sequence $x(D)$, and the second term, the quotient of the feed forward $d(D)$ and the feed back polynomial $g(D)$, also multiplied by $x(D)$, provides for the parity outputs. More conveniently, the polynomials are usually given in octal form.

It is known that the asymptotic performance of a convolutional code can be improved by maximising the free distance of its distance spectrum. However, due to the presence of the interleaver it is extremely difficult to find the distance spectrum for the TC. This is overcome by averaging the weight spectrum over all interleavers using the notion of an "uniform interleaver"[8]. The basic idea is to combine the distance spectra of the two CEs to create a super spectra which contains all possible combinations of all the paths. This technique shows that there is a need to alter the usually desired characteristics of the CEs in order to suit the characteristics of the TC.

By choosing recursive CE, any weight one information sequence can not be a ST sequence due to the infinite impulse response of the recursive structure, and we are to consider higher input weights. Due to the interleaver embedded into the TC structure, most of the low weight ST sequences will have a chance of being broken up to produce a high weight at the second CE. Also, heavier input weights, after permuting, will have a much higher chance of being broken up than those of lower weight. It is well known that the minimum information weight in the error events of a RSCC is $w_{\min} = 2$, and this particular input weight is most contributory one for the bit error probabilities ranging from 10^{-3} down to 10^{-10}.

Hence the problem of finding good codes for TCs lies in finding RSCCs that have maximum output weight for weight two input sequences which define a figure of merit for TCs named as effective free distance:

$$d_{\text{free,eff}} = 2 + 2z_{\min} \tag{1}$$

where $z_{\min}$ is the weight of minimum weight parity sequence generated by RSC CE with a weight 2 input. So, a TC_1 composed by two equal $\text{CE}_1 = [1, 5/7]_8$ will have $d_{\text{free,eff}} = 10$, while TC_2 composed by two $\text{CE}_2 = [1, 7/5]_8$ will have $d_{\text{free,eff}} = 8$ and the "good choice" is quite obvious. The higher order polynomials, however, have much more different polynomials of the same order among which many of them have the same value of respective $d_{\text{free,eff}}$. For example a 32-state CE has 15 different polynomials yielding the same value of $z_{\min}$. The "good choice" then should search for maximising $d_{\text{free,eff}}$ for minimum weight parity sequence generated by a weight 3 input, and so on. It appears that the number of nearest neighbours, defined as the number of paths having the same effective distance, is also an important parameter which should be minimised. It is also noticed that the order of importance of different parameters varies...

Having also in mind the above mentioned unrealisable uniform interleaver, the simulation of TC's BER and FER performances remains as a unique model to evaluate both the CEs and the interleaver "good choices" to suit one another. The method in this article offers another possibility.

III. Density Evolution Model

The optimum decoding of TCs is the maximum likelihood (ML) decoding algorithm applied to the TC trellis structure. However, due to the interleaver embedded into the TE's structure, the trellis will have an extremely large number of states thus making the whole ML decoding process almost unrealisable in practice. Since the TC is a concatenation of component codes, a more practical solution is to sequentially decode the component codes in a iterative fashion using one decoder at a time for each code. A simple SISO (Soft-In Soft-Out) maximum a posteriori (MAP) decoder which minimise the probability of bit error appears to be the best solution for component decoders (CDs). The MAP algorithm provide as an output a real number which is a measure of the probability of error in decoding a particular bit. This extra information termed as extrinsic information, λ_i, can be passed as input to the second CDs allowing it to create its own extrinsic to be passed to the first CD in the next iteration. The process is than iterated until reaching a satisfactory degree of confidence regarding the received noisy examples contained in sequence of length N, as shown on Fig. 2.

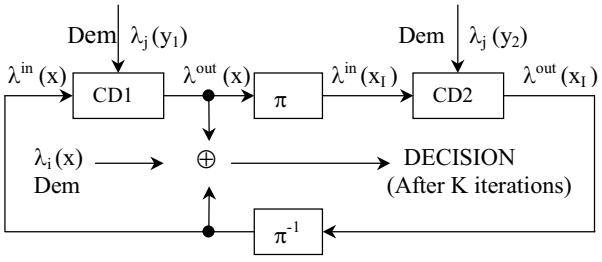


Fig. 2. Iterative TD with two MAP CDs

Such iterative decoder can be considered as a non-linear dynamical feedback system. Extrinsic information messages $\{\lambda_i\}$ are passed from one to the other constituent decoder. The message λ_i measures the log-likelihood ratio for the i-th bit based on input massages $\{\lambda_j\}$ from all other bits but the i-th. So, if we assume that the all-zero codeword is transmitted (with BPSK modulation corresponds to transmission of "+1"s on the channel) then a positive value of the extrinsic information, $\lambda_i > 0$, for each i, will represent a favourable evidence toward determining the true value of the i-th bit.

When the interleaver on Figs. 1 and 2 is large and random, the extrinsics λ_i are independent and identically distributed with probability density function $f(\lambda)$. As shown in [4], this pdf is consistent ($\lambda = \log[f(\lambda)/f(-\lambda)]$), its mean,

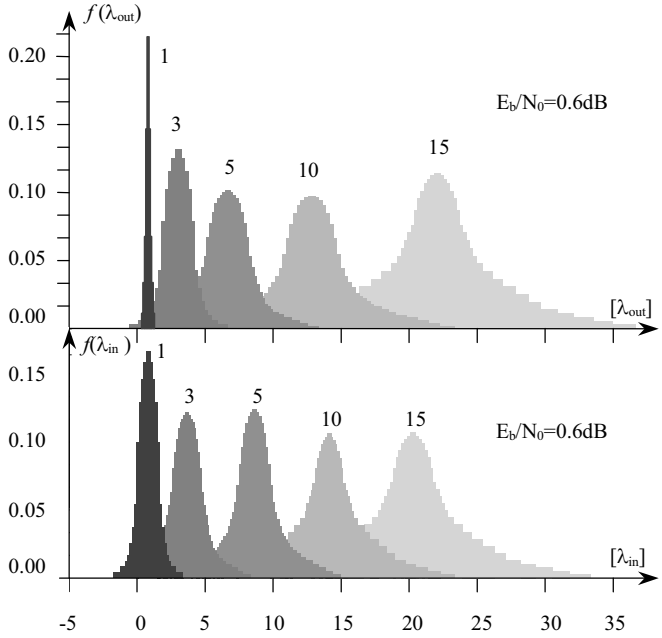


Fig. 3. Evolution of the input and output extrinsics

$\mu = \mathrm{E}(\lambda)$, is discrimination between the two densities $f(\lambda)$ and $f(-\lambda)$, and the error probability, $e = \mathrm{Prob.}\{\lambda < 0\}$, can be evaluated as $e = \mathrm{E}\{1/(1 + e^{|\lambda|})\}$. Computed histograms of the λ_{in} and λ_{out} extrinsics at the input and output of a SISO MAP decoding module for a 4 states, rate 1/3, $[1,5/7]_8$ turbo code are plotted on Fig. 3 for a number of iterations. As it can be seen, the empirical probability densities $f(\lambda_{\mathrm{in}})$ and $f(\lambda_{\mathrm{out}})$ evolve with successive decoder iterations from narrow densities concentrated nearby $\lambda = 0$, to broader Gaussian-shaped densities with increasing means as the iterations continue. Ignoring some irregularities at the beginning of the process, this probability density function can be approximated by a Gaussian density in which case its statistics depend on two parameters: its mean $\mu = \mathrm{E}(\lambda)$ and its variance $\sigma^2 = \mathrm{Var}(\lambda)$.

A signal-to-noise ratio for such random variable can be defined as $\mathrm{SNR} = \mu^2/\sigma^2$ but since it is both Gaussian and consistent, then $\sigma^2 = 2\mu$ and, consequently, $\mathrm{SNR} = \mu/2$. This evaluation gives the best approximation of the empirically measured variance. Now we can observe the input and output SNRs for each decoder denoted as $\mathrm{SNR1_{in}}$, $\mathrm{SNR1_{out}}$, $\mathrm{SNR2_{in}}$, and $\mathrm{SNR2_{out}}$, at each iteration as shown on Fig. 4. A non zero E_b/N_0 from the channel helps CD1 to produce a non zero $\mathrm{SNR1_{out}}$ for the extrinsic information despite starting with $\mathrm{SNR1_{in}} = 0$.

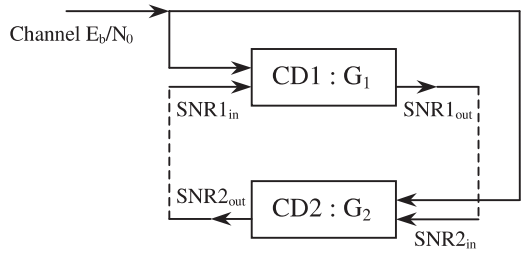


Fig. 4. Analytical density evolution model

So, for given value of E_b/N_0 the output SNR of each CD is a non-linear function of its input, denoted as G1 for CD1 and G2 for CD2. Thus we have:

$$\mathrm{SNR1_{out}} = \mathrm{G1}(\mathrm{SNR1_{in}}, E_b/N_0) \quad (2)$$

$$\mathrm{SNR2_{out}} = \mathrm{G2}(\mathrm{SNR2_{in}}, E_b/N_0) \quad (3)$$

From the Fig. 4 it follows that $\mathrm{SNR2_{in}} = \mathrm{SNR1_{out}}$, so we have

$$\mathrm{SNR2_{out}} = \mathrm{G2}(\mathrm{G1}(\mathrm{SNR1_{in}}, E_b/N_0), E_b/N_0) \quad (4)$$

The G1 and G2 functions can be evaluated either directly from the histogram of output λ's from the previous decoder or to generate input λ's from the consistent Gaussian density with mean μ and variance 2μ. In our simulations we use the former model and SNRs are computed from the actual λhistograms as $\mathrm{E}\{\lambda\}/2$.

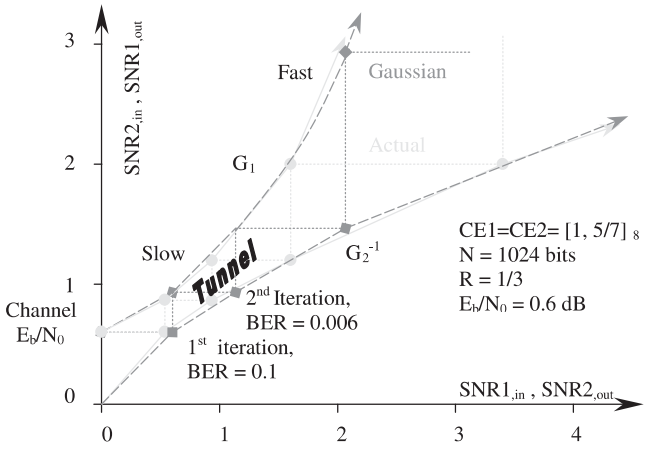


Fig. 5. Iterations and convergence of $[1, 5/7\text{-}5/7]_8$ TC

The decoder convergence is assessed by tracking the evolution of the extrinsic information's SNRs in each half iteration. As it is shown on Fig. 5, the analytical model is to plot the output SNR of CD1 versus its input, and the input SNR of CD2 versus its output SNR. For this case we followed the extrinsics evolution of the memory 2, rate 1/3, [1,5/7] turbo code, at $E_b/N_0 = 0.6$ dB.

Both curves for G_1 and G_2^{-1} obtained from actual density evolution are just sequences of discrete points joined by linear interpolation to give estimates at intermediate points. Actually, the Fig. 5 shows the progress of the decoder's iterations. The improvement in the SNR of the extrinsics and corresponding BER follows a staircase path reflecting at right angles between the G_1 and G_2^{-1} curves.

The steps are large when the bounding curves are far apart, and small when they are close together in which case the improvement in BER slows down , for many iterations are required to get through the narrow iterative decoding tunnel between the curves. If the iterative decoding process successfully passes through the tunnel, the convergence becomes rapid as the curves part more at the higher SNRs.

Since the curves are a reflection of the density evolution of the extrinsics we believe that the manner in which the curves

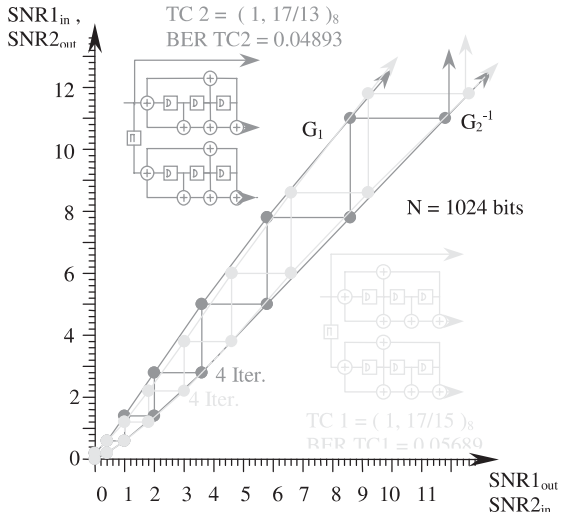


Fig. 6. Comparison at $E_b/N_0 = 0.4$ dB

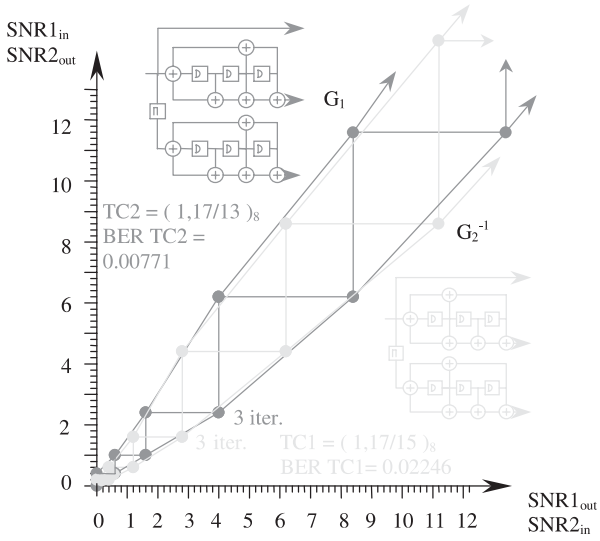


Fig. 7. Comparison at $E_b/N_0 = 0.5$ dB

part might be interpreted as a response of the component encoder's quality. So, through a simulation process we tested a number of CEs known as a "good choice" and compared the results for several E_b/N_0 values with the results of the CEs which are not classified as a "good choice" with respect to their distance properties.

The Figs. 6, 7, 8 and 9 represent test-comparison between TC1 with two CE1 = $[1,17/15]_8$ and TC2 with a pair of CE2 = $[1,17/13]_8$, the later being the best among the all possible, rate 1/2, memory 3, constituent RSCCs. It can be noticed on Fig. 6 that both TC1 and TC2 have the same properties nearby the waterfall region because the TC's performances at low E_b/N_0 values are mainly governed by the interleaver gain. By increasing the E_b/N_0 value the curves go apart and it becomes quite clear that TC2 is the better choice which is also confirmed by corresponding BER values.

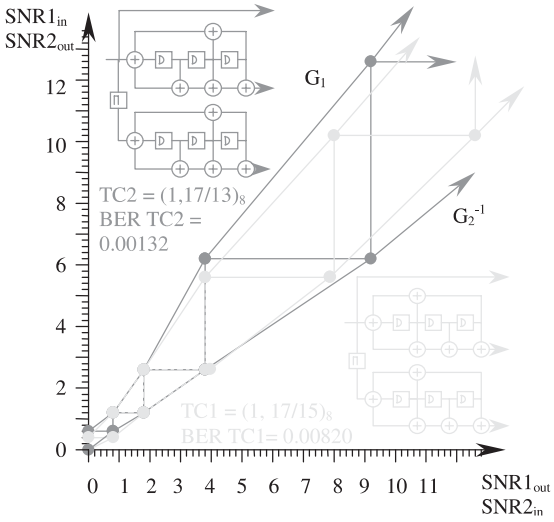


Fig. 8. Comparison at $E_b/N_0 = 0.6$ dB

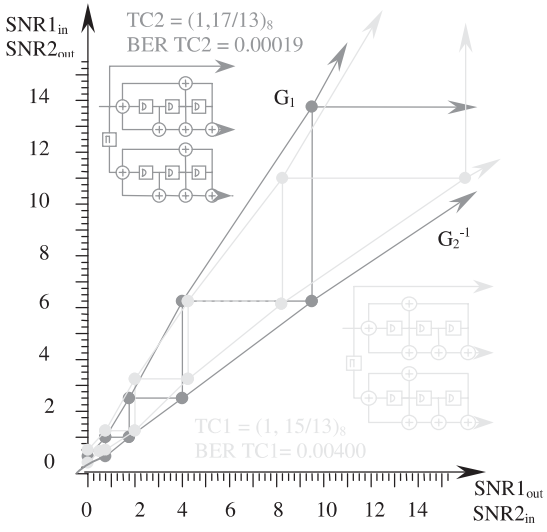


Fig. 9. Comparison at $E_b/N_0 = 0.7$ dB

IV. Conclusion

Basically, we introduced the main problems encountered in turbo codes design process. Then we adopted a model to trace the density evolution of the extrinsics and explored the possibility of its application in the evaluation process of a turbo code constituent encoders for several E_b/N_0 values. The simulations carried out for a "good" and a "not so good" CEs confirmed our expectations. During the simulations we used the same type and the length of the interleaver. It is quite obvious that we might have kept the CEs as constant and to vary the type or the length of interleaver. Also, by using this density evolution model of a tested turbo code as a benchmark one can evaluate any turbo or turbo-like structure.

References

[1] C.Berrou, A.Glavieux and P.Thitimajshima, “Near Shannon Limit Error Correcting Coding: Turbo Codes”, *Proceedings 1993 IEEE Int.Conf. on Comm*., pp.1064-1070, Geneva, May 1993.

[2] L.Bahl, J. Cocke, F. Jelinek and J. Raviv, “Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate”, *IEEE Transactions on Information Theory*, vol.IT-20, 1974.

[3] T.Richardson, R.Urbanke, “The Capacity of Low Density Parity Check Codes Under Message Passing Decoding”, *IEEE Transac. on Information Theory*.

[4] T.Richardson, A.Shokrollahi and R.Urbanke, “Design of probably Good LDPC Codes”, *IEEE Transactions on Information Theory*.

[5] S. ten Brink, “Convergence of Iterative Decoding”, *Electronics Letters*, vol.35, May 24, 1999.

[6] D. Divsalar, S. Dolinar, and F. Pollara, “Iterative Turbo Decoder Analysis Based on Density Evolution”, *TMO Progress Report*, pp.42-144, February 15, 2001.

[7] S. Benedetto and G. Montorsi, “Average Performance of Parallel Concatenated Block Codes”, *El. Letters*, 31(3): 156-158, Feb 1995.

Invariant Subspaces in Third-Order Digital Filters with Two's Complement Overflow

Cvetko D. Mitrovski[1]

Abstract – **In this paper we propose a general approach for locating the invariant subspaces of some classes of third order digital filters with two's complement overflow. The proposed approach is based on the analysis of the projection of the points of the trajectories on adequately chosen directions in the filter phase space.**

Keywords – **Digital filter, two's complement overflow, attractor, invariant space, phase space, eigenvalue, and eigenvector**

I. Introduction

Nonlinearities in digital filters and their influences on their dynamics have attracted considerable attention in the last fifteen years. The very first works in this field [1,2], show that even the simple systems like second order digital filters can exhibit complex behavior due their two's complement overflow nonlinearity. This phenomenon is manifested by existence of various types of trajectories in the digital filter's phase space which depend both, on the filter parameters, and by their initial starting points.

In the farther papers, special attention was attended on the third and the higher order digital filters operating out of their stability region [3,4]. In these papers it was noticed that the trajectories of the digital filters usually non-uniformly visit various parts of the phase space, and in some cases they visit regularly only some invariant-subspaces of the phase space.

In this paper we concentrate on the third order digital filters with two's complement overflow nonlinearity, operating outside of their linear stability region. For this type of filters we develop a new, geometry oriented approach for localization of the invariant subspaces visited by their trajectories. The proposed approach is based on the analyses of the projection of the trajectories of the digital filter on suitably chosen directions in its phase space, which depend on the filter parameters.

The material in this paper is organized as follows. The piece wise linear model of the third order digital filter is given in Section II. In Section III, we develop expression for the projection of the k-th iteration of a map on an arbitrary vector in the phase space, and determine the criteria when the trajectories of the digital filters visit finite number of planes into the phase space. In Section IV we consider situation when the Jacobijan of the map has at least one eigenvalue inside and one outside of the unit circle, and determine the conditions when the attractors of the map are localized in some compact parts (subspaces) of the phase spaces. In Section V we illustrate and discuss our results and at last, in Section VI we give some conclusions.

[1]Cvetko D. Mitrovski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: cvetko.mitrovski (auklo.edu.mk

II. Piece-Wise Linear Model

The behavior of zero input, third order digital filter with 2's complement overflow can be described by the following system of difference equations

$$\begin{aligned} x_1(k+1) &= x-2(k) \\ x_2(k+1) &= x_3(k) \\ x_3(k+1) &= f(a_1x_1(k)+a_2x_2(k)+a_3x-3(k)) \end{aligned} \quad (1)$$

where: $f(\cdot)$ is an almost odd function defined by:

$$f(x) = x-2s \quad \text{for} \quad -l+2s \geq x < l+2s; \quad (2)$$

a_1, a_2, a_3 are filter parameters; and $x_1(k)$, $x_2(k)$ and $x_3(k)$ are the filter internal states.

The behavior of the filter, can be also described by the piece-wise linear map $\mathbf{F}(x(k)) : \mathbf{I}^3 \rightarrow \mathbf{I}^3$,

$$\mathbf{x}(k+1) = \mathbf{F}(f(\mathbf{a}^T\mathbf{x}(k)) = \mathbf{A}\mathbf{x}(k)+\mathbf{b}s \text{ for } \mathbf{x}(k) \in \mathbf{I}_s^3 \quad (3)$$

where:

$$\mathbf{x}(k) = \begin{vmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{vmatrix}; \quad \mathbf{a} = \begin{vmatrix} a_1 \\ a_2 \\ a_3 \end{vmatrix}; \quad ||\mathbf{a}|| = |a_1|+|a_2|+|a_3|;$$

$$\mathbf{A} = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a_1 & a_2 & a_3 \end{vmatrix}; \quad \mathbf{b} = \begin{vmatrix} 0 \\ 0 \\ 2 \end{vmatrix};$$

$$\mathbf{I}^3 = \{\mathbf{x} = [x_1, x_2, x_3]^T : -1 \leq x_i < 1,\ i = 1,2,3\};$$

$$\mathbf{I}_s^3 = \{\mathbf{x} : -2s-1 \leq \mathbf{a}^T\mathbf{x} < -2s+1,\ \mathbf{x} \in \mathbf{I}^3, s \in Z^*\};$$

$$Z^* = \{s_{\min}, s_{\min}+1, ..., 0, 1, ..., s_{\max}\};\ s_{\max} = -s_{\min};$$

and

$$s_{\max} = \begin{cases} \left[\frac{||\mathbf{a}||+1}{2}\right]; & \text{for } ||\mathbf{a}|| \neq 2p+1 \\ \left[\frac{||\mathbf{a}||+1}{2}\right]-1; & \text{for } ||\mathbf{a}|| = 2p+1 \end{cases}; \ p = 0, 1, ...$$

This means that $\mathbf{I}^3$ is divided into $2s_{\max}+1$ subspaces $\mathbf{I}_s^3$, $s \in Z^*$, separated by $2s_{\max}$ parallel planes π,

$$\pi = \{\mathbf{x} : \mathbf{a}^T\mathbf{x} = 2s-1, s \in \{Z^* \backslash s_{\max}\}\}, \quad (4)$$

(as illustrated in Fig. l).

By using k recursive iterations in (3), one can obtain the expression for the k-th iteration of the map

$$\mathbf{x}(k) = \mathbf{A}^k\mathbf{x}(0) + \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j}\mathbf{b}s_j, \quad (5)$$

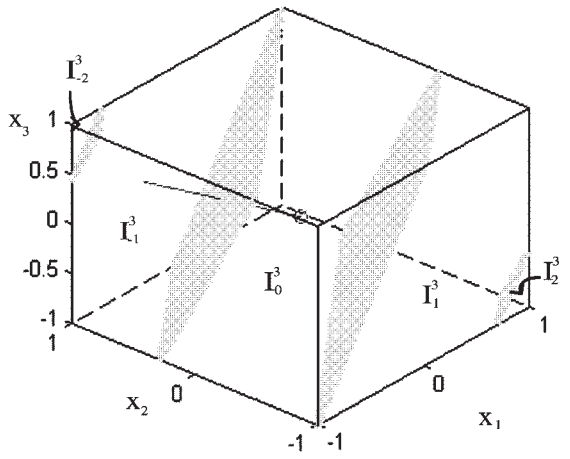


Fig. 1. Subspaces of the filter phase space I^3 for the filter with parameters $\mathbf{a} = (a_1, a_2, a_3) = (-1.17, 1.57, 0.6)$ (eigenvalues of $\mathbf{A}$ are $\lambda_1 = 1$, $\lambda_2 = -1.3$, $\lambda_3 = 0.9$)

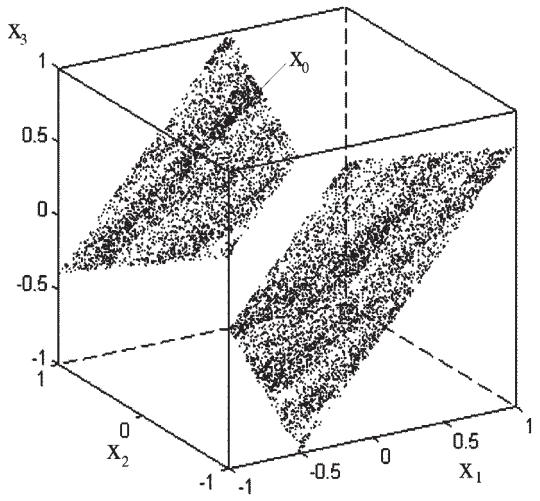


Fig. 2. Trajectory of the filter with parameters $\mathbf{a} = (a_1, a_2, a_3) = (-1.17, 1.57, 0.6)$ (eigenvalues of $\mathbf{A}$ are $\lambda_1 = 1$, $\lambda_2 = -l.3$, $\lambda_3 = 0.9$), starting from $\mathbf{x}_0 = [-0.34, 0.23, 0.7]^T$

in which s_j $(j = 0, l, ..., k-l)$ is an integer that corresponds to the index of the subspace $\mathbf{I}^3_{s_j}$ visited by the j-th iteration. Hence, each trajectory of the map (3), starting from any initial point $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{x}_0 \in \mathbf{I}^3$, generates infinite symbolic sequence $s_0 s_1 ... s_j ...$ composed of the indexes of the visited subspaces.

III. Invariant Planes

The characteristics of each trajectory, given by Eq. (5), depend on both, the initial point $\mathbf{x}(0)$ and the properties of the map which are determined by the Jacobijan matrix of the map $\mathbf{F}(\mathbf{x})$, $D(\mathbf{F}(\mathbf{x})) = \mathbf{A}$.

We analyze the behavior of the map (3) under assumption that the matrix $\mathbf{A}$ has full rank, $\text{rank}(\mathbf{A}) = 3$, and that it has 3 distinct eigenvalues λ_1, λ_2 and λ_3 with corresponding eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$. Without any constrains on the eigenvalues of $\mathbf{A}$, the projection of the k-th (k-arbitrary non-negative integer) iteration of the map on a direction defined by an arbitrary unit vector $\mathbf{v}$ in $\mathbf{R}^3$, $\mathbf{v} \in \mathbf{R}^3$, is

$$\mathbf{v}^T \mathbf{x}(k) = \mathbf{v}^T \mathbf{A}^k \mathbf{x}_0 + \mathbf{v}^T \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j} \mathbf{b} s_j, \qquad (6)$$

$$\mathbf{v}^T \mathbf{x}(k) = ((\mathbf{A}^k)^T \mathbf{v})^T \mathbf{x}_0 + \sum_{j=0}^{k-1} ((\mathbf{A}^{k-1-j})^T \mathbf{v})^T \mathbf{b} s_j. \qquad (7)$$

If $\mathbf{v}$, is an unit eigenvector of $\mathbf{A}^T$ that corresponds to the eigenvalue λ_j, $(\mathbf{A}_i^T \mathbf{v}_i = \lambda_i \mathbf{v}_i;\ i = l, 2, 3)$, then the last equation becomes

$$\mathbf{v}_i^T \mathbf{x}(k) = \lambda_i^k \mathbf{v}_i^T \mathbf{x}_0 + \mathbf{v}_i^T \mathbf{b} \sum_{j=0}^{k-1} \lambda_i^{k-1-j} s_j. \qquad (8)$$

By denoting $\mathbf{v}_i^T \mathbf{x}_0 = \alpha_i(\mathbf{x}_0)$ and $\mathbf{v}_i^T \mathbf{b} = \beta_i$, the last equation can be rewritten as

$$\mathbf{v}_i^T \mathbf{x}(k) = \alpha_i(\mathbf{x}_0) \lambda_i^k + \beta_i \sum_{j=0}^{k-1} \lambda_i^{k-1-j} s_j. \qquad (9)$$

This equation is crucial for our further analysis of the invariant subspaces of the map (3).

If $\lambda_1 = 1$, then the last equation becomes

$$\mathbf{v}_1^T \mathbf{x}(k) = \alpha_1(\mathbf{x}_0) + \beta_i \sum_{j=0}^{k-1} s_j. \qquad (10)$$

Since $\mathbf{x}(k) \in \mathbf{I}^3$ and $s_0 + s_1 + \cdots + s_k$ is an integer, it implies that $s_0 + s_1 + \cdots + s_k$ must be a finite integer for each k. Therefore, one can conclude that $\mathbf{x}(k)$ belongs to a finite set of parallel planes Ω_1 normal on $\mathbf{v}_1^T$ [5], defined by

$$\Omega_1(\mathbf{x}_0) = \{\mathbf{x} : \mathbf{v}_1^T \mathbf{x}(k) = \alpha_1(\mathbf{x}_0) + \beta_i \sum_{j=0}^{k-1} s_j,\ k = 0, 1, ...\}. \qquad (11)$$

Hence, for each initial point $\mathbf{x}_0$, we can determine the invariant set of points

$$\Theta_1(\mathbf{x}_0) = \{\Omega_1(\mathbf{x}_0) \cap \mathbf{I}^3\} \qquad (12)$$

that satisfy $\mathbf{F}^k(\Theta_1(\mathbf{x}_0)) \subseteq \Theta_1(\mathbf{x}_0)$; $k = 0, 1, 2, ...$ It means that each trajectory that starts from any point of this set of parallel planes remains on them. This property is illustrated in Fig. 2 in which we present the first 20 000 points of the trajectory of the filter (defined by the eigenvalues $\lambda_1 = 1$, $\lambda_2 = -1.3$ and $\lambda_3 = 0.9$ of the matrix $\mathbf{A}$), starting from $\mathbf{x}_0 = [-0.34, 0.23, 0.7]^T$.

Similar result could be obtained for $\lambda_1 = -1$. In this case Eq. (8) becomes

$$\mathbf{v}_1^T \mathbf{x}(k) = (-1)^k \alpha_1(\mathbf{x}_0) + \beta_1 \sum_{j=0}^{k-1} (-1)^{k-1-j} s_j. \qquad (13)$$

Since $\mathbf{x}(k) \in \mathbf{I}^3$, it implies that $\mathbf{x}(k)$ belongs to a finite set of parallel planes $\Omega_2(x_0) \cup \Omega_3(x0)$ normal on $\mathbf{v}_1^T$, where

$$\Omega_2(\mathbf{x}_0) = \{\mathbf{x} : \mathbf{v}_1^T\mathbf{x} = \alpha_1(\mathbf{x}_0) + \beta_1 \sum_{j=0}^{k-1}(-1)^{k-1-j}s_j,\ k-\ odd\}; \quad (14)$$

$$\Omega_3(\mathbf{x}_0) = \{\mathbf{x} : \mathbf{v}_1^T\mathbf{x} = -\alpha_1(\mathbf{x}_0) + \beta_1 \sum_{j=0}^{k-1}(-1)^{k-1-j}s_j,\ k-\ even\}. \quad (15)$$

Hence, for each initial point $\mathbf{x}_0$, we can determine the invariant set of points

$$\Theta_2(\mathbf{x}_0) = \{(\Omega_2(\mathbf{x}_0) \cup \Omega_3(\mathbf{x}_0)) \cap \mathbf{I}^3\} \quad (16)$$

that satisfy $\mathbf{F}^k(\Theta_2(\mathbf{x}_0) \subseteq \Theta_2(\mathbf{x}_0)$; $k = 0, 1, 2, ...$

If $\mathbf{A}$, has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1$, then the invariant space of the map (3) becomes a set of parallel lines collinear with the vector $(\mathbf{v}_1^T \times \mathbf{v}_2^T)$, that intersects $\mathbf{I}^3$

$$\Lambda(\mathbf{x}_0) = \{\Omega_1(\mathbf{x}_0) \cap (\Omega_2(\mathbf{x}_0) \cup \Omega_3(\mathbf{x}_0))\} \cap \mathbf{I}^3. \quad (17)$$

IV. Invariant Subspaces

It is obvious that the behavior of the trajectories of the map (3), depend on the eigenvalues of the Jacobijan matrix $\mathbf{A}$. If $\mathbf{A}$ has at least one eigenvalue out of the unit circle then each trajectory of the map is either periodic or chaotic. If the trajectories are chaotic then they are all attracted by attractors which are very strangely positioned in the phase space.

In this section we analyze the subspaces of the phase space in which are placed all attractors in cases when the Jacobijan matrix $\mathbf{A}$ of the map has at least one eigenvalue in the interior of the unit circle, say $|\lambda_2| < 1$, and one eigenvalue out of the unit circle, say $|\lambda_3| > 1$.

In this case, if $\mathbf{v}_2$ is the eigenvector of $\mathbf{A}^T$ that corresponds to λ_2, the Eq. (8) becomes

$$\mathbf{v}_2^T\mathbf{x}(k) = \lambda_2^k\alpha_2(\mathbf{x}_0) + \beta_2 \sum_{j=0}^{k-1}\lambda_2^{k-1-j}s_j. \quad (18)$$

For sufficiently large k, Eq. (l8) could be further reduced to

$$\mathbf{v}_2^T\mathbf{x}(k) = \beta_2 \sum_{j=0}^{k-1}\lambda_2^{k-1-j}s_j, \quad (19)$$

since $|\lambda_2| < 1$ and $\lim_{k\to\infty}\alpha_2(\mathbf{x}_0)\lambda_2^k = 0$.

The last equation can be rearranged in the following form

$$\frac{\mathbf{v}_2^T\mathbf{x}(k)}{\beta_2} - s_{k-1} = \lambda_2(s_{k-2} + s_{k-3}\lambda_2 + \cdots + s_0\lambda_2^{k-2}). \quad (20)$$

By using the fact that $-s_{\max} \le s_j \le s_{\max}$, we obtain

$$\left|\frac{\mathbf{v}_2^T\mathbf{x}(k)}{\beta_2} - s_{k-1}\right| < |\lambda_2|\,|s_{\max}|\frac{1-|\lambda_2|^{k-1}}{1-|\lambda_2|} < |s_{\max}|\frac{|\lambda_2|}{1-|\lambda_2|} = R(\lambda_2), \quad (21)$$

from which we conclude that the projection of the k-th iterate of the map (3) on the unit vector $\mathbf{v}_2^T$ belongs to the interval

$$\mathbf{X}_{s_{k-1}} = \Big(\beta_2 s_{k-1} - |\beta_2 R(\lambda_2)|,\ \beta_2 s_{k-1} + |\beta_2 R(\lambda_2)|\Big), \quad (22)$$

$(s_{k-1} \in [-s_{\max}, s_{\max}])$, determined by the index of the subspace visited in the $(k-l)$-st iteration of the map. Since the number of subspaces $\mathbf{I}^3_{s_j}$ is $2s_{\max}+1$, the number of intervals (22) is also $2s_{\max}+1$. If $R(\lambda_2) < 0.5$, then there will be some gaps between those intervals. Therefore, all the attractors of the map (3) will belong to the union of $2s_{\max}+1$ disjointed subspaces of $\mathbf{I}^3$, $(\Psi_s \subset \mathbf{I}^3)$, defined by

$$\Omega_4 = \bigcup_{s=-s_{\max}}^{s_{\max}} \Psi_s, \quad (23)$$

where $\Psi_s(\mathbf{x}) = \{\mathbf{x} : d_l(s) < \mathbf{v}_2^T\mathbf{x} < d_h(s);\ \mathbf{x} \in \mathbf{I}^3\}$; $s_{k-1} \in [-s_{\max}, s_{\max}]$, $d_l(s) = \beta_2 s - |\beta_2 R(\lambda_2)|$ and $d_h = \beta_2 s + |\beta_2 R(\lambda_2)|$.

It means that each point of any trajectory of the map, after certain iteration, will belong to some parts of the phase space that are bounded by the planes $\gamma_l = \{\mathbf{x} : \mathbf{v}_2^T\mathbf{x} = d_l(s)\}$ and $\gamma_h = \{\mathbf{x} : \mathbf{v}_2^T\mathbf{x} = d_h(s)\}$ normal on $\mathbf{v}_2^T$.

In this case, the set Ω_4 doesn't depend on the initial condition $\mathbf{x}_0$ and it is definitely the invariant set of the phase space since $\mathbf{F}^k(\Omega_4) \subseteq \Omega_4$ for any k.

V. Simulations

In this section we illustrate our analytical results by functions of the histograms of the projections of the trajectories of two suitably chosen third order digital filters that belong to the observed classes of filters.

In Fig. 3, we present the histograms of the projections of the trajectory of the filter with parameters $\mathbf{a} = (-1.77, 1.57, 0.6)$ (starting from $\mathbf{x}_0 = [-0.34, 0.23, 0.7]^T$) on the eigenvectors of the transpose of the Jacobijan of the map that describes the behavior of the filter. The Jacobijan matrix has the following eigenvalues: $\lambda_1 = l$, $\lambda_2 = 1.3$, and $\lambda_3 = 0.9$.

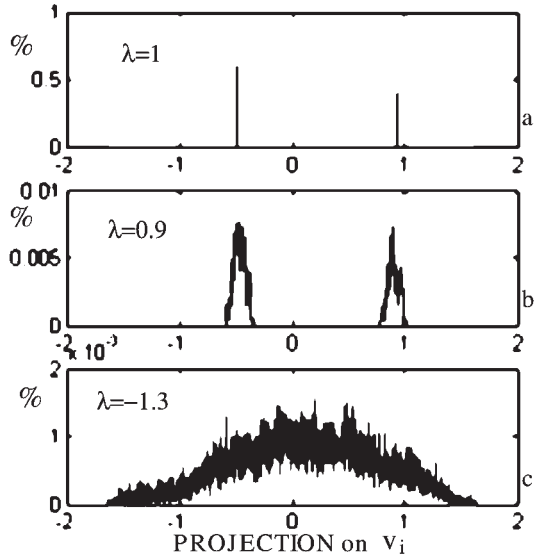


Fig. 3. Histograms of the projections of the first 20 000 points of the trajectory of the filter on the eigenvectors of $\mathbf{A}^T$, binned into 1001 containers: a) projection on $\mathbf{v}_1$, b) on $\mathbf{v}_2$, c) on $\mathbf{v}_3$

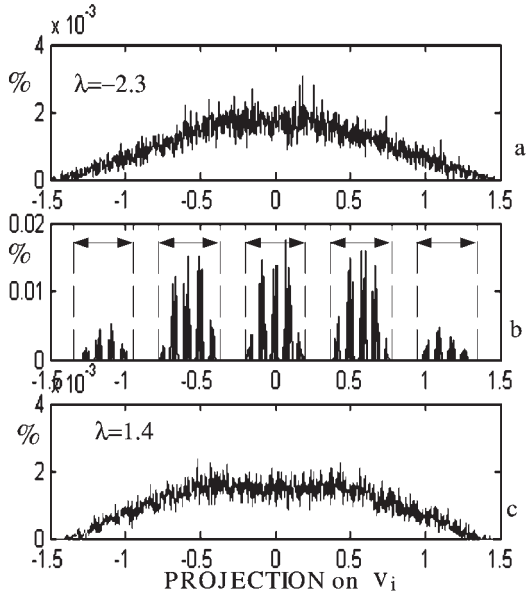


Fig. 4. Histograms of the projections of the first 20 000 points of the trajectory of the filter with eigenvalues of $\mathbf{A}$: $\lambda_1 = -2.3$, $\lambda_2 = 0.15$, $\lambda_3 = 1.4$ on the eigenvectors of $\mathbf{A}^T$, binned into 1001 containers: a) projection on $\mathbf{v}_1$, b) on $\mathbf{v}_2$, c) on $\mathbf{v}_3$

The histogram of the projection of the trajectory on the eigenvector $\mathbf{v}_1$ of $\mathbf{A}^T$ that correspond to $\lambda_1 = 1$ is discrete with two clearly defined picks that correspond to projection of the planes (Fig. 2) visited by the trajectory.

The histogram of the projection on the eigenvector $\mathbf{v}_2$ of $\mathbf{A}^T$ that corresponds to $\lambda_2 = 0.9$, $(|\lambda_2| < 1)$ is localized in two narrow isolated intervals, while the histogram of the projection of the trajectory on the eigenvector $\mathbf{v}_2$ of $\mathbf{A}^T$ that corresponds to $\lambda_3 = -1.3$ is non-uniformly speeded all over the projection domain.

Although this filter does not satisfy the constrain $R(\lambda_2) < 0.5$, the projection of its trajectory on $\mathbf{v}_2$ is concentrated on a narrow region due to the nature of the admissible symbolic sequences that correspond to the points of the phase space. Numerous simulations have shown us that the width of the isolated nonzero intervals of the histogram will start to spread (and eventually it will begin to fractalize) by decreasing the module of the eigenvalue $|\lambda_2|$ toward zero. When the condition $R(\lambda_2) < 0.5$ is reached, the trajectory of the filter enters the region Ω_4 and remains there.

In Fig. 4 we present the histograms of the projections of the trajectory of a filter described by a map which Jacobijan have the following eigenvalues: $\lambda_1 = -2.3$, $\lambda_2 = 0.15$ and $\lambda_3 = 1.4$.

In this case the condition $R(\lambda_2) < 0.5$ is satisfied. Therefore the histogram of the projection of the trajectory on $\mathbf{v}_2$ is located in five intervals $\mathbf{X}_s$, $s = -2, -l, 0, l, 2$, with gaps among them as illustrated in Fig. 4b. As $R(\lambda_2) \to 0$, the width of the intervals with nonzero histogram functions starts to decrease (and fractalize) in favor of the gaps between them.

The nature of the histograms of the projections of trajectories on eigenvectors that correspond to eigenvalues that are out of the unit circle is out of the scope of this paper.

VI. Conclusions

In this paper we have analytically located the invariant subspaces of some classes of third order digital filters with two's complement overflow by using the eigenvalues of the Jacobijan which describes the behavior of the filter. In cases of real eigenvalues that belong to the unit circle, the invariant subspaces are sets of parallel planes determined by the starting point of the trajectories of the filter. In case of eigenvalues with sufficiently small modules, the invariant subspaces are set of slices of the phase (each bounded by two parallel planes) which are independent of the starting point of the trajectories.

References

[1] L. Chua and T. Lin, "Chaos in digital filters", *IEEE Trans. Circuits and Systems*, Vol 35 pp 648-658, June 1988.

[2] L. Chua and T. Lin, "Fractal pattern in second order nonlinear digital filters", *Int J. Circuit Theory and Application*, Vol 18, pp 541-550, 1990.

[3] L. Chua and T. Lin, "Chaos and Fractals From Third-Order Digital Filters", *Int. J. Circuit Theory and Applications*, vol 18, pp 241-255,1990.

[4] C. Mitrovski and Lj. Kocarev, "Nonlinear phenomena in Third order digital filters", Seville Spain, NDES-1996, pp 405-409, June 1996.

[5] C. Mitrovski, Lj. Kocarev and N. Jonovska, "On a clas of n-th order digital filters operating outside the region of stability", *Int. Journal of Circuit Theory and Applications*. Vol 26, pp 199-205. February, 1998.

Inverse Hausdorf LC Filters

Peter S. Apostolov[1]

Abstract – **This paper presents a method for synthesis of inverse filters with Hausdorf-type of transmission characteristics. The frequency characteristics of the filters are determined and a comparison with inverse Chebyshev filters is done.**

Keywords – **approximation, Hausdorf polynomial, synthesis, LC filter, frequency characteristic**

I. Introduction

Hausdorf filters are implemented via an approximation of the "shifted" Delta-function with a Hausdorf polynomial of the following type:

$$P_n(\omega) = \varepsilon T_n\Big(\frac{2\omega}{2-\alpha\varepsilon}\Big), \tag{1}$$

where T_n is a Chebyshev polynomial, ε is the best approximation of a "shifted" delta function, α is parameter, defining the hold bandwidth of the filter, ω is the angle frequency. Low-pass non-inverse filters synthesized using this type of approximation have parameters identical to the Chebyshev filters' parameters. The pass bandwidth of these filters is narrowed by a coefficient equal to one half of the Hausdorf distance $\alpha\varepsilon$ [3].

Given the filter order n and passband ripple DA [dB], the Hausdorf space ε and the argument α could be found via the following equations [2]:

$$\varepsilon = \sqrt{10^{0.1DA}-1}, \tag{2}$$

$$\alpha\varepsilon = 2\frac{\cosh\Big[\frac{1}{n}\text{arccosh}\Big(\frac{1}{\varepsilon}\Big)\Big]-1}{\cosh\Big[\frac{1}{n}\text{arccosh}\Big(\frac{1}{\varepsilon}\Big)\Big]+1}. \tag{3}$$

The inverse low-pass Hausdorf filters are of interest from the syntheses point of view. They have a maximally flat passband and an even ripple stopband.

II. Synthesis Implementation

The synthesis has the following prerequisites: filter order n, cut-off frequency f_c and its transmission function attenuation $k = \sqrt{10^{0.1DA}-1}$, stopband frequency f_s and the Hausdorf space $\alpha\varepsilon$.

The Hausdorf space $\alpha\varepsilon$ can be derived by calculating the transmission function of an inverse Chebyshev filter of the same order for the stopband frequency:

$$DS_{ch} = \log\Big[k^2\cosh^2\Big(n\text{arccosh}\frac{f_s}{f_c}\Big)\Big]. \tag{4}$$

Next, the ε of its low-pass filter prototype is determined as:

$$\varepsilon = \frac{1}{\sqrt{10^{0.1DS}-1}}$$

and from equation (3) $\alpha\varepsilon$ is derived.

The square of the transmission function module is derived from the characteristic function, which in this case is the Hausdorf polynomial:

$$|A|^2 = \frac{k^2T_n^2\Big(\frac{2-\alpha\varepsilon}{2\omega}\Big)}{1+k^2T_n^2\Big(\frac{2-\alpha\varepsilon}{2\omega}\Big)}. \tag{5}$$

The two functions can be described as relations between the three polynomials $e(s)$, $p(s)$ and $q(s)$ of the complex frequency $s = j\omega$. The polynomial $e(s)$ is a strict Hurvitz polynomial and its roots are the fundamental frequencies of the filter, the roots of $p(s)$ are extreme frequencies for which the transmission function has infinite attenuation. The answer to the synthesis problem is usually found by determining two of the polynomials, the third polynomial is the result of the Feldtkeller equation:

$$e(s)e(-s) = p(s)p(-s) + q(s)q(-s). \tag{6}$$

The roots of $e(s)$ and $p(s)$ are derived as follows:

$$S_p = \frac{\Big(1-\frac{\alpha\varepsilon}{2}\Big)}{(Q_1+jQ_2)}, \tag{7}$$

where

$$\begin{aligned} Q_1 &= -\sin\Big(\frac{2i-1}{n}\frac{\pi}{2}\Big)\sinh\Big[\frac{1}{n}\text{arcsinh}\Big(\frac{1}{k}\Big)\Big];\\ Q_2 &= \cos\Big(\frac{2i-1}{n}\frac{\pi}{2}\Big)\cosh\Big[\frac{1}{n}\text{arcsinh}\Big(\frac{1}{k}\Big)\Big];\\ S_n &= \frac{j\Big(1-\frac{\alpha\varepsilon}{2}\Big)}{\cos\Big(\frac{2i-1}{n}\frac{\pi}{2}\Big)}, \quad (i = 1 \div n) \end{aligned} \tag{8}$$

The value of the transmission function, when the cut-off frequency and the stopband frequencies are known, is:

$$DS_h = -20\log\sqrt{\frac{k^2\cosh^2\Big[n\,\text{arccosh}\frac{f_c(1-\alpha\varepsilon/2)}{f_s}\Big]}{1+k^2\cosh^2\Big[n\,\text{arccosh}\frac{f_c(1-\alpha\varepsilon/2)}{f_s}\Big]}}. \tag{9}$$

The value of the elements is calculated using the methodology described in [4] by transforming the variable s into a new variable z:

$$z^2 = 1 + \frac{\omega_c^2}{s^2} \tag{10}$$

[1]Peter S. Apostolov is with the Institute for Special Technical Equipment-MI, E-mail: p_apostolov@abv.bg

Based on this methodology, two different software programs for filter calculation are offered in [3], called APPROX and LC. When entering the input data, the frequencies resulting from (8) need to be used.

III. Frequency Characteristics of an Inverse Hausdorf Filter

A. Magnitude (amplitude) response

Using the synthesis method described above, a low-pass Hausdorf filter of third order ($n = 3$) has been calculated assuming that the cut-off frequency is 10 KHz, the stopband frequency is 15 KHz, 0.3 dB attenuation at f_c, input and output resistance of 1 Ω. The electrical circuit is shown in Fig. 1, and a computer simulation of the magnitude response of the filter is shown in Fig. 2. The magnitude response of an inverse Chebyshev Filter with the same input specification is shown in Fig. 3. The comparison of the two shows that the Hausdorf filter has a steeper slop in $(f_c \div f_\infty)$ interval. In this case the attenuation of the stopband frequency is greater than 4.2 dB. The maximum in the passband is greater than 1.85 dB. This is due to the fact that the extreme frequencies of the inverse Hausdorf Filter are $(1-\alpha\varepsilon/2)$ times lower than these of the inverse Chebyshev filter, as seen in equation (8).

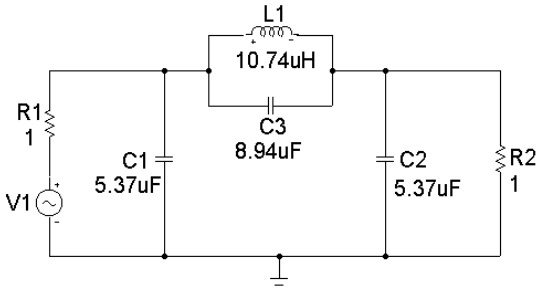


Fig. 1.

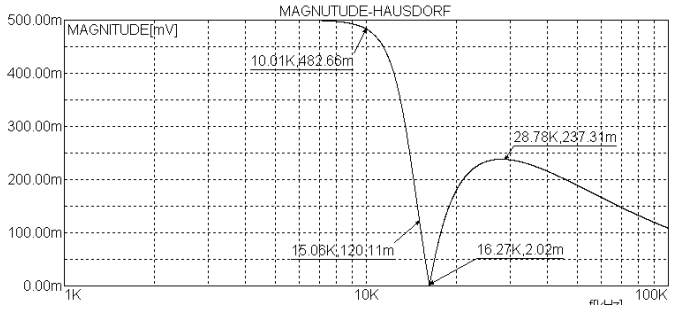


Fig. 2.

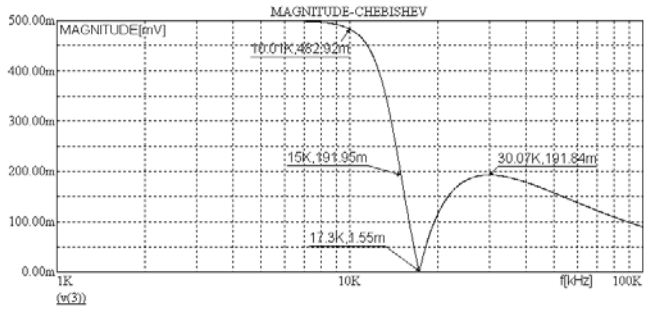


Fig. 3.

B. Phase-frequency response

Polynomial $e(s)$ is determined by solving the equation (7). Shown as a rational function, after substituting $s = j\omega$, it is broken into real e_R and imaginary e_I polynomials. The same is done for the polynomial $p(s)$, derived in (7). The Phase-frequency Response is:

$$\varphi(\omega) = \arctan\frac{p_I}{p_R} - \arctan\frac{e_I}{e_R}. \qquad (11)$$

The phase characteristics of Hausdorf and Chebyshev filters are shown on Figs. 4 and 5. They point out that the Hausdorf filters have worse linearity – it is 5.36% for the marked frequencies.

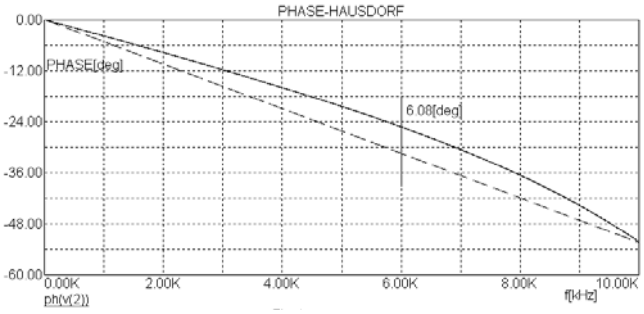


Fig. 4.

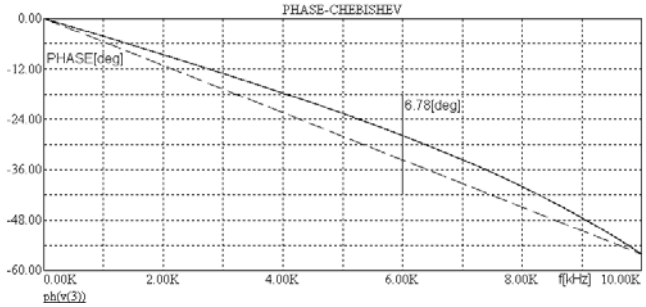


Fig. 5.

C. Group delay time (GDT)

GDT is calculated from:

$$t_{gr} = \mathrm{Re}\Big[\frac{1}{e(s)}\frac{de(s)}{ds} - \frac{1}{p(s)}\frac{dp(s)}{ds}\Big]. \qquad (12)$$

Figs. 6 and 7 show the GDT graphs of the filters described in the previous paragraph. It follows from the graph that the Hausdorf filter has on overall lower values for t_{gr}, meaning that it has lower reactivity. Comparing the two graphs, it is calculated that the Hausdorf filter graph is 13.6% more non linear.

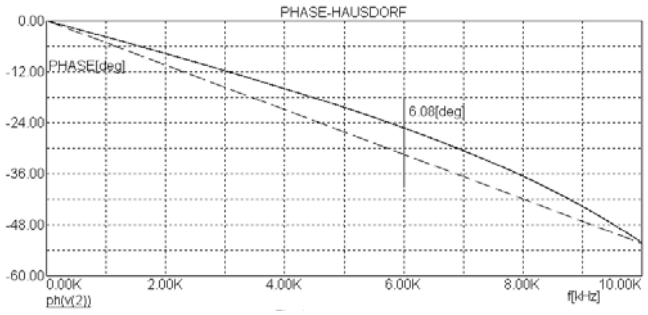


Fig. 6.

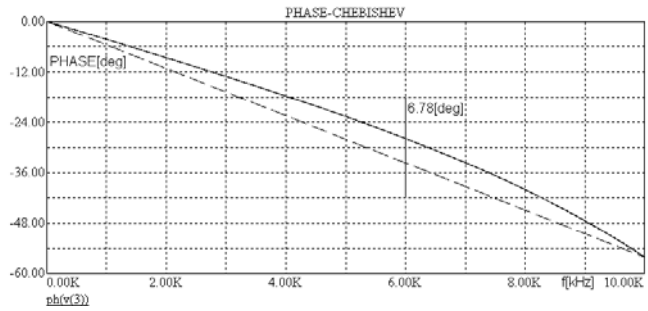


Fig. 7.

IV. Conclusion

The specific characteristics for the Hausdorf filters show that they can be used in circuits requiring high attenuation for frequencies close to the cut-off frequencies. Their lower reactivity is a prerequisite for their use in cases when a greater correspondence in the shape of the input and output signal is required.

References

[1] L. Chua and T. Lin, “Chaos in digital filters”, *IEEE Trans. Circuits and Systems*, Vol 35 pp 648-658, June 1988.

[2] L. Chua and T. Lin, “Fractal pattern in second order nonlinear digital filters”, *Int J. Circuit Theory and Application*, Vol 18, pp 541-550, 1990.

[3] L. Chua and T. Lin, “Chaos and Fractals From Third-Order Digital Filters”, *Int. J. Circuit Theory and Applications*, vol 18, pp 241-255,1990.

[4] C. Mitrovski and Lj. Kocarev, “Nonlinear phenomena in Third order digital filters”, Seville Spain, NDES-1996, pp 405-409, June 1996.

[5] C. Mitrovski, Lj. Kocarev and N. Jonovska, “On a clas of n-th order digital filters operating outside the region of stability”, *Int. Journal of Circuit Theory and Applications*. Vol 26, pp 199-205. February, 1998.

Transitional Characteristics of the Loudspeaker Systems

Ekaterinoslav S. Sirakov[1], Atanaska A. Angelova[2]and Georgi K. Evstatiev[3]

Abstract – **In the work, a block scheme of a digital system for researching pulse characteristics of an electrodynamic loudspeaker with a direct radiating is considered. The given measuring the system, she displays for demonstrating: the frequency response, the pulse reverberation, the time of increment, the time of fading. These characteristics are particularly important for the audio (the sound) quality of the loudspeaker.**

Keywords – **Audio, Loudspeaker, Pulse (transitional) characteristics, Cumulative spectral decay.**

I. Introduction

The dynamic transitional characteristic of the spectrum of the sound pressure created by a loudspeaker or a Loudspeaker system with a pulse signal, enables us to read the so-called impediment of the resonances (delayed).

In fact, this can not be read by the conventional frequency characteristics (response) and curve the phase. These resonances are a result of the own resonances of the tweeters, midrange and woofer; of the own resonance frequency of the box size; of the box corners; of the special location of the driver in the box; of the influence of the separating filters and the correcting chains, etc.

According to many experts these results correlate with the subjective perception of the spatial characteristic of the sound.

The pulse $g(\tau)$ is measured in order to directly evaluate the bendings in the transitory area of the signals of a radiating loudspeaker.

The pulse characteristic is the response of the system under the influence of a signal – delta a function (unit function) with zero initial conditions.

II. Block Diagram

Through power amplifier 2, the pulse generator 1 or the PC (See Fig. 1) apply to the measured loudspeaker/loudspeaker system 3 a series of rectangular pulses with the width of $2 \div 200$ μs, repetition period less than 10 Hz and amplitude of $10 \div 100$ V (See Fig. 2 and 3).

The signal from the microphone or the microphone preamplifier 6 is applied to the sound card 5 and across the analogue to digital converter it is applied to the personal computer 9 (See Fig. 1). The $16 \div 24$ bits, ADC (analogue to digital converter) must be used to provide dynamic range of up to $86 \div 120$ dB.

When Qts is greater than 0.5 [1], the response is oscillatory with increasing values of Qts contributing increasing amplitude and decay time. That's why the value of Qts=0.5 is critically-damped alignment.

The input voltage V(Vad:+)/1.3 MegV is rectangular volt-

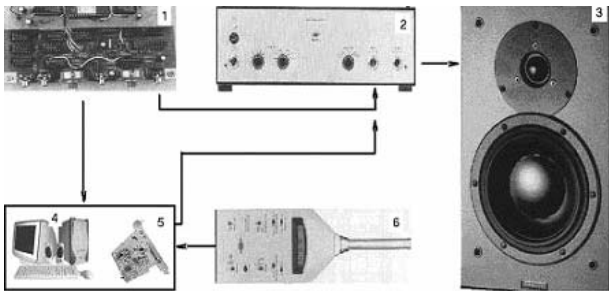


Fig. 1. Block diagram of a system for measuring the characteristics of the Loudspeaker systems.

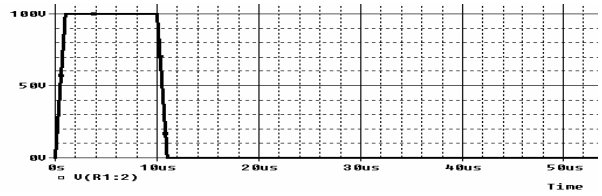


Fig. 2. Rectangular pulse with a width of 10 μs, a time of increment and fading 1 μs, maximum amplitude 100 V.

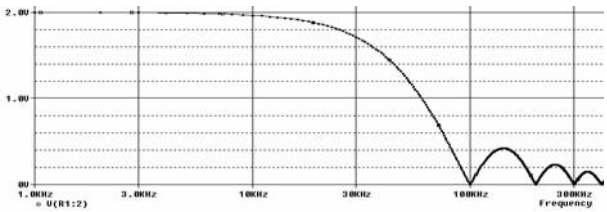


Fig. 3. Spectrum of the rectangular pulse with a width of 10 μs.

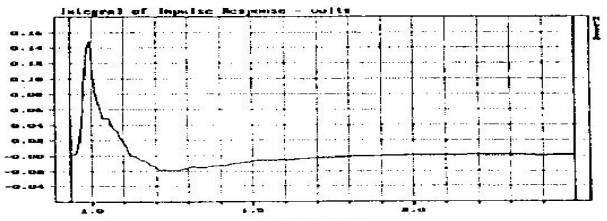


Fig. 4. Transitional characteristic of a loudspeaker D-28AF. (the response is made available by the producer DYNAudio)

[1]Ekaterinoslav S. Sirakov is with the Department of Radio engineering, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria, E-mail: katio@mail.bg

[2]Atanaska A. Angelova is with the Department of Radio engineering, Faculty of Electronics, Technical University-Varna, Studentska Street 1,Varna 9010, Bulgaria, E-mail: lz4hi@yahoo.com

[3]Georgi K. Evstatiev is with the Department of Radio engineering, Faculty of Electronics, Technical University-Varna Studentska Street 1, Varna 9010, Bulgaria, E-mail: evstatg@mail.bg

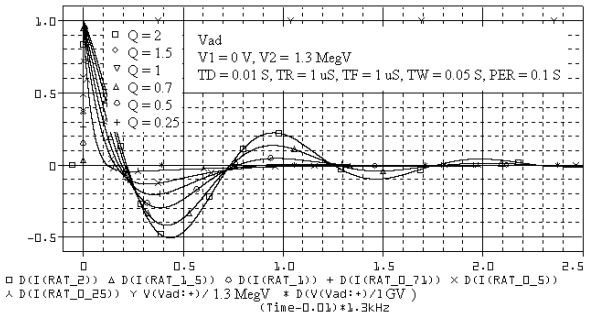


Fig. 5. Normalized transitional characteristic of the loudspeaker D-21AF for parameter Q.

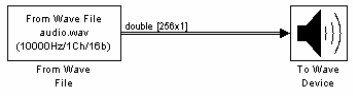


Fig. 6. Pulse Generator in SimuLink.

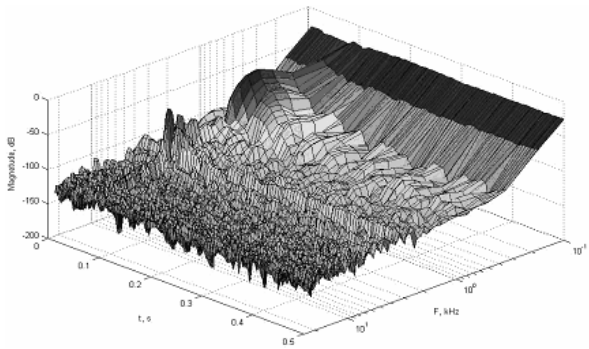


Fig. 7. Dynamic transitional characteristic of the spectrum of the sound pressure, created by the Loudspeakers system with a pulse signal. (Commutative spectral decay).

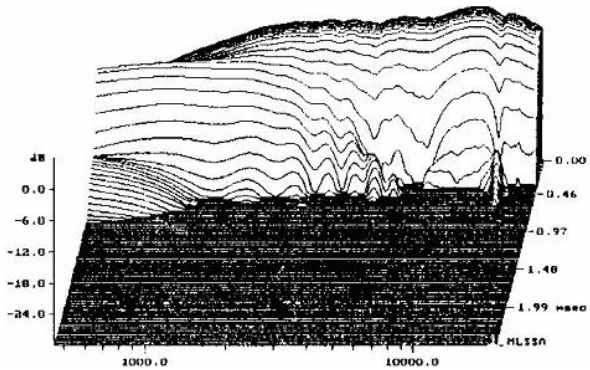


Fig. 8. Dynamic transitional characteristic of the spectrum of create by the loudspeaker D-28AF sound pressure with a pulse signal. (the response is made available by the producer DYNAudio)

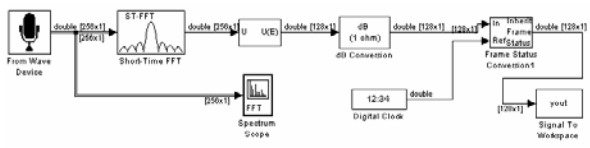


Fig. 9. Measure the transitional characteristic the loudspeakers with the program SimuLink.

age with a time of fading (these definitions are for the simulations in the program PSpice) TR=TF=1 uS, its derivative D(V(Vad:+)/1G) is delta a function (unit function) [2].

III. Block Diagram in Sumulink and Real Concrete Measured Results

All values in the spectrum are formed in a three - dimensional array. All data from the array are processed and displayed for demonstrating with the help of a suitable algorithm.

On the three axises is: the time in seconds, the frequency in kHz and the level of the sound pressure in dB.

```
out=[];
t=[];
out(:,:)=yout(:,1,:);
t=tout(:,:)';
surf(t,y,out);
set(gca,'YDir','reverse') set(get(gca,'XLabel'),'String','t, s')
set(get(gca,'YLabel'),'String','F, kHz')
set(get(gca,'ZLabel'),'String','Magnetude, dB')
set(gca,'YScale','Log')
```

IV. Conclusion

The block diagram offered successfully demonstrates the transitional characteristics of the sound system, and the software doesn't require a concrete SoundBlaster; it is only limited concerning the SoundBlaster used.

In future the test rectangular pulse could be generated by an external device, while a DSP (Data Signal Processor) would be able to digitally process the results, thus replacing the SoundBlaster partially or completely. The increase of the signal to noise ratio is achieved by a series of pulses with accumulation and processing of the results. (An example: in case the pulses are 64, the signal to noise ratio can be increased with 18 dB)

References

[1] R. H. Small, "Closed-Box Loudspeaker Systems Part I: Analysis," J. Audio Eng. Soc.,, vol. 20, Number 10, 1972 Dec.
[2] Allan V. Oppenheim, Al. S. Uilski and J. . Jang, "Signals and Systems", "Tehnika", Sofia, 1993.
[3] D. Campbell, "MLSSA vs. sine testing: a tutorial on methodology" Speaker Builder, Volume13, Number 2, 1992.
[4] Irina A. Aldishina, A. G. Voishcilo, "High-Fidelity loudspeaker systems and loudspeakers", Moscow, Radio and svias, 1985.
[5] Jim Moriyasu, "A Study of Midrange Enclosures" Speaker Builder, the Loudspeaker Journal, Volume 21, Number 8, December 2000
[6] Heinz Scbmitt "ATB-Precision, das Profisystem mit intutiver Bedienbarkeit", Klang+Ton, Februar/Marz, 2, 2002.
[7] Rob Baum, "SLAB Technology Flat Diagrams Loudspeakers", Voice Coil, Volume 15, Issue 3, January 2002.
[8] http: //www.microsim.com/, /www.pspice.com/
[9] http://www.mathworks.com/products/dspblockset/
[10] Brochures issued by DynAudio, Morel, Fokal Audax

Examination and Analysis of Psychoacoustic Models in Transparent Perceptual Audio Compression

Angel R. Kanchev

Abstract – **Different psychoacoustic models for transparent perceptual audio compression concerning fastness and calculation complexity are examined. Encoders using these models are compared.**

Keywords – **Audio coding, filter bank, psychoacoustic model, calculation complexity, time delay.**

I. Introduction

Masking properties of the human auditory system allow lossy audio encoding with no hearing quality loss (i.e. transparent encoding). The mathematical models using some of those properties to determine sound's audibility are called psychoacoustic models. In this article, a determination of computation complexity and time delay for encoders with different psychoacoustic models is made (delay caused by buffering - not by calculations). The analysis method is described in point II. The encoders to be compared are presented in point III. These encoders are chosen because they are optimal by some of the comparing parameters. The results of the examination and analysis are given in point IV.

II. Examination and Analysis Description

Input audio signal is considered to be one-channel (mono), 16-bit/sample, sampled with frequency 44.1 kHhz.

The complexity is given as a number of DSP operation per one input sample for a generalized DSP. The number of DSP operations is calculated as the number of operations multiplied by the number of clock ticks for each operation to be done. Multiplication, addition and multiplication with addition are considered to be one clock tick operations. Division is 5..10 cycles, log – 5..15. Radix 4 FFT is considered to be 8000 cycles for 256-point input and 32000 for 1024 point input. Divided by the number of samples FFT is 31.25 cycles per one input sample in both cases.

The formulas for psychoacoustic model using Signal To Mask Ratio (SMR) are given in point IIIA. This model (with modifications) is used in all examined encoders. There are analytical models that are not presented here because they have too big computational cost or their parameters are not optimal (although the hearing quality given by some of them is much better). These are: model with perceived loudness (N') [1]; with specific partial loudness (N_S') [3]; with Just-Noticeable Level of Difference (JNLD) [1] and with Just-Noticeable Distortion (JND) [1]. Example encoders using such models are MASCAM, MUSICAM, OCF, PXFM and ASPEC – [11]. Actually, encoders using Filter Banks (FB) only are examined here (although not all encoders are using FB).

The delay is the biggest delay caused by buffering and filter bank processing. It has no connection with the time necessary for calculations. The second one is determined by the number of DSP operations (and DSP's clock frequency).

III. Encoders with Filter Banks

A. Common – NMR Calculation

The purpose of a psychoacoustic model is to calculate the Noise to Mask Ratio (NMR) using some masking model of the human auditory system. The encoding, which uses psychoacoustic models, is inaudible when the maximum NMR is negative or zero dB. Non-linear loudness scales used in the calculations are phon [1] and sone [5]. Non-linear frequency scales are Bark – $z(f)$ [4] and Equivalent Rectangular Bandwidth Scale – ERBS(f) [2,3].

1. Frequency distribution of the signal level $L(k)$ determination. The input samples sequence $s(n)$ with number of bits per sample b is normalized (Eq. (1)) and transformed with FFT with length N (Eq. (2)). To avoid spectral leak caused by the finite sum ($N < \infty$) Hann window – Eq. (3) is used.

$$x(n) = \frac{s(n)}{N\left(2^{b-1}\right)} \tag{1}$$

$$L(k) = PN + 10\log_{10}\left|\sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi kn}{N-1}}\right|^2, \ 0 \leqslant k \leqslant \frac{N}{2} \tag{2}$$

$$w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N-1}\right)\right] \tag{3}$$

PN – power normalization term = 90 dB. The index k determines corresponding Fourier spectral line with frequency $f(k)$. Bark and ERB scales are indexed buffers: $z(k)$ and ERBS(k).

2. The threshold in quiet L_{Tq} is taken from buffer L_{Tq}(k).

3. Each masker's threshold level L_T calculation:

$$L_T(k, k_c) = L(k) + SF(k, k_c) + MI(k_c), \ \text{dB} \tag{4}$$

In Eq. (4) $SF(k, k_c)$ is a "Spreading" Function simulating the fall-off (in dB) of the masking curve for sine tone masker with frequency k_c. $MI(k_c)$ is a Masking "Index" (in dB) – correction caused by the frequency width of the masker.

Optimization formulas in calculation of $SF(k,k_c)$:

$$SF(k,k_c)=SF(\Delta z,z_c)= \begin{cases} 17(\Delta z+1)-(0.4L(z_c)+6), & -3\leqslant\Delta z<-1 \\ (0.4L(z_c)+6)\Delta z, & -1\leqslant\Delta z<0 \\ -17\Delta z, & 0\leqslant\Delta z<1 \\ -(\Delta z-1)(17-0.15L(z_c))-17, & 1\leqslant\Delta z<8 \end{cases} \quad (5)$$

In Eq. (5):

$$\Delta z=z(k_c)-z(k);\ z_c=z(k_c) \quad (6)$$

Optimized calculation of $MI(k_c)$:

$$MI(k_c)=\alpha MI_T(k_c)+(1-\alpha)MI_N(k_c),\ \text{dB} \quad (7)$$

$$MI_T(k_c)=-6.025-0.275z(k_c)\ ,\text{dB} \quad (8)$$

$$MI_N(k_c)=-2.025-0.175z(k_c)\ ,\text{dB} \quad (9)$$

In Eqs. (7)-(9): MI_T is tonal index, MI_N is noise index, $\alpha\in[0;1]$ – "tonality" factor (constant for each critical band [1])

$$\alpha=\min\left(\frac{SFM}{-60},1\right);\quad SFM=10\log_{10}\left(\frac{G_m}{A_m}\right) \quad (10)$$

$$\alpha=-0.3-0.43\log_{10}(e)\in[0;1] \quad (11)$$

SFM – Spectral Flatness Measure; e – relative prediction error (for models with prediction); G_m and A_m are geometric and arithmetic means of $L(f)$ in a critical band.

4. Temporal masking – backward masking is seldom taken into account so forward masking is examined here: small

$$L_{TF}(t,T_m,k,k_c)= =L_T(k,k_c)\left[1.0-\frac{1}{1.35}\arctan\left(\frac{13.47t}{T_m^{0.25}}\right)\right],\ \text{dB} \quad (12)$$

T_m – masker duration, s; t – time after the end of masker, s. For Eq. (12) to be correct $t<t_m=0.3307T_m^{0.25}$ should be satisfied. Up until t_m seconds after each masker, L_{TF} is summed to the current L_T and the cumulative level $L_{T\Sigma}$ is determined. In most psychoacoustic models $L_{T\Sigma}\equiv L_T$.

5. Excitation level $E(k)$:

$$E(k)=10\log_{10}\left\{\left[\sum_{k'=0}^{N-1}\left(10^{L_{T\Sigma}(k',k)/10}\right)^p\right]^{1/p}\right\},\text{dB} \quad (13)$$

$p\in[0.2;0.3]$ for high quality models and p=1 for fast calculations. For optimization purposes in MPEG [6], [7] the sum is over "detected" masker frequency indexes only.

6. NMR calculation:

Global threshold level is L_{TG}:

$$L_{TG}(z(k))=10\log_{10}\left(10^{\frac{L_{Tq}(z(k))}{10}}+10^{\frac{E(z(k))}{10}}\right),\ \text{dB} \quad (14)$$

$$SMR(k)=L(k)-L_{TG}(z(k))\ ,\ \text{dB} \quad (15)$$

$$NMR(k)=SMR(k)-SNR,\ \text{dB} \quad (16)$$

B. Encoder with Linear Filter Bank with IIR Filters

The encoder with cochlear filter bank gives the best quality [8]. It consists of Low-Pass (LPF) and High Pass (HPF) filter pairs – Fig. 1. The section S_k is with center frequency of its amplitude response f_c(k). Numerous sections for one stage (concerning the decimation) are necessary (Fig. 2).

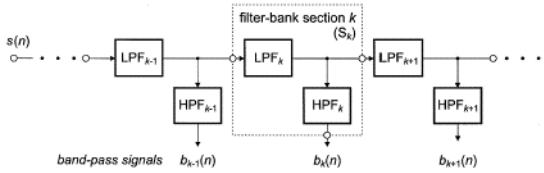


Fig. 1. Filter bank structure [8]

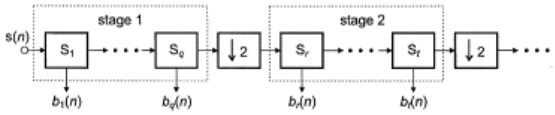


Fig. 2. Down sampling scheme [8]

Each section's center frequency can be determined by Eq. (17) for k=1,2,3,…:

$$f_c(k+1)=\begin{cases} 0.5^{1/N_s},\ f_c(k)\geqslant 500 \\ f_c(k)-22.4,\ f_c(k)<500 \end{cases},\ Hz \quad (17)$$

f_c(1) and N_s depend on sampling frequency (f_s): for f_s=44.1 kHz, f_c(1)=20948 Hz, N_s=15. The desired amp120litude frequency response of one band centered at f $_c$ for $f_c\geqslant 500$ Hz is:

$$|H(f)|=\left|\frac{1}{1+\left(\frac{f}{f_c}\right)^{S_{LP}}}\cdot\frac{\left(\frac{f}{f_c}\right)^{S_{HP}}}{1+\frac{j}{4}\left(\frac{f}{f_c}\right)^{\frac{S_{HP}}{2}}-\left(\frac{f}{f_c}\right)^{S_{HP}}}\right| \quad (18)$$

$$S_{LP}=\frac{25}{20\log_{10}(1.2)};\quad S_{HP}=\frac{8}{20\log_{10}(1.2)} \quad (19)$$

$$j=\sqrt{-1} \quad (20)$$

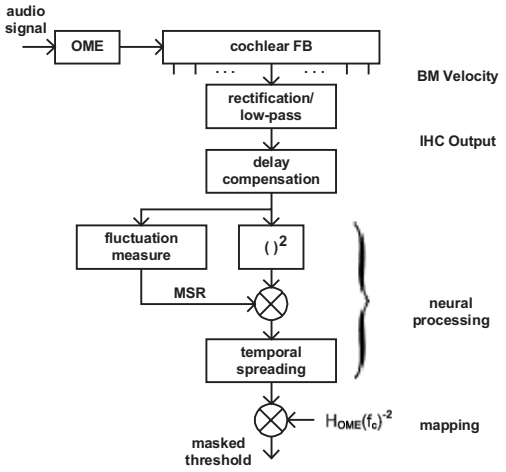


Fig. 3.

For f_c <500 Hz the desired response is a replica of the filter response closest to, but not less than a center frequency of 500 Hz shifted on a linear frequency scale.

When using cochlear filter bank the psychoacoustic model is simplified (Fig. 3).

OME – Outer- and Middle Ear transfer filter with amplitude response $H_{OME}(f)$ – see Fig. 4 [8].

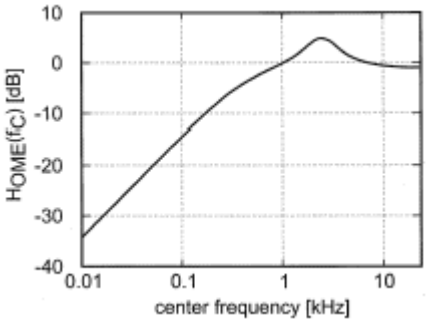


Fig. 4. OME filter amplitude response [8]

BM – Basilar Membrane; IHC – Inner-Hair Cells effect. The cut-off frequency of the second order low-pass filter is:

$$f_{LP} = \begin{cases} f_c, & f_c < 300 \\ 300\left(\frac{f_c}{300}\right)^{0.25}, & f_c \geqslant 300 \end{cases} \tag{21}$$

The delay compensation is at most 10 ms (Fig. 5) [8].

The fluctuation measure corresponds to unpredictable tonality index (1-α) – Eq. (10). MSR – Masker to Signal Ratio. The temporal spreading is for backward and forward (Eq. (12)) masking.

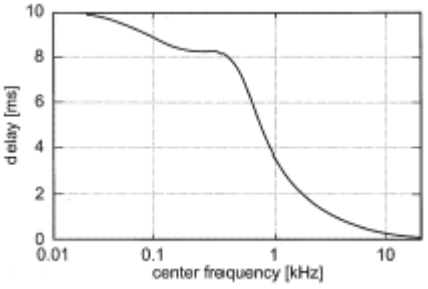


Fig. 5. Filter bank time delay [8]

C. *Encoder with Linear Filter Bank with γ – Tone Filters*

An easier implementation (and optimal encoder according quality/complexity ratio) is achieved with γ –tone filter banks [9].

$$\begin{aligned} |H\gamma(f, f_c)| &= \frac{1}{\left(1 + \left(\frac{f - f_c}{kERB(f_c)}\right)^2\right)^{n/2}}; \\ k &= \frac{2^{n-1}(n-1)!}{\pi(2n-3)!!} \end{aligned} \tag{22}$$

$H\gamma(f, f_c)$ – amplitude frequency response; f_c – center frequency; n – filter order (usually is 4).

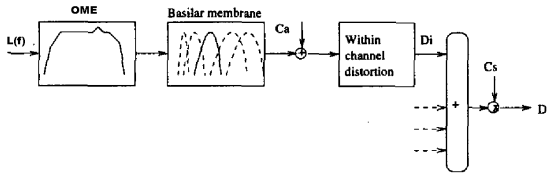


Fig. 6. General structure of the model [9]

In Fig. 6 Basilar membrane is simulated via γ – tone filter bank, C_a is the absolute threshold noise power, C_s – integral ear detectability, D_i– "specific" and D – total distortion detectability (L_T analogue).

Lets present the input signal L(f) as

$$L(f) = m(f) + r(f) \tag{23}$$

$m(f)$ is the masker and $r(f)$ – the distorted signal.

$$D(m, r) = \sum_f |r(f)|^2 v^2(f) \tag{24}$$

$v(f)$ represents the masking curve in linear scale (like SF+MI in dB scale):

$$v^2(f_m) = C_s \hat{T} \sum_i \frac{|H_{OME}(f_m)|^2 |H\gamma(f_m, f_i)|^2}{\sum_f |H_{OME}(f)|^2 |H\gamma(f, f_i)|^2 |m(f)|^2 + C_a} \tag{25}$$

f_i is the center frequency of the i-th filter.

$$\hat{T} = min\left(\frac{T}{T_{300ms}}, 1\right) \tag{26}$$

$\hat{T}$ – effective duration; $T_{300\ ms}$ represents 300 ms segment duration, T – relevant segment duration.

Coefficients C_a and C_s satisfy:

$$\begin{aligned} & C_a = C_s \hat{T} \sum_i |H\gamma(f_{1kHz}, f_i)|^2 \\ & \frac{1}{C_s} = \hat{T} \sum_i \frac{|H_{OME}(f_{1kHz})|^2 |H\gamma(f_{1kHz}, f_i)|^2 A_{53}^2}{|H_{OME}(f_{1kHz})|^2 |H\gamma(f_{1kHz}, f_i)|^2 A_{70}^2 + C_a} \\ & A_{53} = 2.10^{-7}\ \text{W / m}^2\ (8, 93.10^{-3}\ \text{Pa}) \\ & A_{70} = 10^{-5}\ \text{W / m}^2\ (6, 325.10^{-2}\ \text{Pa}) \end{aligned} \tag{27}$$

D. *Encoder with Wavelet Packet Decomposition Via FIR Filters*

The optimal encoder considering complexity and delay is the encoder with wavelet filter bank[10].

In [10] encoder with a wavelet packet filter bank is presented. Fig. 7 depicts encoder's decomposition tree (the decoder part is analogous).

Each lattice section is a N-th order FIR filter realizing Daubechies wavelet (Fig. 8).

$$A_m(z) = A_{m-1}(z) - \frac{\gamma_m}{z} B_{m-1}(z) \tag{28}$$

$$B_m(z) = \gamma_m A_{m-1}(z) + \frac{1}{z} B_{m-1}(z) \tag{29}$$

$$A_0 = \left(1 - \frac{\gamma_0}{z}\right) x \tag{30}$$

$$B_0 = \left(\gamma_0 + \frac{1}{z}\right) x \tag{31}$$

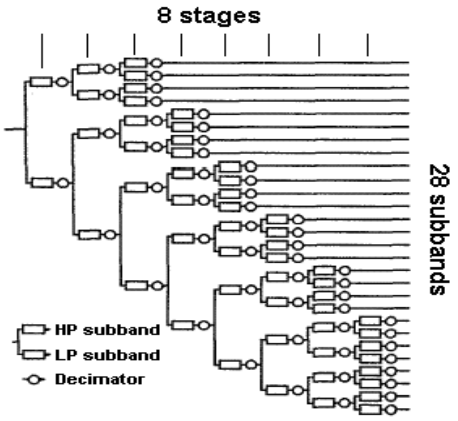


Fig. 7. Wavelet packet decomposition tree [10]

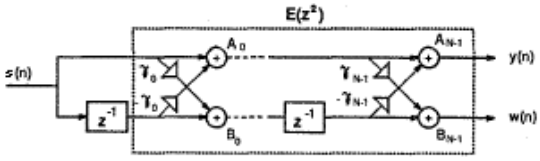


Fig. 8. Lattice element (filter) [10]

An optimization when N is even can be made because $\gamma_m = \gamma_{m-1}$, m=1,3,..,N-1. The number of sections in Fig. 8 will be cut by half and decimation position will change – see [10].

The psychoacoustic model uses Eqs. (4)–(9) with MI_N only. Determination of noise maskers is the same as in MPEG Psychoacoustic model I.

IV. Results and Conclusions

As a basis for comparison, MPEG Layer I with Psycho model I encoder is used (the filter bank is 32-band, polyphase). In Table 1 the encoders presented in chapter III are compared.

Table 1. Perceptual encoders comparison

Q	Encoder	DSP op.	delay, ms	bands
1	Cochlear FB	550	21.6	103
2	γ-tone FB	200-250	≈ 310	48
3	Wavelet packet FB	120-140	93	28
4	MPEG	400-500	11.6	32

The first column is a Quality order value (determined by descriptions in [8–10]). It is determined by the closeness of the FB bands to the critical bands and model's precision – no subjective hearing quality test is made. The third column contains number of DSP operations per one input sample (complexity value). "Delay" is the sum of the maximum buffer size in samples divided by 44100 Hz and FB time compensations. Delay under 100ms allows interactive real-time encoding.

For conclusions see Table 2.

Table 2. Conclusions

Encoder	Comment
Cochlear FB	Best quality; optimum quality / delay
γ-tone FB	Optimum quality / complexity
Wavelet packet FB	Optimum complexity + delay
MPEG	Least delay

References

[1] Zwicker, E.; Fastl, H.: Psychoacoustics, Facts and Models. Berlin; Heidelberg: Springer Verlag, 1990.

[2] Moore, B.C.J.; Glasberg, B.R.: Suggested Formulae For Calculating Auditory-Filter Bandwidths And Excitation Patterns. Journal of the Acoustical Society of America, Vol. 74 (3), September 1983, pp. 750-753.

[3] Moore, B.C.J.; Glasberg, B.R.; Baer, Th .: A Model For The Prediction Of Thresholds, Loudness, And Partial Loudness. Journal of the Audio Engineering Society, Vol. 45 (4), April 1997, pp. 224-240.

[4] Zwicker, E.; Terhardt, E.: Analytical Expressions For Critical Bandwidth As A Function Of Frequency. Journal of the Acoustical Society of America, Vol. 68 (5), November 1980, pp. 1523-1525.

[5] Stevens, S.S.: A Scale For The Measurement Of A Psychological Magnitude: Loudness. Psychological Review, Vol. 43, 1936 pp. 405-416.

[6] ISO/IEC 11172-3 – MPEG 1 Audio part, Psychoacoustic models 1 and 2

[7] ISO/IEC 13818-3 – MPEG 2 Audio part, Psychoacoustic models 1 and 2 extensions

[8] Baumgarte F.: Improved Audio Coding Using a Psychoacoustic Model Based on a Cochlear Filter Bank, IEEE, Vol. 10, No. 7, October 2002.

[9] Van de Par, S.; Kohlrausch A.: A New Psychoacoustical Masking Model For Audio Coding Applications, IEEE 0-7803-7402-9/02, 2002

[10] Black M.; Zeytinoglu M.: Computationally Efficient Wavelet Packet Coding Of Wide-Band Stereo Audio Signals, IEEE 0-7803-2431-5/95, 1995

[11] Painter T.; Spanias A.: A Review Of Algorithms For Perceptual Coding Of Digital Audio Signals, IEEE 0-7803-4137-6/97, 1977

Extremums of the Average Output Dissipation of a Class AB Audio Amplifier

Ekaterinoslav S. Sirakov[1] and Ivan Lazarov[2]

Abstract – **Theoretical analysis of the output dissipation of an audio amplifier operating in class AB. Extremums (maximal and minimal values) of the average output dissipation of an output stage.**

Keywords – **Audio, amplifier, average dissipation, class AB, extremums.**

I. Introduction

The output stage is based on push-pull circuit with complementary power bipolar transistors, complementary power MOSFET's transistors or power vacuum tubes. Fig. 1 shows popular realization with complementary power MOSFET's transistors [1÷4].

The analysis is done at a constant sinusoidal signal or at smooth change of frequency and amplitude.

The theoretical analysis and the computer simulation are performed for an active load (rated impedance).

The average dissipation output can be obtained by integrating the instantaneous dissipation output.

In class AB and output level of $m = V_0/V_{ss} = 0 \div 1$ and $v = V_{Rt}/V_{ss}$, when integrating
from: $-\arcsin(v/m)$ $(t_0 = -\phi_{V_{Rt}} = -\arcsin(v/m))$
to $\pi + \arcsin(v/m)$ $(t_1 = \pi + \phi_{V_{Rt}} = \pi + \arcsin(v/m))$:

$$P_{d(AV)} = \frac{1}{2\pi}\int\limits_{-\phi_{V_{Rt}}}^{\pi+\phi_{V_{Rt}}} P_{d(inst)}\, d\varpi t = 1 + \frac{2}{\pi} a\sin\left(\frac{v}{m}\right) + \frac{2}{\pi}\sqrt{m^2 - v^2} - \frac{v}{\pi}\sqrt{m^2 - v^2} - \frac{m^2}{\pi} a\sin\left(\frac{v}{m}\right) - \frac{m^2}{2} \quad (1)$$

In class B $(v = V_{Rt}/V_{SS} = 0)$:

$$P_{d(AV)} = \frac{V_{ss} V_O}{\pi R_L} - \frac{V_O^2}{4R_L} = \frac{2}{\pi} m - \frac{1}{2} m^2 \,, \quad (2)$$

In class A and in class AB, but at $m = V_0/V_{ss} < v = V_{Rt}/V_{ss}$:

$$P_{d(AV)} = \frac{1}{2\pi}\int\limits_{0}^{2\pi} P_{d(inst)}\, d\varpi t = 2v - m^2 \,, \quad (3)$$

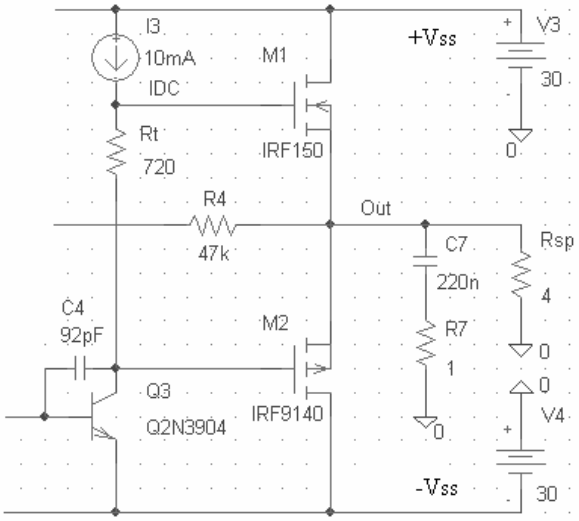


Fig. 1. Power-board (output stage) schematic diagram

II. Analysis of the Output Dissipated by the End Transistors at a Resistive Load

The first derivative of the expression for the average dissipation output in class B ($P_{d(AV)} = 2m/\pi - m^2/2$) is a straight line, trace 3 $(2/\pi - m)$, which crosses the ordinate

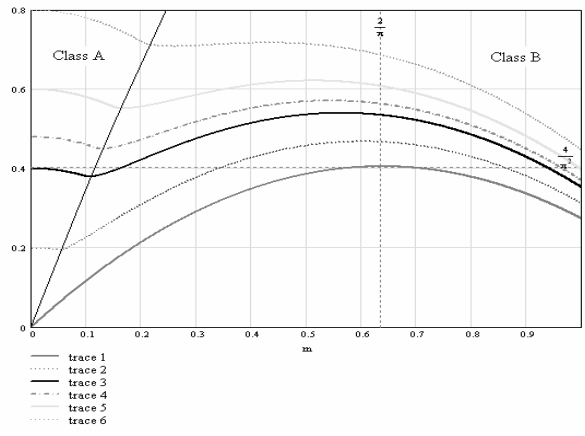


Fig. 2. Average dissipation output of both transistors M_1 and $M_2 P_{d(AV)}$ in class AB and B in case with a resistive load.
trace 1 – a curve of the average output dissipation of both transistors of the output stage working in class B.
trace 2÷6 – curves of the average output dissipation of both transistors of the output stage working AB.

[1]Ekaterinoslav S. Sirakov is with the Technical University-Varna, Faculty of Electronics, Department of Radio engineering, Studentska Street 1, Varna 9010, Bulgaria E-mail: katio@mail.bg

[2]Ivan Lazarov is with the Technical University-Varna, Faculty of Electronics, Department of Mathematic, Studentska Street 1, Varna 9010, Bulgaria E-mail: ivan@mail.bg

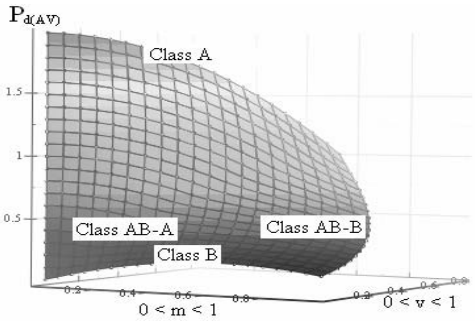


Fig. 3. 3 D - Average dissipation output in class A, AB and B

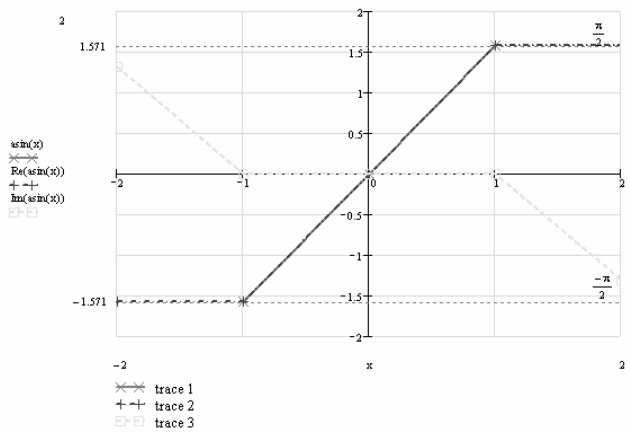


Fig. 4. Real and Imagine part of the mathematical function $\arcsin(x)$.
trace 1 – in this area $\arcsin(x)$ accepts real values only. This is a line determined by the points with co-ordinates $-1, -\pi/2$; $0, 0$ ($v = 0$ – class B) and $1, \pi/2$. This is the area of class AB, but at $m > v$ - the output stage operates with big signal, i.e. class B.
trace 2 – the real part of $Re[\arcsin(x)]$. Out of the area of trace 1 $Re[\arcsin(x)]$ accepts the value of $\pm\pi/2$ for $|\pm x| \geqslant 1$ correspondingly. This is the area of class AB, but at $m < v$ – the output stage operates with small signal, i.e. class A.
trace 3 – the imaginary part of $\arcsin(x)$.

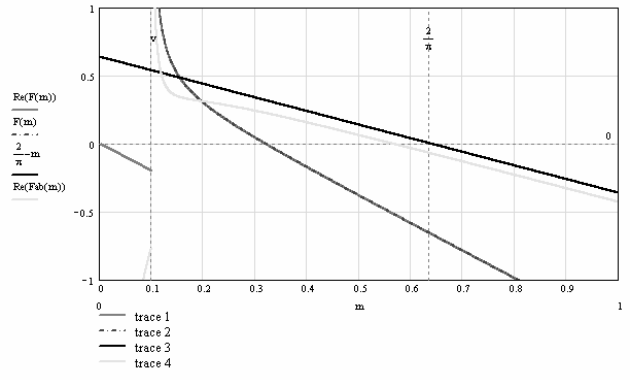


Fig. 5. Curves of the first derivative of the average power dissipation $P_{d(AV)}$ in class AB and B at a resistive load.

at $m = 2/\pi = 0.637$,– an extremum.

In class AB crosses the ordinates twice (two peaks) and is broken for the minimum.

III. Conclusions

Fig. 2 shows the extremums of the average power dissipation of the two push-pull arms of the output stage operating in class AB.

The average power dissipation in class AB there is maximum and minimum values depending on the level of the signal $m = 0 \div 1$, where the parameter is the bias voltage determining the working point of the end transistors.

The first minimum describes the curve which starts from the center of the co-ordinate system $0, 0$ ($v = 0$ - class B) and end at point $1, 1$ ($v = 1$ - class A).

This area characterize working amplifiers in class AB when they pass from class A to class B and conversely, where the parameters are the signal level (variable $m = 0 \div 1$) and the bias voltage determining the class of working (parameter $v = 0 \div 1$).

The two maximums for class AB correspond to the two areas of operation: area of class A (to the left of the ordinate) and area of class B (above the abscissa).

The first maximum with parameter $v = 0 \div 1$, describes a segment which coincides with the ordinate (area of operation in class A).

The second maximum, with parameter $v = 0 \div 1$, describes an arc (area of working in class B). The beginning of this curve you can see the maximum of the average dissipation output in class B, with co-ordinates $2/\pi, 4/\pi^2$. The end of the arc shows the extremum in class A$(0, 2)$.

The average dissipation output is minimum in this area at maximum output voltage (parameter $m = 1$,v).

References

[1] Sirakov Ek.S., "Instantaneous and Average Output Dissipation by an Output Stage Operating in Class AB" 2002, XXXVII International Scientific Conference on Information, Communication and Energy System and Technologies ICEST 2002**,** 2÷4 October 2002**,** Volume 1, pp. 93 - 96.

[2] Benjamin Er., "Audio Power Amplifiers for Loudspeaker Loads" /Dolby Lab, Ins., San Francisco, CA 94103, USA, *J.Audio Eng. Soc.*, Vol. 42, No. 9, pp. 670 - 683, Sept., 1994.

[3] Sirakov E. S., "Universal audio power amplifier model in Spice form", *International Scientific Conference "Communication, Electronic and Computer System 2000"*, pp. 167-172, Sofia 18.05.2000.

[4] Sirakov Ek.S., "Models of push-pull transistor power amplifier", 11^{th} *International Scientific Conference, "RADIOELECTRONICA 2001"*, pp. 330-333, Brno, Czech Republic, May 10-11, 2001

Fast CELP Code Book Search Method

Sn.Pleshkova-Bekjarska[1], M.Momchedjikov[2]

Abstract – **In this paper it is present a method for fast search in stochastic code book for CELP speech coding. It is found that it is possible to reduce the mathematical complexity in the minimization of error and to proposed an efficient and much fastest method, with some predefined error value.**

Keywords – **Speech coding, CELP, Code book, Search methods.**

I. Introduction

The analysis of the basic mathematical functions in CELP speech coding showns, that the difficulties are in the realization of algorithms of code book search [1]. The most of the efforts are directed for creating of effectives methods in the limits of federal standards FS1016 for improving the speech of code book search [2, 3].

In this article it is proposed a method for fast code book search in CELP speech coding, which is based on mathematical complexity estimation, and on an assumption of minimum quality degradation of decoding speech.

II. Basic Algorithm

The algorithm of excitation signal d_{ij} calculation from code book in CELP speech coding used the minimization of error e_{ij}defined as the difference between real speech signal S^r and synthesized signal S^s_{ij} (Fig. 1).

The minimization of error e_{ij} is shown in fig.1 as a block **min**, in which it is estimation the value of error e_{ij} for minimum and the index i and j are changed iteratively for choosing the next excitation signals b_i and c_i collecting in stochastic SCB and adaptive ACB code book, respectively. The current value of error e_{ij} is calculated from the expression:

$$e_{ij} = W.e'_{ij} \ , \tag{1}$$

where e'_{ij} is the error from the evaluation:

$$e'_{ij} = S^r - S^s_{ij} \, , \tag{2}$$

in which there are not involved the perceptual characteristics of human - ear;

W – matrix, describing the perceptual human – ear filter characteristics :

$$W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ w_2 & w_1 & \dots & 0 \\ . & . & \dots & 0 \\ w_{SF} & w_{SF-1} & \dots & 0 \end{bmatrix} \ , \tag{3}$$

where SF is the number of samples in a subframe (SF=60).

From (1) and (2) it can see, that in the execution of iterations the synthesized signals S^s_{ij} is changed and is defined from the expression:

$$S^s_{ij} = F.d_{ij} \ , \tag{4}$$

where d_{ij} is the excitation signals b^g_i and c^g_j from stochastic and adaptive code book, respectively:

$$d_{ij} = b^g_i + c^g_j \ ; \tag{5}$$

F – matrix, describe the linear prediction filter coefficients a, defined current frame of real speech signals S^r :

$$F = \begin{bmatrix} f_1 & 0 & \dots & 0 \\ f_2 & f_1 & \dots & 0 \\ . & . & \dots & 0 \\ f_{SF} & f_{SF-1} & \dots & f_1 \end{bmatrix} \ . \tag{6}$$

Each from the excitation signals b^g_i and c^g_j are extracted from the collection of these signals in stochastic and adaptive code book, respectively (b_i and c_j), and are multiplied with the current gains for current iteration g_i and g_j:

$$b^g_i = g_i.b_i \tag{7}$$

$$c^g_j = g_j.c_j \ . \tag{8}$$

From (1) and with respect of (2) and (7) it is seen, that in the general case the number of iterations for minimization of error e_{ij} dependent from the range of indexes i and j:

$$i = 1, \ 2, \ 3, \ \dots \ , \ n_{SCB} \tag{9}$$

$$j = 1, \ 2, \ 3, \ \dots \ , \ n_{ACB} \ , \tag{10}$$

where n_{SCB} and n_{ACB} are respectively the number of excitation signals in stochastic and adaptive code books (n_{SCB} =512 and n_{ACB} =256).

If a full search algorithm is used for error e_{ij} minimization in both stochastic and adaptive code books in each moment, then the number of iterations is extremely grown, because it is involved all possible combinations between each two excitation signals b_i and c_j and g_i, g_j from stochastic and adaptive code books. The main part of the time in CELP coding algorithms is spend for minimization of error e_{ij}.

III. Fast Code Book Search Method

An essential decrease of time for minimization of error can be realized, it if is assumed a sequential implementation of search algorithm. For example if the search is made first for adaptive code book, and then for stochastic, the expression (4) and (5) gives:

$$S^s_{ij} = F.\left(b^g_i + c^g_j\right) + S_z \ , \tag{11}$$

[1]Sn.Pleshkova is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail:snegpl@vmei.acad.bg.

[2]M.Momchedjikov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail:snegpl@vmei.acad.bg.

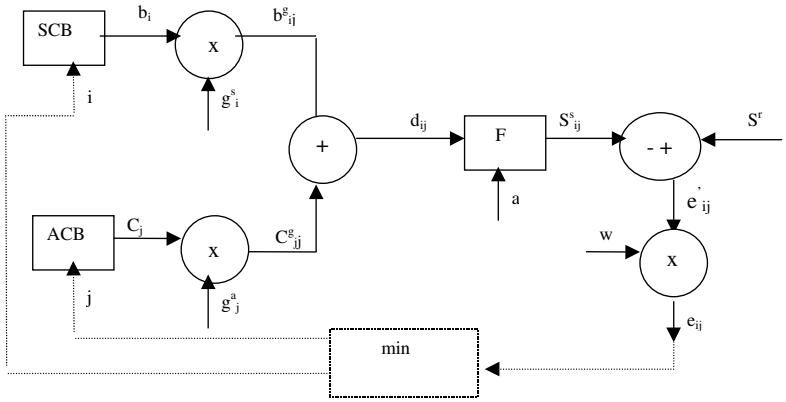


Fig. 1.

where S_z is zero input response of the linear prediction filter F.

The expression (1) can be changed as:

$$e_{ij} = W.S^r - W.F.b_i^g - W.F.c_j^g - W.S_z \ . \qquad (12)$$

If in (12) the minimization is only for stochastic code book, then c_j^g is calculated and it is assumed constant. Then (12) can be transformed as:

$$e_{ij} = S^t - L.b_i^g \ , \qquad (13)$$

where: $S^t = W.S^r - W.F.c_j^g - W.S_z$ is part of expression (12), which rest unchanged for stochastic code book minimization (j=const; i=1, 2, …, n_{SCB}; $L = W.F$ – matrix from multiplication of W and F).

Matrix L is defined from (3) and (6), describing W and F, which are lower triangular. Matrix L is too a lower triangle matrix:

$$L = \begin{vmatrix} l_{11} & l_{12} & l_{13} & \dots & l_{1,SF} \\ l_{21} & l_{22} & l_{23} & \dots & l_{2,SF} \\ . & & & & \\ . & & & & \\ l_{SF,1} & l_{SF,2} & l_{SF,3} & \dots & l_{SF,SF} \end{vmatrix} , \qquad (14)$$

where

$$l_{ij} = 0 \ \text{ for } j > i \text{ and } i = 1 \div SF \ . \qquad (15)$$

From (14) and (15) it is possible to make assumption, that the number of calculations:

$$N_L = SF.SF \qquad (16)$$

in (13) decrease two time:

$$N'_L = \frac{N_L}{2} \ . \qquad (17)$$

A supplement analysis of lower triangular part of matrix L shows, that most of the members l_{ij} of lowest triangle are nearly zeros and cal be replaced by zeros:

$$l_{ij} \approx 0 \ \text{ for } j < \theta \text{ and for } i > \theta \ . \qquad (18)$$

This analysis and the expression (18) shows, that the matrix L can be transformed as a band matrix:

$$L = \begin{vmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ l_{\theta,1} & l_{\theta,2} & l_{\theta,3} & \dots & 0 \\ 0 & l_{\theta+1,2} & \dots & & \\ . & \dots & & & \\ 0 & \dots & l_{SF,\theta} & l_{SF,\theta+1} \dots & l_{SF,SF} \end{vmatrix} . \qquad (19)$$

From expressions (19) it is possible to calculate the value of a supplementary decreasing of calculations in (13):

$$N''_L = N'_L - \frac{(SF-\theta)^2}{2} \ . \qquad (20)$$

This results are made only from analysis of matrix L in expression (13). It is possible to make an analysis of expressions of square error follow from expressions (13) and (17):

$$\vec{e_i} \,.\, \vec{e_i'} = \vec{S^t} \,.\, \vec{S^{t'}} - 2 \,.\, \vec{S^t} \,.\, g_i \,.\, \vec{B_i'} + g_i^2 \,.\, \vec{B_i'} \,.\, \vec{B_i} \ . \qquad (21)$$

In first member in expression (21) didn't depend from index i of code book and is constant in the time of minimization. Then it is possible to transform the algorithm for finding the minimum of expression (21) as a finding of maximum of the rest part of (21):

$$mx = 2 \,.\, \vec{S^t} \,.g_i \,.\, \vec{B_i'} - g_i^2 \,.\, \vec{B_i'} \,.\, \vec{B_i} \ . \qquad (22)$$

In the same time it is possible to make the partial minimization of (21) with respect to g_i:

$$\frac{\partial \vec{e_i}}{\partial g_i} = -2 \,.\, \vec{S^t} \,.\, \vec{B_i'} + 2 \,.\, g_i \,.\, \vec{B_i'} \,.\, \vec{B_i} = 0 \ , \qquad (23)$$

and to define g_i as:

$$g_i = \frac{\vec{S^t} \,.\, \vec{B_i'}}{\vec{B_i'} \,.\, \vec{B_i}} \ . \qquad (24)$$

With substitution of (24) in (22) it is possible to find:

$$mx = \frac{\left(\vec{S^t} \,.\, \vec{B_i'}\right)^2}{\vec{B_i'} \,.\, \vec{B_i}} \ . \qquad (25)$$

In the expression (25) $\vec{S^t} . \vec{B_i}$ is correlation between real signal $\vec{S^t}$ and excitation signal $\vec{B_i}'$, in relation of energy $\vec{B_i}' . \vec{B_i}$ of excitation signal $\vec{B_i}$ from stochastic code book.

The decrease of calculation in (21) with respect of (25) for minimization of error e_i follows from substitution of minimization of e_i with searching maximum of a part of expression (21). This substitution give some advantages for practical implementation of minimization algorithm, but still in (25) the calculations are relatively complex. It is possible to analyzes expression (25) and to shows that the essential calculations are in numerator and enumerator make normalization of the expression.

It is possible to taking only the numerator from the equation (25) and to use only the absolute correlation:

$$\vec{S^t} . \vec{B_i}' = \vec{S^t} . L' . \vec{b_i}' = SL . \vec{b_i}' , \qquad (26)$$

where $SL = \vec{S^t} . L'$ did'nt depend from index i and can be recalculated for all iterations of minimization.

IV. Conclusion

The proof of this possibility for simplification of algorithm of minimization can be made first with a statistical analysis and estimation of numerator of expression (25) and with analysis of decoded speech signal error with using expression (26) as a simplification of (25). The results from this analysis are the object of a next work.

References

[1] Schoreder, M.R., and Atal, B.S., "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates," ICASSP, pp.937-940, 1985.

[2] FS1016 National Standard, USA.

[3] Campell, J.R., Jr., Tremain, T.E., and Welch, V.C., "The Federal Standard (FED-STD) 1016 4800 bps CELP Voice Coder," Digital Signal Processing I, pp.145-155, 1991

A Method for Fast Index Lookup in Databases through Linear Approximation Window

Krassimir Tzvetanov[1], Michael Momchedjikov[2]

Abstract – **in this paper we present a fast and simple algorithm for fast timestamp-to-sequence-number matching based on the linear window approximation.**

Keywords – **database, search, lookup, approximation**

I. Introduction

Have you ever had a database that gets some 15×10^6 records per week? Did you ever try to make a simple search in it? The following method presents a very efficient way of improving performance by locating the indexed sequential number of the record(s) your looking for.

In large databases it is necessary to index data so that you queries can run faster. Unfortunately many times this is not possible if the data you are trying to index is very diverse and not correlating. In the traffic analysis system presented in [1], there were collected about 1.5-2 million records per day. Every record keeps an indexed sequential number, a time/date field and other not essential to the algorithm data. The time/data field expresses the count of seconds that have passed since the epoch in seconds since Jan 1 1970 seconds [2]. Keeping in mind that a day has 86400 and the database collects 1.5 million records per second it is not practical to index by that field. The first big problem we encounter is the size of the index itself. It's easy to calculate the for only one week we will have 604 800 groups of indexes and each one of them will hold average of 18 different records. 18 out of millions is negligible number it's not worth indexing. Over a half a million groups in an index is something not easy to search through (consider multiplying that by the number of weeks you plan to keep accounting information for). The time for searching through the index file is considerable. There is even no way to fit the index into the memory so it has to, always, be searched from the disk.

The experiment showed that on MySQL[3] database the index created on the timestamp field takes 16-18% of the database itself.

II. Presenting the Method

In this paper we show a method that allows significantly faster lookup of the value that must be indexed. It helps you find the sequence number of the record in the database based on the timestamp of the record. It's obvious that the sequence number multiplied by the size of the record gives you the exact location of the record.

$$O = I_r * R_s \ , \tag{1}$$

where O is the offset in the file, I_r is the sequence of the record and R_s is the size of the record.

If the traffic accounting records were inserted into the database with a constant speed, it would also be possible to use that formula. Unfortunately this is not possible since traffic has very different characteristics. On Figure 1 the number of records received on an average work and weekend days of the university network is shown. As it can be seen during work days the distribution is far from linear.

(It is also interesting to note that if you have data from very long time period (over 10 days) the daily oscillations will be invisible in the first few steps, even though they will become a cause of errors in the next steps. In my further research we are considering dynamically changing the distribution law used for approximation.)

In this case linear[3] approximation will yield very incorrect results which will increase search time, etc.

At the same time the distribution is linear for an infinitely short time frame. Therefore the smaller the time frame is – the closer to linear the distribution will be.

You can also consider that if you constantly decrease the time frame (i.e. search window) you'll be getting better results.

This is what exactly the algorithm does. It makes its first linear approximation with a window the size of the whole database. As expected, the result will be far from precise. However, this information is sufficient so that we can narrow the search window for the next step. The new window is calculated on the basis of the approximated value. If the new window doesn't contain the value we are looking for, then we rerun this step using a window with borders: the old value of the border and the new value (the one that is not correct). This window approximately has the same size of the one, which we would have used if there had been no error.

The approximation is done in the following manner: first, we find the average records per hour (2):

$$R_r = (R_t - W_{ti})/(W_{tx} - W_{ti}) \ , \tag{2}$$

where R_r is the record rate (record per second), W_{ti} and W_{tx}

[1]Krassimir Tzvetanov holds Bs from the Faculty of Communictions and Communications Technologies, Technical University, Sofia; Email: krassi@tu-sofia.bg

[2]Michael Momchedjikov is with the Faculty of Communictions and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria; Email: mom@tu-sofia.bg

[3]If linear approximation is replaced by approximation following the distribution law of the variable we use. The results should be much better. This is a focus of future work.

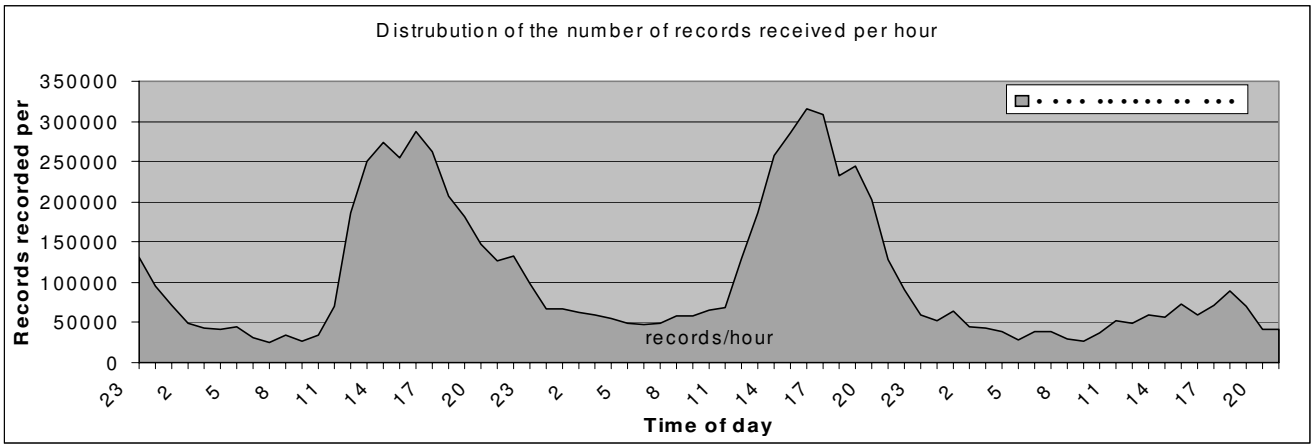


Fig. 1. Distribution of the number of records received per hour

are the lower and upper time borders of the window. R_t is the time/date of the record we are looking for.

The next step is to approximate the sequence number (I_a)(3) based on the this rate:

$$I_a = W_{ii} + (W_{ix} - W_{ii}) * R_r \ , \qquad (3)$$

where W_{ii} and W_{ix} are the sequence numbers of the lower and upper borders.

As a next step, we calculate the new window borders (4)(5). Note that it is not symmetric around the approx. value.

$$B_{ln+1} = \max(I_a - (I_a - W_{ii}) * 0.15, B_{ln}) \qquad (4)$$

$$B_{un+1} = \min(I_a + (W_{ix} - I_a) * 0.20, B_{un}) \qquad (5)$$

Where B_l and B_u denote the lower and upper border. Where n refers to the current step and $n+1$ to the next step of the algorithm.

We repeat this procedure recursively, making smaller and smaller windows – therefore achieving better approximation. There is a minimum distance below which is better to continue searching consequently without approximation. This value is based on the speed of the CPU, disk array, and network activity. It is difficult to calculate this value precisely. However the experiment shows that this value varies between 200 and 500 records.

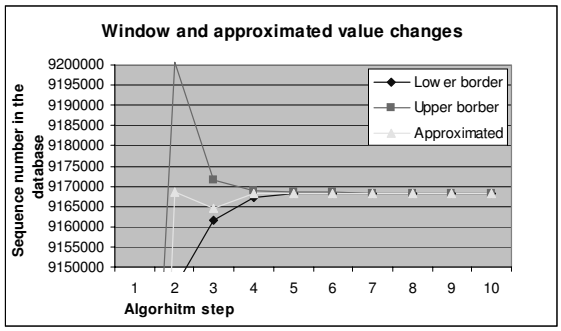


Fig. 2. Window and approximated value changes

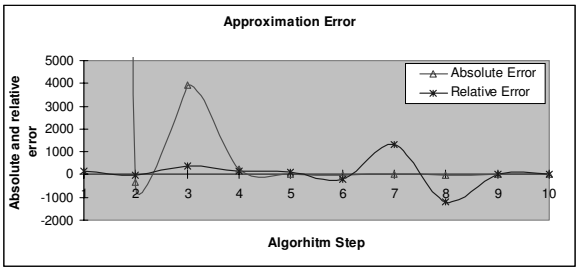


Fig. 3. Approximation Error

III. Results from an Experiment Conducted in TU-Sofia

Below are included graphs based on data taken from the university network during 5 consecutive days and a test run of the program. (Both workdays and weekends are included to make distribution more complex; the data is about 12 million records).

On Figure 2 you can see how both the approximated value and the window change. Both reach close proximity really fast.

On Figure 3 you can see the absolute and relative error during each step of the algorithm. The line, marked with $\triangle$, represents the absolute error in prediction. It goes down exponentially. (The value after the first step, not shown on the graph, is 236 128 records absolute error).

Note that the minimum distance value of the algorithm is achieved at step 6 (or 7). This means that the algorithm is more time consuming than regular sequential search from this moment on.

The line, marked with $*$, represents the relative error (absolute error/how wide the search window is). As you can see on the first step the relative error is small because we have data from several days which hides the daily variations.

When we continue after the 6th step, the algorithm starts to work inefficiently. This is because the rules for calculating the window are not symmetric. This helps while working with large data arrays but makes the algorithm unstable

when working with small amounts of data. This is not really an issue since there is one more reason that justifies why we should stop before that point. Even more, those quotients can be changed.

On the test system the proposed algorithm finds the record about 35-40 times faster than the one without an index (the test database includes only some 10 million records). If you have more records in the database – you'll get better results.

Because the method of searching through the database without an index is sequential, the time increases for the newer records linearly because they occupy the last positions in the database. The search time of proposed method does not depend on the location of the record in the database. It depends solely on the number of records in the database but not in a linear way. The search time grows very slowly related to the increase (of records) in number.

IV. Further Research

As already noted in our future work we plan to find ways to improve the approximation method. It is interesting to replace the linear approximation with approximation with normal distribution. It is also interesting research ways to dynamically change the approximation method.

Acknowledgements

We would like to express our special thanks to the Ministry of Education and Science for their continuing support.

References

[1] Krassimir Tzvetanov, 2003, System for Traffic Analysis and Accounting in IP Networks

[2] http://cr.yp.to/proto/utctai.html

[3] MySQL documentation: http://www.mysql.com/doc/en/

Multi-Layer Watermarking Aimed at Closed Information Systems

Roumen Kountchev[1], Vladimir Todorov[2] and Roumiana Kountcheva[3]

Abstract – **A new method for image watermarking, aimed at documents archiving is presented in this paper. This specific application is very useful in all cases, when large amounts of original paper documents have to be stored for many years, in accordance with financial lows. The substitution of these originals with their electronic copies ensures easy access and security. The new watermarking method is based on the Inverse Difference Pyramid (IDP) decomposition, performed in the image spectrum.**

Keywords – **Information assurance, Watermarking, Image compression, Web-based information systems.**

I. Introduction

In correspondence with the financial lows' requirements, all companies are obliged in their everyday practice to store large amounts of original paper documents (invoices, etc.). As it is known, the long-term storage of paper documents creates many troubles for the owner, most of which are connected to the fact, that they require too much physical space. Most up-to-date computer technologies permit the creation of electronic copies of the documents and their archiving and saving on diskettes, optical disks, etc., which in result makes their storage much easier. To do this, the documents must be scanned and thus obtained digital images - compressed and saved. The most frequently used compression methods are based on the standards JPEG or JPEG2000. The traditional approach in the big enterprises is to store the compressed images on CD, using WARM (Write-Once-Read-Many) technologies. In some cases electronic signature is used as well, but this is not enough for the secure document storage: as it is known, there are many software products, which permit easy image editing. This requires additional security levels in the archived image files, in order to ensure their original contents untouched and to prove any kind of un-authorized access. Part of these problems could be solved using image watermarking [1-3]. In some applications, the watermark extraction is performed without comparison with the original image, and in other cases, the original must be available. The second case is suitable for documents archiving, because the original image will be easily available, when requested, and the comparison will ensure the watermark extraction [4,5]. A new method for multi-layer image watermarking, ensuring the document content authenticity, is proposed in this paper. The watermarking is performed using a new method for image decomposition, called Inverse Difference Pyramid (IDP) [6]. In this case, the decomposition is performed in the image spectrum area, and the image content is processed in consecutive layers with increasing resolution. This approach permits the insertion of different watermark in every layer. Any change in the watermark image, noticed after the extraction, is a proof for un-authorized image content editing. To perform the watermarking the original paper documents have to be scanned and saved as bmp or JPEG files, and after that - watermarked.

[1]Roumen Kountchev is with the Faculty of Communications and Communications Technologies, Technical University of Sofia, Kliment Ohridsky 8, 1000 Sofia, Bulgaria, E-mail1: rkountch@tu-sofia.bg

[2]Vladimir Todorov is with T&K Engineering Co., Sofia 1712, P.O.Box.12, Bulgaria. E-mail:vtodorov@yahoo.com

[3]Riumiana Kountcheva is with T&K Engineering Co., Sofia 1712, P.O.Box.12, Bulgaria. E-mail: rkountcheva@yahoo.com

II. Basic Principles of the IDP Decomposition

In the general case, every digital image $[B(2^n)]$, consisting of $2^n \times 2^n$ pixels could be represented with inverse difference pyramid (IDP) [6] in accordance with the expression:

$$[B(2^n)] = [\tilde{B}_0(2^n)] + \sum_{p=1}^{n-1} [\tilde{E}_{p-1}(2^n)] + [E_{n-1}(2^n)] \ , \quad (1)$$

where $p = 0, 1, 2, \ldots, n$ is the number of the pyramidal decomposition level.

The first component $[\tilde{B}_0(2^n)]$ in Eq. (1) corresponds with the level $p = 0$ and defines the coarse image approximation as a result, obtained after applying the inverse orthogonal transform of the truncated image spectrum $[\tilde{S}_0(2^n)]$ with the matrix $[T_0(2^n)]$ with size $2^n \times 2^n$, i.e.:

$$[\tilde{B}_0(2^n)] = [T_0(2^n)]^{-1}[\tilde{S}_0(2^n)][T_0(2^n)]^{-1} \ . \quad (2)$$

The coefficients of the spectrum matrix $[\tilde{S}_0(2^n)]$ are defined with the expression:

$$\tilde{s}_0(u,v) = m_0(u,v)s_0(u,v), \text{ for } u,v = 0,1,..,2^n-1. \quad (3)$$

Here the elements $m_0(u,v)$ of the matrix-mask $[M_0(2^n)]$ define the position of the retained coefficients from the two-dimensional image spectrum $[S_0(2^n)]$:

$$m_0(u,v) = \begin{cases} 1 \text{ if } (u,v) \in V_0; \\ 0 - \text{in other cases,} \end{cases} \quad (4)$$

where V_0 is the area of the retained spectrum coefficients $s_0(u,v)$. The corresponding spectrum matrix $[S_0(2^n)]$ is obtained in result of applying the direct orthogonal transform for $[B(2^n)]$ with the matrix $[T_0(2^n)]$, i.e.

$$[S_0(2^n)] = [T_0(2^n)][B(2^n)][T_0(2^n)]. \quad (5)$$

In the decomposition levels $p = 1, 2, ..., n-1$ from Eq. (1) the corresponding component is defined as:

$$[\tilde{E}_{p-1}(2^n)] = \begin{bmatrix} [\tilde{E}^1_{p-1}(2^{n-p})] & [\tilde{E}^2_{p-1}(2^{n-p})] & \cdots & [\tilde{E}^{2^p}_{p-1}(2^{n-p})] \\ [\tilde{E}^{2^p+1}_{p-1}(2^{n-p})] & [\tilde{E}^{2^p+2}_{p-1}(2^{n-p})] & \cdots & [\tilde{E}^{2^{p+1}}_{p-1}(2^{n-p})] \\ \cdots & \cdots & \cdots & \cdots \\ [\tilde{E}^{4^p!-2^p+1}_{p-1}(2^{n-p})] & [\tilde{E}^{4^p!-2^p+2}_{p-1}(2^{n-p})] & \cdots & [\tilde{E}^{4^p}_{p-1}(2^{n-p})] \end{bmatrix} \quad (6)$$

The sub-matrices $[\tilde{E}^{k_p}_{p-1}(2^{n-p})]$ for $k_p = 1, 2, .., 4^p$ of the matrix $[\tilde{E}_{p-1}(2^n)]$ are obtained as a result of its quad tree division in 4^p equal parts. Each sub-matrix in Eq. (6) has size $2^{n-p} \times 2^{n-p}$ and is defined in similar way as shown in Eq. (2).

$$[\tilde{E}^{k_p}_{p-1}(2^{n-p})] = [T_p(2^{n-p})]^{-1}[\tilde{S}^{k_p}_p(2^{n-p})][T_p(2^{n-p})]^{-1}, \quad (7)$$

where

$$\tilde{s}^{k_p}_p(u,v) = m_p(u,v)s^{k_p}_p(u,v) \quad (8)$$

for $u, v = 0, 1, .., 2^{n-p} - 1$.

The elements of the matrix-mask $[M_p(2^{n-p})]$ of the retained coefficients are defined with:

$$m_p(u,v) = \begin{cases} 1 & \text{if } (u,v) \in V_p; \\ 0 - & \text{in other cases.} \end{cases} \quad (9)$$

Here V_p is the area of the retained spectrum coefficients $s^{k_p}_p(u,v)$, whose matrix $[S^{k_p}_p(2^{n-p})]$ is obtained after applying a direct orthogonal transform on the difference matrix $[E_{p-1}(2^{n-p})]$, using the transform matrix $[T_p(2^{n-p})]$:

$$[S^{k_p}_p(2^{n-p})] = [T_p(2^{n-p})][E^{k_p}_{p-1}(2^{n-p})][T_p(2^{n-p})] \quad (10)$$

In Eq. (10) the term $[E_{p-1}(2^{n-p})]$ is defined as:

$$[E_{p-1}(2^{n-p})] = \\ = \begin{cases} [B(2^n)] - [\tilde{B}_0(2^n)] \text{ for } p = 1; \\ [E_{p-2}(2^{n-p})] - [\tilde{E}_{p-2}(2^{n-p})] \text{ for } p = 2,..,n\text{-}1. \end{cases} \quad (11)$$

The remaining component, corresponding to the last level $p = n$ in Eq. (1), is presented with the difference:

$$[E_{n-1}(2^n)] = [E_{n-2}(2^n)] - [\tilde{E}_{n-2}(2^n)] \, . \quad (12)$$

In the case, when the decomposition from Eq. (1) is truncated up to the component r, the so-called "truncated" IDP is obtained, presented by the expression:

$$[\hat{B}(2^n)] = [\tilde{B}_0(2^n)] + \sum_{p=1}^{r} [\tilde{E}_{p-1}(2^n)] \quad (13)$$

In the image frequency domain the IDP pyramid consists of the spectrum coefficients $s^{k_p}_p(u,v)$, for $k_p = 1, 2, .., 4^p$. For the level p their number is as follows:

$$M(p) = 4^p M_p = 4^p \sum_{u=0}^{2^{n-p}-1} \sum_{v=0}^{2^{n-p}-1} m_p(u,v) \quad (14)$$

for $p = 0, 1, 2, .., n-1$.

Then the total number of coefficients for IDP pyramid with r levels is:

$$M_\Sigma(r) = \sum_{s=0}^{r} 4^s M_s = \\ = \sum_{s=0}^{r} \sum_{u=0}^{2^{n-s}} \sum_{v=0}^{2^{n-s}} 4^s m_s(u,v) \;\; \text{for} \;\; r < n \, . \quad (15)$$

In particular, for $M_0 = M_1 = ... = M_r$=4 one can get from Eq. (15):

$$M_\Sigma(r) = \sum_{s=0}^{r} 4^{s+1} = \frac{4^2}{3}(4^r - \frac{1}{4}) \approx \frac{1}{3}4^{r+2} \, . \quad (16)$$

In case, when the matrix $[B(2^n)]$ represents one image block with size $N \times N$ pixels, the total number of coefficients for all blocks is correspondingly:

$$M_\Sigma = \frac{N^2}{4^n} M_\Sigma(r) = N^2 \sum_{s=0}^{r} \sum_{u=0}^{2^{n-s}} \sum_{v=0}^{2^{n-s}} 4^{s-n} m_s(u,v) \, . \quad (17)$$

Then for $M_0 = M_1 = ... = M_r$=4, Eq. (17) is producing $M_\Sigma = (N^2/3)4^{r-n+2}$.

After a lossless compression for all coefficients' values $s^{k_p}_p(u,v)$, from pyramid level p a corresponding binary massif $\{X_p(r) : r = 1, 2, .., l_p\}$ with length l_p, is obtained. At the beginning of this massif $\{X_p(r)\}$ is inserted a special header, H_p. It contains information about the pyramid level p, the elements of the matrix-mask $[M_p]$, the kind of the orthogonal transform, the arrangement of coefficients in the massif, etc.

III. Image Watermarking Based on IDP

The decomposition, represented by Eq. (1), permits the insertion of different watermark in every level of the IDP pyramid. For this, elements $Z_p(r)$ for the level p of the data massif, are represented as follows:

$$Z_p(r) = \begin{cases} X_p(r) \text{ for } r = 1, 2, .., l_p; \\ W_p(r) \text{ for } r = l_p + 1, l_p + 2, .., l_p + l_{wp}. \end{cases} \quad (18)$$

Here $W_p(r)$ is the compressed data with length l_{wp} representing the watermark in the level p, obtained after applying the password, Y_p. The password itself is a code, with length $l^y_p \leqslant l_p$. The number N_{wp}, corresponding to the watermark for the level p, is defined with the equation:

$$N_{wp} = l_p \oplus Y_p = \sum_{i=0}^{l^y_p - 1} (l^p_i \oplus y^p_i) \, , \quad (19)$$

where l^p_i and y^p_i are the corresponding i-th digits of l_p and Y_p, and with "$\oplus$" is noted the operation "exclusive OR". The number N_{wp}, corresponding to the watermark, is inserted in the header H_{wp} of the compressed data, W_p(r).

The watermark image (prepared in advance) is compressed with lossless compression. For this reason it is suitable to use relatively small watermark (256x256 or 512x512 pixels) and in the process of decomposition to apply it on the document image as many times, as necessary, to cover it. The IDP decomposition permits watermark images with size 512x512 pixels to be compressed losslessly in a file with size 500-600 Bytes, depending on image contents. This size is negligible, compared to the size of the digital image of the paper document. Even after compression is used, the corresponding file is very large, because the quality of the restored document image must be good, and the compression ratio could not be high. The following (higher) pyramid level creates a new

data sequence $Z_p(r)$, where another watermark image could be inserted. Then all the information is arranged in one common massif. The data, obtained in result of the image compression, and the inserted watermark, is saved and stored, or sent to the receiver in accordance with the application, using the existing standard communication nets. The watermark insertion in the image data is equivalent to the creation of additional level in the pyramid.

IV. Image Decoding and Watermark Visualization

The decoding is done, performing the already described operations in reverse order:

- The components $X_p(r)$ and $W_p(r)$ are extracted from the massif of the received data $Z_p(r)$ and the watermark image $W_p(i,j)$ for the level p is restored;
- The information $X_p(r)$, obtained in result of the lossless compression, is decoded, and the values of coefficients $s_p^{k_p}(u_r, v_r)$, are calculated;
- The model of the sub-image $[\tilde{B}_0]/[\tilde{E}_{p-1}^{k_p}]$ is calculated, using inverse orthogonal transform, as presented with Eqs. (2) and (6);
- The values of the elements $\hat{B}_w(i,j)$ of the restored image are calculated. This image can contain two (or more) watermarks in the lower IDP pyramid levels, in correspondence with the expression:

$$[\hat{B}_w(i,j)] = [\tilde{B}_0(i,j) + W_0(i,j)] + [\tilde{E}_0(i,j) + W_1(i,j)] + \sum_{p=2}^{r} [\tilde{E}_{p-1}(i,j)] \quad (20)$$

In the already described watermarking method the IDP decomposition starts from the lowest level, where the square, corresponding with one sub-image is 16x16 or 8x8 pixels (resp. n=4 or n=3). In this image layer the document owner can insert his (or her) own watermark. Example image for original image document is shown in Fig. 1. This watermark is visible, and it is the first authentic watermark, inserted in the document image. It corresponds to the "public" document watermark. The result is shown in Fig. 2 - the example text image with inserted visible watermark. The same watermark could be made invisible as well. This is performed, choosing a watermark in which the brightness value is smaller than the sensitivity threshold of the human visual system.

The choice of a visible or invisible watermark is a result of a decision, made by the document owner, and usually is the same in all documents, created by him. The size of the watermark image usually is smaller then that of the document and in the process of the image recovering it is applied as many times, as necessary, to cover all the document image.

The next level of the decomposition (the higher pyramid level) is the place to insert the second watermark – usually, invisible. In the method, offered here, this watermark depends on the image contents, and correspondingly – on the

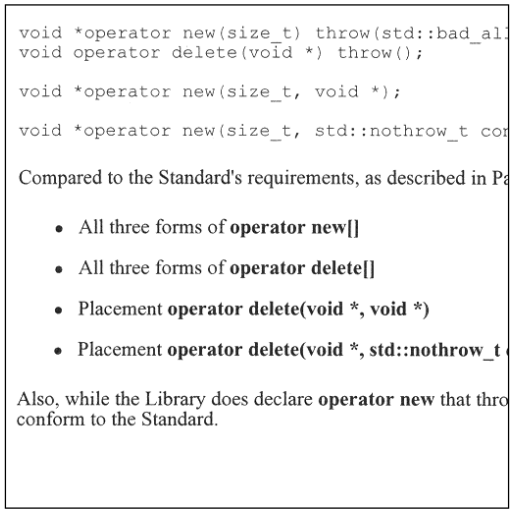


Fig. 1. Original text image

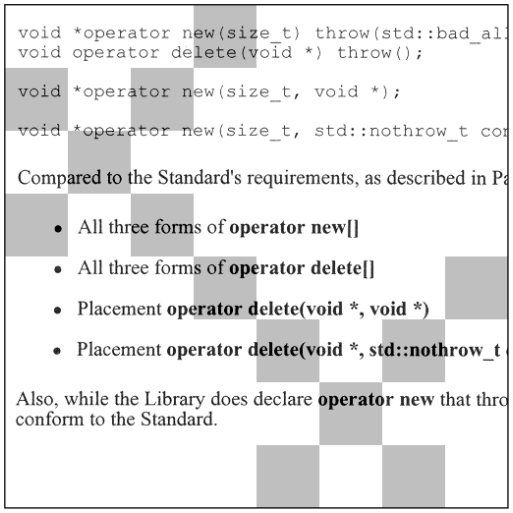


Fig. 2. Watermarked Text image

result, obtained after the compression of the data from the second pyramid level. As watermark images could be used all two-dimensional Walsh-Hadamard functions, or another images, included in the special image library. The choice of the watermark is performed, after the compressed data is obtained, and the number of the compressed bits is known, Eqs. (12) and (13). The fact, that the watermark is invisible, and the requirement to use a password ensures, that the possession of the decompression software will not permit the watermark visualization and the existence of watermark will not be known. The watermark visualization will be possible only if the password is available. The analysis of the image header (in case, that somebody knows the algorithm in detail and is able to analyze it) will show that the image contains inserted invisible watermarks, but the extraction will be impossible without the password.

In case of un-authorized image contents change (using special software for image editing) the visible watermark should be changed as well, but the invisible one will not be

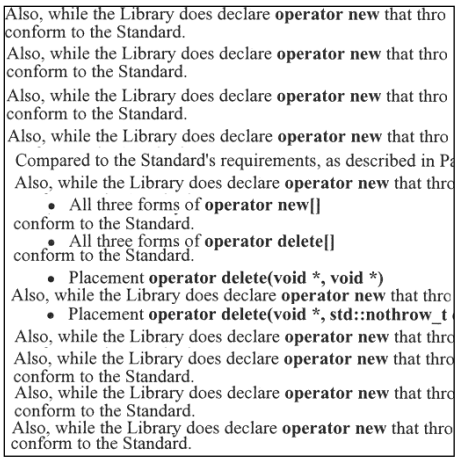


Fig. 3. Original text image

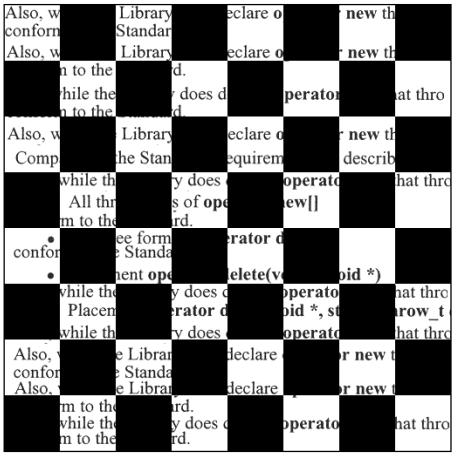


Fig. 4. Watermarked text image with masking watermark

touched. There are two reasons for this. The first is that in result of the use of the algorithm for lossless coding of the equal symbols, and of the adaptive modified Huffman code, the length of the compressed file could not be calculated or defined in advance. The data, obtained in result of the compression, are changed in accordance with the image contents and any change in this contents results in change of the file length l_p. Because of this peculiarity of the algorithm, the first indication that the image had been edited is the change in the compressed data length. If the password is known, the new value of the data file length together with the password, will point at another watermark image from the library after the processing. Even the document owner does not know this, because the process of choosing the invisible watermark image is performed automatically. The same approach is used in the next, higher pyramid level, where another invisible watermark is inserted. The watermark visualization is performed in a way, similar to visualization of the classic paper watermarks: for them we use light with higher brightness, and in our case, algorithm for sharpening the brightness transitions is used. All watermarks are visualized layer by layer, until all of them are processed.

In some cases the watermark could be used to cover (hide) the original image, or a selected part of it. Example for such application is shown in Figs. 3 and 4. In this case, instead of an invisible watermark, is applied a watermark, which destroys the original image. This example illustrates that the text information could not be used until the watermark is successfully removed. The recovery of the original document is possible only if the password is available.

All watermarks must answer the following requirements:

- The watermark image must be a drawing, consisting of relatively large figures (8x8 pixels, or larger), with constant brightness. Such image is resistant against JPEG compression with high compression ratio.
- In order to ensure that the watermark is invisible, its brightness value must be relatively small. In the cases, when it is applied on parts of the original document, where the total brightness is high, the brightness value of the watermark image should be under the sensitivity threshold of the human visual system. In other cases, where the document brightness is lower, the brightness value of the watermark image could be higher. Such approach requires the creation of intelligent, adaptive algorithm, which in most cases would make its application difficult. In result, it is easier to use watermarks with small brightness value, which could be applied over the whole document and remain invisible.

The basic application of the described method is aimed at closed information systems for saving and storage the images of the electronic copies of documents with paper originals. This application area defines some restrictions:

- The watermark extraction requires a password;
- The method requires the document creator to have a library of watermark images, one or more of which to be inserted in the document image in correspondence with the described algorithm;
- The volume of the compressed data, representing the original document, is increased with the watermark insertion, and in result, the compression ratio is reduced.

V. Basic Method Advantages

1. The knowledge of the algorithm and the usage of decompression software do not permit the watermark extraction. This is possible only if the password is available;

2. Any change, noticed in the visualized invisible watermark, proves un-authorized access and image contents editing;

3. The fact, that the inserted watermark becomes a part of the image, makes it resistant against another compression methods, cropping, shifting, resizing, etc.;

4. The ability to insert more than one watermark in a single image increases the image contents security.

VI. Conclusion

The IDP decomposition permits easy watermark insertion in the consecutively processed image layers with increasing res-

olution. In result, the image editing and the watermark extraction are very complicated and together with the requirement to know the password and to have access to the image library of the document creator, the un-authorized access becomes practically impossible. The method could be used in PC image processing, Windows environment.

References

[1] J. Tzeng, W. L. Hwang, I. L. Chern. Enhancing Image Watermarking Methods with/without Reference Images by Optimization on Second-order Statistics. IEEE Transactions on Image Processing, Vol. 11, No. 7, July 2002, pp. 771-782.

[2] J. Brassil, S. Low, N. Maxemchuk, L.O'Gorman. Electronic Marking and Identification Techniques to Discourage Document Copying. IEEE J. Select. Areas Commun., Vol. 13, Oct. 1995, pp. 1495- 1504.

[3] W. Szepansky. A Signal Theoretic Method for Creating Forgery-proof Documents for Automatic Verification. Carnahan Conf. On Crime Countermeasures, Lexington,KY,1979,pp.101-109.

[4] B. J. Falkowsky, Lip-San Lim. Image Watermarking Using Hadamard Transforms. Electronics Letters, 3^{rd} February 2000, Vol. 36, No. 3, pp. 211-213.

[5] M. Hartung, M. Kutter. Multimedia Watermarking Techniques. Proceedings of the IEEE, Vol. 87, No. 7, July 1999, pp. 1079-1086.

[6] R. Kountchev, V. Haese-Coat, J. Ronsin. Inverse Pyramidal Decomposition with multiple DCT. Signal Processing : Image Communication, Vol. 17, January 2002, pp. 201-218.

Investigation of Handwriting General Features in Case of Neurological Diseases

Elena Kalcheva[1], Georgi Gluhchev[2]

Abstract – **A method for objective assessment of some neurological diseases (ND) is developed. It is based on the automatic processing and analysis of images of a handwritten script belonging to patients with ND and healthy individuals. In this method connected components are extracted, filtered and analyzed both for the whole script and for each of the automatically separated text-lines. Eight common geometric features describing size, shape and orientation of the handwriting are computed and statistically investigated. The vertical stroke size, aspect ratio and mean slope's angle revealed differences between ND-patients and those from the control group. Also the average stroke width and average stroke number relative to the row length can be used to separate the clinical Parkinsonians from the others. Nevertheless additional features and data are required to obtain reliable evaluation of the handwriting changes due to neurological diseases.**

Keywords – **Neurological diseases, Document image processing, Feature selection, Discriminant ability**

I. Introduction

The disturbances and dysfunction in the human motor control system often represent certain indications of neurological diseases (ND). It is well known that the early diagnosis of many diseases assures a good therapeutic effect in most cases.

In [3] the authors propose several instruments for identification of patients at risk and at early stages of the Parkinson's disease (PD). They suggest some modern invasive techniques, like PET and SPECT studies of the dopaminergic system, analysis of transcranial ultrasonic images, MBIG scintigraphy etc., for recognizing the early stages of the PD. Also, some tests for evaluation of motor abnormalities concerning the velocity, reaction time and precision of the movements are discussed.

The exact study of the effects of the pharmacotherapy on the motor behavior (or so called 'pharmacodynamics') is also very important to determine the appropriate treatment (the amount, timing and spacing of doses) for each patient according to the disease progression and his daily-life requirements [6]. Therefore a simple non-invasive tool for evaluation of the pharmacodynamics for the individual patient is necessary.

In this regard, knowing that motor abnormalities influence the ability to control the movements of the hands, a fine analysis of the handwriting may be useful for objective assessment of the ND. There are some reports on applying handwriting-based techniques to appraise movement disorders. Some of the approaches use only standard neurophysiology tests [5,10]. In [4] the authors combine the Test of Motor Impairments with a SPECT-analysis to examine the possible correlation between long-time treatment and the kinematic parameters of handwriting defects in patients with Wilson's disease.

Moreover, the recent data suggest that handwriting may reveal significant changes many years before the clinical onset of the ND.

In this paper we present a method for objective assessment of some neurological diseases. It is based on the automatic processing and analysis of images of handwritten script belonging to patients with ND and healthy individuals.

II. Automatic Script Processing

In our experiments we use Handwriting Test Form to collect scripts. The scripts consist of a printed sentence in Bulgarian language and the writers are asked to write it in the blank zone below without any requirements about the interpretation. The documents are scanned at 300 dpi and stored as gray-level images.

A. Preprocessing

Most of the algorithms of document analysis and recognition use pre-processing stages to enhance the images. That facilitates the further extraction of relevant information and also increases the accuracy of the automatic measurements. The presence of possible random noise in the scanned images may impede the automatic processing. Therefore in our examinations a 3×3 low-pass filter is applied to remove the noise, consisting of isolated small groups of dark pixels.

At the next step we convert the gray-level images to binary ones using an adaptive thresholding technique [16]. The method is noise resistant and does not depend on the object's area and shape.

B. Connected Components Labeling

Detection of Connected Components (CCs), or blobs, between pixels in binary images is a fundamental step in segmentation of objects and regions within the image. A unique value is assigned to each of them, which allows to separate CC from other blobs.

A CC is denoted as a set of black pixels where each pixel has at least one black 8-neighbor. The original algorithm for

[1]Elena Kalcheva is with the IIT-BAS, Acad. G. Bonchev str., Bl29A, 1113 Sofia, Bulgaria, E-mail: elena@iinf.bas.bg

[2]Georgi Gluhchev is with the IIT-BAS, Acad. G. Bonchev str., Bl29A, 1113 Sofia, Bulgaria, E-mail: gluhchev@iinf.bas.bg

CCs extraction and labeling was developed by Rosenfeld and Pfaltz in 1966 [14]. It performs two passes through the binary image. In the first pass the image is processed from left to right and top to bottom to generate the labels for each pixel and all the equivalent labels are stored in a pair of arrays. In the second pass each label is replaced by the label assigned to its equivalence class. Several papers describe the modifications of the algorithm where the authors try to solve the problem with large images at the second pass where the equivalence arrays become unacceptably huge [9,13]. Another group of algorithms generates Bounding Boxes (BBs) of the Connected Components using the connectivity between the segments of black pixels. Here the binary image is scanned line by line and the arrays of black segments are composed. Then the connectivity between these black segments is examined for each pair of adjacent scanlines. BBs are extended to include all black pixels connected to the previous scanline [2].

In our examinations a simple procedure for Connected Components labeling is used. The binary image is scanned from left to right by rows. The CCs of each row are detected and labeled consecutively using 8-connectivity. The background is assigned 0 and never changes since it is not associated with any other region. Then the equivalent components of adjacent rows are merged and the image is relabeled to obtain a consecutive set of labels for the CCs. The procedure is fast and does not require much space in memory.

A data structure is appointed to each CC. It holds the number of black pixels, the coordinates of the corresponding Bounding Box, the width and height of the CC, the coordinates of its mass center etc. This structure facilitates the manipulation of the components as objects and their further analysis.

After finding all the CCs in the script unusually small and too large components are filtered out. To remove small objects that bring little information we use an absolute criteria – all the components with height or width less than 3 pixels ($\boldsymbol{h}_i < \boldsymbol{3}$ or $\boldsymbol{w}_i < \boldsymbol{3}$, $\boldsymbol{i=1,N}_{cc}$) are eliminated. Then we filter out the CCs with a height greater than twice the mean components height ($\boldsymbol{h}_i > \boldsymbol{2h}_{i\_\boldsymbol{av}}$) which are considered as 'big' since they are connected across more then one text line. Thus, in our further examinations we deal with the remaining CCs.

C. Text-Lines Segmentation

A text has a linear structure and the physical components corresponding to this linear structure are the text-lines. There are a lot of methods that are successfully applied to the detection of text-lines in printed documents [7,17]. But the layout variability of handwritten materials makes it difficult to obtain a reliable baseline for each of the text-lines. Two basic groups of methods for segmentation of the scripts are presented in the literature – the projection profile techniques [11,12,15,18] and the methods that use CCs analysis [1,8].

The first group of methods is based on the analysis of the regularity of peaks and valleys in the horizontal projection profile of the image. These peaks and valleys correspond respectively to the text-lines and the spaces between them. But the presence of large skew angles (above $\pm 5^\circ$) and overlapping regions (where lines are not parallel) can significantly deteriorate the segmentation results. So the authors often perform prior normalization or use some additional assumptions about the text.

The second group of methods is based on the extraction and clustering of the Connected Components or their corresponding Bounding Boxes. Some authors use a nearest-neighbor technique to merge adjacent components from a same line. Also, Hough transform applied over the set of CCs is used to detect fluctuating baselines and sloped remarks between them [8]. Other authors apply a graph representation where the CCs/BBs are the vertices of the graph, which edges are the lines joining each pair of objects. A Minimum-spanning tree for such graph is computed and the text is segmented using split or merge operations over the branches of the graph according to different rules [1].

Our approach to separate text-line is a simple combination of projection profile technique and CCs analysis. Since the CCs have been already extracted and filtered we compute the horizontal projection profile of their gravity centers. Then the histogram is smoothed and the gaps in the projection profile are taken as rough separators between the lines. The CCs which centers of gravity fall between these borders are attached to the corresponding line. The next step is to obtain the reliable baseline for each of the sets of separated CCs. The straight line which best fits to a given set of points is the regression line determined via the least-squares technique that minimizes the fitting error. Coefficients $\boldsymbol{a}$ and $\boldsymbol{b}$ of the line:

$$y = a + bx \tag{1}$$

are computed as follows:

$$b = \frac{N \sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \tag{2}$$

$$a = \frac{1}{N}\left(\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i\right) \tag{3}$$

where (x_i, y_i), $i = 1, N$ is a point from the set of N points.

The original image of handwritten script and the result image after preprocessing and baseline detection steps are shown on Fig. 1a, b.

III. General Features of Handwritten Text

The preprocessed images of the handwritten scripts are used as input data for extraction of general features at both line and page level.

At the line level we compute several common geometrical parameters that describe size, orientation, shape and baseline deviations of the handwriting. The data structure assigned to each of the CCs at the preprocessing stage facilitates the extraction of some size-related characteristics. We use the height ($\boldsymbol{h}_i$) and the width ($\boldsymbol{w}_i$) of the i^{-th} CC to form the ***aspect ratio***:

$$ar_i = \frac{w_i}{h_i}\,, \tag{4}$$

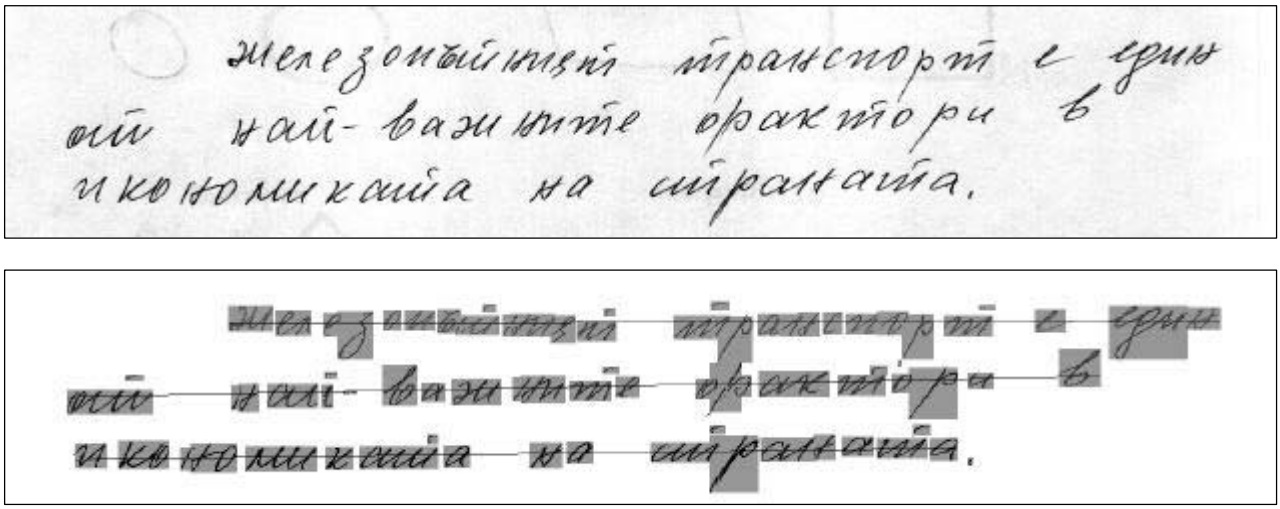


Fig. 1. The original image (a) and the results from the connected components and baselines detection (b).

where ***i*** is the number of the corresponding CC.

The larger ratio the higher probability that a few characters are included in the component. Thus, the ***aspect ratio*** directly gives information about the consistency of the handwriting.

Another feature that fits to the characteristics of different handwritings is the ***stroke to area ratio***. It is defined as:

$$sar_i = \frac{n_i}{w_i h_i} \ , \tag{5}$$

where $\boldsymbol{n}_i$is the number of the black pixel in the i^{-th} CC. The ***stroke to area ratio*** describes the density of the black pixels within the zone of CC.

These four parameters - $\boldsymbol{h}_i$**,** $\boldsymbol{w}_i$**,** $\boldsymbol{ar}_i$ and $\boldsymbol{sar}_i$ are averaged over the number of the CCs ($\boldsymbol{N}_l$) of each text-line (***l***) and the obtained characteristics $\boldsymbol{w\_av}_l$**,** $\boldsymbol{h\_av}_l$**,** $\boldsymbol{ar\_av}_l$ and $\boldsymbol{sar\_av}_l$ are used in our examinations. Also, the ***stroke frequency in a row*** is computed as the number of CCs relative to the length of the corresponding baseline:

$$sf_l = \frac{N_l}{Lenght_l} \tag{6}$$

$\boldsymbol{Lenght}_l$ represents the Euclidean distance between the leftmost and the rightmost black pixel of the CCs in the line.

Additional information about the handwritten scripts gives the ***slope angle*** of each text row. Once the baselines of text are obtained it is easy to calculate each slope angle as:

$$slope_l = \arctan\left(b_l\right) \ , \tag{7}$$

where b_l is the coefficient from the equation of the regression line Eq. 1.

A certain measure for baseline variability can be defined as:

$$f_l = \frac{S_l}{Lenght_l} \ , \tag{8}$$

where:

$$S_l = \frac{\sum\limits_{i=1}^{N_l} |a_l + b_l x_i - y_i|}{\sqrt{b_l^2 + 1}} \tag{9}$$

is the square root of the residual sum for each regression line and represents the sum of the deviations of the gravity centers of the CCs from the corresponding baseline.

At the page level, all of these parameters are normalized by the number ***R*** of the text-rows in each document. (Same notations but with the capital letters are used for the parameters of the whole script). Also, the average distance between lines is computed as:

$$D\_av = \frac{\sum\limits_{l=1}^{R-1} (a_{l+1} - a_l)}{R-1} \tag{10}$$

Here a_l is the vertical intercept of the beginning of the corresponding baseline.

IV. Experiments and Results

To evaluate the influence of the ND on the handwriting general characteristics several experiments have been carried out. We use the script materials collected from 7 Parkinsonians (Group A), 18 patients with other neurological diseases (Group B) and 10 healthy individuals (Control group C).

The discriminant ability of eight common features of handwriting at line and page level is examined in three experiments:

Experiment 1: A comparison between Control group and the whole group of patients with neurological diseases.

Experiment 2: A comparison between Control group and the group of patients with Parkinson's disease.

Experiment 3: A comparison between the group of Parkinsonians and the group of patients with other neurological diseases.

The two-sided t-criterion is used to estimate the significance of every feature in the experiments. We assume a normal distribution with unknown values of the mean and standard deviation for each parameter in the groups.

The results from the experiments include the evaluation of the minimal and maximal values of the corresponding parameter, its mean and standard deviation, and t-statistics value.

The features of the handwritten scripts that have the greatest values of t are summarized in Table 1. At the page level, vertical stroke size (H), stroke area ratio (*SAR*) and mean slope's angle (*Slope*) reveal differences between the ND-patients and the healthy persons. Moreover, the *SAR* and F parameters show a relatively high t-value in the Experiment 3 where the PD-patients are compared to the other ND-patients.

The tremor, due to certain ND influences the writing control and especially at the beginning and the end of the process. That is confirmed by the line-level analysis where the corresponding features (ar_av and f) have comparatively large t-values for both the first and the last text-lines. Also the slope angle in the first row proves to be a significant feature in the comparison between the ND-patients and the healthy individuals.

Table 1. Summarized results from the experiments

Experiments	Line level features			Page level features
	Line 1	Line 2	Line 3	
Experiment 1 (AB/C)	f, slope, h_av, ar_av	h_av	f, sf, ar_av	H, SAR
Experiment 2 (A/C)	f, ar_av, slope, h_av	ar_av	sf, ar_av, f, sar_av	SAR, Slope
Experiment 3 (A/B)	f, ar_av	ar_av, f, sar_av, sf, w_av	sf ,sar_av, slope,w_av ar_av	SAR, F

V. Conclusion

The motor abnormalities due to neurological diseases influence the ability of patients to control the precise hand movements. Thus, a fine analysis of handwriting could be used as a reliable tool for assessment and evaluation of the ND.

A method for automatic processing of handwriting is proposed to extract general features of handwritten scripts belonging to patients with ND and healthy persons. The statistical significance of 8 parameters is examined at both page and line level using t-test. Most of the features show relatively high discriminant ability in the three implemented experiments.

Nevertheless, the future work must be concentrated on the inner-level analysis of handwritten materials, concerning words and characters, to evaluate the changes in the particular characteristics of handwriting due to neurological diseases.

References

[1] Abuhaiba I. S. I., S. Datta, M. J. J. Holt, *Line extraction and stroke ordering of text pages*, Proc. of the Third Int. Conference on Document Analysis and Recognition (ICDAR), pp. 390-393, Montreal, Canada, August 1995

[2] Amin A., St. Fischer, T. Parkinson, R. Shiu, *Fast algorithm for skew detection*, In Proc. of SPIE'96, 1996

[3] Becker G. et al., *Early diagnosis of Parkinson's disease*, Journal of Neurology, Volume 249, Issue s03, pp iii40-iii48, 2002

[4] Hermann W. et al., *Correlation between automated writing movements and striatal dopaminergic innervation in patients with Wilson's disease*, Journal of Neurology, Volume 249, Issue 8, pp 1082-1087, 2002

[5] Holmen K., K.Ericsson, Y.Forsell, B.Winblad, *Human-Figure-Drawing (HFD) in the screening of different types of dementia in old age*, In Proc. of Sixth Int. Conf. on Handwriting and Drawing, pp. 264-266, Paris, 1993

[6] Jose L.Contreras-Vidal, *Towards non-invasive diagnostic tools and biological markers for Parkinson's disease: A commentary*, Bulletin of the International Graphonomics Society, BIGS, Volume 14, N: 1, pp. 4-5, 2000

[7] Liang J., I. Philips, R. Haralick, *A statistically based, highly accurate text-line segmentation method*, In Proc. of Fifth Int. Conference on Document Analysis and Recognition (ICDAR), pp. 551-554, Bangalore, India, Sept. 20-22, 1999,

[8] Likforman-Sulem L., A. Hanimyan, C. Faure, *A Hough based algorithm for extracting text lines in handwritten documents*, In Proc. of the Third Int. Conference on Document analysis and recognition (ICDAR), pp. 774-777, Montreal, Canada, August 1995

[9] Lumia R., *A new connected components algorithm for virtual memory computers*, IEEE Trans. on CVGIP, Vol. 23, pp. 287-300, 1983

[10] Mai N., C.Marquardt, *Analysis of handwriting movements in brain damaged patients*, In Proc. of Sixth Int. Conf. on Handwriting and Drawing, pp. 240-242, Paris, 1993

[11] Marti U., H. Bunke, *Handwritten Sentence Recognition*, In Proc. of the 15^{th} Int. Conf. on Pattern Recognition (ICPR), Volume 3, pp. 467-470, Barcelona, Spain, 2000

[12] Marti U., H. Bunke, *A full English sentence database for off-line handwriting recognition*, In Proc. of the Fifth Int. Conf. on Document Analysis and Recognition (ICDAR'99), pp. 705 – 708, Bangalore, 1999

[13] Park J.-M, C. Looney, H.-C. Chen, *Fast connected component labeling algorithm using a divide and conquer technique*, In Proc. of the 15^{th} Int. Conference on Computers and Their Applications (CATA), pp 373-376, New Orleans, Louisiana, USA, March 29-31, 2000

[14] Rosenfeld A., J.L. Pfaltz, *Sequential operations in digital processing*, JACM, Vol.13, pp. 471-494, 1966

[15] Said H.E.S., G. Peake, T. Tan, K. Baker, *Writer identification from non-uniformly skewed handwriting images*, On-line Proc. of the Ninth British Machine Vision Conference, 1998

[16] Shapiro V.A., P.K. Veleva, V.S.Sgurev, *An adaptive method for image thresholding*, Proc. of 11th Int. Conf. on Pattern Recognition, pp. 696-699, Hague, 1992

[17] Waked B., S. Bergler, C. Y. Suen, S. Khoury, *Skew detection, page segmentation and script classification of printed document images*, IEEE Int. Conf. on SMC (SMC'98), pp. 4470-4475, San Diego, California, Oct. 1998

[18] Wienecke M., G. A. Fink, G. Sagerer, *Experiments in unconstrained off-line handwritten text recognition*, In Proc. of the 8th Int. Workshop on Frontiers in Handwriting Recognition (IWFHR), Ontario, Canada, August 2002

Simulation of MPEG Codec for Moving Pictures

Rumen P. Mironov[1] and Lidia A. Nikolova[2]

Abstract – **A method for objective assessment of some neurological diseases (ND) is developed. It is based on the automatic processing and analysis of images of a handwritten script belonging to patients with ND and healthy individuals. In this method connected components are extracted, filtered and analyzed both for the whole script and for each of the automatically separated text-lines. Eight common geometric features describing size, shape and orientation of the handwriting are computed and statistically investigated. The vertical stroke size, aspect ratio and mean slope's angle revealed differences between ND-patients and those from the control group. Also the average stroke width and average stroke number relative to the row length can be used to separate the clinical Parkinsonians from the others. Nevertheless additional features and data are required to obtain reliable evaluation of the handwriting changes due to neurological diseases.**

Keywords – **image processing, video coding, simulation.**

I. Introduction

One of the fastest developing technologies in the field of video and audio compression is the MPEG standard [1]. It was initially oriented towards using in the digital television, but afterwards is approved in the personal computers and is included in the software packages of the operating systems as an optimal tool for interactive video, multimedia and graphics applications. In connection with this a number of substandards for video and audio compression were developed as Px64 (H.261, H.263), MJPEG, MPEG-1 to 7 [1-3].

In the presented paper a software model of coder/decoder based on MPEG-2 [4] standard is described, which is designed to work in large number of application areas.

II. Mathematical Description

In MPEG-2 [4] standard there is indexing through levels, which form approximated borders between process power, based on image size and profiles, which limit the algorithm properties.

On Fig. 1 the principle block scheme of MPEG-2 coder is depicted [1], where the different blocks perform the following functions:

A. Discrete Cosine Transform and Inverse Discrete Cosine Transform

The forward and inverse discrete cosine transform of fragments with size of 8×8 is described with the following equations:

$$F(u,v) = \frac{C(u)C(v)}{4} \sum_{i=0}^{7} \sum_{j=0}^{7} f(i,j) \times \cos\left[\frac{(2i+1)u\pi}{16}\right] \cos\left[\frac{(2j+1)v\pi}{16}\right] \quad (1)$$

$$f(i,j) = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C(u)C(v)F(u,v) \times \cos\left[\frac{(2i+1)u\pi}{16}\right] \cos\left[\frac{(2j+1)v\pi}{16}\right] \quad (2)$$

where: $C(u), C(v) = \begin{cases} 1/\sqrt{2} & u,v = 0 \\ 1 & u,v = 1,2,\ldots,7 \end{cases}$.

B. Quantisation

The standard permits an equal quantisation step for each DCT coefficient, where the step size can vary for each coefficient and for each macroblock, according to the equation $ss = qf[m,n].qs$. The factor qf depends on the position of the coefficients of the block, and the factor qs is the basic quantisation step. By default there are weighted matrixes $qf[m,n]$ defined in MPEG used for interframe and intraframe blocks.

For intraframe blocks the DC coefficient is proportional to the mean value of the spatial block and it is not quantised with value from the weighted matrix.

C. Variable Length Coding

Experimentally is proven that after quantisation a large number of zeroes are obtained in the blocks of DCT coefficients, especially in the high spatial frequencies. In MPEG this is realized by conversion of the two-dimensional area with size 8x8 in one-dimensional sequence by using zigzag scanning and applying variable length coding on it, which is more effective for long series of zeroes.

In some cases MPEG-2 gives the opportunity for alternative vertical scanning which is often more effective especially applied to the half size frame based DCT.

The basic algorithm for coding uses modified Huffman code. In the standard is permitted using the complement modified Huffman code, which is specially optimized and created for improving the intraframe blocks, which is called alternative intra Huffman code.

D. Motion Estimation

The MPEG standard does not define the method for motion estimation, but methods based on block correspondence

[1]Rumen P. Mironov is with the Faculty of Communication Techniques and Technology, Technical University of Sofia, Kl.Ohridski 8, 1000 Sofia, Bulgaria, E-mail: rpm@vmei.acad.bg

[2]Lidia A. Nikolova is with the Institute of Information Technologies, Bulgarian Academy of Sciences, 1113 Sofia, "Acad. G. Bonchev" Str., Bl.2, E-mail: lnikolova@iit.bas.bg

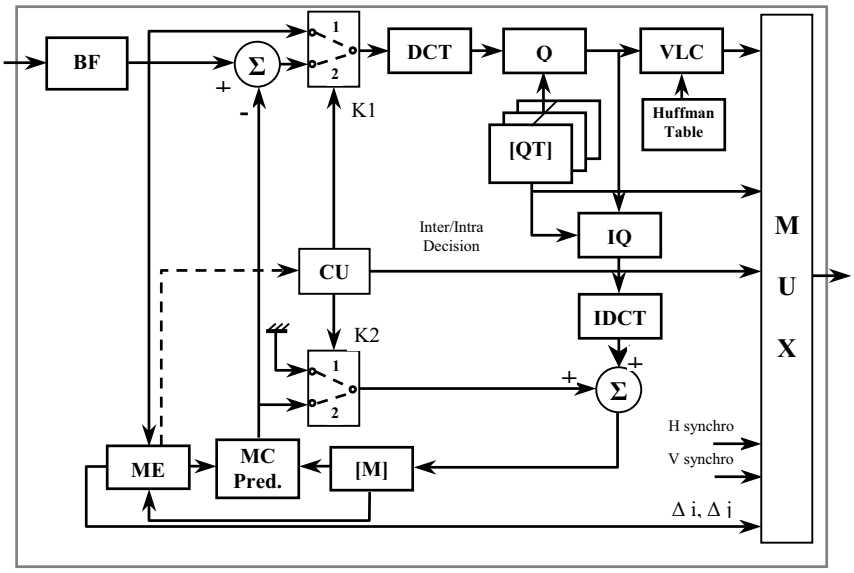


Fig. 1. Block scheme of MPEG-2 coder

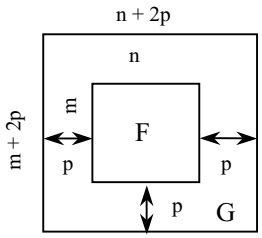


Fig. 2. Motion estimation macroblocks

are preferred. In them block with size $[m \times n]$ is compared to the ones connected with him, in search area with size $[(m + 2p) \times (n + 2p)]$ from previous or following frame. For typical MPEG system the correspondence block (or macroblock) is with size 16×16 pixels ($m = n$=16) and parameter =6, as depicted on Fig. 2, where F is the macroblock in the current frame, while G is the search area in the previous (or following) frame.

In the literature some basic algorithms for estimation of the motion vector are described as: algorithm for full search; three step search algorithm; algorithm with two dimensional logarithm search; algorithm for searching with connected direction; parallel hierarchical one dimensional algorithm for searching and algorithm for modified classification with the pixel differentiation (algorithm with layer structure).

All of them are using different estimation functions [3].

Mean Absolute Difference (MAD), defined as:

$$MAD(dx, dy) = \\ = \frac{1}{mn} \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} |F(i,j) - G(i+dx, j+dy)| \,, \quad (3)$$

where: $F(i,j)$ is macroblock from the current frame, $G(i,j)$ represent the same macroblock from a previous or a following frame, and (dx, dy) – vector representing the search area, limited in the interval $\{-p, +p\}$.

Mean Square Difference (MSD), which is defined as:

$$MSD(dx, dy) = \\ = \frac{1}{mn} \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} [F(i,j) - G(i+dx, j+dy)]^2 \,. \quad (4)$$

Cross Correlation Function (CCF), defined as:

$$CCF(dx, dy) = \\ = \frac{\sum_i \sum_j F(i,j) G(i+dx, j+dy)}{\left[\sum_i \sum_j F^2(i,j)\right]^{\frac{1}{2}} \left[\sum_i \sum_j G^2(i+dx, j+dy)\right]^{\frac{1}{2}}} \,. \quad (5)$$

MAD estimation function is appropriate for video applications, because is easy for hardware implementation. The other two estimation functions MSD and CCF might be more effective but are difficult for hardware implementation. For simplifying the calculation complexity of MAD, MSD and CCF estimation functions a simple criterion for correspondence is proposed – pixel differential classification (PDC). PDC criteria is defined as:

$$PDC(dx, dy) = \sum_i \sum_j T(dx, dy, i, j), \quad (6)$$

where:

$$T(dx, dy, i, j) = \begin{cases} 1, & |F(i,j) - G(i+dx, j+dy)| \leqslant t \\ 0, & |F(i,j) - G(i+dx, j+dy)| > t \end{cases}$$

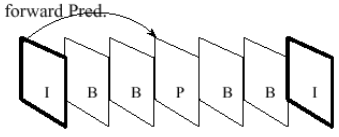


Fig. 3. P-picture coding

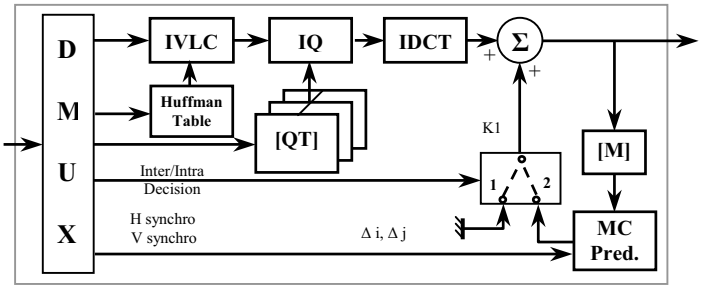


Fig. 4. Block scheme of MPEG-2 decoder

is the binary representation of pixel difference.

In this way each pixel from the macroblock is classified either as a pixel with correspondence (T=1) or as a pixel for which there is no correspondence (T=0). The block, for which PDC function has minimum, is chosen as the block with the maximum correspondence.

E. Intraframe and Interframe coding

In the MPEG standard three types of frames are defined: intra, predicted and bi-directional.

The I-pictures are coded intraframe and represent potential points for random access in compressed video date.

The P-pictures coded using forward prediction, based on the nearest previous I- or P- frames. As the I-, the P-frames are used as predicted reference frames for the next B- and P-frames as is shown on Fig. 3. The difference is that the P-frames using motion compensation reach better compression from the possible one for I frames.

The B-pictures are coded by the bi-directional prediction, using as reference both following and previous frames, as is shown on Fig. 5. The B-pictures deliver the best compression and don't distribute errors, as they are never used as reference frames. The bi-directional prediction leads to minimizing the noise effect by two frames averaging.

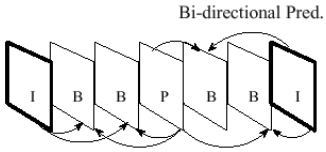


Fig. 5. B-picture coding

On the Fig. 4 [1] the block scheme of the corresponding decoder, which blocks have analogy functions is depicted.

III. Experimental Results

The developed software MPEG-2 codec is realized in correspondence with the basic block schemes, depicted on Fig. 1 and Fig. 4 and support levels LOW and MAIN in the MAIN profile. The different blocks from the schemes are realized as subroutines in the environment MATLAB 5.3 for Windows 9x/2000 platform.

The module which realizes the whole work of the coder and combines all other blocks is ***CODMPG()***. It supports two realizations, which differ by the way of coding of B frames according to: algorithm for coding with bi-directional motion estimation and the algorithm for coding using motion vector interpolation. The other modules are the following:

- $B2 = dctcode(I, mask)$ – forward DCT and quantisation;
- $I2 = dctdecode(B, mask, aman)$ – IDCT and dequantisation;
- $[vectx, vecty] = blkmatch(X1, X2, Mval, Nval, searchlimit)$ – motion estimation, which includes two sub-functions for motion vector estimation using the algorithm for full searching and three step search algorithm;
- $[fcod, Prec] = pframe(frm, Irec, vectx, vecty, Nval, Mval, nx, ny)$ – used for full coding and reconstruction of P frames;
- $[fcod, Irec] = iframe(Itemp)$ – coding and reconstruction of I frames.

The module realizing the whole work and combining all the other blocks is ***DECODMPG()*** and contains the following modules:

- $I2 = dctdecode(B, mask, aman)$ – IDCT and dequantisation;
- $Prec = pframedec(Irec, fcod, vectx, vecty, Nval, Mval, nx, ny)$ – decoding of P-frames;
- $Brec = bframedec(Irec, Prec, Pxy1, Pxy2, Nval, Mval, nx, ny)$ – decoding of B-frames;

The software modeling is performed for video sequences of 13 frames in BMP format, displayed on Fig. 7, where the separate images are ordered in the following order: $\mathbf{I_1}$, $\mathbf{B_1}$,

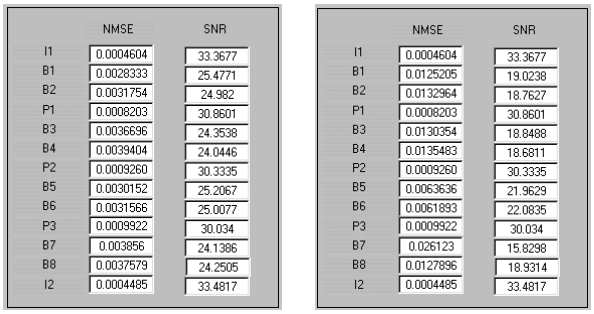


Fig. 6. Results for NMSE and SNR for two algorithms

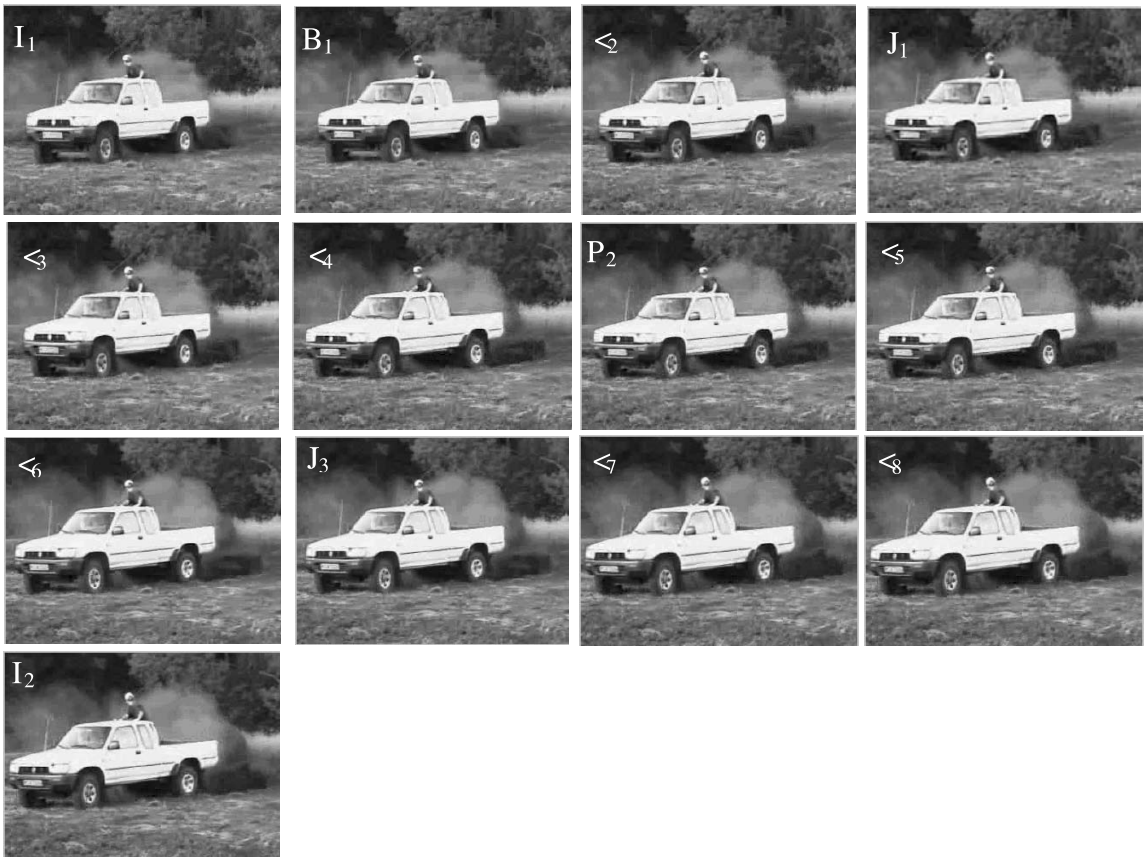


Fig. 7. Original test sequence with 13 frames

$\mathbf{B_2}$, $\mathbf{P_1}$, $\mathbf{B_3}$, $\mathbf{B_4}$, $\mathbf{P_2}$, $\mathbf{B_5}$, $\mathbf{B_6}$, $\mathbf{P_3}$, $\mathbf{B_7}$, $\mathbf{B_8}$, $\mathbf{I_2}$. For testing of the developed software package a PC Pentium MMX 200 Hz / 96 MB RAM with Windows 98 OS is used. The quality estimation is performed subjectively and objectively by estimation of NMSE and SNR for each of the frames. The obtained results are displayed in the output tables, shown on Fig. 6.

The coding time for the first testing sequence is 24 min 47 sec and the decoding time is 8 min 41 sec. The compression coefficient is 40.5175. The coding time for the second sequence is 9 min 8 sec and the decoding time is 8 min 41 sec. The compression coefficient is 41.0985.

IV. Conclusion

Two methods for obtaining the motion vectors of B frames are tested, in order to compare the quality of the video sequence, processed by the coder, calculation complexity and time for coding. From the obtained experimental results is made a conclusion for the advantage of bi-directional motion estimation for B frames for fast changing scenes, by increasing the calculation complexity of the coding process. The method with interpolation is effective for slow changing scenes, for which small distortion in quality is observed, but the number of calculation operations is substantially decreased.

The developed codec is used in laboratory work on the disciplines: “Image and Signal Processing” and “Audio and Video Communication on Internet” and in the experimental work in laboratory “ESVI” in Technical University of Sofia.

References

[1] J.Gibson, T.Berger, T.Lookabaugh. *Digital Compression for Multimedia*. Morgan Kaufmann Publishers, 1998.

[2] J.Ohm. “Encoding and Reconstruction of Multiview Video Objects”, IEEE Signal Processing Magazine, May, 1999.

[3] N. Beser, J. Hopkins. Image Compression and Packet video.

[4] T.Sikora. “MPEG Digital Video-Coding Standards”, IEEE Signal Processing Magazine, September 1997.

[5] ISO/IEC. “MPEG-2 Description”, ISO/IEC JTC1/ SC29/ WG11, July 1996.

Analogue Neural Network Comparative Simulation by Means f MATLAB and PSPICE Software Products

Liliana Docheva[1], Aleksander Bekiarski[2]

Abstract – **The most popular method used for the purpose of investigating a neural network's capability is that of simulation. While such a simulation ignores the parallelism issues inherent in neural networks, it nevertheless provides us to verify the proposed neural network's performance in an empirical setting rather then from a theoretical standpoint. In this paper, a comparison is made between the MATLAB and PSPICE simulation results.**

Keywords – **Neural networks, Analogue methods, Analogue models.**

I. Introduction

The simulation is the most popular method used for the purpose of investigating a neural network's capability. While such a simulation ignores the parallelism issues inherent in neural networks, it nevertheless provides us to verify the proposed neural network's performance in an empirical setting rather then from a theoretical standpoint. The aim of this paper is to investigate an analogue neural network cell propriety and a comparison is made between the MATLAB and PSPICE simulation results [2-5].

II. Matlab model

In last twenty years many researchers investigated a neural network computational model. Figure 1 depicts a simple model of neural network cell [1]. It consist of two building blocs. The first of them is an adder. It sum up the weighted input $\mathbf{w}_{ij}\mathbf{X}_i$. The second bloc transfers the sum which is the only argument of the transfer function **f**. The transfer function, typically is a step function or a sigmoid function, that takes the argument **h** and produces the output **Y**.

$$Y_j = f\left(\sum w_{ij}X_j + b\right) = f\left(h_j\right) \tag{1}$$

The transfer function that we use is a sigmoid function:

$$f\left(h_i\right) = \frac{1}{1+e^{-h_i}}\ . \tag{2}$$

III. Pspice Model

A neural network cell presented as equivalent electrical scheme is shown on Fig. 2 [2-5]. The scheme include independent voltage source **E**, independent current source **I**, capacitor **C**, resistors $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$, linear voltage controlled current sources, and nonlinear voltage controlled current source.

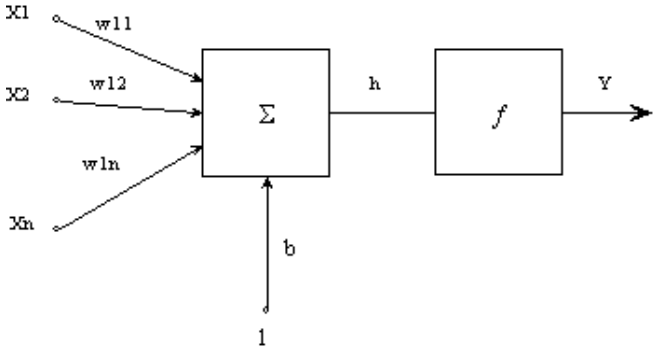


Fig. 1. A simple model of neural network cell.

The input signals are presented to the neural network cell as voltages $\mathbf{V}_1$, $\mathbf{V}_2$, …,$\mathbf{V}_n$, which are converted into currents. Summation of the product of the weight and input occurs by injecting a current, proportional to the synaptic weight, into a summing node for the duration of each input signal.

The adder is realized by means of a resistor $\mathbf{R}_2$. The

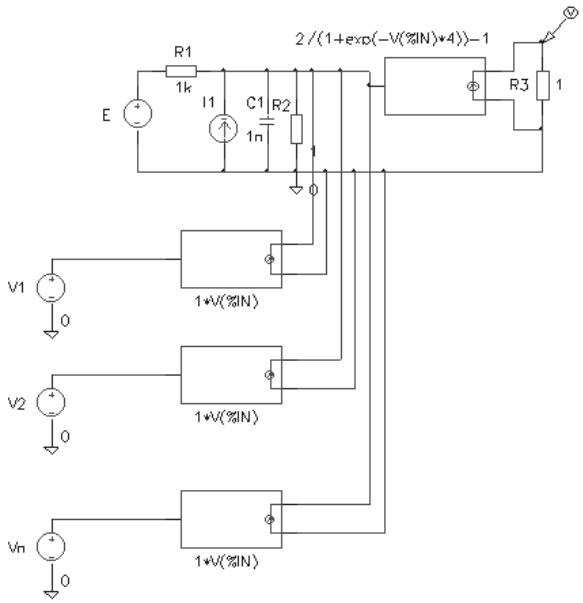


Fig. 2. A neural network cell presented as equivalent electrical scheme.

[1]Liliana Docheva is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: docheva@vmei.acad.bg

[2]Aleksanser Bekiarski is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: aabbv@vmei.acad.bg

currents produced by the linear voltage controlled current sources are passed into the summing resistor producing summation through Kirchhoff's current law.

In this paper are investigated static characteristics of the neural network cell. Then the neuron body voltage $\mathbf{U}_{z,j}$ is:

$$U_{z,j} = (I_1 + I_2 + ... + In + I) .R2 , \quad (3)$$

where **I** correspond to the bias (**b**).

With nonlinear voltage controlled current source is realized the sigmoid transfer function and the output voltage is:

$$U_y = \frac{1}{1 + e^{-U_{z,j}}} . \quad (4)$$

IV. Simulations Results

The mathematical model (Eq. 2) is modified in order to obtain output voltage range of -1 V to 1 V.

$$U_y = \frac{2}{1 + e^{-4U_{z,j}}} - 1 \quad (5)$$

The simulation MATLAB result in case of **b=0, $\mathbf{w}_{ij}$=1** and the input signal variation range is of -1 to 1 with variation step 0.1 is shown on Fig. 3. The simulation PSPICE result in case of $\mathbf{I}_1$**=0, $\mathbf{w}_{ij}$=1** and input signal variation range is of -1 V to 1 V with variation step 0.1 V is shown on Fig. 4.

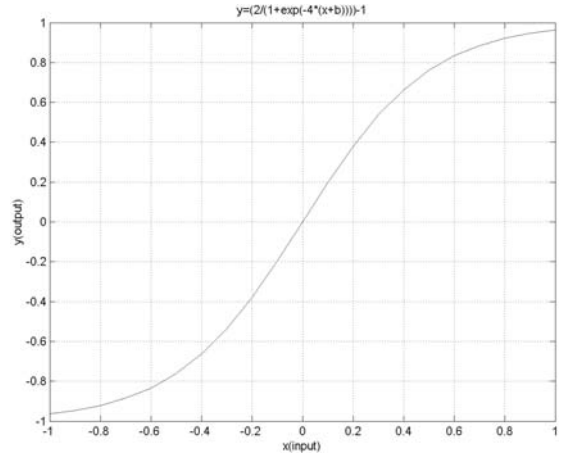


Fig. 3. The simulation MATLAB results **b=0, $\mathbf{w}_{ij}$=1**.

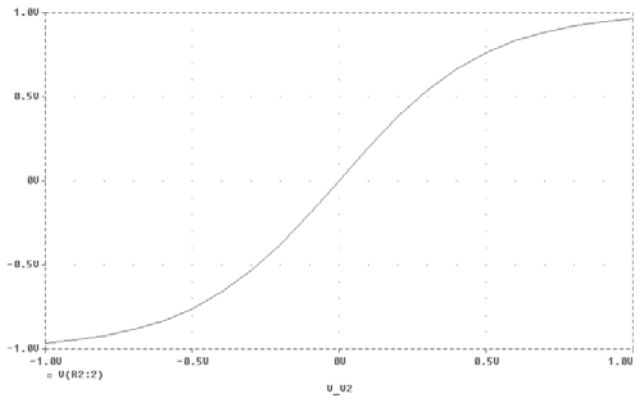


Fig. 4. The simulation PSPICE results $\mathbf{I}_1$**=0, $\mathbf{w}_{ij}$=1**.

The simulation MATLAB results in case of $\mathbf{w}_{ij}$**=1**, input signal variation range is of -1 to 1 with variation step 0.1 and **b** variation range is of -1 to 1 with variation step 0.5 are shown on Fig. 5. The simulation PSPICE results in case of $\mathbf{w}_{ij}$**=1**, input signal variation range is of -1 V to 1 V with variation step 0.1 V and $\mathbf{I}_1$ variation range is of -1 A to 1 A with variation step 0.5 A are shown on Fig. 6.

The simulation MATLAB results in case of **b=0**, input sig-

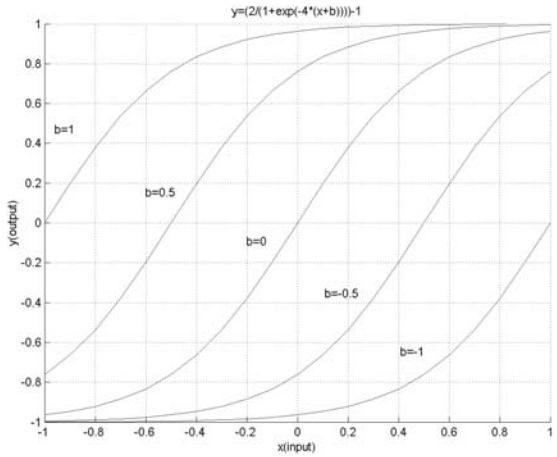


Fig. 5. The simulation MATLAB results **b=var, $\mathbf{w}_{ij}$=1**.

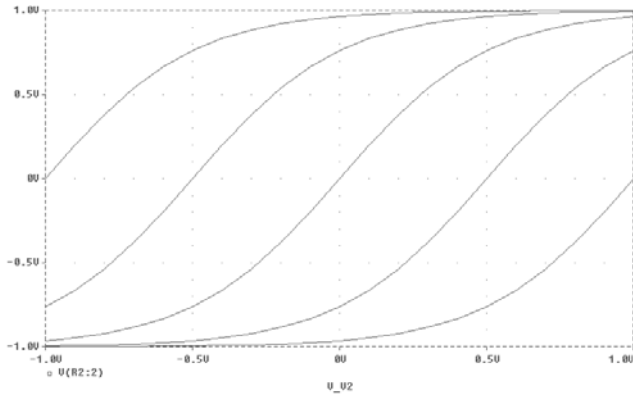


Fig. 6. The simulation PSPICE results $\mathbf{I}_1$**=var, $\mathbf{w}_{ij}$=1**.

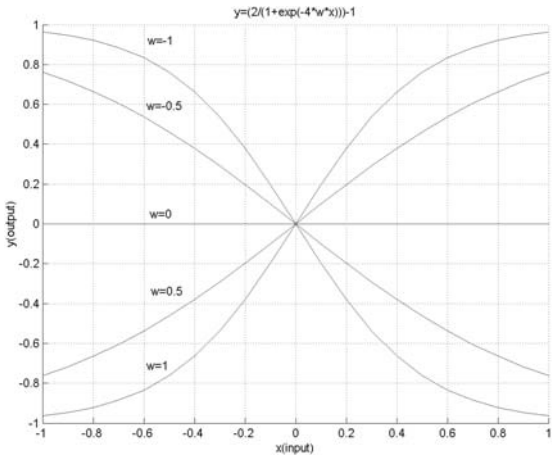


Fig. 7. The simulation MATLAB results **b=0, $\mathbf{w}_{ij}$=var**.

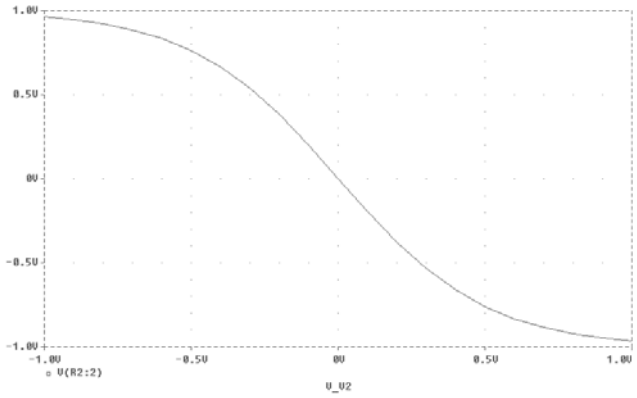


Fig. 8. The simulation PSPICE results $\mathbf{I}_1$**=0,** $\mathbf{w}_{ij}$**=-1**.

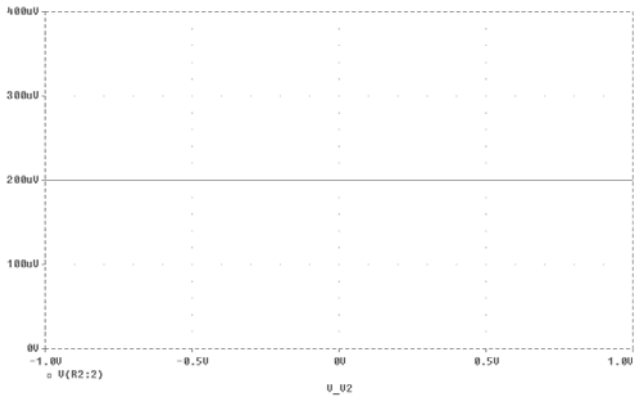


Fig. 9. The simulation PSPICE results $\mathbf{I}_1$**=0,** $\mathbf{w}_{ij}$**=0**.

nal variation range is of -1 to 1 with variation step 0.1 and **w** variation range is of -1 to 1 with variation step 0.5 are shown on Fig. 7. The simulation PSPICE results in case of $\mathbf{I}_1$**=0**, input signal variation range is of -1 V to 1 V with variation step 0.1 V and $\mathbf{w}_{ij}$ is respectively -1, 0, 1 are shown on Fig. 8, Fig. 9 and Fig. 10.

As shown of the presented above investigation (from Fig. 3 to Fig. 10) the analogue neural network cell model behavior is quite as the MATLAB mathematical model performance. Therefore the investigated analogue neural network cell model can be used to build neural networks.

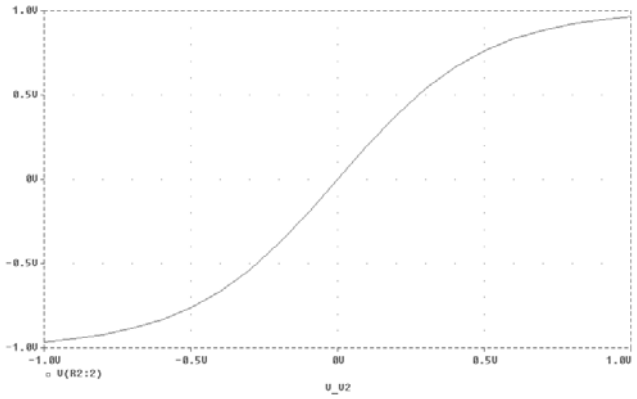


Fig. 10. The simulation PSPICE results $\mathbf{I}_1$**=0,** $\mathbf{w}_{ij}$**=1**.

V. Conclusion

In this article we presented a comparison between the MATLAB and PSPICE simulation results of an analogue neural network cell models. It is shown how the bias variation and the weigh variation influence over the output signal. Since, as the analogue neural network cell model behavior is quite as the MATLAB mathematical model performance, the investigated analogue neural network cell model can be used for neural networks building.

References

[1] T. Kirova, *Neural Networks. Main arhitectures and learning algorithm*, Softeh, Sofia 1995.

[2] A. Bekiarsky, L. Docheva, "Neural Network design through analog methods", *Communication, electronic computer systems*, Vol. 1, pp. 112-117, Sofia 2000.

[3] Docheva L., A. Bekiarsky, "Neural Networks modeling through analog equivalent scheme", *Energy and information systems and technologies*, Vol. 2, pp. 471-475, Bitola 2001.

[4] L.O. Chua and L. Yang, "Cellular Neural Networks: Theory", *IEEE Transactions Circuit and systems*, CAS-35, pp. 1257-1272, 1988.

[5] L.O. Chua and L. Yang, "Cellular Neural Networks: Applications", *IEEE Transactions Circuit and systems*, CAS-35, pp. 1257-1272, 1988.

Image Compression with Recursive IDP Algorithm

Roumen Kountchev[1]

Abstract – **A recursive algorithm for compression of halftone and color images, in correspondence with the method for Inverse Difference Pyramid decomposition (IDP), based on two-dimensional orthogonal transforms, is described in the paper. The basic steps of the algorithm for coding and decoding are defined, and the corresponding block diagram is synthesized. Some results of the model of the method behavior and comparison with the standard JPEG are given also.**

Keywords – **image compression, inverse difference pyramid decomposition, truncated orthogonal transform.**

I. Introduction

The basic requirements towards image compression algorithms are the big compression ratio, the high quality of the restored image and the low computational complexity. The algorithm for still images compression, based on the Inverse Pyramid Decomposition (IDP) method [1], presented here, suits these requirements to a high degree. The method efficiency is evaluated comparing it with the wavelet decomposition using a bank of 3/5-tap digital filters, used in the JPEG 2000 standard [2]. Possible applications of the offered algorithm are evaluated on the basis of the comparison results.

II. IDP Coding

The algorithm for recursive IDP coding of halftone digital images with orthogonal transforms is presented with the block diagram in Fig. 1. The orthogonal transform, used here, is the two-dimensional Discrete Cosine Transform (DCT) [1]. The coding of the original digital image is done performing the following steps:

Step 1: The Matrix $[B(i,j)]$ of the digital halftone image is divided in K sub-images, $[L(i,j)]$ each with size $2^n \times 2^n$ elements, where n is in the interval $n = 3, 4, 5$.

Step 2: The elements of the matrix $[L_{k_p}(i,j)]$ with number k_p=1,2,..,4^pK are defined for the level $p = 0, 1, .., P-1$ ($1 < P \leqslant n$) of the IDP pyramid of levels, in correspondence with:

$$L_{k_p}(i,j) = \begin{cases} B_{k_0}(i,j) & \text{for } p = 0; \\ E_{k_{p-1}}(i,j) & \text{for } p = 1, 2, ..., P-1, \end{cases} \quad (1)$$

k_p=1,2,.., 4^pK, i,j=0,1,2,.. 2^{n-p}-1.

Here $B_{k_0}(i,j)$ is the (i,j) pixel from the sub-image with number k_0=1,2,..,K in the zero level (p=0) of the pyramid, which coincides with the pixel $B(i,j)$ from the original image, and $E_{k_{p-1}}(i,j)$ is respectively the pixel (i,j) from the difference image with number k_p in the pyramid level p, for p=1,2,…,$P-1$.

Step 3: The matrix of the sub-image $[L_{k_p}(i,j)]$ is transformed, using the so called "truncated" two-dimensional orthogonal transform of the kind, selected in advance, like DCT, WHT, Haar, etc. The coefficients of the corresponding matrix-transform are defined with the expression:

$$s_{k_p}(u,v) = \begin{cases} s_{k_p}(u_r, v_r) & \text{for} \quad m_p(u,\, v) = 1; \\ 0 & \text{for} \quad m_p(u,\, v) = 0, \end{cases} \quad (2)$$

$u, v = 0, 1, .. , 2^{n-p} - 1$ and p=0,1,..,P-1, where

$$s_{k_p}(u_r, v_r) = \frac{1}{4^{n-p}} \sum_{i=0}^{2^{n-p}-1} \sum_{j=0}^{2^{n-p}-1} L_{k_p}(i,j)\, t_p(i,j,u_r,v_r)$$

for r=1,2,...,R_p; $m_p(u,v)$ are the elements of the binary matrix-mask $[M_p]$ with size $2^{n-p} \times 2^{n-p}$ for the level p, which defines the position of the "retained" coefficients $s_{k_p}(u_r, v_r)$; $R_p = \sum_{i=0}^{2^{n-p}-1} \sum_{j=0}^{2^{n-p}-1} m_p(i,j)$ – the number of the "retained" spectrum coefficients in the level p, selected in advance in the interval $1 \leqslant R_p < 4^{n-p}$; $t_p(i,j,u_r,v_r)$ - the pixel (i,j) from the basic image (the kernel of the selected orthogonal transform) with spatial frequency (u_r, v_r) in the level p.

The elements $m_p(u,v)$ in the Eq. (2) are defined with the following four steps:

3.1. Calculation of the modules of the spectrum coefficients of every sub-image in the pyramid level p:

$$|s_{k_p}(u,v)| = \frac{1}{4^{n-p}} \left| \sum_{i=0}^{2^{n-p}-1} \sum_{j=0}^{2^{n-p}-1} L_{k_p}(i,j)\, t_p(i,j,u,v) \right| \quad (3)$$

for $u, v = 0, 1, 2, .., 2^{n-p} - 1$.

3.2. Calculation of the modules of the coefficients of the mean transform for the level p:

$$|\bar{s}_{k_p}(u,v)| = \frac{1}{4^pK} \sum_{k_p=1}^{4^pK} |s_{k_p}(u,v)| \quad (4)$$

3.3. Arrangement of the "mean" modules in uniformly decreasing order:

$$|\bar{s}(u_1,v_1)| \geqslant |\bar{s}\,(u_2,v_2)| \geqslant \ldots \geqslant |\bar{s}(u_{R_p}, v_{R_p})|; \quad (5)$$

3.4. Definition of the elements m_p(u,v) in accordance with Eq. (5) as follows:

$$m_p(u,v) = \begin{cases} 1 \text{ for } u{=}u_r \text{ and } v{=}v_r \text{ for } r{=}1, 2, .., R_p; \\ 0 - \text{in all other cases.} \end{cases} \quad (6)$$

Step 4: The quantizated value of every retained spectrum coefficient is defined:

[1]Roumen K. Kountchev is with the Faculty of Communications and Communications technologies, Technical University of Sofia, Kliment Ohridsky 8, 1000 Sofia, Bulgaria; E-mail: rkountch@vmei.acad.bg

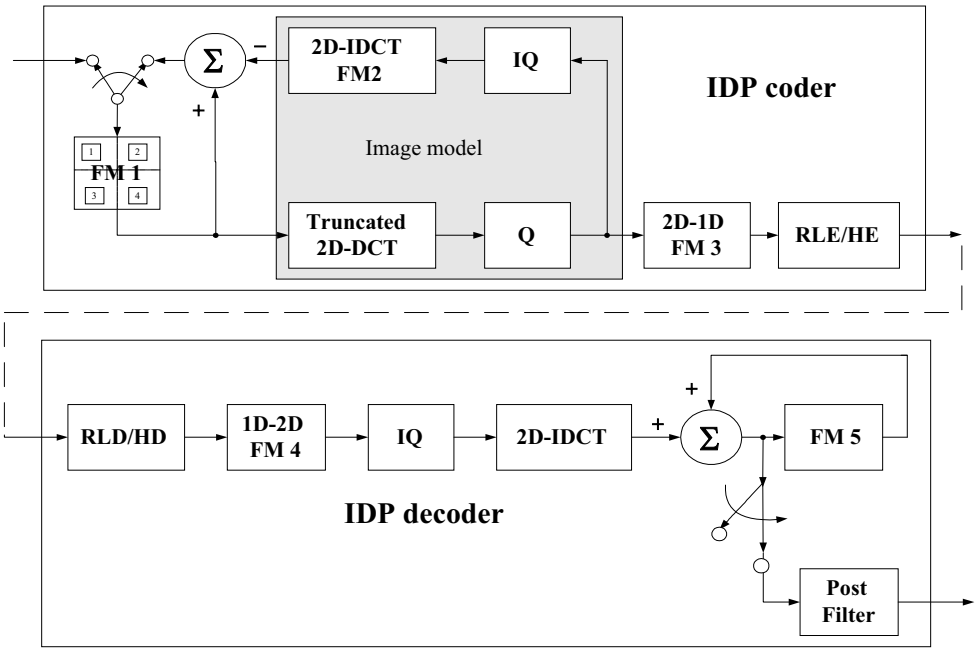


Fig. 1. Block-diagram of recursive IDP coder

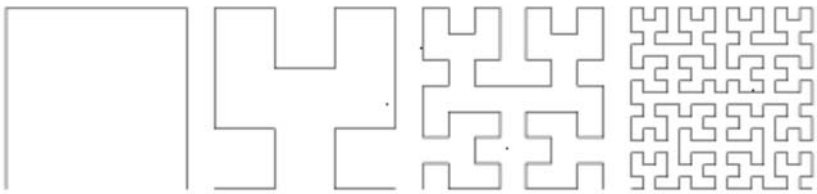


Fig. 2. Recursive Hilbert scan for $n = 2, 3, 4$

$$s^q_{k_p}(u_r, v_r) = [s_{k_p}(u_r, v_r) / \Delta_p(u_r, v_r)]_{\text{integer}} \quad (7)$$

for r=1,2,.., R_p, where $\Delta_p(u_r, v_r)$ is an element of the quantizating matrix $[Q_p]$ for the level p, selected in advance on the basis of experimental research on the influence of the quantization error on the restored image quality, $[*]_{integer}$ – operator, defining the integer part of the number in the brackets.

Step 5: The dequantizated value of every quantizated spectrum coefficient is defined:

$$s'_{k_p}(u_r, v_r) = s_{k_p}(u_r, v_r).\Delta_p(u_r, v_r) \quad (8)$$

Steps 6: The approximating model $\tilde{L}_{k_p}(i, j)$ for the sub-image k_p in pyramid level p is defined, using the inverse orthogonal transform:

$$\tilde{L}_{k_0}(i, j) = \tilde{B}_{k_0}(i, j) = \sum_{u_r} \sum_{v_r} s'_{k_0}(u, v) t_0^{\text{in}}(i, j, u, v) \quad (9)$$

for p=0 and $i, j = 0, 1, .., 2^n - 1$,

$$\tilde{L}_{k_p}(i, j) = \tilde{E}_{k_p}(i, j) = \sum_{u_r} \sum_{v_r} s'_{k_p}(u, v) t_p^{\text{in}}(i,\ j,\ u, v) \quad (10)$$

for $p = 1, .., P - 1$ and $i, j = 0, 1, .., 2^{n-p} - 1$.

Here $t_p^{in}(i, j, u, v)$ is the kernel of the selected inverse orthogonal transform for the level p.

Step 7: The elements of the difference image k_p in the pyramid level p are defined:

$$E_{k_p}(i, j) = \begin{cases} B_{k_0}(i, j) - \tilde{B}_{k_0}(i, j) \text{ for } p = 0; \\ L_{k_{p-1}}(i, j) - \tilde{L}_{k_{p-1}}(i, j) \text{ for } p = 1, 2, .., P-1. \end{cases} \quad (11)$$

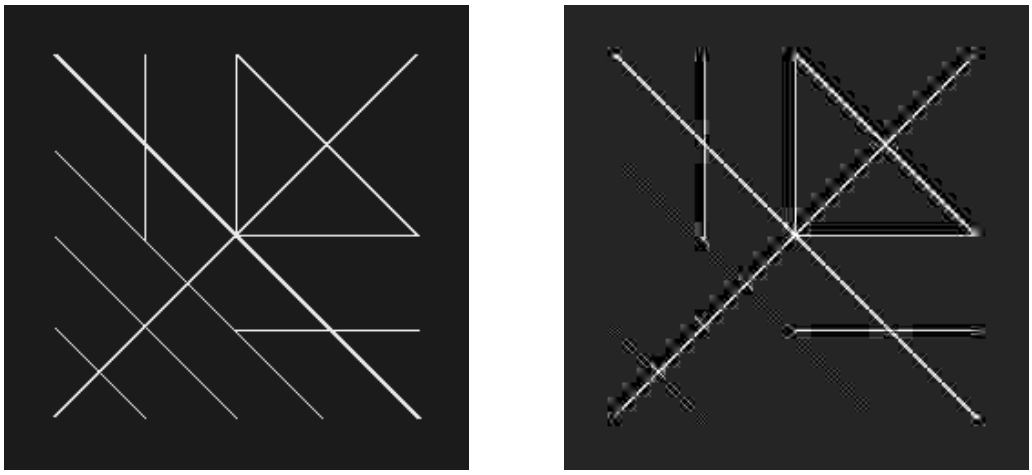
when $i, j = 0, 1, .., 2^{n-p} - 1$.

Step 8: The coefficients $s^q_{k_p}(u_r, v_r)$ from all sub-images in the pyramid level p are arranged in R_p two-dimensional massifs in accordance with their spatial frequency (u_r, v_r) for r=1,2,..,R_p.

Step 9: Every two-dimensional massif of spectrum coefficients is converted into one-dimensional, using recursive Hilbert scan [3], shown in Fig. 2. Thus obtained one-dimensional massifs from every IDP level are arranged as one common sequence. At the beginning of this sequence is inserted the header, which contains information about the elements of the mask $[M_p]$, the number of the selected matrix $[Q_p]$, the values of R_p and , the kind of the orthogonal transform for every pyramid level, etc.

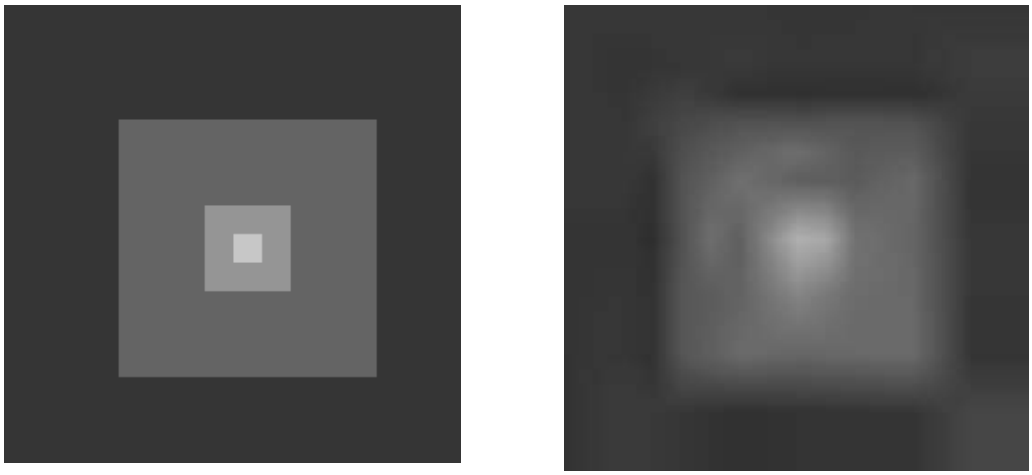
10.1. Adaptive coding of the lengths of the series of equal symbols (RLE);

10.2. Adaptive coding with modified Huffman code (HE). In result is obtained a compressed data file of the image data, which could be transferred via the communications channel or saved in a PC memory, depending on the IDP algorithm.



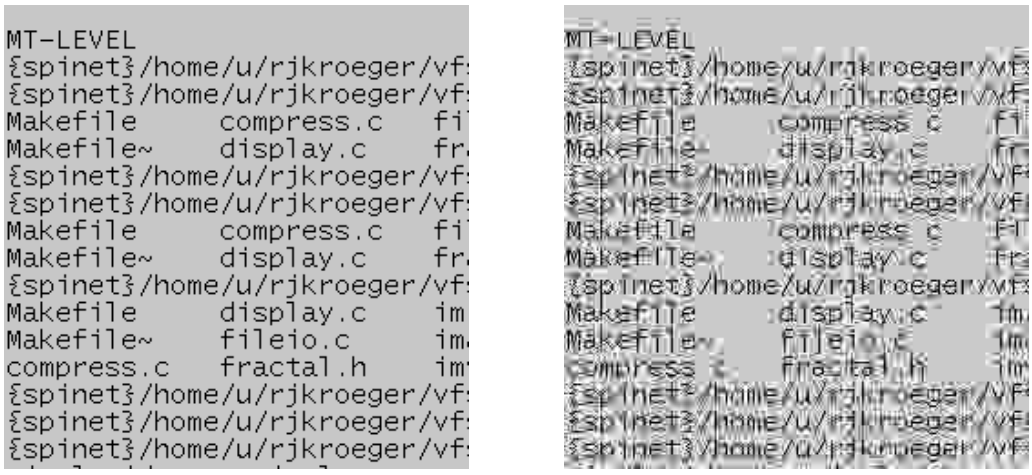
IDP compression 38,8:1 with PSNR = f JPEG2000 compression 38,5:1

Fig. 3. Test image “Crosses”, 256x256, 8 bpp



IDPcompression 468:1 with PSNR = f JPEG2000 compression 468:1

Fig. 4. Test image “Squares”, 256x256, 8 bpp



IDP compression 13.2:1 with PSNR = f JPEG2000compression 13,2:1

Fig. 5. Test image “Text”, 256x256, 8 bpp

III. IDP Decoding

The decoding of the compressed image data in accordance with the general recursive IDP algorithm is performed following the steps below (Fig. 1):

Step 1: The Huffman codes and the lengths of the series of equal symbols are decoded (RLD+HD decoding);

Step 2: The coefficients $s^q_{k_p}(u_r, v_r)$ are dequantizated, in accordance with Eq. (8);

Step 3: The approximation model of the sub-image $\tilde{L}_{k_p}(i,j)$ is calculated using the inverse orthogonal transform, Eqs. (9)-(10);

Step 4: The elements $B'(i,j)$ of the restored image are calculated:

$$B'_k(i,j) = \tilde{B}_{k_0}(i,j) + \sum_{p=1}^{P-1} \tilde{E}_{k_{p-1}}(i,j) \qquad (12)$$

for $i,j = 0,1,..,2^n-1$ and $k = 1,2,..,K$.

Step 5: The decoded image is post-filtrated adaptively meaning the values of the pixels from the both sides of the borders of the sub-images $[L(i,j)]$ with size $2^n \times 2^n$. The border pixels are positioned in bands (respectively rows and columns) with width equal to two pixels.

IV. Coding of Color Images

The coding of color images (written in format 4:4:4), based on the described algorithm was performed, applying it on the matrix of every primary color component: R,G,B. In order to obtain higher compression ratio, the R,G,B components of every pixel (i,j) were transformed in Y,Cr,Cb and the 4:4:4 format was converted into 4:2:0, in correspondence with the ITU Recommendation 601-R:

$$\begin{bmatrix} Y(i,j) \\ Cb(i,j) \\ Cr(i,j) \end{bmatrix} = \begin{bmatrix} 0,2990 & 0,5870 & 0,1140 \\ -0,1687 & -0,3313 & 0,5000 \\ 0,5000 & -0,4187 & -0,0813 \end{bmatrix} \begin{bmatrix} R(i,j) \\ G(i,j) \\ B(i,j) \end{bmatrix} \qquad (13)$$

$$\begin{bmatrix} R(i,j) \\ G(i,j) \\ B(i,j) \end{bmatrix} = \begin{bmatrix} 1,0000 & 0,0000 & 1,4020 \\ 1,0000 & -0,3441 & -0,7141 \\ 1,0000 & 1,7720 & 0,0000 \end{bmatrix} \begin{bmatrix} Y(i,j) \\ Cb(i,j) \\ Cr(i,j) \end{bmatrix}$$

When the image format is changed from 4:4:4 into 4:2:0, the horizontal and the vertical size of the matrices $[Cb]$ and $[Cr]$ becomes two times smaller in respect to the matrix $[Y]$. This is carried out, using the mean value of every four neighboring pixels in $[Cb]$ and $[Cr]$ in correspondence with the relations:

$$\begin{aligned} Cr_0(i,j) &= \frac{1}{4}[Cr(i,j-1) + Cr(i+1,j-1) \\ &\quad + Cr(i,j) + Cr(i+1,j)] \\ Cb_0(i,j) &= \frac{1}{4}[Cb(i,j-1) + Cb(i+1,j-1) \\ &\quad + Cb(i,j) + Cb(i+1,j)], \end{aligned} \qquad (14)$$

where $Cr_0(i,j)$ and $Cb_0(i,j)$ are respectively the elements of the matrices $[Cr_0]$ and $[Cb_0]$, whose size is two times smaller than that of $[Cb]$ and $[Cr]$.

Every one of the components Y, Cr_0, Cb_0 is processed, applying the already described general algorithm IDP for coding of halftone images.

The decoding of the compressed color images is performed with the following two operations:

- Double increasing the size of the recovered (decompressed) matrices $[\hat{C}r_0]$ and $[\hat{C}b_0]$, using 2D zero-order interpolation, for the calculation of matrices $[\hat{C}r]$ and $[\hat{C}b]$;
- Inverse transform in correspondence with Eq. (13) for the components $\hat{Y}$, $\hat{C}r$, $\hat{C}b$ of every pixel, for the calculation of the corresponding primary color components, $\hat{R}, \hat{G}, \hat{B}$.

V. Results of the Modelling

Some results of the modeling of the described IDP algorithm and from its comparison with the JPEG 2000 standard are illustrated with the images shown in Figs. 3-5 and the data in Table 1 [1]. The results show that the IDP algorithm has some advantages over JPEG 2000 in the lower computational complexity and gives higher image quality for same compression ratio for text images and drawings. For texture images the algorithm, used in the JPEG 2000 standard, ensures higher compression ratio.

Table 1. Comparison of the computational complexity of decompositions IDP and WP (JPEG 2000)

	Operation			
	Sums per pixel in	Sums per pixel in	Multiplies per pixel in	Multiplies per pixel in
Decomposition	the coder	the decoder	the coder	the decoder
IDP	14,31	7,15	2,18	1,09
WP (JPEG 2000)	31,5	36,75	21	21

VI. Conclusion

A new image compression algorithm based on the IDP decomposition with 2D orthogonal transform, and suitable for processing of halftone and color images has been developed and presented in this work. It was firmly demonstrated, that the algorithm should be preferred in the following two applications:

- For compression of graphics and text images;
- For real-time intra-frame compression of TV images.

References

[1] R. Kountchev, V. Haese-Coat, J. Ronsin. Inverse Pyramidal Decomposition with multiple DCT. Signal Processing: Image Communication, Elsevier, Vol. 17, February 2002, pp. 201-218.

[2] M. Rabbani, R. Joshi. An Overview of the JPEG 2000 Still Image Compression Standard. Signal Processing: Image Communication, Elsevier, Vol. 17, Jan. 2002, pp. 3-48.

[3] J. Quinqueton, M. Berthod. A locally adaptive Peano scanning algorithm.IEEE Trans. Pattern Analysis and Machine Intelligence, No 3, 1981, pp. 403-412.

One Realization of Inexpensive System for News Contribution from Corresponding Offices to Central TV Broadcasting Station

Miloje Nesic[1], Aleksandar Spasic[2] and Dragan Jankovic[3]

Abstract – **One inexpensive realization of network for exchanging of news between TV stations, as well as system for news production is presented in this paper.**

Keywords – **Video material exchange**

I. Introduction

This project is realized on demand of Association of Independent Electronic Media (ANEM), the biggest non-government radio and television network in Serbia and Montenegro, which originally consisted of 16 TV and 28 radio broadcasters. After the events influenced on the fall of Milosevic regime and "Serbian smooth revolution" held on October 5 Th, 2000, more than 80 radio and television broadcasters became affiliate member of ANEM. ANEM strongly supported this project by all means.

System is simplified version of the system designed and partially realized during the ruling of Slobodan Milosevic and high constraints are related mainly on constant depraving, banning and financial exhausting of ANEM stations.

Basic user's demand was to provide reliable news packages exchange in broadcast quality (full PAL).

Among the others, this project proposed application of digital standards for video acquisition, non-linear video editing and communications in low-budget TV stations members of ANEM.

Also, the devices used for realization of this project had to be in customer (i.e. inexpensive) instead professional quality.

This project is realized on demand of Association of Independent Electronic Media (ANEM), the biggest non-government radio and television network in Serbia and Montenegro, which originally consisted of 16 TV and 28 radio broadcasters. After the events influenced on the fall of Milosevic regime and "Serbian smooth revolution" held on October 5 Th, 2000, more than 80 radio and television broadcasters became affiliate member of ANEM. ANEM strongly supported this project by all means.

[1]Miloje Nesic is with Radio Television Pirot, Trg pirotskih oslobodilaca 30, 18300 Pirot, Serbia & Montenegro. Miloje Nesic is a chairman of ANEM Technical Committee. E-mail: miloje.nesic@sloboda-pirot.co.yu

[2]Aleksandar Spasic is with Agency for Computer Engineering "String", Bore Stankovica 26, 18300 Pirot, Serbia & Montenegro. Aleksandar Spasic is a member of ANEM Technical Committee. E-mail: aspasic@string.co.yu

[3]Dragan Jankovic is with Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia & Montenegro, E-mail: gaga@elfak.ni.ac.yu

System is simplified version of the system designed and partially realized during the ruling of Slobodan Milosevic and high constraints are related mainly on constant depraving, banning and financial exhausting of ANEM stations.

Basic user's demand was to provide reliable news packages exchange in broadcast quality (full PAL).

Among the others, this project proposed application of digital standards for video acquisition, non-linear video editing and communications in low-budget TV stations members of ANEM.

Also, the devices used for realization of this project had to be in customer (i.e. inexpensive) instead professional quality.

II. System Overview

The 9 biggest and most influential TV broadcasters members of ANEM and ANEM Central Office in Belgrade participated in this project. One of the above mentioned broadcasters is B92, located in Belgrade, with national terrestrial covering and satellite covering.

The TV stations involved in this project are located in biggest regional community centers in Serbia.

Fig. 1. Corresponding offices (TV stations) location

The locations and list of participating stations is given on Table 1.

Table 1. Information about participating TV stations

TV Station	Location	Distance (km)	Connection type
B92 (main)	Belgrade	-	DSL, ISDN
Urbans PG	Novi Sad	75	Broadband cable
TV Nis	Nis	240	ISDN/DSL
TV Kragujevac	Kragujevac	140	ISDN
TV 5	Uzice	180	ISDN
TV Cacak	Cacak	140	ISDN
TV Kraljevo	Kraljevo	170	ISDN
TV Trstenik	Trstenik	210	ISDN
TV Pirot	Pirot	310	ISDN

III. Workstation Overview

Realized system consists of 9 completely configured workstations. Each workstation has to provide autonomous work in video acquisition, non-linear digital video editing and exchanging news.

The parts of each workstation are the digital systems for collecting video information and non-linear digital video editing ‘and systems for communications.

Workstation consists of the digital camcorder, digital tape, PVM monitor, specially configured PC-based computer, ISDN public network connection (or DSL network connection or broadband cable connection) and cabling.

The heart of workstation is PC-based computer equipped with non-linear digital video editing adapter card, i-Link (IEEE 1394) adapter card and ISDN network adapter. As an alternatives, DSL adapter or broadband modem are used. Workstation block-diagram is showed in Fig. 2.

IV. News Editing and Package Contributing Methodology

In accordance with BBC news standard norm, the duration of each news package is approximately one minute and 45 seconds.

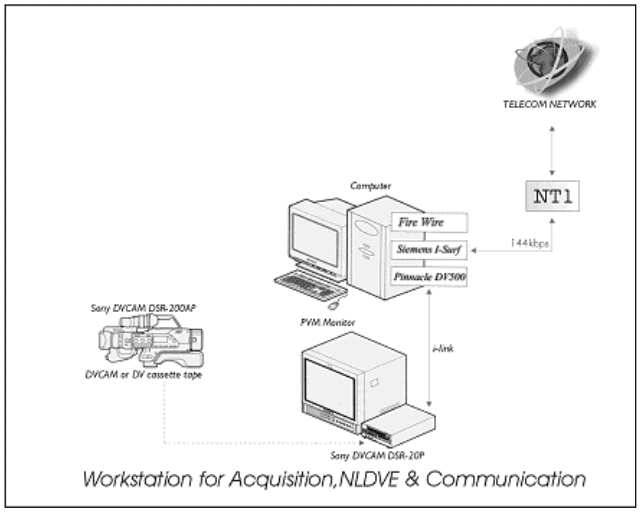


Fig. 2. Workstation block diagram

Video information collected by the ENG team are putting in the computer by FireWire (i-Link, IEEE 1394).

Non-linear digital video editing is making using the Pinnacle miro DV 500 adapter card and the suitable software (Adobe Premier).

Package is compressing using the MPEG4 software compressor (DivX Codec). Duration of compressing needed for one minute of the video material is approx. one minute.

Public ISDN services and File Transfer Protocol are using for the transmission of the news package from the workstation located in the TV station to the FTP Server computer located in B92 Central Office. It is assumed that ISDN or DSL or broadband cable connection to local Internet Service Provider (ISP) can be used.

It is possible to connect directly to FTP server located in main station using public ISDN connection.

Direct cross-connection between two of the participating stations is also possible. One station can upload/download files with video material directly to/from local FTP server located in some other station. Public ISDN or other Internet connection is using.

Duration of decompression needed for one minute of the video material is approx. one minute.

V. Statistics

Number of sended news packages per station, in average, was 5 per month. Average size of compressed file was 33.7 Mb. It means that average size of the news package was 19.2 Mb per minute.

In practice, one minute of the video material is transmitted using ISDN connection for approx. 21 minutes. The duration of transmission of one minute of video package is shown in Table 2.

Table 2. Duration of transmission

ISDN	Broadband	DSL
128K/s	256K/s	1M/s
21 min	10 min	2 min

VI. Conclusion

System for news video package exchange, reliable and available for low-budget local TV broadcasting stations is considered in this paper.

System is completely based on digital solutions that are available on the market.

Further plans are:

- Incrementally growing of the system (i.e. involving other ANEM stations and corresponding offices);
- Improving the system’s performances (throughput and compressing/decompressing techniques).
- Designing the system for management and automatic production of metadata related on video material.

References

[1] EBU Status Report, *"Networked Television Production-Compression issues"*, 1996.

[2] Wickelgren I.J. *"The facts about FireWire"*, IEEE Spectrum, April 1997, pp 19-25.

[3] *"Pinnacle Systems: DV 500"*, www.pinnaclesys.com/dv500/.

[4] Eutelsat-Internet via satellite, available from Eutelsat.

Extending Database Technology to Support Location-Based Service Applications

Dragan H. Stojanović[1] and Slobodanka J. Djordjević-Kajan[2]

Abstract – **Location-based service applications are based on mobile objects and management of their continuously changing locations. The goal of work presented in this paper is to provide extended database support for such applications, by defining mobile objects data model and an SQL extension, based on widely accepted OGC and ISO TC 211 specifications.**

Keywords – **Database, Location-Based Services, Mobile objects**

I. Introduction

Advances in wireless communication technologies, mobile positioning and Internet-enabled mobile devices, like smart phones and PDA, have given rise to a new class of location-based applications and services. Location-based services (LBS) deliver geographic information and geo-processing power to the mobile/static users in accordance with their current location and preferences, or location of the static/mobile objects of their interests. LBS are specialized, multi-tiered, component Web GIS applications, which can be published, located and invoked across the wired/wireless Web [7]. Such services like automatic vehicle location, fleet management, tourist services, transport management, traffic control and digital battlefield, are all based on mobile objects and management of their continuously changing locations. Thus, LBS applications require database and application support to model and manage mobile objects in both database and application domain and to support querying on the motion properties of the objects [8,9].

II. Mobile Point Objects in LBS

LBS are multi-tiered Internet GIS applications with architecture presented in Fig 1. The location of mobile information appliances can be determined by using GPS or mobile network triangulation. They can report their location to the LBS server through a wireless interface, or their location can be obtained through ground-based radars or satellites. At the LBS server, the data is processed and services, based on such data, are provided to the users. Mobile users may also represent mobile objects of interest to other users of the particular service. Mobile objects in LBS are characterized by point geometry and to the rest of the paper the term mobile object pertains to point object. According to T. Brinkhoff in [1] mobile real-world point objects always move along a path in a transportation network. Vehicles, trains, boats and passengers move following a particular network (roads, railways, rivers, pedestrian tracks). Air traffic also follows a (3-D) network of air corridors. Knowing their destinations, mobile objects often use fast and/or shortest paths depending on the cost criteria (time and/or distance).

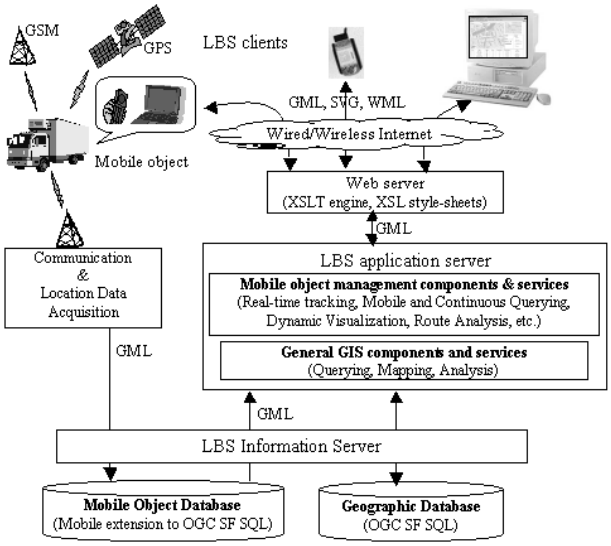


Fig. 1. The LBS architecture

The application scenario for LBS that involves mobile objects both as the users of the service and as tracked objects is as follows [8]. The mobile object registers for the certain location-based service connecting to the LBS server by sending the starting location (coordinates of the starting point or address), ending location, eventually set of points of interest that it is going to visit along its route and current speed (intended average speed). The LBS server retrieves to the mobile object its route and uncertainty threshold [9]. Based on the route and the current/average speed defined, an approximation of the expected motion of the object from the last registered location to the (near) future can be determined. The uncertainty threshold specifies the responsibility of the mobile object to send the location update to the LBS server if its current location is deviated from its expected location by defined uncertainty threshold. With every location update, the mobile object must also update the current/average speed to enable prediction of its future motion till the next location

[1]Dragan H. Stojanović is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Serbia and Montenegro, Email: dragans@elfak.ni.ac.yu

[2]Slobodanka J. Djordjević-Kajan is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Nis, Serbia and Montenegro, Email: sdjordjevic@elfak.ni.ac.yu

update. Performing such scenario, mobile object generates its certain trajectory, as an approximation of its motion in space and time. The trajectory of the mobile object is a polyline in three-dimensional space (two-dimensional space and time) represented as a sequence of points (x_i, y_i, t_i). The reason it is only an approximation is that the object does not move in straight lines at constant speed. The number of points along the trajectory is proportional to the accuracy of such approximation. An additional parameter mt_i for every point defines the type of motion during period $[t_i, t_{i+1}]$. The four motion types are defined: *punctual* – the mobile object isn't tracked and its location isn't defined during certain time period; *stepwise* – the mobile object does not move during the time period; *linear* – the mobile object move along the straight line from (x_i, y_i) to (x_{i+1}, y_{i+1}), and at constant speed, and *interpolated* – the speed of the mobile object during defined time period isn't constant and must be defined by an interpolation function.

III. Modeling Mobile Objects in the Argonaut Framework

The standard-based component framework, named ARGONAUT, represents a suite of mobile object data modeling and management components [8]. The foundational data model of the ARGONAUT framework was developed to support conceptual modeling and querying of changing properties of geographic features. The ARGONAUT data model is extensible, object-oriented, and specified using UML class diagram notation. We base our modeling approach on the comprehensive framework of data types and rich algebra of operators defined in [3] but extending their approach to the object-oriented modeling paradigm, and representation of the current as well as the future motion of the mobile objects. Also, the main emphasis we put on the mobile point objects moving on the transportation network [1,8].

The ARGONAUT data model extends the OGC Simple Features model, also adopted by ISO TC 211 [4]. The standard defines abstract *Geometry* class, and its hierarchy of specialized geometric classes (*Point, LineString, Polygon, MultiPoint,* etc.). The attributes and operations defined within geometry classes support specification of topological relations, direction relations, metric operations and operations that support spatial analysis (point-set operations) on appropriate geometry types. Time dimension of a mobile object is specified through *TimeObject* class hierarchy, defined in accordance with ISO TC 211 Temporal Schema [5] (*TimeInstant, TimePeriod TimeDuration, MultiTimeInstant* etc.). The *TimeObject* class includes attributes and operations for specifying topological relations, metric operations and point set operations on time dimension.

The base class for introducing mobility of features and continuous change of their geometric properties is an abstract *MobileGeometry* class, as the root of the extensible hierarchy of classes for specifying mobile geometries (Fig. 2). For every class in Geometry class hierarchy an appropriate class for the representation of the mobile geometry is defined (*MobilePoint, MobileLineString, MobilePolygon, MobileMultiPoint,* etc.). Any of these classes appropriately restrict the Geometry class aggregated within the *MotionSlice* class (≪restriction≫ stereotype). Being a specialization of the *Geometry* class, a *MobileGeometry* and its specialized classes can be treated in the same way as any other geometric object, i.e. it can represent the geometric property of any feature which is dynamic in nature and participate in all geometric operations and relations. An instance of the *MotionSlice* class, aggregated by the *MobileGeometry* class with multiplicity 0..n, represents the registered location of a mobile geometry, by containing a geometry value (instance of *Geometry* class), the valid time of such value (instance of the *TimeInstant* class) and the motion type (value of enumeration type *MotionType*), which describes the way geometry changes between two successively registered geometries. The ARGONAUT data model provides four enumerated values for motion types, and those are: *Punctual, Stepwise, Linear* and *Interpolated*. The first three motion types don't require any additional information to be included in the data model, but for the *Interpolated* motion type, a reference to the *Interpolation* class, with defined interpolation parameters, must be specified. The motion of a mobile object also causes continuous change of its non-spatial properties like distance from some static or a mobile object or topological relation between two mobile objects, as elaborated in [2]. Using the same approach the ARGONAUT data model defines *MobileBoolean* and *MobileDouble* classes, by inheritance from *Boolean* and *Double* base classes respectively and aggregation of appropriate *MotionBoolSlices* or *MotionDoubleSlices* classes.

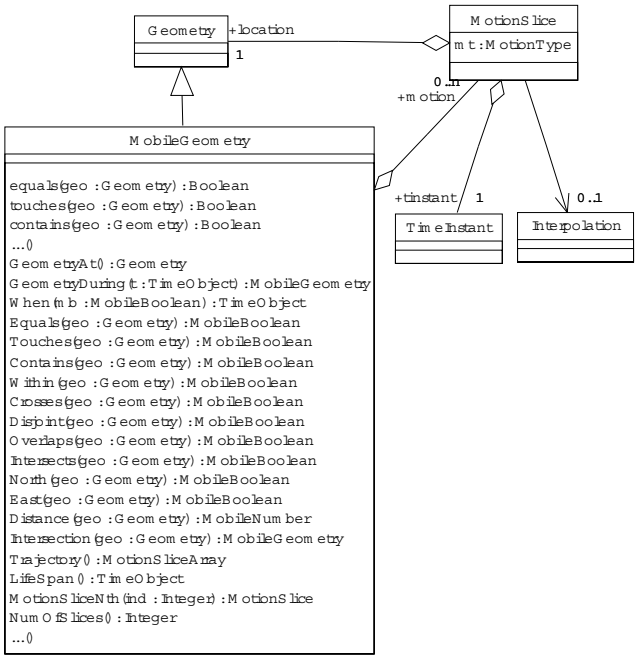


Fig. 2. The foundation of the ARGONAUT data model

The ARGONAUT data model overrides topological relations, direction relations, metric operations and spatial analysis operations inherited from the base *Geometry* class [4].

These relations and operations take mobile or non-mobile geometry argument and generate non-mobile results (Boolean, Double, Geometry, etc.). Specialized topological and direction relations, like *touches* and *contains*, have single argument of type *Geometry* (which could also be *MobileGeometry*) and return Boolean true value indicating that such relations are satisfied during the lifespan of the *MobileGeometry* argument(s) according to temporal aggregation defined in [2]. Such operations correspond to spatio-temporal predicates. Metric and spatial analysis overridden operations can not be considered as predicates, and for *MobileGeometry* argument operation is delegated by default to the geometry value at the current time instant defined using *GeometryAt(Now)* operation. The ARGONAUT data model also defines mobile variants of mentioned non-mobile relations and operations, named with starting capital letter in Fig. 2. Such operations correspond to the temporally lifted operations that handle non-mobile or mobile geometry arguments and generate mobile result, since for mobile argument(s) different results are generated in the course of time. The motion type associated with the mobile result depends on the operator or relation. In general, it is not sufficient to inherit a motion type of a mobile argument. For example, for mobile points characterized by linear motion type, a quadratic interpolation method is needed to represent the mobile result of distance operation.

To support the querying of mobile objects following operations are defined. The *Lifespan* operation returns time object defining the whole life span of the mobile object. The aforementioned *GeometryAt* operation returns geometry value at specific time instant. In order to restrict the sequence of motion slices of a mobile object according to specific time object, *GeometryDuring* operation is defined. That operation restricts the mobile object to only those motion slices from the sequence that belong to the specified time object given as an argument. The *When* operation returns the time object during which the mobile object satisfies the criteria specified by a mobile Boolean argument. The *Route* operation returns *Geometry* result representing the path traversed by the mobile geometry. All these operations are overridden in specific mobile classes inherited from the *MobileGeometry* class. To support manipulation of mobile properties of type *MobileBoolean* and *MobileDouble*, predicates and operations (*and, not, max, add*, etc) and their temporally lifted counterparts (*And, Not, Max, Add*, etc.), are defined accordingly [8].

Modeling mobile point objects that move continuously over a predefined network infrastructure, as existed in LBS, are provided by the *MobilePoint* class (Fig. 3). To determine the current location of a mobile point object, based on the last location update, the attribute *speed* is defined. If a mobile object does not follow the predefined route, the attribute *direction* must be defined accordingly. To specify a predefined route of the mobile point, the *MobilePoint* class is associated to the *Route* class which aggregates ordered set of *RoadSegment* class objects with OGC SF *LineString* geometry, and contains starting and destination points that lays on the first/last road segment respectively (Fig. 3). The *Route* class defines operations for specifying point along the route on specific distance from defined point (*PointOnDistance*) and the route distance between two points on the route (*RouteDistance*). The *GetRoute* method enables retrieval of *LineString* object, which represents the route geometry, while the *RoadSeg* operation retrieves the road segment along the route, which contains the specified point. The *MobilePoint* class inherits the *MobileGeometry* class and thus inherits and overrides aforementioned spatio-temporal operations and relations. The *Route* operation defined in *MobileGeometry* class is redefined in inherited classes, because for different mobile geometry types, the trajectory operation returns specific geometric result. Such operation applied to mobile line string object yields *Polygon* object as result, and for mobile point object with associated linear motion function returns a *LineString* geometric object.

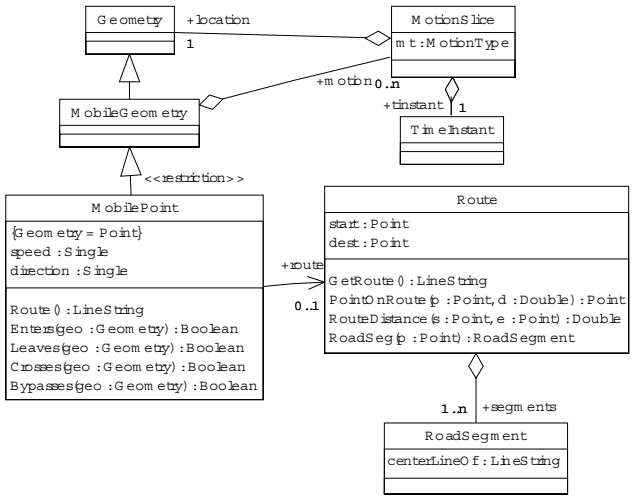


Fig. 3. Modeling mobile points on the transport network

The model defines relations arisen from the motion of mobile point over mobile or static polygonal area, such as *Enters, Leaves, Crosses, Bypasses*, etc. as proposed in [2] (Fig. 3). We define those relations for a mobile/static point and a polyline using the same semantics and definitions given in [2] for a mobile point and a region. Those operations are particularly useful in querying mobile points moving on network paths. Such relations can be defined by successive application of several basic mobile relations and enable examination of changing spatial relations over time by simply sequencing basic spatio-temporal predicates. Thus mobile object enters in the area of static or mobile polygonal object during given time period, if it was outside of the polygon at the beginning of the period (Disjoint relation), then at certain time instant was at the border of the polygon (touches relation) and is within the region to the rest of the time period (Within relation). Similar definitions hold for *Leaves* (inverse of *Enters*), *Crosses* and *Bypasses* predicates.

IV. Implementation in an Object-Relational DBMS

The implementation of proposed mobile object data model in an object-oriented application and a database is straightforward using defined UML class diagrams. The data model implementation in (object-) relational database domain is based

on definition of user-defined data types and operations within Object-Relational DBMS and appropriate mobile extension of OGC Simple Features for SQL specification [4]. If the LBS server is based on a relational DBMS, support for mobile object data management is integrated in LBS application components. User-defined data types and user-defined functions are defined in terms of the SQL DDL statement CREATE TYPE [6] as follows:

```
CREATE TYPE MotionSlice AS OBJECT
(slice#  NUMBER, geometry Point, timeinst DATE, mt MotionType);

CREATE TYPE Motion AS TABLE MotionSlice;

CREATE TYPE MobilePoint AS OBJECT
(GID  NUMBER, speed NUMBER, direction NUMBER, motion
Motion,
MEMBER FUNCTION GeometryAt(tTimeInstant) RETURN Point
… );

CREATE TYPE Ambulance AS OBJECT
(ID NUMBER, name CHAR(64), driver CHAR(64), type CHAR(256),
locations MobilePoint);

CREATE TYPE Taxicab AS OBJECT
(ID NUMBER, company CHAR(64), driver CHAR(64), type
CHAR(256), locations MobilePoint);

CREATE TYPE Street AS OBJECT
(ID NUMBER, name CHAR(64), centerLineOfLineString);
```

The operations defined for those types are implemented as user defined functions that can be applied to user defined data types. The query facility of SQL is provided by the well-known SELECT-FROM-WHERE clause. User defined operations on user defined data types can be also included as predicates and functions within SELECT and WHERE clause and embedded in an SQL statement. Thus, mobile spatio-temporal queries can be specified and processed, like:

1. Select ambulances that are within 2 km around of my current address:

```
select amb.id, amb.name, amb.driver
from Ambulance amb
where
(amb.locations.GeometryAt(NOW)).Distance(Point(xref,yref))  <=
2000
```

We assume that the LBS application provide geocoding possibilities that convert location expressed as address in point value with exact coordinates xref and yref. Thus *GeometryAt* operation returns the location of a mobile point at the specified time instant (NOW defines the current time). Operation *Distance* returns the distance between point values and the whole predicate in the WHERE clause specifies only those ambulances whose determined distance is less then 2000 meters.

2. Return the length of path of truck "BioExport-1" in kilometers and time:

```
select (t.troute.Trajectory()).length(), (t.troute.Lifespan()).duration()
from Truck t
where t.name="BioExport-1"
```

Trajectory operation returns *LineString* object and the *length* operation calculates length of that line string. Similarly, *Lifespan* operation returns *TimeObject* object, and *duration* operation applied on *TimeObject* argument returns its duration.

3. Return the position of a taxicab "Banker-17" if it enters the "Cara Dusana" street in last 5 minutes:

```
select t.locations.GeometryAt(NOW)
from Taxicab t, Streets
where t.name = "Banker-17" and s.name="Cara Dusana" and
    (t.locations.GeometryDuring(TimePeriod(NOW -
300,NOW))).Enters(s.centerLineOf)
```

Operation *Enters* is applied to the *MobilePoint* object whose motion is restricted by time period value which correspond to the "last 5 minutes" expression, and the *LineString* object representing the center line property of the street object.

V. Conclusion

The main contribution of this paper is the object-oriented data model and SQL extension for representation and querying mobile objects. It represents the foundation of the ARGONAUT component framework for development of location-based services that involve mobile objects. Such framework can be easily and effectively integrated with any OGC-compliant object-relational DBMS to provide mobile object data management and querying capabilities. We are currently developing a prototype application for tourist and business guiding based on proposed data model and the ARGONAUT component framework implemented over Web architecture.

Acknowledgement

The research was partially supported by the project "Geographic Information System for Local Authorities based on Internet/WWW Technologies", funded by Ministry of Science, Technology and Development, Republic of Serbia, and Municipality of Nis, Contract No. IT.1.23.0249A.

References

[1] T. Brinkhoff, "A Framework for Generating Network-Based Moving Objects", *GeoInformatica Journal*, Vol. 6, No. 2, June 2002, pp. 153-180.

[2] M. Erwig and M. Schneider, "Spatio-Temporal Predicates", *IEEE Trans. On Knowledge and Data Engineering*, Vol 14, No. 4, July/Aug. 2002, pp. 881-901.

[3] R. H. Gueting, M. H. Boehlen, M. Erwig, C. S. Jensen, N. A. Lorentzos, M. Schneider and M. Vazirgiannis, "A Foundation for Representing and Querying Moving Objects", *ACM-Transactions on Database Systems Journal*, Vol. 25, No. 1, 2000, pp. 1-42.

[4] ISO/ TC 211 Geographic Information/Geomatics, ISO 19125 - Geographic information - Simple feature access. Oct. 2000.

[5] ISO/ TC 211 Geographic Information/Geomatics, ISO 19108 - Temporal Schema. Oct 2000.

[6] Oracle 9i Enterprise Edition, Oracle Documentation Library, Oracle Corporation, http://technet.oracle.com, 2002.

[7] D. Stojanović and S. Djordjević-Kajan, "Location-based Web service for tracking and visual route analysis of mobile objects", *Proc. of Yu INFO 2002*, Kopaonik, 2002.

[8] D. Stojanović and S. Djordjević-Kajan, "Modeling Mobile Point Objects in Location-based Services", *Int. Journal Facta Universitatis: Mathematic & Informatics*, accepted for publication, 2003.

[9] O. Wolfson, "Moving Objects Information Management: The Database Challenge", 5th Workshop on Next Generation Information Technologies and Systems (NGITS 2002) Israel, 2002.

Retrieval by Spatial Similarity in Image Databases

Mariana Stoeva[1], Slava Jordanova[2]

Abstract – **This paper presents our work in the area of image retrieval in Image Databases for images saved by spatial similarity of domain objects location. We propose a geometric based structure which makes possible the extraction of directional spatial relations among domain objects that are invariant with respect to transformations. We introduce an algorithm for spatial similarity retrieval in Image Database.**

Keywords – **Image databases, Query by index content, Spatial reasoning, Similarity retrieval**

I. Introduction

Similarity based retrieval of images is an important task in many image database application. A major class of users' requests is the one that requires retrieval of those images in databases that are spatially similar to the query images. A survey of existing approaches to the problem of spatial image retrieval and their limitations is described in [9,16]. The Query by Pictorial Example (QPE) [4] philosophy expresses the objects and the spatial relations to be retrieved through a symbolic image which serves as a query and which is matched against the images in the database. Then, a query is an iconic image itself, represented by the same method used to describe an iconic index. As a result, a variety of approaches have been proposed witch use objects and spatial relationships to describe the visual content of an image.

In real-world database applications, the rotation invariance is a basic issue because each image is captured and stored in agreement with a viewpoint which is implicitly dependent on an outside viewer who establishes a fixed scanning direction. Any translation of this viewpoint and/or any rotation of the image affects the direction relations between each pair of objects. In the recent literature, several approaches can be found whose aim is to provide a solution to the rotation invariance of the conventional indexing methodologies based of symbolic projections [11,12,9,16]. However, in real application, it would also to be able to find the images in the database that present a given pattern, even if it appears mirror reflected.

In this paper we propose a geometric based structure which makes possible the extraction of spatial relations among domain objects that are invariant with respect to transformation such as translation, rotation, scaling, reflection, view point change as well as arbitrary compositions of these transformations. We introduce an algorithm for spatial similarity retrieval I Image Database, presented as SIMR$_{R\theta}$ that recognizes transformed images and sub-images.

[1]Mariana Stoeva is with the Technical University Varna, Department of Computer Systems and Technologies, Bulgaria, E-mail: mariana_stoeva@hotmail.com

[2]Slava Jordanova is with the Technical University Varna, Department of Computer Systems and Technologies, Bulgaria, Email: sl@windmail.net

The paper is organized as follows: In Section 2 we present the symbol description used by us as well as the approximations used for computing the spatial relationships among objects. In Section 3, for the purpose of similarity, we introduce descriptions of spatial relations, of spatial relation similarities, as well as similarity measure between two images and also an algorithm presented as SIMR$_{R\theta}$ that recognizes transformed images and sub-images. We publish a part of the experiments made and their results in Section 4. In Section 5 some conclusions are shown and further work aims are planned.

II. Image Description and Approximations Used

In most applications the objects in the images have not an exactly fixed shape. The performance of any spatial investigation on the exact location and shape of each object domain is very expensive. Due to this an initial approximation or filtration is used. The most commonly used objects shape approximation to the purpose of image retrieval in Image Databases (ID) is the Minimal Boundary Rectangle (MBR – the smallest rectangle that bounds the object). Whatever the approximation approach is, it cannot influence the method of indexing in ID [15].

The use of MBR for shape representation allows obtaining of spatial relations that are invariant with respect to translation and scaling, but variable with respect to rotation and reflection transformations.

In our work we are searching such an approximate objects shape representation that would be invariant with respect to arbitrary compositions of transformations and at the same time would store the information that is necessary for accounting the spatial relations among extended objects in an image.

We assume that an image I consists of n domain objects denoted as O_j. The symbolic description of image I is stored in Image Database. Let C_j is an object centroid and C_I is the centroid of the image obtained from the centroids of the objects contained in the image. The description presents an image as $I = ((O_j, (O_j.x_{cj}, O_j.y_{cj}), ((O_j.x_{jl}, O_j.y_{jl}), 1 \leq l \leq 4), 1 \leq j \leq n)$, where O_j is a name of the object, $(O_j.x_{cj}, O_j.y_{cj})$ are coordinates of the centroid C_j of object O_j, $((O_j.x_{jl}, O_j.y_{jl}), 1 \leq l \leq 4)$ are the coordinates of the typical for the object shape coordinates of four points from the object external contour in Cartesian coordinate system with initial point C_I. The typical points are vertices of an approximating object shape tetragon. We assume that the domain objects are stored in lexicographic /alphabetic/ order on the object names. The notation O_I is used to refer to a set of objects in image I.

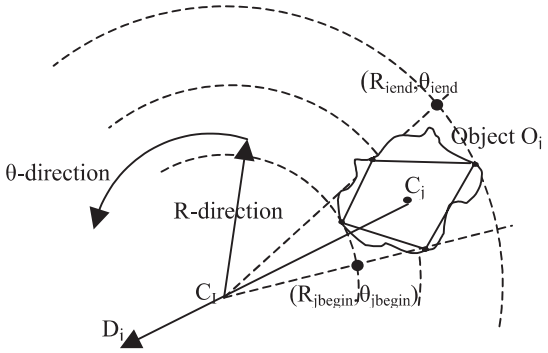


Fig. 1. Illustration of the approximations used

The domain objects of each image are stored in ID by their symbol names, centroid coordinates, and four characterizing the object shape points, whose determination is invariant with reference to transformations. The determination of the four typical for the object shape points used in approximate representation of each object is submitted to the consistent execution of the next requirements in the corresponding priority order" to be independent of transformations, to represent the shape in the form of tetragon whose centroid is identical with the object centroid, and whose area is maximally close to that of the object.

The implementation of these requirements in the typical points determination imposes the use of criteria treating the point distance from the external contour to the object centroid C_j, the length of the maximal segments determined by the cross-section of the external contour by axes passing through the centroid C_j, as well as the angle between these axes. These criteria are presented in details in [17].

The extended object O_j shape approximation built on stored information for each object has the view of a ring sector obtained from concentric circles around image C_I centroid. The ring boundaries are determined by the minima and maxima of the polar coordinates of the points from the tetragon determined by the stored four typical object points $((O_j.x_{jl}, O_j.y_{jl}),\ 1 \leq l \leq 4)$. The ring sector initial point has polar coordinates $(R_{\text{begin}}, \theta_{\text{begin}})$ and the final point has polar coordinates. $(R_{\text{end}}, \theta_{\text{end}})$. The description of object O_j takes the form $(O_j, (R_{cj}, \theta_{cj}),\ (R_{j\text{begin}}, \theta_{j\text{begin}}),\ (R_{j\text{end}}, \theta_{j\text{end}}),\ 1 \leq j \leq n))$, where (R_{cj}, θ_{cj}) are the polar coordinates of the centroid C_j, $(R_{j\text{begin}}, \theta_{j\text{begin}})$, $(R_{j\text{end}}, \theta_{j\text{end}})$ – the origin and the end of the object approximating ring sector. An example of ring sector approximating object O_j shape and described by its four typical points is depicted in Fig. 1. Objects centroids polar coordinates and the polar coordinates of the initial and the final points of the approximating them ring sectors are used for obtaining the spatial relations of each object with the other image objects.

III. Spatial Similarity Image Retrieval in Image Database

Following the QPE philosophy the query is processed by matching the obtained from its description spatial relations with these of the stored in the database images. Investigations concerning similarity retrieval methodologies can be found in [3,10]. We use a directional approach to determine the spatial relationships between each match of domain objects in an image. We use two directions corresponding to the directions in a polar coordinate system whose centre is the image centroid. The linear scanning R-direction begins from the image centroid C_I and corresponds to the concentric circles going from the origin to outside. The circle θ-direction corresponds to a trace swiping anticlockwise around the origin. The directions are shown in Fig. 1. Allen's well-known 13 types of spatial relationships [1] whose defining is adapted to a polar coordinate system are obtained in determining the relationships between each object match in the sense of conditions for their initial and final points in each separate direction. In both directions a total of 169 spatial relationships arise between two objects in two dimensions through which the spatial content of an image can be suitably represented. The spatial relations among the image objects are computed by using the polar coordinates of the initial $(R_{\text{begin}}, \theta_{\text{begin}})$ and the final $(R_{\text{end}}, \theta_{\text{end}})$ points of object approximating ring segments whose centre is the object centroid.

Definition 1. A triple like $(O_j\, \gamma\, O_i)$ is called an atomic spatial relation, where O_j, $O_i \in O_I$ are object names and $\gamma \in \{<, >, |, ", =, [, (,],), /, \setminus, \%, \#\}$ is an operator.

We use the notation $(O_j\, \gamma_R\, O_i)$ to indicate that the pair of objects (O_j, O_i) belongs to relation γ in R-direction and $(O_j\, \gamma_\theta\, O_i)$ to indicate that the pair of objects (O_j, O_i) belongs to relation γ in θ-direction.

The defining of binary spatial relations is identical to those in [1] adapted to a polar system and it is presented in [16]. The 7 known symbols for spatial operators used by Lee and Hsu [13,14] are utilized and four new symbols are introduced. In the linear direction R the analogy of spatial relations defining with the one used in orthogonal model is complete due to the linear type of the direction. To achieve invariance in determining the spatial relations for each object match with reference to the transformations rotation and reflection in circle direction θ it is necessary the angle coordinates to be recomputed. This is necessary due to the existing peculiarity of the circle direction (its relative beginning and end are identical). The re-computation stores the proximity of points and corresponds to the real ordering of the initial and final object points in θ-direction.

We provide the independence of the spatial relations in θ-direction of the leading object O_j and the other objects O_i by re-computing their initial coordinates $(\theta_{i\text{begin}})$ with reference to an origin point D_j. The point D_j is at a distance of π from the value θ_{cj}. For object O_j the relative circle origin in θ-direction is point D_j $(\theta_{cj} \pm \pi)$. Thus the coordinate θ_{cj} of object O_j centroid remains always in the relative middle of the numerical axis. The coordinates of the end points $(\theta_{i\text{end}})$ are computed with reference to the recomputed origin and the stored distance between the initial $(\theta_{i\text{begin}})$ and the final $(\theta_{i\text{end}})$ coordinates. The relative origin for computing the spatial relations of object O_j in circle direction θ in the example at Fig. 1 is denoted by D_j. We hold without formal proof, that with thus recomputed initial and final points of the ring sector that describes the domain object, the relations

remain invariant with reference to rotation transformation. In case of reflection transformation the relations in θ-direction may be easily obtained by changing each operator with corresponding inverse operator.

Since the query processing goal is to retrieve those images from the ID that contain almost the same objects as contained in the query, and have similar spatial ordering. Our similarity measure has to take into account the similarities between the spatial operators for each relation.

Definition 2. Let the objects O_j and O_i belong to image $Q(O_j, O_i \in O_Q)$ and to image $I(O_j, O_i \in O_I)$. Let the object match $((O_j, O_i))$ in image Q belongs to the spatial relation $\gamma(O_j\, \gamma\, O_i)$ and the object match (O_j, O_i) in image I belongs to the spatial relation $\gamma'(O_j\, \gamma'\, O_i)$. We define the similarity between the spatial relations in both images Q and I for the object match (O_j, O_i) as $\text{sim}_{ji}(\gamma, \gamma') = t$, where $t \in [0, 1]$, is the similarity between the spatial operators γ and γ'. We denote the similarity between the relations in both the images Q and I for the object match (O_j, O_i) in R-direction as $\text{sim}_{ji}(\gamma_R, \gamma'_R)$ and as $\text{sim}_{ji}(\gamma_\theta, \gamma'_\theta)$ in θ-direction.

We adopt the interval neighborhood graph as in [16], that defines the distances among spatial operators. According the definition of the interval neighborhood graph given by Freksa in [7], two projection relationships are neighbors if they can be transformed into one another by continuous deforming the projections. The similarity values $t = \text{sim}(\gamma, \gamma')$ between each operator match we use, are shown in [16]. In case of multiple instances of one object, the greatest value of similarity $\max\{\text{sim}_{ji}(\gamma, \gamma')\}$ between the operators that describe it is used. To this aim we introduce a formula that measures the similarity degree between an image and a query in terms of objects and spatial relationships and expresses it as a value in the range [0,2]. According our understanding for similarity and our desire the method to be independent of human interpretation, we put forward a formula for similarity evaluation that assesses the similarity between the common for both the images objects and their corresponding atomic relations.

Definition 3. Let the query image be Q and the Image Database image is I. We define the similarity distance between Q and I by equation (1),

$$\text{sim}(Q, I) = \frac{1}{m}\left(n + \frac{1}{2n}\sum_{j}^{n}\sum_{i}^{n}\text{sim}_{ji}(\gamma_R, \gamma'_R) + \text{sim}_{ji}(\gamma_\theta, \gamma'_\theta)\right) \quad (1)$$

where $m = ||O_Q||$ is the number of objects in the query image, $n = ||O_Q \cap O_I||$ is the number of the common for both the images objects, $\text{sim}_{ji}(\gamma_R, \gamma'_R)$ is the spatial similarity between the images Q and I for the object match (O_j, O_i) in R-direction, and $\text{sim}_{ji}(\gamma_\theta, \gamma'_\theta)$ is the spatial similarity between the images Q and I for the object match (O_j, O_i) in θ-direction.

An image I from the Database answers the query Q only if it contains all the query objects with the same relations as in image Q. The more objects from Q are contained in image I, the higher the similarity degree. This formula returns a value of the order of [0,2] by computing the correlation between the number of objects that are retrieved together with the similarity value of their atomic relations and the number of objects demanded by the user and their atomic relations. We expect a value of 2 when the input images are identical, otherwise the function will take a lower values. These values are proportional to the degree of disagreement in the spatial relationships between the corresponding objects in the input images.

We present a spatial similarity algorithm, that not only recognizes transformation variants of an image but also recognizes sub-images (or transformation variants of sub-images) of the query image in the Image Database.

◇ Q and I are the query image and the Database image.

Algorithm $\text{SIM}_{R\theta}(Q = ((O_j, (O_j.x_{cj}, O_j.y_{cj}), ((O_j.x_{jl}, O_j.y_{jl}), 1 \le l \le 4)), 1 \le j \le m),\ I = ((O_j, (O_j.x_{cj}, O_j.y_{cj}), ((O_j.x_{jl}, O_j.y_{jl}), 1 \le l \le 4)), 1 \le j \le v))$

1 $Q' \leftarrow ((Q.O_j, (Q.O_j.x_{cj}, Q.O_j.y_{cj}), ((Q.O_j.x_{jl}, Q.O_j.y_{jl}), 1 \le l \le 4)), 1 \le j \le m) | O_j \in O_Q \cap O_I$
2 $I' \leftarrow ((I.O_j, (I.O_j.x_{cj}, I.O_j.y_{cj}), ((I.O_j.x_{jl}, I.O_j.y_{jl}), 1 \le l \le 4)), 1 \le j \le m) | O_j \in O_Q \cup O_I$
3 $n \leftarrow |O_{Q'}| = |O_{I'}|$
4 $m \leftarrow |O_Q|$
5 $Q'_{R\theta} \leftarrow ((O_j, (R_{cj}, \theta_{cj}), (R_{j\text{begin}}, \theta_{j\text{begin}}), (R_{j\text{end}}, \theta_{j\text{end}}), 1 \le j \le n)$
6 $I'_{R\theta} \leftarrow ((O_j, (R_{cj}, \theta_{cj}), (R_{j\text{begin}}, \theta_{j\text{begin}}), (R_{j\text{end}}, \theta_{j\text{end}}), 1 \le j \le n)$
7 sim $\leftarrow 0$
8 **for** $j \leftarrow 1$ **to** n **do**

◇ Accumulating object factor

9 sim $\leftarrow$ sim $+ 1$

◇ Accumulating spatial factor

10 **for** $i \leftarrow 1$ **to** n **do**

◇ The atomic spatial relations between objects O_j and O_i in image Q'

11 $O_j \gamma_R O_i \leftarrow (Q'_{R\theta})$
12 $O_i \gamma_\theta O_j \leftarrow (Q'_{R\theta})$

◇ The atomic spatial relations between objects O_j and O_i in image I'

13 $O_j \gamma'_R O_i \leftarrow (I'_{R\theta})$
14 $O_i \gamma\_\theta O_j \leftarrow (I'_{R\theta})$

◇ Defining the spatial similarity values between the atomic relations of images Q' and I' for the object match (O_j, O_i) in R-direction and θ-direction

15 $\text{sim}_{ji}(\gamma_R, \gamma'_R) \leftarrow ((O_j \gamma_R O_i), (O_j \gamma'_R O_i))$
16 $\text{sim}_{ji}(\gamma_\theta, \gamma'_\theta) \leftarrow ((O_j \gamma_\theta O_i), (O_j \gamma'_\theta O_i))$
17 sim $\leftarrow$ sim $+ (\text{sim}_{ji}(\gamma_R, \gamma'_R) + \text{sim}_{ji}(\gamma_\theta, \gamma'_\theta))/(2n)$
endfor
endfor
18 sim $\leftarrow$ sim$/m$
return sim
end $\text{SIM}_{R\theta}$

Fig. 2. Spatial similarity algorithm

The proposed algorithm $\text{SIM}_{R\theta}$ is further described. The formal definition of $\text{SIM}_{R\theta} : Q \times I \rightarrow [0, 2]$, where Q and I are the symbol presentations of the query and the Database images, respectively. The algorithm is shown in Fig. 2. Lines beginning with the symbol ◇ indicate comments. Certain lines are numbered for comment convenience.

The algorithm has computational complexity n^2, where n is the number of the common for Q' and I' objects. The computational complexity is determined by the time necessary

for the performance of the body of the imbedded for loops (lines $8 \div 17$). This complexity is the same as the complexity of algorithms that take in account orthogonal relations of domain objects [16].

IV. Experiments

For the purpose of the experiment we use a test image collection of 10 original images and 9 variants of each of these images. Variants 1-6 are generated by applying transformation operators to all domain objects, while for variants 8-9 the operators are applied to a subset of domain objects. The transformation operators used in producing the variants of the original images include compositions of the transformations (translation, scaling, rotation, reflection) with different length.

The experimental evaluation is done in two parts. In the first part the focus is on understanding the correspondence between the similarity computed by $\text{SIM}_{R\theta}$ and that to the intuitively expected. In the second part the robust behavior of $\text{SIM}_{R\theta}$ algorithm in combined transformations is studied. The idea of using expert-provided rank ordering of images, elative to each query in given set of test queries, in quantifying the retrieval quality is presented in [8] and [9]. This quantification is based on a measure referred to as the R_{norm} [8].

Fig. 3. . Test images for intuitive understanding of SIMR? similarity computation

To establish whether $\text{SIM}_{R\theta}$ rank ordering of Database images with respect to a query image agrees with our intuitive visual rank ordering, we consider the images shown in Fig. 3. We consider each of these images as the query and compute its similarity with all the other images by $\text{SIM}_{R\theta}$ and by an expert. The $\text{SIM}_{R\theta}$ rank ordering confirms to the intuitively expected rank ordering. The research for robust behavior includes two sets consisting of one original image and its all 9 variants. Then we consider again each image in the set as the query. The experiment shows that the $\text{SIM}_{R\theta}$ algorithm performs our expectations.

V. Conclusions

This paper proposes a geometric based structure which makes possible the extraction of directional spatial relations among domain objects that are invariant with respect to transformations. The presented spatial similarity retrieval distance and algorithm solve a class of tasks namely contain-based spatial similarity research. The similarity distance and the algorithm are not influenced by possible image transformations of the query and database, moreover, they catch also transformed sub-images. The algorithm temporary complexity is n^2, where n is the number of objects that are common for both the query and database images. The algorithm is robust in the sense that it can recognize translation, scale, and rotation variant images and the variants generated by an arbitrary composition of these three geometric transformations. The effectiveness and efficiency of the spatial similarity retrieval algorithm are evaluated by using an expert-provided rank ordering of a test collection with respect to a set of test queries. Our efforts for the future point to improving the algorithm effectiveness and reducing the error inserted by the used approximations.

References

[1] J. F Allen, “Maintaining Knowledge about temporal Intervals”, Comm. ACM, vol. 26 (1983) 832-843

[2] C.C. Chang, C. F. Lee, “Retrieve Coordinates Oriented Symbolic String for Spatial Relationship Retrieval”, Pattern Recognition, vol. 28 no.4 pp. 563-570 1995

[3] C.C. Chang, S.Y. Lee, “Retrieval of Similar Pictures in Pictorial Database”, Pattern Recognition, vol. 7 (1991) 675-680

[4] N. S. Chang, K.S. Fu, “ Query-by-pictorial-example”, IEEE Trans. Soft. Eng., vol. 6 (1980) 519-524

[5] S.K. Chang, E. Yungert, “Pictorial Data Management based upon the Theory of symbolic Projection”, J. Visual Language and Computing, vol. 2 no3 pp.195-215 1991

[6] N.S. Chang, K.S Fu, “Query-by-pictorial-example”, IEEE Trans. Soft. Eng., vol. 6 pp.519-524 1980

[7] C. Freksa, “Temporal Reasoning Based on Semi-Intervals”, Artificial Intelligence, vol. 54 no.1-2 pp.199-227 1992

[8] V.Gudivada, V. Raghavan “Design and Evaluation of Algorithms for Image Retrieval by Spatial Similarity”, ACM Trans. Information Systems, vol. 13 no.1 pp.115-144 1995

[9] V. Gudivada, “$\theta\Re$-string: A geometry-based representation for efficient and Effective Retrieval of Images by Spatial Similarity”, IEEE Trans. KDE, vol. 10, no.3, pp.504-511, 1998

[10] P.W. Huang, Y.R. Jean, “Using 2d C+-Strings as Spatial Knowledge Representation for Image Database Systems”, Pattern Recognition” vol. 27 no.9pp. 1249-1257 1994

[11] E. Jungert, “Rotation Invariance in Symbolic Projection as a Means for Determination of Binary Object Relations”, Proc. Workshop (QUARDET ’93) pp.503-512 1993

[12] E. Jungert, “ Rotation Invariance in Symbolic Slope Projection as a Means to Support Spatial Reasoning”, Proc. Third Int’l Conf. Automation, Robotics and Computer Vision (ICARCV’95) pp.364-368 1994

[13] S.Y. Lee, F.J. Hsu, “2D C-string: A new Spatial Knowledge Representation for Image Database System”, Pattern Recognition, vol. 23 no.10 pp.1077-1088 1990

[14] S.Y. Lee, F.J. Hsu, “ Pictorial Algebra for Spatial Reasoning of Iconic Images Represented 2D C-string”, Pattern Recognition, vol. 12 pp.425-435 1991

[15] B. C. Ooi, “ Efficient Query Processing in Geographic Information Systems”, Edited by G. Goos, G, Hartmanis, J. (eds.) Springer-Velag 1990

[16] G. Petraglia,, M. Sebillo , G. Tucci G. Tortora, “Virtual Images for Similarity Retrieval in Image Databases”, IEEE Trans. on knowledge and data engineering vol. 13 no.6 pp.951-967 2001

[17] M. Stoeva, Sl. Jordanova, “ Similarity Shape Retrieval from Image, Database Proc. XXXVII International scientific conference on information, communication and energy systems and technologies 1 (ICEST’ 2002) pp.283-287 2002

Application of Distributed Search in Databases for Web Services

Elena Ivanova[1]

Abstract – **Distributed search in WAN environment is considered Databases, located in different places and hosts are targeted by web applications. Standardisation issues are under considerations.**

Keywords – **Database, Internet, SOAP, PHP,MySQL**

I. Introduction

Communication and information exchange is the need of the today's world of extreme competition on the business front. This need for information exchange brings in another need to make this information selectively visible, and its visibility to be changed on-the-fly.

In reply to these problems occurs the idea of Web Services. It is not, of course, the first solution to the problem. RMI (Remote Method Invocation), CORBA (Common Object Request Broker Architecture) and others also address the same problem space.

In general, Web services are modular applications or functions, which are generally independent and self-describing, that can be discovered and called across the Internet or an enterprise intranet [2].

Web Services is based on the already existing and well-known HTTP protocol, and uses XML as the base language. This makes it a very developer-friendly service system. However, most of the other technologies such as RMI and CORBA involve a whole learning curve.

II. Classification

Web Service implementations can be organized into the following three basic models [2]:

1. Enterprise Application Integration – it is the use of Web Services within a single organization. According to this first level, the firm may develop libraries of reusable programming building blocks to speed application development. Such Web Services may be components of company-specific program logic or commonly used application functions, such as currency conversions or date calculation (one-tier model-see Fig. 1).

2. Single Partner Integration – the next level extends application integration reach beyond the enterprise. Web services are shared between organizations that likely have formal partnerships. Components of core business applications are exposed as Web services and shared, which facilitates inter-organizational collaboration.

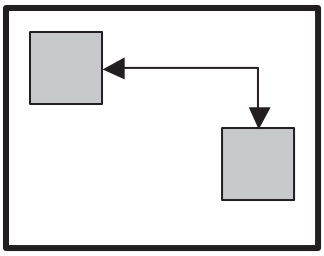


Fig. 1. One-tier model

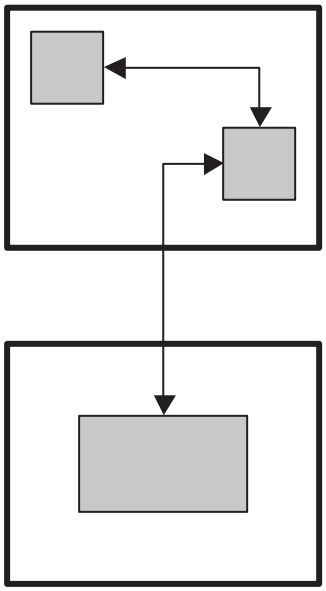


Fig. 2. Two-tiers model

Applications may be constructed using multiple Web services, from various sources, which work together regardless of where they actually reside or how they were implemented (two-tiers model-see Fig. 2).

3. Multiple Partner Integration-the third level. It is an advanced evolution of the previous model and requires the most complex levels of application collaboration. Application integration is extended to and coordinated with multiple business partners. This may entail integrating simple information sources, such as weather forecasts, sports scores, horoscopes, or complex, critical capabilities, such as credit card verification or user authentication services. The web services themselves may be exposed between trusted business partners or discovered in services directories (three-tiers model-see Fig. 3).

[1]Elena Ivanova is with Institute of Computer and Communication Systems - Bulgarian Academy of Sciences, Acad. G. Bonchev bl.2, 1113 Sofia, Bulgaria, e-mail: e_ivanova@hsh.iccs.bas.bg

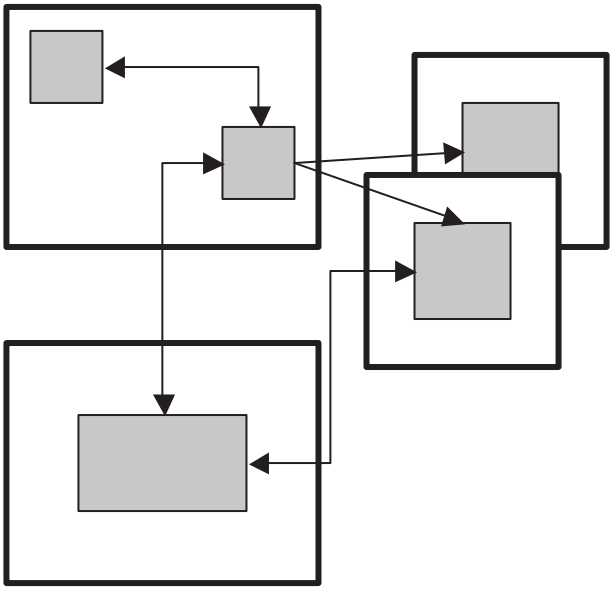


Fig. 3. Three-tiers model

III. Architecture and technologies

The applications implement Web services are based on service-oriented architecture. This architecture includes[1]:

1 a standard way for communication
2 a uniform data representation and exchange mechanism
3 a standard meta language to describe the services offered
4 a mechanism to register and locate Web Services-based applications (Fig. 4).

The mapping between the different layers of the Web Services architecture and the technologies used is listed in Table 1.

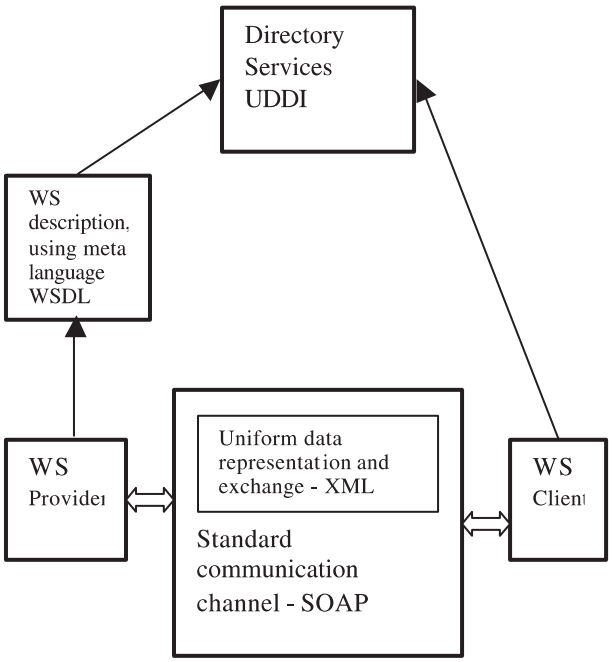


Fig. 4. Architecture of Web Services

XML is used in the Web Services architecture as the format for transferring information/data between a Web Services provider application and a Web Services client application.

Extended Markup Language (XML) is a meta language that has a well-defined syntax and semantics. The syntax and semantics "self describing" features of XML make it a simple, yet powerful, mechanism for capturing and exchanging data between different applications

Table 1.

Layer	Technology
Uniform data representation and exchange	XML
Standard communication channel	SOAP
Standard meta language to describe the services offered	WSDL
Registering and locating Web Services	UDDI

The Simple Object Access Protocol (SOAP) is the channel used for communication between a Web Services provider application and a client application. The simplicity of SOAP is that it does not define any new transport protocol; instead, it re-uses the Hyper Text Transfer Protocol (HTTP) for transporting data as messages. SOAP is the method for sending messages across different modules. This is similar to how someone communicates with the search engine that contains an index with the Web sites registered in the index associated with the keywords.

Web Services provider applications advertise the different services they provide using a standard meta language called the Web Services Description Language (WSDL). WSDL is based on XML and uses a special set of tags to describe a provided Web service and where to locate it. Client applications obtain information about a Web service prior to accessing and using a Web service of a Web Service provider. WSDL is the method through which different services are described in the UDDI. This maps to the actual search engine in our example.

Web Services application providers are listed in a registry of service providers using UDDI. Similarly, client applications locate Web Services application providers using UDDI. Like in the case of WSDL, UDDI also is based on XML. This is analogous to the index service for the search engine,

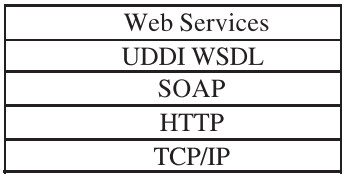


Fig. 5.

in which all the Web sites register themselves associated with their keywords.

These technologies form The Web Services Protocol Stack (Fig. 5).

IV. Application

A. Description of the Database

Our distributed database, named LearningDB, consist of information about students, teachers and other academic data in one university. The tables of the database can be stored on different machines and can be managed from different database servers. We use MySQL Server and ODBC as a server to Access' tables. The used tables are listed bellow.

Field	Type
usr	Varchar(40)
pass	Varchar(40)

Fig. 6. Table admin

Field	Type
Fid	Int(10)
FName	Varchar(40)
Phone	Varchar(40)

Fig. 7. Table faculty

Field	Type
Sid	Int(10)
LastName	Varchar(40)
FirstName	Varchar(40)
Fid	Int(10)

Fig. 8. Table students

Field	Type
Tid	Int(10)
LastName	Varchar(40)
FirstName	Varchar(40)
Subject	Varchar(40)
Fid	Int(10)

Fig. 9. Table teachers

B. Structure of the Application

The main goal of our application is to obtain a selective search in the database. This functionality is implemented by software decision, based on the SOAP protocol and realized by PHP programming language. The structure of the application is shown on Fig. 10.

SoapServer, the program that client communicates with by using SOAP protocol, is located on the machine (called end point), where the relevant database is stored. The program-client and the program SoapServer is using two base classes, named class.soap_client and class.soap_server, developed by

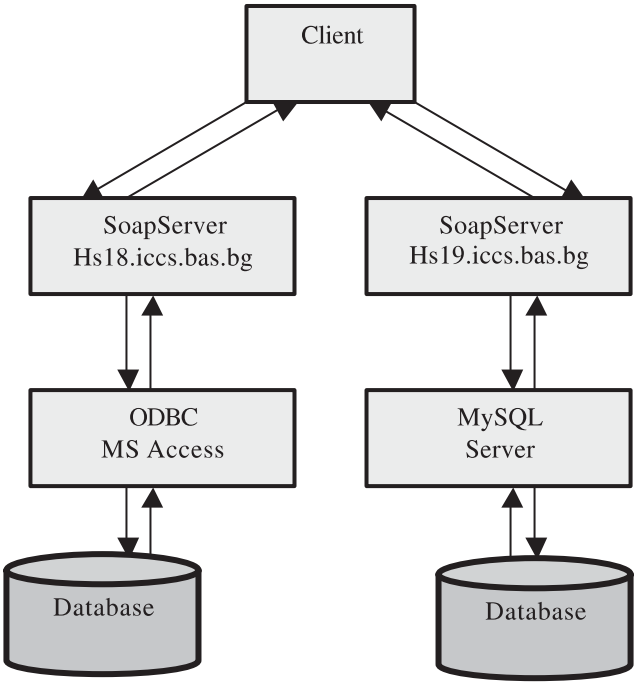


Fig. 10. Structure of the application

Nusphere corp. [1]. The other used class DBFunctions describes the connection to Database Server and the manipulations on the data, included in Database. Because of the two kinds of databases-MySQL and MS Access, two classes DBFunctions are presented according to them. In SoapServer program two functions - "List" and "Search", are developed as web services. Apart from this, the user must have a valid username and password in order to use these services. The table *admin* comprise authentication information.

C. Software components of the application

1 Apache web server (http://www.apache.org) is an open-source HTTP server, developed by Apache Software Foundation, for various operating system, such as UNIX and WINDOWS.

2 MySQL Server (http://www.mysql.org) is an open-source database server, developed, distributed and supported by MySQL AB.

3 ODBC (Object DataBase Connectivity) is a programming interface that enables applications to access data in database management systems that use SQL as a data access standard. In this case the database is associated with DSN-Data Source Name.

4 PHP (http://www.php.net), a widely-used Open Source server-side scripting language, is especially suited for Web development and can be embedded into HTML.

5 The admin.php script, that is used for authentication of the users. The source code bellow is the base part of the program.

```
    // Connect to the Database
    if  (!($link=mysql_pconnect("localhost","usr",  "pass")))
{
        DisplayErrMsg()) ;
        exit() ;}
    // Select the Database
    if (!mysql_select_db("LearningDB", $link)) {
        DisplayErrMsg(sprintf()) ;
            exit() ;}
    $query="select * from admin where usr='$username' ";
    //Execute the query
    $result=mysql_query($query);
    $line=mysql_fetch_array($result);
    while ($line[pass]!=$password){
          DisplayErrMsg("Invalid input data")) ; exit();}
```

6 Two php classes soap_client and soap_server, developed by Nusphere corp. (http://www.nusphere.com) .

7 The main client script is search_soap.php, by whish the user can obtain the selective search into the database. The base code of the program is listed bellow.

```
    //Pass the relevant end-point
    $SoapServerURL="http://".$URL[j]."/Server/
SOAPServer.php";
    // instantiate soap client object
    $soapclient = new soapclient("$SoapServerURL");
    //Invoke the service
    if        ($hits        =        $soapclient->call("search",
array("pattern"=>$pattern),   "urn:nusphere-web-services"))
{ foreach($hits as $data){
       $result                .=                "$data[FirstName]
$data[LastName]br>";}}
    else {
        // report error
         print "error:<br>";}
```

8 The SoapServer program, which is used for communication with the client on one hand, and with the database server on the other hand.

The application, named LearningPro, can be reach at http://hs18.iccs.bas.bg .

```
    //Call the add_to_map() method for each "service"
(function)
    // Write method
    // This method returns an associative array of all the
records
    $server->add_to_map(
          "List",       // function name
          array(),                  // array of input types
          array("array")            // array of output types);
    function List(){
          global $contact;
          $contacts = array();
          $Sql = "SELECT LastName, FirstName,Subject
            FROM   teachers
            ORDER BY LastName";
            $ResultList = $contact->DB_Query($Sql);
          while ($Row = mysql_fetch_assoc($ResultList)){
                $contacts[]                =                new
soapval("contact","SOAPStruct",$Row);
          }
          return $contacts;  }
    // This method accepts a pattern as argument and
    // returns associative array of matching items
    $server->add_to_map(
          "search",       // function name
          array("string"),       // array of input types
          array("array")            // array of output types
    );
    function search ($pattern){
          global $contact;
          $matches = array();
          $Sql = "
          SELECT      FirstName,      LastName,      Subject,
FName,Phone
          FROM   teachers, faculty
          WHERE  teachers.Fid = faculty.Fid
               AND (FirstName regexp '$pattern'
                  OR LastName regexp '$pattern'
                  OR Subject regexp '$pattern'
                  OR Fname regexp '$pattern')
          ORDER BY teachers.LastName";
            $ResultList = $contact->DB_Query($Sql);
          while ($Row = mysql_fetch_assoc($ResultList)){
                $matches[] = $Row;
          }
          return $matches; }
    // Call the service method to initiate transaction
    // and send response
    $server->service($HTTP_RAW_POST_DATA);
```

V. Conclusion

The application of distributed search in databases, based on SOAP protocol and develop as a web service, was presented. Selective search in two kind of databases was explained and proposed.

References

[1] www.developer.com.

[2] www.nusphere.com

[3] Lee Anne Phillips, *Using XML*, QUE, 2000.

[4] Christopher Scollo, Jesus Castagnetto, Deepak Veliath, Harish Rawat, Sascha Schumann, *Professional PHP Programming*, Wrox, December 1999

[5] Paul Dubois, *MySql*, New Riders, 2000.

Distributed Search in WAN Located Databases (Repositories) Using Z39.50 Protocol

Krasimir Trichkov[1]

Abstract – **This standard specifies a client/server based protocol for Information Retrieval. It specifies procedures and structures for a client to search a database provided by a server, retrieve database records identified by a search, scan a term list, and sort a result set. The protocol addresses communication between corresponding information retrieval applications, the client and server.**

Keywords – **Internet, Z39.50, Zebra, Yaz, Zap, Php, Database.**

I. Introduction

Z39.50 [1] protocol specifies formats and procedures governing the exchange of messages between a client and server enabling the client to request that the server search a database and identify records which meet specified criteria, and to retrieve some or all of the identified records. This standard, ANSI/NISO Z39.50-1995 [2,3], *Information Retrieval (Z39.50) Application Service Definition and Protocol Specification*, is one of a set of standards produced to facilitate the interconnection of computer systems. It is positioned with respect to other related standards by the Open Systems Interconnection (OSI) basic reference model (ISO 7498). This standard defines a protocol within the application layer of the reference model, and is concerned in particular with the search and retrieval of information in databases.

II. Basics of the Protocol

The client may initiate requests on behalf of a user; the protocol addresses communication between corresponding information retrieval applications, the client and server (which may reside on different computers); it does not address interaction between the client and user.

Z39.50 provides the following basic capabilities, all of which are supported in Z39.50 as well. The client may send a search, indicating one or more databases, and including a query as well as parameters which determine whether records identified by the search should be returned as part of the response. The server responds with a count of records identified and possibly some or all of the records. The client may then retrieve selected records. The client assumes that records selected by the search form a "result set" (an ordered set, order determined by the server), and records may be referenced by position within the set.

Optional capabilities include:

1. The client may specify an *element set* indicating data elements to retrieve in cases where the client does not wish to receive complete database records. For example, the client might specify "If 5 or less records are identified, transmit 'full' records; if more than 5 records are found, transmit 'brief' records".
2. The client may indicate a *preferred syntax* for response records, for example, USMARC [4,5].
3. The client may *name* a result set for subsequent reference.
4. The client may *delete* a named result set.
5. The server may impose *access control* restrictions on the client, by demanding authentication before processing a request.
6. The server may provide *resource control* by sending an unsolicited or solicited status report; the server may suspend processing and allow the client to indicate whether to continue.

Fig. 1 explains distributed search.

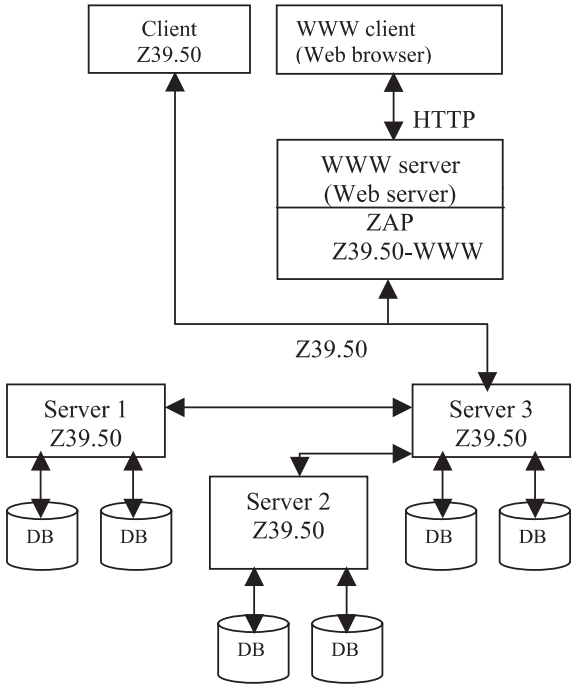


Fig. 1. Distributed search

[1]Krasimir Trichkov is with Institute of Computer and Communication Systems - Bulgarian Academy of Sciences, Acad. G. Bonchev bl.2, 1113 Sofia, Bulgaria, e-mail: krasi@hsi.iccs.bas.bg

This Z39.50 network was implemented in E-culture portal, which is part of the REGNET project (www.regnet.org). REGNET aims to set up a functional Network of Cultural Service Centers through Europe which will provide IT-services dedicated to cultural heritage organizations and it is funded by European Commission. For software development are used Apache web server [10] under Linux Operation System [11] (www.apache.org), Zebra Server [6] and Zap module [6]. To exchange data between REGNET centers is used Z39.50 protocol. Sample record that uses in Regnet is presented below.

Table 1. Example of record

<gils>
<Title> Spirit </Title>
<Creator> Romil Kalinov </Creator>
<Contribution> UBA </Contribution>
<Date> 2001.10.29 </Date>
<Description> Spirit, 2001 </Description>
<Identifier> 1 </Identifier>
<Type> Image </Type>
<Language> en </Language>
<Subject> Mixed technique </Subject>
<Publisher> 2001.11.06 </Publisher>
<Format> jpeg </Format>
<Source>Romil Kalinov , Sofia, Spirit, 2001, Mixed technique, 66 x 48 cm, Price: $ 135
Dimension 501x 709pixels, resolution 72 pixels/inch
</Source>
<Relation>
http://www3.iccs.bas.bg/RecordsUBA/romil.jpg
</Relation>
<Coverage> Contemporary Bulgarian Art
</Coverage>
<Rights> UBA </Rights>
</gils>

III. Record Types

Indexing is a per-record process, in which either insert/modify/delete will occur. Before a record is indexed search keys are extracted from whatever might be the layout the original record (sgml,html,text, etc..). The Zebra system currently supports two fundamental types of records: structured and simple text. To specify a particular extraction process, use either the command line option -t or specify a recordType setting in the configuration file.

The Zebra system is designed to support a wide range of data management applications. The system can be configured to handle virtually any kind of structured data. Each record in the system is associated with a *record schema* which lends context to the data elements of the record. Any number of record schema can coexist in the system. Although it may be wise to use only a single schema within one database, the system poses no such restrictions.

Records pass through three different states during processing in the system.

1. When records are accessed by the system, they are represented in their local or native format. This might be SGML or HTML files, News or Mail archives, MARC records. If the system doesn't already know how to read the type of data you need to store, you can set up an input filter by preparing conversion rules based on regular expressions and possibly augmented by a flexible scripting language (Tcl). The input filter produces as output an internal representation:

2. When records are processed by the system, they are represented in a tree-structure, constructed by tagged data elements hanging off a root node. The tagged elements may contain data or yet more tagged elements in a recursive structure. The system performs various actions on this tree structure (indexing, element selection, schema mapping, etc.),

3. Before transmitting records to the client, they are first converted from the internal structure to a form suitable for exchange over the network - according to the Z39.50 standard.

The RecordType parameter in the *zebra.cfg* file, or the $-t$ option to the indexer tells Zebra how to process input records. Two basic types of processing are available - raw text and structured data. Raw text is just that, and it is selected by providing the argument *text* to Zebra. Structured records are all handled internally using the basic mechanisms described in the subsequent sections. Zebra can read structured records in many different formats. How this is done is governed by additional parameters after the "grs" keyboard, separated by "." characters.

Four basic subtypes to the *grs* type are currently available:

grs.sgml - This is the canonical input format – described below. It is a simple SGML-like syntax.

grs.regx.*filter* – This enables a user-supplied input filter. The mechanisms of these filters are described below.

grs.tcl.*filter* – Similar to grs.regx but using Tcl for rules.

grs.marc.*abstract syntax* – This allows Zebra to read records in the ISO2709 (MARC) encoding standard. In this case, the last parameter *abstract syntax* names the .abs file (see below) which describes the specific MARC structure of the input record as well as the indexing rules.

IV. Canonical Input Format

Although input data can take any form, it is sometimes useful to describe the record processing capabilities of the system in terms of a single, canonical input format that gives access to the full spectrum of structure and flexibility in the system. In Zebra, this canonical format is an "SGML-like" syntax [6,7].

To use the canonical format specify grs.sgml as the record type.

Consider a record describing an information resource (such a record is sometimes known as a *locator record*). It might contain a field describing the distributor of the information resource, which might in turn be partitioned into various fields providing details about the distributor, like this in Table 2:

Table 2. Fields details

```
<Distributor>
<Name> ICCS </Name>
<Organization> ICCS </Organization>
<Street-Address>Sofia, Bulgaria</Street-Address>
<City> Sofia </City>
<Zip-Code> 1113 </Zip-Code>
<Country> Bulgaria </Country>
<Telephone> (359 2) 9792774 </Telephone>
</Distributor>
```

The keywords surrounded by <...> are *tags*, while the sections of text in between are the *data elements*. A data element is characterized by its location in the tree that is made up by the nested elements. Each element is terminated by a closing tag - beginning with </, and containing the same symbolic tag-name as the corresponding opening tag. The general closing tag - <>/ - terminates the element started by the last opening tag. The structuring of elements is significant. The element *Telephone*, for instance, may be indexed and presented to the client differently, depending on whether it appears inside the *Distributor* element, or some other, structured data element such a *Supplier* element.

V. Record Root

The first tag in a record describes the root node of the tree that makes up the total record. In the canonical input format, the root tag should contain the name of the schema that lends context to the elements of the record. The following is a GILS record that contains only a single element (strictly speaking, that makes it an illegal GILS record, since the GILS profile includes several mandatory elements - Zebra does not validate the contents of a record against the Z39.50 profile, however - it merely attempts to match up elements of a local representation with the given schema) – Table 3:

Table 3. Match up elements

```
<gils>
<title> Geometry of motion </title>
</gils>
```

VI. Variants

Zebra allows you to provide individual data elements in a number of *variant forms*. Examples of variant forms are textual data elements which might appear in different languages, and images which may appear in different formats or layouts. The variant system in Zebra is essentially a representation of the variant mechanism of Z39.50.

The following is an example of a title element which occurs in two different languages – Table 4.

Table 4. Different languages

```
<title>
<var lang lang "eng">Geometry of motion </>
<var lang lang "bg">Text in Bulgarian</>
</title>
```

The syntax of the *variant element* is <var class type value>. The available values for the *class* and *type* fields are given by the variant set that is associated with the current schema .Variant elements are terminated by the general end-tag </>, by the variant end-tag </var>, by the appearance of another variant tag with the same *class* and *value* settings, or by the appearance of another, normal tag. In other words, the end-tags for the variants used in the example above could have been saved.

Variant elements can be nested – Table 5.

Table 5. Variant of elements

```
<title>
<var lang lang "eng"><var body iana "text/plain">
Geometry of motion
</title>
```

Associates two variant components to the variant list for the title element. Given the nesting rules described above, we could write Table 6.

Table 6. Variant list for the title element

```
<title>
<var lang lang "eng"><var body iana "text/plain">
Geometry of motion
</title>
```

The title element above comes in two variants. Both have the IANA body type “text/plain”, but one is in English, and the other in Danish. The client, using the element selection mechanism of Z39.50, can retrieve information about the available variant forms of data elements, or it can select specific variants based on the requirements of the end-user [8,9].

VII. Exchange Formats

Converting records from the internal structure to en exchange format is largely an automatic process. Currently, the following exchange formats are supported:

1. GRS-1. The internal representation is based on GRS-1/XML, so the conversion here is straightforward. The system will create applied variant and supported variant lists as required, if a record contains variant information.
2. XML [12,13]. The internal representation is based on GRS-1/XML so the mapping is trivial. Note that XML

schemas, preprocessing instructions and comments are not part of the internal representation and therefore will never be part of a generated XML record. Future versions of the Zebra will support that.

3. SUTRS. Again, the mapping is fairly straightforward. Indentation is used to show the hierarchical structure of the record. All "GRS" type records support both the GRS-1 and SUTRS representations.

4. ISO2709-based formats (USMARC, etc.). Only records with a two-level structure (corresponding to fields and subfields) can be directly mapped to ISO2709. For records with a different structuring (eg., GILS), the representation in a structure like USMARC involves a schema-mapping, to an "implied" USMARC schema (implied, because there is no formal schema which specifies the use of the USMARC fields outside of ISO2709). The resultant, two-level record is then mapped directly from the internal representation to ISO2709.

5. Explain. This representation is only available for records belonging to the Explain schema.

6. Summary. This ASN-1 based structure is only available for records belonging to the Summary schema - or schema which provide a mapping to this schema (see the description of the schema mapping facility above).

7. SOIF. Support for this syntax is experimental, and is currently keyed to a private Index Data OID (1.2.840.10003.5.1000.81.2). All abstract syntaxes can be mapped to the SOIF format, although nested elements are represented by concatenation of the tag names at each level.

VIII. Software Components

A. Zebra. (http://www.indexdata.dk/zebra/).

Zebra is a fielded free-text indexing and retrieval engine with a Z39.50 frontend. You can use any compatible, commercial or freeware Z39.50 client to access data stored in Zebra. Zebra may be used free-of-charge in non-profit applications by non-commercial organizations. Zebra is a high-performance, general-purpose structured text indexing and retrieval engine. It reads structured records in a variety of input formats (eg. email, XML, MARC) and allows access to them through exact Boolean search expressions and relevance-ranked free-text queries. Zebra supports large databases (more than ten gigabytes of data, tens of millions of records). It supports incremental, safe database updates on live systems. You can access data stored in Zebra using a variety of Index Data tools (eg. YAZ and PHP/YAZ) as well as commercial and freeware Z39.50 clients and toolkits.

B. Yaz (http://www.indexdata.dk/yaz/).

The YAZ toolkit offers several different levels of access to the Z39.50 and ILL protocols. The level that you need to use depends on your requirements, and the role (server or client) that you want to implement.

C. Zap (http://www.indexdata.dk/zap/).

ZAP is a module which allows you to build simple WWW interfaces to Z39.50 servers. ZAP hides most of the complexity of session management, parallel searching. The integration of system into the popular Apache server offers several advantages to the operators and users of the software, including simplified maintenance of the Module, and improved performance. However, it is also possible to run the software as a CGI-script if required.

This is free software (open source) that can works on various operating systems (as Windows and Linux) and various Web Servers (as Apache and IIS).

IX. Conclusion

The essence and functional possibilities on communication protocol Z39.50 was presented. Definite are special futures of the protocol and its application for information search in distributed databases. Definitely are software component of the protocol. Proposed the decision for works with distributed and heterogeneous databases using Z39.50. The protocol is platform and software independent.

This protocol is applied for Bulgarian partnership in the international project REGNET – REGional NETwork (Regional Networks of Culture Heritage).

References

[1] http://www.loc.gov/z3950/agency
[2] http://www.niso.org/z3950.html
[3] http://www.ansi.org
[4] http://www.loc.gov/marc
[5] http://lcweb.loc.gov/marc/umb/um07to10.html
[6] http://www.indexdata.dk
[7] http://www.gils.net
[8] http://www.amico.org
[9] http://dublincore.org
[10] Peter Wainwright, *Professional Apache*, Wrox, November 1999
[11] David Pitts, Bill Ball, et al., *Red Hat Linux 6*, Sams, 1999.
[12] Alex Homer, *XML in IE5 Programmer's Reference*, Wrox Press, 1999.
[13] Lee Anne Phillips, *Using XML*, QUE, 2000.

Internet Databases Using SOAP Protocol and XML Standard

Martin Tsenov[1]

Abstract – **The internet database application of SOAP protocol using XML is considered. Two methods of hierarchical system: two tiers and three tiers models are presented and the communication between client and server depends on XML and SOAP are explained.**

Keywords – **Database, Internet, SOAP, Xml**

I. Introduction

SOAP (Simple Object Access Protocol) is a simple, lightweight protocol for structured and strong-type information exchange in a decentralized, distributed environment. The protocol is based on XML (eXtensible Markup Language) and consists of three parts:

1. An envelope which describes the contents of the message and how to use it
2. A set of rules for serializing data exchanged between applications
3. A procedure to represent remote procedure calls, that is, the way in which queries and the resulting responses to the procedure are represented.

Similar to object distribution models (IIOP, DCOM...), SOAP can call methods, services, components and objects on remote servers. However, unlike these protocols, which use binary formats for the calls, SOAP uses text format (Unicode), with the help of XML to structure the nature of the exchanges.

SOAP can generally operate with numerous protocols (FTP, SMTP, POP...), but it is particularly well suited to the HTTP protocol. It defines a reduced set of parameters which are specified in the HTTP header, making it easier to pass through proxies and firewalls.

XML (Extensible Markup Language) is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere. For example, computer makers might agree on a standard or common way to describe the information about a computer product (processor speed, memory size, and so forth) and then describe the product information format with XML. Such a standard way of describing data would enable a user to send an intelligent agent (a program) to each computer maker's Web site, gather data, and then make a valid comparison. XML can be used by any individual or group of individuals or companies that wants to share information in a consistent way.

XML is "extensible" because, unlike HTML, the markup symbols are unlimited and self-defining. XML is actually a simpler and easier-to-use subset of the Standard

Generalized Markup Language (SGML), the standard for how to create a document structure. It is expected that HTML and XML will be used together in many Web applications.

II. Physical Implementation

A. Theoretical Part

The main aim of the e-Shop is to obtain a selective search of the offered items in the REGNET system. The functionality of the system is realized by software decision, based on the SOAP protocol and XML [1] standards. SOAP (Simple Object Access Protocol) is a simple, lightweight protocol for structured and strong-type information exchange in a decentralized, distributed environment. The protocol is based on XML (eXtensible Markup Language).

SOAP messages are structured using XML. Within the framework of the remote procedure call (RPC), it represents the parameters of the methods, the return values and any potential error messages linked to the processes.

Coding SOAP messages in XML [2] enables universal communication between applications, services and platforms via the Internet. In order to do this, SOAP makes use of the descriptive nature of the XML language, thus transforming the content into an application.

In more technical terms, just as with an XML fragment, SOAP messages make references to different namespaces, enabling the content to be validated. They must therefore include a call to SOAP namespaces, making it possible to define and specify the use of standard tags in the message and to ensure compatibility between SOAP versions. As soon as a SOAP message is received, the SOAP tags are validated, as are the tags that express the subject of the message. If it fails, an error is generated (http://www.w3.org/TR/SOAP/).

Soap thus defines two namespaces:

http://schemas.xmlsoap.org/soap/envelope/ for the envelope

http://schemas.xmlsoap.org/soap/encoding/ for the coding

[1]Martin Tsenov is with Institute of Computer and Communication Systems - Bulgarian Academy of Sciences, Acad. G. Bonchev bl.2, 1113 Sofia, Bulgaria, e-mail: mcenov@hsh.iccs.bas.bg

```
xml_begin="<?xml version=\"1.0\"?>\r\n".
        "<SOAP-ENV:Envelopexmlns:SOAP-
ENV=\"http://schemas.xmlsoap.org/soap/envelope/\"
xmlns:xsd=\"http://www.w3.org/2001/XMLSchema\"
xmlns:xsi=\"http://www.w3.org/2001/XMLSchema-
instance\"
xmlns:SOAP-
ENC=\"http://schemas.xmlsoap.org/soap/encoding/\"
xmlns:si=\"http://soapinterop.org/xsd\"
SOAP ENV:encodingStyle=\"
http://schemas.xmlsoap.org/soap/encoding/\">\r\n".
        "<SOAP-ENV:Body>\r\n";
$xml_end="</SOAP-ENV:Body>\r\n".
       "</SOAP-ENV:Envelope>\r\n";
```

Deploying SOAP over HTTP makes it possible to use the SOAP decentralization method in the well-used environment of HTTP. Using SOAP over HTTP also enables resources already present on the Web to be unified by using the natural request/response mode of HTTP. The only constraint is that a SOAP message via HTTP must use the MIME type “text/xml”. Fig. 1 presented the example SOAP network.

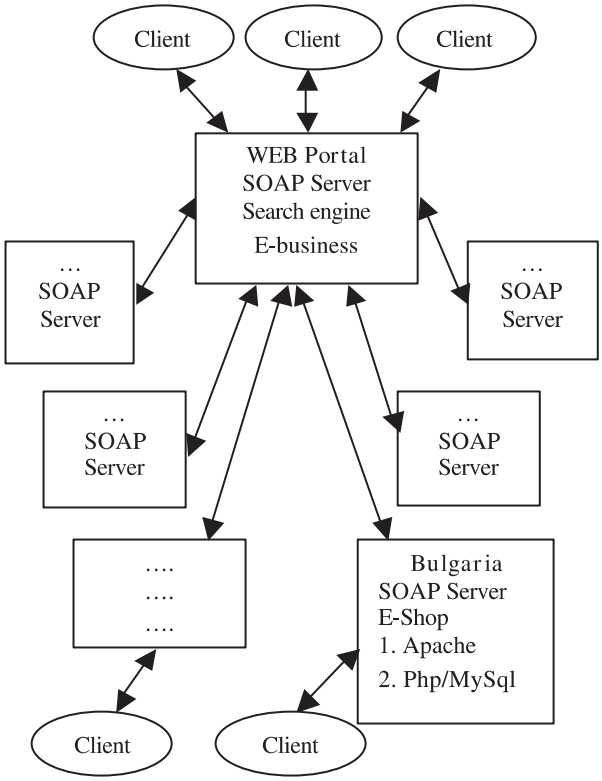


Fig. 1. Structure of the example SOAP network

B. *Structure of the E-business Center*

This Soap network was implemented in E-business portal, which is part of the REGNET project (www.regnet.org). REGNET aims to set up a functional Network of Cultural Service Centers through Europe which will provide IT-services dedicated to cultural heritage organizations and it is funded by European Commission. For software development are used Apache web server [3] under Linux Operation System [4] (www.apache.org). For web programming language we used PHP [5] and MySql Server [6] for database. To exchange data between REGNET centers is used SOAP protocol. Sample source code is presented above.

```
//SOAP-request creating
 $xml_body="<?xml version=\"1.0\"?>\r\n".
        "<SOAP-ENV:Envelopexmlns:SOAP-
ENV=\"http://schemas.xmlsoap.org/soap/envelope/\"
xmlns:xsd=\"http://www.w3.org/2001/XMLSchema\"
xmlns:xsi=\"http://www.w3.org/2001/XMLSchema-
instance\"
xmlns:SOAP-
ENC=\"http://schemas.xmlsoap.org/soap/encoding/\"
xmlns:si=\"http://soapinterop.org/xsd\"
SOAP-
ENV:encodingStyle=\"http://schemas.xmlsoap.org/soap/en
coding/\">\r\n".
        "<SOAP-ENV:Body>\r\n".
        "<xml_body>\r\n".
        " <action>Search</action>\r\n".
          " <request>\r\n";
//$xml_body.="  <pricetype>test</pricetype>\r\n";
if (isset($term))
   $xml_body.="
<prodName>".$term."</prodName>\r\n";
if($term==""$xml_body.="
<prodName>%</prodName>\r\n";
 if ($prices!=")
   $xml_body.="  <price>".$prices."</price>\r\n".
          "  <tup>".$tup."</tup>\r\n";
 if ($category_id!=0)
 $xml_body.="
<category_id>".$category_id."</category_id>\r\n";
 $xml_body.=" </request>\r\n".
        "</xml_body>\r\n".
        "</SOAP-ENV:Body>\r\n".
        "</SOAP-ENV:Envelope>";
 $xml_query=$xml_begin.$xml_body.$xml_end;
 //$xml_body=";
```

At Fig. 1 and Fig. 2 are presented two methods of hierarchical system: two tiers and three tiers models for communications between client and E-business center.

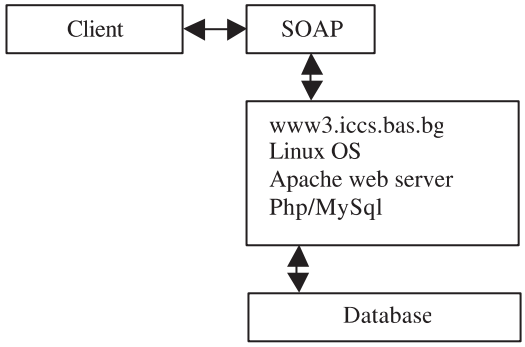


Fig. 2. Two tiers model

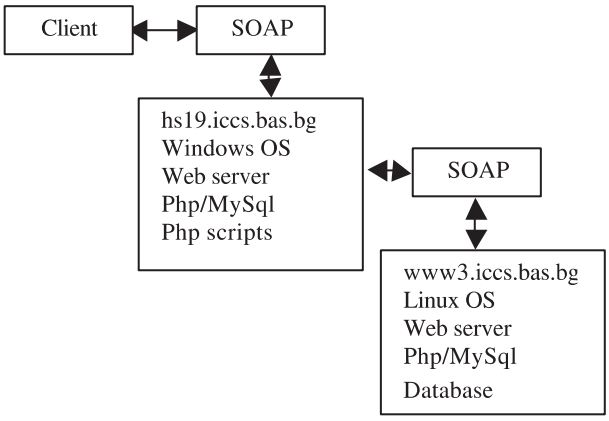


Fig. 3. Three tiers model

C. Software Components.

Apache web server [7]. The Apache HTTP Server Project is an effort to develop and maintain an open-source HTTP server for various modern desktop and server operating systems, such as UNIX and Windows NT. The goal of this project is to provide a secure, efficient and extensible server which provides HTTP services in sync with the current HTTP standards.

Php. Php [8] is a widely-used Open Source general-purpose scripting language that is especially suited for Web development and can be embedded into HTML.

MySql database. For database is used MySql Server [9]. The MySQL database server embodies an ingenious software architecture that maximizes speed and customizability. Extensive reuses of pieces of code within the software and an ambition to produce minimalist but functionally rich features have resulted in a database management system unmatched in speed, compactness, stability and ease of deployment. The unique separation of the core server from the table handler makes it possible to run MySQL under strict transaction control or with ultra fast transaction less disk access, whichever is most appropriate for the situation.

The structure of tables and field of the database are presented at Figures above: Table 1: item_shop database table, Table 2: transaction database table.

Table 1.

Field	Type
item_no	varchar(20)
item_type	varchar(40)
title	varchar(60)
artist	varchar(60)
price	float
text	text
picture	blob
picture1	varchar(60)

Table 2.

Field	Type
order_no	int(11)
user_id	varchar(20)
item_no	varchar(20)
quantity	int(11)
date	date
status	varchar(20)

III. Structure of the Application

The SOAP communication module and server side programs are developed in PHP web programming language [5]. The descriptions of the function are presented above.

SOAP communication module – This module provide relation and links between clients, web portals and E-business centers. It based on SOAP protocol, defined by W3 consortium [10]. The simple source code is presented above.

```
//SOAP-request creating
…
$payload[]="HTTP/1.0 200 OK\r\n";
$payload[]="Status: 200\r\n";
$payload[]="Server: SOAPx4 Server v0.5\r\n";
$payload[]="Connection: Close\r\n";
$payload[]="Content-Type: text/xml; charset=UTF-8\r\n";
$payload[]="Content-Length:
".strlen($xml_query)."\r\n\r\n";
            reset($payload);
            foreach($payload as $hdr)
             {
               header($hdr);
              }
 print $xml_query;
…
//SOAP-response creating
```

Register – this function realizes the registration of the user and saves their data in the database server MySql. The information, required from the user is name, surname, user id, password, e-mail, country. The simple source code is presented above.

```
<?php
require 'functions.php';
// Check if all the form entries are entered, if any form
entry
// is missing then send, error message page
if ( (trim($form_name)=="")|| (trim($form_password)=="")
|| (trim($form_user_id)=="")|| (trim($form_password)=="")
||(trim($form_password1)=="")||
(trim($form_email_id)=="")||(trim($form_phone)=="")||
(trim($form_city)==")" || (trim($form_country)=="")) ) {
  header("Location:error2.htm");   exit();
} else if ($form_password != $form_password1) {
   // If both the passwords are not the same then generate
error message
 header("Location:error3.htm");
   exit(); } else {
```

```
// Open a persistent connection with the Database
 if (!($link = mysql_pconnect ($DB_SERVER,
$DB_LOGIN, $DB_PASSWORD))) {
    DisplayErrMsg(sprintf("internal error %d:%s\n",
      mysql_errno(), mysql_error()));
    exit() ; }
 // Create the user record
  $balance = 0.00;
  if (!($newresult = mysql_db_query($DB, "INSERT
INTO user_profile
       (name,user_id,password,city,
       country,email_id,phone_number,
       account_balance) VALUES
       ('$form_name','$form_user_id','$form_password',
       '$form_city',
       '$form_country',
       '$form_email_id', '$form_phone',
        $balance)"))) {
    DisplayErrMsg(sprintf("internal error %d:%s\n",
                    mysql_errno(), mysql_error()));
    exit() ; }
  // If Registration Successful, then display, else error

header("Location:registration_success.htm");
    exit(); }  ?>
```

Search – this function realize a selective search of the offered items in the e-Shop. The data of the items are saved in the distributed database and used SOAP communication modul. The searching can be restricted by few categories: type, title, artist and price. The simple source code is presented above.

```
/SOAP-request sending
  function send ($soap_data,$path,$server)
  {
   global $outgoing_payload;
   $incoming_payload='';
   $action='urn:soapBI';
   $port='80';
   $fp = fsockopen($server,$port,$errno,$errstr,3);
   $outgoing_payload =
          "POST ".$path." HTTP/1.0\r\n".
          "User-Agent: SOAPx4 v0.5\r\n".
          "Host: ".$server."\r\n".
          "Content-Type:text/xml\r\nContent-Length:
               ".strlen($soap_data)."\r\n".
          "SOAPAction: \"$action\""."\r\n\r\n".
               $soap_data;
  // send
…
mysql_connect($host,$user,$userpass)
    or die("Connect failed");
 if(mysql_select_db($base)==FALSE)
   {   …   }
```

Items – whit this function can be listed the whole items catalogue.

Today Shop – this function is the shop cart of the user.

The e-shop can be reach at http://www3.iccs.bas.bg or http://hs19.iccs.bas.bg/Eth_shop/

IV. Conclusion

The application of SOAP protocol in E-comers Solution was presented. Two methods of hierarchical system: two tiers and three tiers models for communication between client and server depend on SOAP protocol were explained and presented. The e-Shop was developed.

References

[1] Alex Homer, *XML in IE5 Programmer's Reference*, Wrox Press, 1999.

[2] Lee Anne Phillips, *Using XML*, QUE, 2000.

[3] Peter Wainwright, *Professional Apache*, Wrox, November 1999

[4] David Pitts, Bill Ball, et al., *Red Hat Linux 6*, Sams, 1999.

[5] Christopher Scollo, Jesus Castagnetto, Deepak Veliath, Harish Rawat, Sascha Schumann, *Professional PHP Programming*, Wrox, December 1999

[6] Paul Dubois, *MySql*, New Riders, 2000.

[7] http://www.apache.org

[8] http://www.php.net

[9] http://www.mysql.org

[10] http://www.w3.org/TR/SOAP/

Analysis of Efficiency of Protocols, Guaranteeing QoS on VoIP Technologies

Nikolai V. Penev[1] and Vassil M. Kadrev[2]

Abstract – **In the packet switching networks of TCP\IP stack of protocols the main problem is to guarantee the quality of services for services provided in real-time. The inherent particularity of these networks is their aptitude to overloading, which brings about essential distortion of packetized voice facilities. On the base of the developed mathematical model, different varieties of protocols of standby have been analyzed. The evaluation of improving the quality of services has been made on one hand, and on the other hand the reducing efficiency of the network has been estimated.**

Keywords – **IP, QoS, Traffic**

I. Introduction

In the Internet layer of the network a mechanism for inverse request in the case of finding mistakes of the accepted information is not provided. Possibility for inverse request is realized in the TCP layer that is inadmissible for voice and video services. It brings the necessity to introduce mechanisms to ensure the quality of services with the realization of VoIP. The essential parameters of quality of services, which affect and which are guaranteed are: end-to-end delay, jitter, probability of the loss of the package on IP layer because of a mistake. These mechanisms present protocols of standby, anti-jitter buffers and algorithms for correcting the mistakes in the receiver without an inverse request.

With the occurrence of the idea about building a digital network with integrated services on the base of IP, the necessity of using methods to ensure the quality of services (QoS) has appeared. The switching units, terminals and centers for operation and control in the IP network are specialized computers that can perform the algorithms of precise information processing, as the mentioned criteria.

Most generally, QoS contains three criteria of quality: delay, jitter and losses of packets. The various kinds of services are critical to various parameters and have different criteria. With its developing as a network with integrated services (approach ISA - Integrated Services Architecture of IETF – Internet Engineering Task Force), the IP network, which has occurred on the purpose of transmitting data without warranties of the particular packets delivery to the receiver, has already involved mechanisms (protocols) guaranteeing the quality of services critical to the mentioned criteria of QoS.

[1]Nikolai V. Penev is with the War Academy, 82 E. Georgiev, Sofia, Bulgaria, E-mail: penevnv@yahoo.com

[2]Vassil M. Kadrev is with the Higher School of Transport, 158 Geo Milev, 1754 Sofia, Bulgaria, E-mail: kadrev @ internet-bg.net

II. Analyses of QoS in IP Network

The mechanisms of providing guaranteed delay of the traffic critical to this criterion, which are known at present, are three: RSVP (Resource Reservation Protocol), MPLS (Multi-Protocol Label Switching) and DS (Differentiated Services). These are mechanisms without which services such as telephone, video, multimedia and interactive data exchange could not be offered in the IP integrated environment with a guaranteed norm of delay. The problems of the jitter in the IP network, toward which the telephone service is particularly critical, can be solved by anti-jitter buffers in the routers, from which it can be read at a constant rate. With a sufficient bit rate of the lines and switch centers, the routers in the IP network, this mechanism is effective, as the increase of the delay with a big volume of anti-jitter buffers has to be taken into consideration. The TCP layer in the IP network guarantees a sure information delivery with which great delay of the feedback and re-transmitting the data mistaken have been introduced. It could result in exceeding the admissible norms of information delay. That can be avoided by introducing FEC (Forward Error Correction) mechanisms performing marking of the mistaken bits by finding noise-resistant code and correlation data analysis in one or a number of packets aiming at approximate restoration of the mistaken bits. Due to the continuous character and the even changes of the telephone and video signals in time FEC guarantees small losses (sound and picture distortion) without introducing additional delay for that purpose. The experience accumulated with the operation of IP networks with mixed integrated traffic shows that the anti-jitter and FEC mechanisms are not the main part for the great “end-to-end” information delay. In order to meet the norms of the delay of the telephone and video services, effective mechanisms for mixed traffic control in the IP network are necessary.

IETF defines two main mechanisms of the IP traffic control guaranteeing QoS (delay) by the priorities of the traffic flows of the various services. They are the mechanisms of integrated services ISA and of DS. Both mechanisms use field TS 8 bits in the title of IP-packet for specifying the priority. In the field of options data contained can give a possibility to introduce a dynamic priority (moment of packet generating, time of packet life).

The main purpose of ISA mechanism is to recognize packages of the priority service flow (requiring little delay) and transmit them without waiting in queues aiming at delay not exceeding the admissible one. This mechanism is specified in protocol RSVP where two classes of integrated services

are defined: GS – Guaranteed Services and CL – Controlled Load.

GS is a protocol for forwarding packets with a reserved traffic capacity that results in little usage of the network resources. Thus GS guarantees little delay and absence of losses (telephony, video, multimedia). It should be outlined as a protocol disadvantage that it throws off the packets after the deadline of their life and arrived after the admissible time of delay.

CL mechanism allows the router-realizing RSVP to process the service package flow as ordinary IP datagrams (Best Effort), to control the service packages and to increase their priority in the network with the increase of the package delay (decrease of the life left) up to the moment, i.e. to decrease their stay in the lines. In the latest versions of CL, with the impossibility to guarantee the admissible delay of the packets, additional sessions for the same service are established, thus the delay sharply drops down to be included in the norms.

The second mechanism of control and guaranteeing the QoS of IP-traffic is DS based on the so-called PHB (Per Hop Behavior) routing. It includes as separate mechanisms EF (Expedited Forwarding) and AF (Assured Forwarding). With this mechanism a number of classes of packets with various admissible delay are defined (according to the services) for which a respective buffer space and rate of forwarding are reserved. In each class of packets there are a number of priorities as the packets of higher priority are kept and the ones of lower priority are thrown off with network overloading. It is necessary to apply DS with using wide-banded services and it depends on the rates of forwarding. The mechanism of detecting and preventing the conjunctions RED (Random Early Detection) is also maintained. It detects the random arrived packets on the base of the analysis of the title part of TCP datagram and throws them off if they are not addressed to the ports of addresses that prevail in the queue of different priorities. Thus the buffers (queues) with different priority in the router are kept semi-filled up with packets that also decrease their delay.

III. Model, Quantitative Ratios and Results

The presence of multi-priority flow of incoming asking for service is common for all mechanisms of the IP traffic control. For an IP network with integrated services including subscribers and transit devices (routers), the delay of the package "end-to-end" for a homogeneous flow of packages (for each service) is determined as T_w:

$$T_w = 2t_{sl} + \sum_{i=1}^{d} (t_w + t_{pr} + t_{tr}) - t_{tr} \ , \qquad (1)$$

where t_{sl} – average time for transfer by subscriber line; d – average number of hops; t_w – average waiting time in the queue of a node; t_{pr} – average processing time of the packets on the router; t_{tr} – average time for transfer by trunk.

The components of T_w: t_{sl} and t_{tr}, which use DSL technology can be assumed as constants and only t_w and $d$remain to influence on T_w.

The paper presents a model GS mechanism of RSVP protocol examined with which RSVP is provided according to the strict requirements to work in real time a guaranteed frequency band, as well as little delay of "end-to-end" packets and absence of packets loss as a result of their arrangement in queues. Due to this, with serving one and the same GS flow, each router in the network (which requires separate functions for that in network control) has to distribute the frequency band and the respective buffer space according to the priority of entering packets of various services. As a great number of traffic flows along the lines of great traffic capacity enter the router, it can be assumed that the input traffic to the router is punch [5]. The particular router is a system of mass service with waiting and service discipline with priorities. Under these conditions T_w can be determined for each service. It is known that teletraffic system M/D/1 is characterized by average time of waiting twice less than that of system M/M/1. The examinations will be carried out using teletraffic system M/M/1 and stipulating that the results with determining the average time of waiting for a system of mass service M/M/1 are the upper limit for the average time of waiting as M/D/1 is in practice. It is possible to determine the dependencies of service quality by the rates of lines, the size of packets and the number of transit sections. The determination of the results of the priority service influence on the quality of service for a service device can be also used for the network as a whole. The position of each package in the queue of the service device will be a variable function of time having in mind the possibility that a packet of higher priority can enter the queue. On one hand, the system of priorities can be defined according to the priority: if it is absolute (fixed) or depends on a given function, and on the other hand – if processing of the packet served is interrupted at the moment of the entrance of a packet of higher priority and later is restored (at the moment of breaking). For a priority system of service (of P priorities as p = 1, 2, 3, ... P) with a fixed priority of serving with the entrance of a packet of a higher priority there is given (with $0 \leqslant \rho < 1$, i.e. absence of losses)[2]:

$$T_{wp} = \begin{cases} \dfrac{\frac{\rho_p}{\mu_p} + \sum\limits_{i=p+1}^{P} \rho_i \cdot \left(\frac{1}{\mu_p} + \frac{1}{\mu_i}\right) + \sum\limits_{i=p+1}^{P} \rho_i \, T_i}{1 - \sum\limits_{i=p}^{P} \rho_i}, & \text{for } p \geqslant j \\ \infty, & \text{for } p < j \end{cases} , \qquad (2)$$

where j is minimum and integer and $\sum\limits_{i=j}^{P} \rho_i < 1$.

Type of services [4] are: voice, video, interactive data (data 1), download (data 2). The parameters of the arriving traffic and of the data processing [4] are shown in Table 1.

Average servicing time $1/\mu$ depends from average size of the packet l and average send rate c. Average value of input traffic ρ depend from average input rate of the packets λ and average servicing time $1/\mu$=const. For input traffic

$$\sum_{i=1}^{4} \rho_i = \rho < 1. \qquad (3)$$

Table 1. Parameters of servicing

Service	Video	Voice	Interactive data	Download
Average servicing time, $1/\mu = const$	$\frac{1}{\mu_1} = \frac{l_1}{c_1}$	$\frac{1}{\mu_2} = \frac{l_2}{c_2}$	$\frac{1}{\mu_3} = \frac{l_3}{c_3}$	$\frac{1}{\mu_4} = \frac{l_4}{c_4}$
Input traffic, ρ	$\rho_1 = \frac{\lambda_1}{\mu_1}$	$\rho_2 = \frac{\lambda_2}{\mu_2}$	$\rho_3 = \frac{\lambda_3}{\mu_3}$	$\rho_4 = \frac{\lambda_4}{\mu_4}$
Priority of servicing, p	4	3	2	1

Priority of servicing increases with p. Send rate of subscriber line is $c_{sl} = 8$ Mbps, send rate of trunk is $c_{tr} = 40$ Mbps, average number of hops is $d = 4$, average processing time on the router t_{pr} is approximately 0.

Values of output data for determining of the servicing parameters [4] are shown in Table 2.

Table 2. Type of traffic sources and parameters of servicing

	Traffic			
	Video	Voice	Interactive data	Download
Average norm of delay, ms	60	160	600	2000
Average norm of delay for one hope, ms	15	40	150	500
Average length of the block, kbit	burst traffic	burst traffic	400	4000
Average length of the packet, kbit	8	20	200	400
Average time for transfer by trunk, ms	0,05	0,5	5	10
Average time for transfer by subscriber line, ms	0,25	2,5	25	50
Total average transfer time end-to-end for d = 4, ms	0,7	7	70	140
Average admissible waiting time in the queue of a node, ms	14.825	38.25	130	410

The average admissible waiting time in the queue of one processing device is the difference between the average admissible delay for each service and the total average transmission time end-to-end with d sections.

Table 3. Structure of traffic sources

Percentage of the arriving traffic, %	Video	Voice	Interactive data	Download
Relation 1	10	5	15	70
Relation 2	20	10	35	35
Relation 3	49	1	35	15
Delay, ms	w4	w3	w2	w1

The results obtained for the delay values depending on the input traffic ρ and at concrete proportion of the input traffic from the separated services with priority processing (Table 3) are shown on Figs 1, 2 and 3.

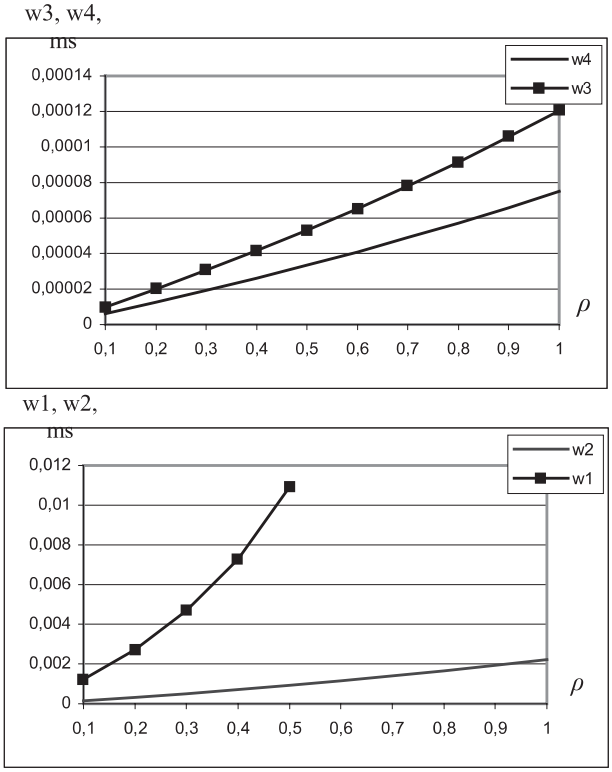


Fig. 1. Delays for relation 1 (Table 3) of input traffic for: video (w4), voice (w3), interactive data (w2), download (w1)

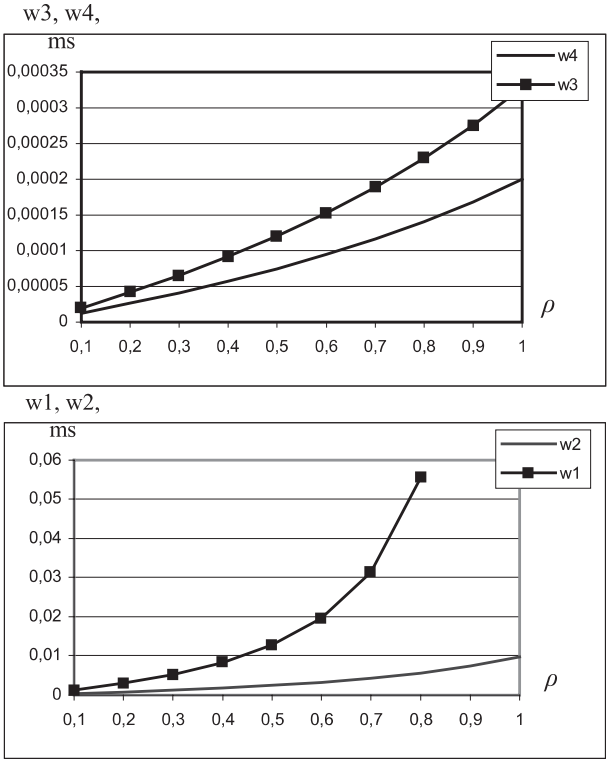


Fig. 2. Delays for relation 2 (Table 3) of input traffic for: video (w4), voice (w3), interactive data (w2), download (w1)

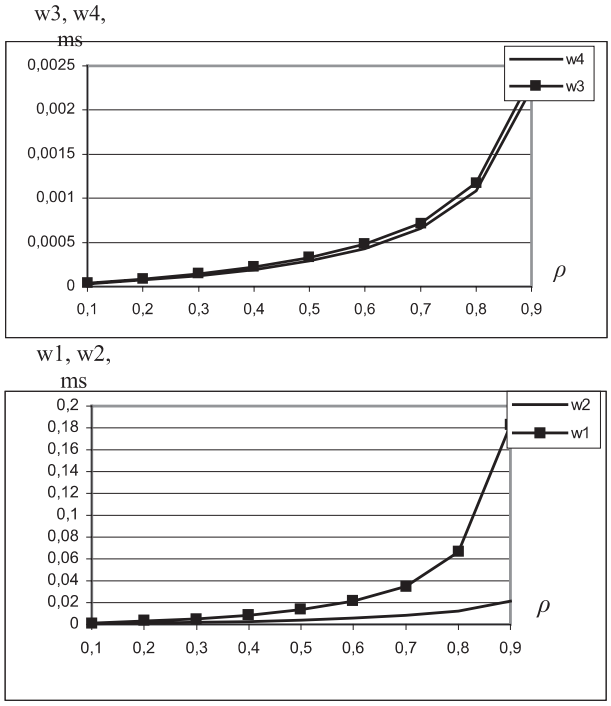


Fig. 3. Delays for relation 3 (Table 3) of input traffic for: video (w4), voice (w3), interactive data (w2), download (w1)

IV. Conclusion

From the results obtained it can be seen that with the given parameters of the lines, devices and the traffic, the delay which the separate services receive is within the norms for a wide range of change for the values of these parameters: the size and the proportions of the arriving traffic (loading of the router) and the size of the packets. This is true when there are faster lines and approximately all the norm of delay may be used for waiting in queues.

By increasing the length of the packets for the traffic with high priority (voice, video), average waiting time sharply decreases. This results quickly not within the norms of delay, for loading of the router over 0,7.

The research is made for unlimited number of nodes as their connection is 30.

The results for mechanism CL of RSVP can be obtained in a similar way with the dynamic priorities of queue servicing.

References

[1] Kleinrock L. Theory of queuing systems – I, Moscow, Mashinostroenie, 1979.

[2] Kleinrock L. Communication Nets (stochastic message flow and delay). Mcgraw-Hill.

[3] Models of teletrafic theory in telecommunications and electronics. Proceedings - A. D. Charkevich, V. A. Garmash. Moscow, Science, 1985.

[4] Parvanova N. Broadband ISDN. NIIS-CENTI, Sofia, 1997.

[5] Alcatel Telecommunications Review, No. 2, 1999.

[6] Gantchev I. Computer Networks and Communications. Plovdiv., University of Plovdiv, 1999.

[7] Boyanov K. and all. Computers networks and Internet. Sofia, CLPOI-BAS, 1998.

[8] www.ietf.org.

Learning of the Artificial Neural Networks with Multilayer Models and Industrial Tasks

Hristo I. Toshev[1], Chavdar D. Korsemov[2] and Stefan L. Koynov[3]

Abstract – **Two approaches of graphical multilayer modeling are introduced - by graphs and by tables. Examples with the architecture and the learning paradigm of the artificial neural networks interpreted with multilayer models are presented. Special attention is dedicated to the formulations of problems about modeling industrial tasks.**

Keywords – **multilayer models, neural networks, architecture, learning**

I. Introduction

The scope of the paper is modelling the learning process of artificial neural networks (ANN) by multilayer models and its implementation for industrial tasks. Also the authors present two approaches of visualizing the multilayer models - by graphs and by tables. Both ways have their advantages and limitations.

The graphical visualization may be applied in every possible case. It is an appropriate base for application of the graph theory, of Petri nets, etc. The visualization scheme may consist of one main multilayer model. Separate points in its layers may be magnified to corresponding multilayer models next to the main model. In this way the main multilayer model is surrounded by a ring of multilayer models *of lower weight* ([2], see also Fig. 1). Instead this paper is a demonstration of the table approach with multilayer models for main features of the ANN. The visualization approach by tables is applicable for *separate* multilayer models in which besides the *layer number* and the *layer name*, the *feature of the layer* defines the layer-characterizing mark (see chapter III for more about the table visualization of multilayer models).

The paper consists of the following chapters. Chapter II formulates the problem about the ANN learning with multilayer models. Chapter III introduces examples of multilayer models of ANN features based on the degree of their sophistication. Chapter IV marks the design of industrial tasks with ANN and multilayer models. Finally the paper resumes the contents with conclusions.

[1]Hristo I. Toshev is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 29A, 1113 Sofia, Bulgaria, E-mail: toshev@iinf.bas.bg

[2]Chavdar D. Korsemov is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 29A, 1113 Sofia, Bulgaria, E-mail: korsemoviinf.bas.bg

[3]Stefan L. Koynov is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 29A, 1113 Sofia, Bulgaria, E-mail: baruch@iinf.bas.bg

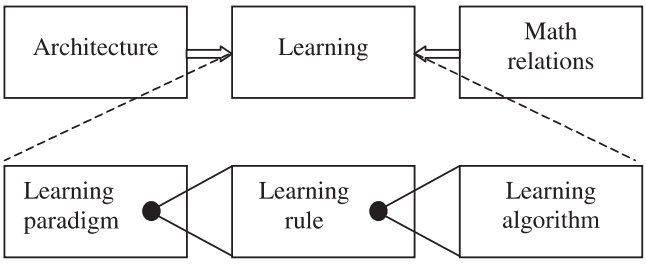


Fig. 1. The ANN learning process with multilayer models

II. The Formulation of the Problem about the Artificial Neural Network Learning with Multilayer Models

ANN learning is the most time-consuming step (if any).

Two determining factors: the goal ANN architecture [1] and the corresponding mathematical description [2] feed the learning process. The architecture is the most directly connected with the physical 'nature' of the ANN while the mathematical formalism present in the discrete time domain the relations between the information streams. The mathematical model is decisive for the ANN learning because it is *defined by the learning rule* but it itself *defines the learning algorithm*.

The ANN learning process consists of the learning paradigm, the learning rule and the learning algorithm. It is influenced by the chosen architecture and the corresponding mathematical formalism.

The learning paradigm presets the possible learning rules which in turn preset the possible learning algorithms. The learning paradigm is the most abstract feature and the learning algorithm is the most concrete one. Starting from left in a direction to the right the learning characteristics become

Table 1. Mathematical formalism for unsupervised learning

UNSUPERVISED LEARNING	**Differential**	**Nondifferential (Signal)**
Hebbian	Differential Hebbian Learning	Signal Hebbian Learning
Competitive	Differential Competitive Learning	ART-Competitive Learning

more concrete and this corresponds to a 'descent' from the peak of a pyramidal structure towards its base; the peak is analogous to the learning paradigm and the base - to the concrete application task. The very 'descent' during the ANN learning corresponds to a draw up to the final ANN design.

The scheme in Fig. 1 presents the links between these aspects in the learning process for ANN with multilayer models.

The mathematical description is of a principal importance for the self-organizing maps when the mathematical formalism may be presented according to Table 1. This example of data representation by tables is similar to the example in [3] when the application of multilayer models is not recommended (the same reference contains more information about details).

Table 2. Most abstract multilayer models of two ANN features

ANN ARCHITECTURE MULTILAYER MODEL		
Layer Name	Layer No.	Feature of the layer
FFNN	1	Linear separability
RNN	2	Sophisticated architectures
HNN	3	FFNN – preprocessors and RNN – the main ANN
FeedForward Neural Networks Recurrent Neural Networks Hybrid Neural Networks		
ANN LEARNING PARADIGM MULTILAYER MODEL		
Layer Name	Layer No.	Feature of the layer
SL	1	Learning modes
UL	2	Input-output transform
HL	3	Combined learning rules (error correction, competitive)
Supervised Learning Unsupervised Learning Hybrid Learning		

III. Examples of Multilayer Models of ANN Features Based on the Degree of Their Sophistication

The following sections present ANN multilayer models in which the layers are numbered from the periphery (layer number 1) to the center (the core) according to the feature sophistication degree.

ANN Learning Paradigm Models by Tables

Additional information about the ANN learning process the reader can find in [3].

Multilayer Models of ANN Architectures and Learning Paradigms by Tables

Detailed description of these models is given in the following sections; see also [3]. The difference between the present paper and [3] is the presence of the *feature of the layer* (see chapter II).

ANN Architecture Multilayer Models by Tables

Table 3 gives a detailed description of the architecture which is analyzed by the authors in [4].

Table 3. ANN types of architectures

RECURRENT NEURAL NETWORKS MULTILAYER MODEL		
Layer Name	**Layer No.**	**Feature of the layer**
SLP	1	Simple architectures
MLP	2	Approximate optimum (Stochastic approximation)
RBF	3	Multivariable interpolation (Statistical approximation)
Single-Layer Perceptron Multilayer Perceptron Radial Basis Functions		
RECURRENT NEURAL NETWORKS MULTILAYER MODEL		
Layer Name	**Layer No.**	**Feature of the layer**
HNN	1	Speed not critical
ARTN	2	Pattern stability
WfdMM	3	Equal dimensions of the input and the output
KSOM	4	Data compression
Hebbian Neural Network Adaptive Resonance Theory Network Willshaw - von der Malsburg Map Kohonen's Self-Organizing Map		

ANN Learning Paradigm Models by Tables

Additional information about the ANN learning process the reader can find in [3].

IV. Industrial Tasks with ANN and Multilayer Models HNN

This chapter illustrates the both approaches of visualizing the multilayer models in the serial production – with graphs and with tables.

The Industrial Task Model Creation – An Environment for Industrial Multilayer Models

This section resumes the sequence of creating multilayer models for application tasks in the industry based on [4,5]: 1) the user gives the designer his task for the production device; 2) the designer analyses the physical nature of the task; 3) a solution is proposed based on the modern scientific paradigms in the area supported by developed mathematical models and oriented towards the existing technologies; 4) the task solution is adapted to concrete industrial producer(s) and a zero series is produced; 5) the designer takes under consideration opinions and remarks from the exploitation of the zero-series device(s) and the technological cycle is corrected in a corresponding way; 6) finally the negotiated industrial

Table 4. ANN supervised and unsupervised learning paradigms

UNSUPERVISED LEARNING MULTILAYER MODEL		
Layer Name	**Layer No.**	**Feature of the layer**
EC	1	Perceptron
HL	2	Linear discriminant analysis
BL	3	Statistical physics (optimization task)
CL	4	ART map, learning vector quantization
Error correction Hebbian learning Boltzmann learning Competitive learning		
UNSUPERVISED LEARNING MULTILAYER MODEL		
Layer Name	**Layer No.**	**Feature of the layer**
EC	1	Multilayer feedforward Sammon's projection
HL	2	Principal-component analysis & associative-memory learning
CL	3	ARTx, SOM, vector quantization
Error correction Hebbian learning Competitive learning		

production series are started. The last two steps may be iteratively repeated creating a sequence of production models which explains the role of the evolutionary and genetic approaches to the design of industrial tasks.

It is clear that the design of such models is complex and sophisticated implying strong penalties for the different production features. The authors chose the multilayer interpretation of industrial tasks due to the distinct hierarchial stratification of the goal task. There are cases which are not recommended for interpretations by multilayer models [3]. Such cases outline the natural bounds of the applicability for multilayer models. Table 5 is complemented by possible applications in other fields, e. g. production control.

Industrial Tasks, ANN and Multilayer Models

[5] investigates an application of multilayer models for modelling serial production of ANN. Fig. 2 and Table 5 introduce multilayer models of applications for production control by ANN. The graphical visualization may be applied in every possible case while the visualization by tables is applicable for *separate* multilayer models.

The serial production multilayer model may be presented not only by tables, but also by graphs [3]. Table 5 is chosen as an approach for this model of serial production for the sake of unity of the presentation style. Briefly said, the serial production model consists of a series of production models – the *generations* which demonstrate concrete tendencies in the *evolution* of the current production model. The mathematical description is given in [5]. It describes the penalty

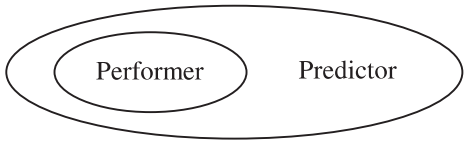


Fig. 2. Multilayer model for prediction and control

functions modeling the search space embedded in the unity global space. The core of the serial production model is the *ANN design model* (ANNDM) which determines the practical application of the ordered by the user concrete production model.

Table 5. Multilayer production models

SERIAL PRODUCTION MULTILAYER MODEL		
Layer Name	Layer **No.**	**Feature of the layer**
Production	1	Production
GEANNDM	2	Generalized Evolutionary ANN Design Model
ANNDM	3	ANN Design Model
MULTILAYER CONTROL MODEL		
Layer Name	**Layer No.**	**Feature of the layer**
P	1	Proportional control
PD, PI	2	Proportional-Differential or Proportional-Integral control
PID	3	Proportional-Integral-Differential control

V. Conclusions

Modelling goal tasks with multilayer models by two different types of presentations is introduced. Special attention is dedicated to the formulation of the ANN-learning problem with such models. Multilayer models of different levels of the ANN architecture and their learning paradigm are presented based on essential features for the levels which in turn may be modeled by multilayer models of lower levels.

References

[1] A. Jain, "Artificial Neural Networks: A Tutorial", *IEEE Computer*, pp. 31-44, March, 1996.

[2] B.Kosko, *Neural networks and fuzzy systems. A dynamic systems approach to machine intelligence, Part 1*, Englewood Cliffs, Prentice Hall, N.Y., 1992.

[3] S. L. Koynov, , Ch. D. Korsemov, Hr. I. Toshev and L. M. Kirilov, "Multi-Layer Models and Learning of The Artificial Neural Networks," *Proceedings of the XXXVII International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST 2002*, 2-4 October 2002, Nish, Yugoslavia, Pp. 113-116, 2002.

[4] S. Koynov, Ch. Korsemov and H. Toshev, "The design of the artificial neural networks as a basis for their generalized evolutionary model", *15th Int. Conf. on Syst. for Automation of*

Engineering and Research SAER'2001,Proc. 21-23 September 2001, Varna - St. Konstantin resort, Bulgaria, pp. 185-190, 2001.

[5] S. Koynov, Ch. Korsemov and H. Toshev, "The artificial-neural-networks-design generalized evolutionary model", *15th Int. Conf. on Syst. for Automation of Engineering and Research SAER'2001,Proc.* 21-23 September 2001, Varna - St. Konstantin resort, Bulgaria, pp. 191-195, 2001.

[6] S. Haykin, *Neural networks*. Englewood Cliffs, Macmillan Publishing Co., 696 p., 1994.

[7] Zb. Michalewicz, "The significance of the evaluation function in evolutionary algorithms", *Evolutionary algorithms*, L. D. Davis, K. De Jong, M. D. Vose, L. D. Whitley (eds.), Springer, pp. 151-166, 1999.

[8] E. Falkenauer, "Applying genetic algorithms to real-world problems", *Evolutionary algorithms*, L. D. Davis, K. De Jong, M. D. Vose, L. D. Whitley (eds.), Springer, pp. 65-88, 1999.

[9] D. Williams and J. B. Gomm, "The introduction of neural network projects in a degree of electrical and electronic engineering", *12th Int. Conf. on Systems for Automation of Engineering and Research SAER'98, Proc*. 19-20 September 1998, Varna - St. Konstantin resort, Bulgaria, pp. 102-106, 1998.

[10] J. R. McDonell, "Training Neural Networks with Weight Constraints", *Proc. of the First Annual Conference on Evolutionary Programming*, D. B. Fogel and W. Atmar (eds.), Evolutionary Programming Society, La Jolla, CA, pp.111-119, 1992.

[11] D.B. Fogel, E. C. Wasson and E. M. Boughton., "Evolving Neural Networks for Detecting Breast Cancer". *Cancer Letters*, v. 96, pp.49-53, 1995.

[12] M., Sipper, E. Sanchez, D. Mange, M. Tomassini, A. Perez-Uribe, and A. Stauffer, "A phylogenetic, ontogenetic, and epigenetic view of bioinspired hardware systems", *IEEE Trans. Evol. Comput.*, V. 1:1, pp.83-97, 1997.

Artificial Neural Networks-Training – Errors, Convergence and Genetic Approaches

Hristo I. Toshev[1], Chavdar D. Korsemov[2] and Stefan L. Koynov[3]

Abstract – **The paper introduces an approach to enhance and speed up the training process of the goal artificial neural networks (ANN). The training and the operation of the ANN is evaluated on the basis of temporally sequential copies of the ANN parameters (generations, offsprings) by means of multilayer and other models. The introduced terminology and mathematical formalism concern the errors, the convergence and the genetic approaches to train ANN.**

Keywords – **multilayer models, neural networks, penalty functions, training.**

I. Introduction

Artificial neural networks (ANN) learning is the most time-consuming step (if any).

Two determining factors: the goal ANN architecture [1] and the corresponding mathematical description [2] feed the learning process. The architecture is the most directly connected with the physical 'nature' of the ANN while the mathematical formalism presents in the discrete time domain the relations between the information streams. The mathematical model is decisive for the ANN learning because it is *defined by the learning rule* but it itself *defines the learning algorithm*.

The ANN learning process consists of the learning paradigm, the learning rule and the learning algorithm. The learning paradigm presets the possible learning rules which in turn preset the possible learning algorithms. The learning paradigm is the most abstract feature and the learning algorithm is the most concrete one. Starting from left in a direction to the right the learning characteristics become more concrete and this corresponds to a 'descent' from the peak of a pyramidal structure towards its base; the peak is analogous to the learning paradigm and the base - to the concrete application task. The very 'descent' during the ANN learning corresponds to a draw up to the final ANN design.

The scheme in Fig. 1 presents the links between these aspects in the learning process for ANN with multilayer models.

The mathematical description is of a principal importance

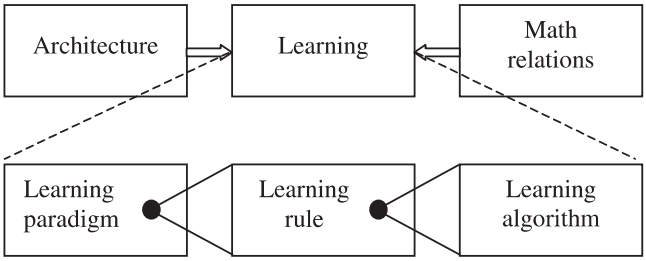


Fig. 1. The ANN learning process with multilayer models

for the self-organizing maps when the mathematical formalism may be presented according to Table 1. The table introduces numbering the layers in the model which is based on the degree of sophistication of the feature for the layer.

Table 1. Most abstract ANN learning process

ANN LEARNING PARADIGM MULTILAYER MODEL		
Layer No.	**Layer Name**	**Feature of the layer**
1	SL	Learning modes
2	UL	Input-output transform
3	HL	Combined learning rules (error correction, competitive)
Supervised Learning Unsupervised Learning Hybrid Learning		

II. ANN Training: Errors, Convergence and Genetic Approaches

In the training phase at an equal predefined number of iterations control copies of the network parameters, the input and output vectors are made. So a series of successive temporal generations of the goal ANN is obtained which is presented in the form of a multilayer model; the interlayer connections are modeled with penalty functions. The definitions which follow below define different types of errors for the artificial neural networks (ANN) training process. They link the ANN with the genetic and evolutionary approaches to the problem of the ANN learning.

DEFINITION 1: Control total error $E_g^{(i)}$ is the ratio of the erroneous outputs (*misses*) $m^{(i)}$ to the correct outputs (*hits*)

[1]Hristo I. Toshev is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 29A, 1113 Sofia, Bulgaria, E-mail: toshev@iinf.bas.bg

[2]Chavdar D. Korsemov is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 29A, 1113 Sofia, Bulgaria, E-mail: korsemoviinf.bas.bg

[3]Stefan L. Koynov is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 29A, 1113 Sofia, Bulgaria, E-mail: baruch@iinf.bas.bg

$h^{(i)}$ for the i-th ANN state generation in the training process, i.e. $E_g^{(i)} = m^{(i)} / h^{(i)}$.

NOTE. The value of $E_g^{(i)}$ usually does not exceed 10^{-2}.

DEFINITION 2: Accumulated total error E_g^i for the last i generations is the sum of the control total errors $E_g^{(i)}$ for the last i generations, i.e. $E_g^i = \sum_i E_g^{(i)}$.

The accumulated total error E_g^i serves both as an indicator of the ANN parameters state evolution and implicitly for the expected number of state offsprings of the being trained ANN.

DEFINITION 3: Error of the trained ANN is any error E_t which does not exceed the allowed operational error of the already trained ANN, i.e. $E_t < E_d$.

COROLLARY 1. The error of the already trained ANN E_t is less than the maximal value of the control total error $\left(E_g^{(i)}\right)_{\max}$ for the last I state generations, i.e. $E_t < \left(E_g^{(i)}\right)_{\max}$.

The two theorems below evaluate the number of iterations necessary to train an ANN and an interval of the estimate for the necessary ANN state offsprings is proposed.

THEOREM 1: Let the abscissa axis is the number of iterations to change the weights between two successive states of the being trained ANN. Then the iteration interval between any two successive states during the training of a single ANN $t_g^{(i)}$ must be less than the product of the overall added in the successive state generations neurons $N^{a(i)}$ by a factor of the number of the weight corrections for the these states $n_g^{(i)} : N^{a(i)} \cdot n_g^{(i)} > t_g^{(i)}$.

Proof: The worst case is when the new state generation has just *one* new added neuron. If this is the case of training for the next successive generations then the iteration interval for adding new neurons will be estimated based on the statistics of the averagely added neurons multiplied by the number of the average weight corrections per one added neuron. ■

Theorem 2: The maximal number of state generations for the training of an ANN I is equal to the product of the overall added in the next state generation neurons $N^{a(i)}$ multiplied by the number of the weight corrections $n_g^{(i)}$ divided by the iteration interval between two successive states of the being trained ANN $t_g^{(i)} : \dfrac{N^{a(i)} \cdot n_g^{(i)}}{t_g^{(i)}} = I \geqslant 1$.

Proof: The proof follows from the previous Theorem 1 by dividing the number of iterations for all new added neurons $N^{a(i)} \cdot n_g^{(i)}$ and the iteration interval between two state generations of the ANN $t_g^{(i)}$. ■

COROLLARY 2: The minimal number of state generations for training a single ANN equals to *one:* $I_{\min} = 1$. It is obtained if the total number of the added neurons in the ANN training process is obtained during the ANN state generation number two $(n = 2)$. It means that $I_{\min} \sim N^{a(i)} = N^{a(2)} = 1$ if $n_g^{(2)} = t_g^{(2)}$: the iteration interval between two successive ANN state generations $t_g^{(i)}$ is adjusted equal to the current number of weight corrections $n_g^{(i)}$ for tuning the ANN to achieve $E_t < E_d$.

The following below definitions link the types of convergence (or divergence) in the automation theory and their analogs in the genetic approach.

DEFINITION 4: Genetic divergence is the tendency the values of the traced parameters to deviate from their average stable states for an offspring series for the species for large values of the number of generations.

COROLLARY 3: The variance and the r.m.s. divergence increase if the genetic divergence increases.

DEFINITION 5: Genetic convergence is the tendency the values of the traced parameters to converge to their average stable states for an offspring series for the species for large values of the number of generations.

COROLLARY 4: The variance and the r.m.s. divergence decrease if the genetic divergence decreases.

DEFINITION 6: Asymptotic genetic convergence implies that the traced parameters are genetically convergent and that their values lie in a predefined small neighborhood around their average values for an offspring series for the species for large values of the number of generations.

III. ANN Training: Multilayer and Other Models

The approach with penalty functions admits an interpretation with multilayer models for exploring the ANN training process; the definitions and the theorems in the previous chapter allow the evaluation of the penalty functions which comprise the multilayer model for the ANN training process. The multilayer approach is already implied by the authors to model a serial industrial application [3].

The series of penalty functions between the separate temporal generations described with the multilayer model naturally converges to a penalty function corresponding to the already trained ANN, so the temporal series of the parameters, of the input and output vectors of the ANN are the analog to the popular mathematical series. The goal of the method is to achieve an estimate of the penalty functions of an arbitrary given sequence number of the ANN state generations by varying the number of iterations between every two state replicas of the being trained ANN on condition that the penalty functions for the first several offsprings of the state are obtained.

The ANN training multilayer model consists of layers which correspond to the successive ANN state parameter records (offsprings). The successive layers may be numbered in two mutually opposite directions: from the periphery to the core or v.v. The default direction of numbering is based on the physical nature of the model [3]; in the ANN training process it is the number of iterations, therefore the greater layer numbers correspond to the successive ANN state generations. Let the whole search space is denoted with S and the feasible subspaces of the solutions are denoted with F. Then the following mathematical description may be formulated for the case of the ANN training process from the point of view with penalty functions:

$$eval\left(\overline{X}\right) = f\left(\overline{X}\right) + \sum_{l=1}^{L} a_l \left[\lambda\left(t\right) \sum_{j=1}^{m} f_j^2\left(\overline{X}\right)\right]^l \qquad (1)$$

Here: $eval\left(\overline{X}\right)$ – feasible and unfeasible solutions if $\overline{X} \in F$ is the optimal solution of the general nonlinear programming model with continuous variables; $f\left(\overline{X}\right)$ – goal function for optimization; $\lambda\left(t\right)$ – updated every generation t in the following way [4]:

$$\lambda\left(t+1\right) = \\ = \begin{cases} \left(1/\beta_1\right) \cdot \lambda\left(t\right), \text{ if } \overline{B}\left(i\right) \in F \text{ for all } t\text{-}k\text{+}1 \leqslant i \leqslant t \\ \beta_2 \cdot \lambda\left(t\right), \text{ if } \overline{B}\left(i\right) \in S - F \text{ for all } t\text{-}k\text{+}1 \leqslant i \leqslant t \\ \lambda\left(t\right), \text{ else} \end{cases}$$

$f_j\left(\overline{X}\right)$ – constraint violation measure for the j-th constraint such that [5]:

$$f_j\left(\overline{X}\right) = \begin{cases} \max\left\{0, g_j\left(\overline{X}\right)\right\}, & \text{if } 1 \leqslant j \leqslant q \\ \left|h_j\left(\overline{X}\right)\right|, & \text{if } q\text{+}1 \leqslant j \leqslant m \end{cases}; \qquad (2)$$

Here $g_j\left(\overline{X}\right) \leqslant 0, \quad j = 1, ..., q$ and $h_j\left(\overline{X}\right) = 0, \quad j = q+1, ..., m$ comprise a set of additional constraints $m \geqslant 0$ the intersection of which with S defines the feasible set F;

l – indicator of the constraint type with upper bound $L = 2$:

$$l = \begin{cases} 1 : \text{inside a given layer (inside an ANN state generation)} \\ 2 : \text{between two layers inside the ANN learning multilayer model (between two successive ANN state generations)} \end{cases}$$

a_l – coefficient array reflecting the weights of the different constraint levels in the formula. It is adjusted heuristically.

Besides multilayer models other types of models can be introduced to model the ANN training process [6]: the homogeneous features (belonging to one class or one group) are ordered in hierarchical systems and the heterogeneous features (belonging to different classes or groups) are clustered in multilayer models; the size of the paper does not allow the detailed introduction of them

IV. Conclusions

The paper presents an approach of the authors to enhance and accelerate the goal artificial neural networks (ANN) training. The training and the operation of ANN are estimated via their successive temporal generations by multilayer and other models. The goal of the method is to achieve an estimate of the penalty functions of an arbitrary given sequence number of the ANN state generation by varying the number of iterations between every two state replicas of the being trained ANN on condition that the penalty functions for the first several offsprings of the state are obtained; an interval of the estimate for the necessary ANN state offsprings is proposed.

References

[1] A. Jain, "Artificial Neural Networks: A Tutorial", *IEEE Computer*, pp. 31-44, March, 1996.

[2] B.Kosko, *Neural networks and fuzzy systems. A dynamic systems approach to machine intelligence, Part 1*, Englewood Cliffs, Prentice Hall, N.Y., 1992.

[3] S. Koynov, Ch. Korsemov and H. Toshev, "The artificial-neural-networks-design generalized evolutionary model", *15th Int. Conf. on Syst. for Automation of Engineering and Research SAER'2001,Proc.* 21-23 September 2001, Varna - St. Konstantin resort, Bulgaria, pp. 191-195, 2001.

[4] J.C.Bean and A. B. Hadj-Alouane, "A dual genetic algorithm for bounded integer programs", Department of Industrial and Operations Engineering, The University of Michigan, TR 92-53, 1992.

[5] Zb. Michalewicz, "The significance of the evaluation function in evolutionary algorithms.",*Evolutionary algorithms*, L. D. Davis, K. De Jong, M. D. Vose, L. D. Whitley (eds.), Springer, pp. 151-166, 1999.

[6] S. L. Koynov, Ch. D. Korsemov, Hr. I. Toshev and L. M. Kirilov, "Multi-Layer Models and Learning of The Artificial Neural Networks", *Proceedings of the XXXVII International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST 2002*, 2-4 October 2002, Nish, Yugoslavia, pp. 113-116, 2002.

[7] S. Koynov, Ch. Korsemov and H. Toshev, "The design of the artificial neural networks as a basis for their generalized evolutionary model", *15th Int. Conf. on Syst. for Automation of Engineering and Research SAER'2001,Proc.* 21-23 September 2001, Varna - St. Konstantin resort, Bulgaria, pp. 185-190, 2001.

[8] S. Haykin, *Neural networks*. Englewood Cliffs, Macmillan Publishing Co., 1994.

[9] E. Falkenauer, "Applying genetic algorithms to real-world problems", *Evolutionary algorithms*, L. D. Davis, K. De Jong, M. D. Vose, L. D. Whitley (eds.), Springer, pp. 65-88, 1999.

[10] D.B. Fogel, E. C. Wasson and E. M. Boughton., "Evolving Neural Networks for Detecting Breast Cancer". *Cancer Letters*, v. 96, pp. 49-53, 1995.

[11] D. Williams and J. B. Gomm, "The introduction of neural network projects in a degree of electrical and electronic engineering", *12th Int. Conf. on Systems for Automation of Engineering and Research SAER'98, Proc*. 19-20 September 1998, Varna - St. Konstantin resort, Bulgaria, pp. 102-106, 1998.

[12] I. Baruch, A. Martinez and B. Nenkova, "Adaptive Neural Control with Integral-Plus-State Action", *Cybernetics and Information Technologies*, Vol. 2, No. 1, Sofia, pp. 37-48, 2002

A Review of Selection, Mutation and Recombination in Genetic Algorithms

Milena Karova[1]

Abstract – **Genetic Algorithms have been applied to a number of optimization problems. This paper presents a new analysis of principal genetic operators: selection, mutation and recombination. New methods are described. Recombination (crossover) is also a multitude of methods to create a coherent set of genes from two parent sets. GA selection perform the equivalent role to natural selection. Mutation enables the GA to maintain diversity whilst also introducing some random search behavior.**

Keywords – **genetic algorithms, selection, mutation, crossover, recombination, inversion, genotype, phenotype.**

I. Introduction

A Genetic Algorithms are a family of computational models inspired by evolution. These algorithms encode a potential solution to a specific problem on a simple chromosome – like data structure and apply a recombination operators to these structures so as to preserve critical information. Genetic algorithms (GA) are often viewed as function optimizers, although the range of problems to which GA have been applied is quite broad.

An implementation of GA begins with a population of (typically random) chromosomes. One then evaluates these structures and allocates reproductive opportunities in such a way that those chromosomes which represent a better solution to the target problem are "given" more chances to "reproduce" than those chromosomes which are poorer solutions. The "goodness" of a solution is typically defined with respect to the current population.

The term Genetic Algorithm has two meanings. In a strict interpretation the genetic algorithm refers to a model introduced and investigated by John Holland (1975) [5]. It is still the case that most of the existing theory for GA applies either solely or primarily to the model introduced by Holland. It comes from the fact that individuals are represented as strings of bits analogous to chromosomes and genes. In addition to recombination by crossover, we also throw in random mutation of these bit-strings every so often.

A GA is any population-based model that uses selection and recombination operators to generate new simple points in a search space. Many GA models have been introduced by researchers largely working from an experimental perspective. Many of these researchers are application oriented and are typically interested in GA as optimization tools.

[1]Milena Karova is with the the department of Computer Science, Studentska 1, Technical University Varna Email: mkarova@ieee.bg

II. Functioning of GA

The fitness or objective function is used to map the individual's bit strings into a positive number which is called the individual's fitness. There are two steps involved in this mapping (however in some problems these two steps are essentially accomplished as one). The first step we will call "decoding" and the second, "calculating fitness". To understand decoding it helps to partition individuals into two parts commonly called the genotype or genome and the phenotype. These terms are borrowed from biology. The genotype, as its name implies, specifically refers to an individual's genetic structure or for our purpose, the individual's bit string(s).

The phenotype refers to the observable appearance of an individual (pheno comes from Greek for "to show" - phainein).

The principle of GA is simple:

1. Encoding of the problem in a binary string.
2. Random generation of a population. This one includes a genetic pool representing a group of possible solutions.
3. Reckoning of a fitness value for each subject. It will directly depend on the distance to the optimum.
4. Selection of the subjects that will mate according to their share in the population global fitness.
5. Genomes crossover and mutations.
6. And then start again from point 3.

The functioning of a GA can also be described in reference to genotype (GTYPE) and phenotype (PTYPE) notions.

1. Select pairs of GTYPE according to their PTYPE fitness.
2. Apply the genetic operators (crossover, mutation. . .) to create new GTYPE.
3. Develop GTYPE to get the PTYPE of a new generation and start again from 1.

III. Selection

Selection is one of the most important elements of all GA's. Selection determines which individuals in the population will have all or some of its "genetic material" passed on to the next generation of individuals. The object of the selection

method employed in a GA is to give exponentially increasing trials to the fittest individuals. The most common way in which this is accomplished is by a technique called "roulette-wheel" selection. As you will see, it is the implementation of roulette-wheel selection which necessitates positive fitness values where higher values indicate greater fitness. Roulette wheel selection gets its name from the fact that the algorithm works like a roulette wheel in which each slot on the wheel is paired with an individual in the population. This is done such that the size of each slot is proportional to the corresponding individuals fitness. It should be obvious then that maximization problems fit directly into this paradigm - larger slot implies larger fitness. Negative values are not allowed because how can you have a slot of negative size?

A common way to implement roulette wheel selection is to:

1. Sum up all the fitness values in the current population, call this value SumFitness. SumFitness is in effect the total area of the roulette wheel.

2. Generate a random number between 0 and 1, called Rand.

3. Multiply SumFitness by Rand to get a number between 0 and SumFitness which we will call RouletteValue (RouletteValue = SumFitnesss x Rand). Think of this value as the distance the imaginary roulette ball travels before falling into a slot.

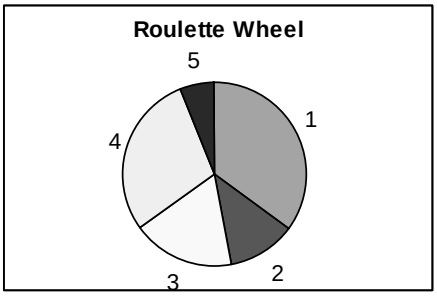
4. Finally we sum up the fitness values (slot sizes) of the individuals in the population until we reach an individual which makes this partial sum greater or equal to RouletteValue [Fig. 1]. This will then be the individual that is selected.

It is not always intuitively obvious that this algorithm actually implements a weighted roulette wheel. To see that it does lets look at some extreme situations. Imagine an individual, I, whose fitness is equal to SumFitness (implying all other individuals have a fitness of zero).

Clearly no matter what number is generated for RouletteValue, I will always throw the partial sum over the top, thus having a selection probability of 1. This corresponds to a roulette wheel with just one slot.

On the other extreme, an individual, i, with fitness zero can never cause the partial sum to become greater than RouletteValue, so it has a zero probability of getting selected. This corresponds to a slot that does not exist on the wheel. All other individuals between these extremes will have a probability of throwing the partial sum over the top that is proportional to their size, which is exactly how we would expect a weighted roulette wheel to behave.

The important quality of all legitimate GA selection techniques is to reward fitter individuals by letting them reproduce more often. This is one of the important ways in which a GA differs from random search.



Individual	Fitness	Slot Size %
1	30	35
2	10	12
3	15	18
4	25	29
5	5	6
SumFittnes:	85	

Fig. 1. Roulette Wheel Selection with Five Individuals of Varying Fitness

IV. Recombination (Crossover)

A. 1-point crossover

The traditional GA uses 1-point crossover, where the two mating chromosomes are each cut once at corresponding points, and the sections after the cuts exchanged. It is here that two individuals selected in the previous step are allowed to mate to produce offspring. Crossover is the process by which the bit-strings of two parent individuals combine to produce two child individuals.

There are many ways in which crossover can be implemented. Some of the ways are broadly applicable to all types of problems and others are highly problem specific. Here we will talk about the most primitive (but also highly effective) form of crossover, single-point crossover [Fig. 2]. Single point crossover starts by selecting a random position on the bit string, called a cut point or cross point. The bits from the left of the cut point on parent1 are combined with the bits from the right of the cut point in parent2 to form child1. The opposite segments are combined to form child2.

Thus child1 and child2 will tend to be different from either of their parents yet retain some features of both. If the parents each had high fitness (which is likely by the fact that they were selected) then there is a good chance that at least one of the children is as fit or better than either parent. If this is the case, then selection will favor this child's s procreation, if not than selection will favor the child's extinction.

There is of course a possibility (albeit small) that most or

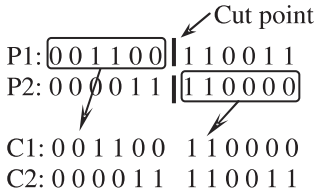


Fig. 2. Example of single-point crossover

all of the crosses produce children of less fitness. To counter this possibility, a parameter, px – the probability of crossover, is introduced. Before crossover is performed a simulated coin is flipped that is biased to come up heads (TRUE) with probability px. If it does, then crossover is performed, if not than the parents are passed into the next generation unchanged. Since without crossover there is no hope for advancement, px is usually high ($0.5 < px < 1.0$).

However, many different crossover algorithms have been devised, often involving more than one point. [2]

B. 2-point crossover

In 2-point crossover, (and multi-point crossover in general), rather than linear strings, chromosomes are regarded as loops formed by joining the ends together. To exchange a segment from one loop with that from another loop requires the selection of two cut points, as shown in Fig. 3.

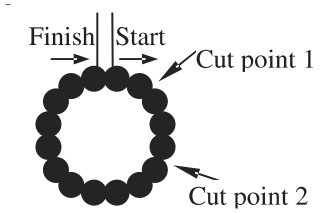


Fig. 3. Chromosome Viewed as Loop

In this view, 1-point crossover with one of the cut points fixed at the start of the string. Hence 2-point crossover performs the same task as 1-point crossover (i.e. exchanged a single segment), but is more general. A chromosome considered as a loop can contain more building blocks – since they are able to "wrap around" at the end of the string. Researchers now agree that 2-point crossover is generally better than 1-point crossover.

C. Uniform crossover

Uniform crossover is radically different to 1-point crossover. Each gene in the offspring is created by copying the corresponding gene from one of the other parent, chosen according to a randomly generated crossover mask. Where there is a 1 in the crossover mask, the gene is copied from the first parent, and where there is a 0 in the mask, the gene is copied from the second parent, as shown in Fig. 4. The process is repeated with the parents exchanged to produce the second offspring. A new crossover mask is randomly generated for each pair of parents.

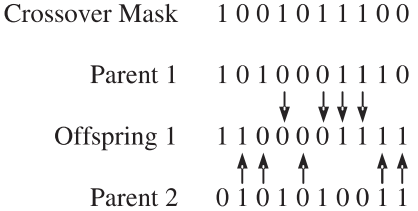


Fig. 4. Uniform Crossover

Offspring therefore contain a mixture of genes from each parent. The number of effective crossing points is not fixed, but will average L/2 (where L is the chromosome length).

D. Which technique is best?

Arguments over which is the best crossover method to use still rage on. Syswerda [7] argues in favor of uniform crossover. Under uniform crossover the number of specify bit values, is equally likely to be disrupted. uniform crossover has the advantage that the ordering of genes is entirely irrelevant. This means that reordering operators such as inversion are unnecessary. GA performance using 2-point crossover drops dramatically if the recommendations of the building block hypothesis are not adhered to. Uniform crossover , on the other hand, still performs well – almost as well as 2-point crossover used on a correctly ordered chromosome. Uniform crossover therefore appears to be robust.

Spears & DeJong [6] are very critical of multi-point and uniform crossover. They stick by the theoretical analyses which show 1- and 2-point crossover are optimal. 2-point crossover will perform poorly when the population has largely converged, due to reduced crossover productivity.

In a slightly later paper DeJong &Spears [6] conclude that modified 2-point crossover is best for large populations, but the increased disruption of uniform crossover is beneficial if the population size is small (in comparison to the problem complexity), and so gives a more robust performance.

Goldberg [2] describes a rather different crossover operator, partially matched crossover (PMX), for use in order-based problems. (In an order-based problem, such as the traveling salesperson problem, gene values are fixed, and the fitness depends on the order in which they appear). In PMX it is not the values of the genes which are crossed, but the order in which they appear. Offspring have genes which inherit ordering information from each parent. This avoids the generation of offspring which violate problem constraints.

V. Mutation

Another important GA operator is mutation. Although mutation is important, it is secondary to crossover. Many people have the erroneous belief that mutation plays the central role in natural evolution. This is simply not the case. The reason is that mutation is more likely to produce harmful or even destructive changes than beneficial ones. An environment with high mutation levels would quickly kill off most if not all of the organisms. In genetic algorithms, high mutation rates cause the algorithm to degenerate to random search [Fig. 5].

Unlike crossover, mutation is a unary operator - it only acts on one individual at a time. As bits are being copied from a

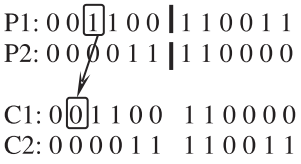


Fig. 5. Example of mutation

parent individual to a child. a weighted coin is flipped, if it comes up TRUE than the bit is inverted before copying. The probability of the simulated coin coming up TRUE is called pm – the probability of mutation. As previously stated pm is small ($0 \leq$ pm ≤ 0.1).

VI. Inversion

Another genetic operator is called inversion. Inversion is not used as often as crossover and mutation in most GA's. Inversion is a process that shifts the locus of one or more gene in a chromosome from one point to another. This does not change the meaning of the genotype in the sense that a genotype before and after inversion will still decode to they same phenotype. If this is true, then why bother with inversion at all? The theory behind inversion is that there are groups of two or more genes in a chromosome that work together to yield a high fitness. If these genes are physically close together than single point crossover is much less likely to disturb these groups. Although this argument seems reasonable, inversion used in practice as achieved very mixed results. This is why many GA's ignore inversion all together.

VII. Conclusion

GA are original systems based on the supposed functioning of the Living. The method is very different from classical optimization algorithms.

- Use of the encoding of the parameters, not the parameters themselves.
- Work on a population of points, not a unique one.
- Use the only values of the function to optimize, not their derived function or other auxiliary knowledge
- Use probabilistic transition function not determinist ones.
- Using selection alone will tend to fill tend the population with copies of the best individual from the population.
- Using selection and crossover operators will tend to cause the algorithms to converge on a good but suboptimal solution.
- Using mutation alone induces a random walk through the search space.
- Using selection and mutation creates a parallel, noise-tolerant, hill climbing algorithm.

References

[1] Goldberg D. L., "Genetic Algorithms in Search, Optimization, and Machine Learning", Addison-Wesley, 1989

[2] Goldberg D, "Web Courses", http://www.engr.uiuc.edu/OCEE, 2000.

[3] Mitchell M., "An Introduction to Genetic Algorithms", Massachusetts Institute of Technology, 1996.

[4] Paechter B., Rankin R., Cumming A., "Timetabling the Classes of an Entire University with an Evolutionary Algorithm", Napier University, Edinburgh, Scotland.

[5] Holland, John H., "Adaption in Natural and artificial systems", the Mit Press, 1992.

[6] Spears, W. M. and DeJong K., An analysis of multi-point crossover", Foundations of Genetic Algorithms, pp. 301-315, Morgan Kaufmann, 1999.

[7] Syswerda, G., Uniform crossover in genetic algorithms, Procedings of the Third International Conference on Genetic Algorithms, pp. 2-9, 1995

[8] Ladd, S. R., Genetic Algorithm in C++, 1999-2000

Attack on the Polyalphabetic Substitution Cipher Using a Parallel Genetic Algorithm

A. Dimovski[1], D. Gligoroski[2]

Abstract – **In this paper is presented an automated attack on the polyalphabetic substitution cipher. The property which make this cipher vulnerable, is that it is not sophisticated enough to hide the inherent properties or statistics of the language of the plaintext. The attack described here effectively reduces the complexity of a polyalphabetic substitution cipher attack to that of a monoalphabetic one, if there is a computer with *B* processing nodes, where *B* is the period of the polyalphabetic substitution cipher.**

Keywords – **Polyalphabetic substitution cipher, Cryptanalysis, Parallel genetic algorithm**

I. Polyalphabetic Substitution Ciphers

The polyalphabetic substitution cipher is a simple extension of the monoalphabetic one. The difference is that the message is broken into blocks of equal length, say B, and then each position in the block $(1, \ldots, B)$ is encrypted (or decrypted) using a different simple substitution cipher key. The block size B is often referred to as the period of the cipher.

An example of a polyalphabetic substitution cipher is shown on Table 1. The block size (i.e., B) is chosen to be three, and Table 1 gives an example key and shows the corresponding encryption, it is clear that the decryption process is reversal of the encryption.

Table 1. Example of the polyalphabetic substitution cipher key and encryption process

KEY:	
Plaintext:	
ABCDEFGHIJKLMNOPQRSTUVWXYZ_	
Ciphertext:	
PQOWIEURYTLAKSJDHFGMZNX_BCV	(Position 1)
LP_MKONJIBHUVGYCFTXDRZSEAWQ	(Position 2)
GFTYHBVCDRUJNXSEIKM_ZAWOLQP	(Position 3)
ENCRYPTION:	
Position:	12312312312
Plaintext:	HOW_ARE_YOU
Ciphertext	RYWVLKIQLJR

The number of possible keys for a polyalphabetic substitution cipher using an alphabet size of 27 and a block size of B is $27!^B$. This is significantly greater than the simple substitution cipher with 27! possible keys, especially when the period B is large. The polyalphabetic substitution cipher is somewhat more difficult to cryptanalyse than the simple substitution cipher because of the independent keys used to encrypt successive characters in the plaintext, but it is still relatively simple to cryptanalyse the polyalphabetic substitution cipher based on the n-gram statistics of the plaintext language.

So, despite the monoalphabetic substitution cipher where every bigram (for example _A) is mapped to the same encrypted bigram each time, this is not the case for the polyalphabetic substitution cipher, where the encrypted value of a bigram is dependent upon two factors: the individual key values and the position of the characters within the block.

The polyalphabetic substitution cipher is simply a number of simple substitution ciphers operating on the different positions within each block. One possible attack strategy, then, is to solve each of the simple substitution ciphers in parallel. Here we will use a parallel genetic algorithm to attack the polyalphabetic substitution cipher.

II. Genetic Algorithms

The genetic algorithm is based upon Darwinian evolution theory. The genetic algorithm is modelled on a relatively simple interpretation of the evolutionary process, however, it has proven to be a reliable and powerful optimisation technique in a wide variety of applications.

Holland [1] in 1975, was first who proposed the use of genetic algorithms for problem solving. Goldberg [2] and DeJong [3] were also pioneers in the area of applying genetic processes to optimisation. Over the past twenty years numerous applications and adaptations of genetic algorithms have appeared in the literature.

Consider a pool of genes that have the ability to reproduce, are able to adapt to environmental changes and, depending on their individual strengths, have varying lifespans. In such an environment only the fittest will survive and reproduce giving, over time, genes that are stronger and more resilient to conditional changes. After a certain amount of time the surviving genes could be considered "optimal" in some sense. This is the model used by the genetic algorithm, where the gene is the representation of a solution to the problem being optimised.

As with any optimisation technique there must be a method of assessing each solution. The assessment technique used by a genetic algorithm is usually referred to as the "fitness function". The aim is always to maximise the fitness of

[1]Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University Arhimedova b.b., PO Box 162, 1000 Skopje, Macedonia adimovski@ii.edu.mk

[2]Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University Arhimedova b.b., PO Box 162, 1000 Skopje, Macedonia gligoroski@yahoo.com

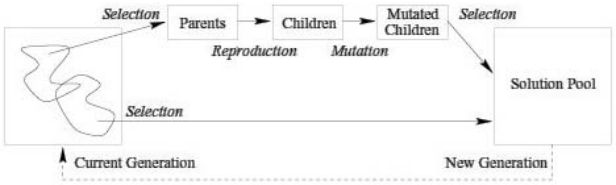


Fig. 1. The Evolutionary Process.

the solutions in the solution pool.

Fig. 1 gives an indication of the evolutionary processes used by the genetic algorithm. During each iteration of the algorithm the processes of selection, reproduction and mutation each take place in order to produce the next generation of solutions. The actual method used to perform each of these operations is very much dependent upon the problem being solved and the representation of the solution.

An algorithmic representation of the genetic algorithm is given in Fig. 2. This description is independent of any solution representation, fitness function, selection scheme, reproduction scheme and mutation scheme. Each of these will be described in detail where the genetic algorithm has been applied.

1. Initialize algorithm variables: *G* the maximum number of generations to consider, *M* the solution pool size and any other problem dependent variables.

2. Generate an initial solution pool containing *M* candidate solutions. This initial pool can be generated randomly or by using a simple known heuristic for generating solutions to the problem at hand. This solution pool is now referred to as the *current* solution pool.

3. For *G* iterations, using the current pool:

   a) Select a breeding pool from the current solution pool and make pairings of parents.
   b) For each parental pairing, generate a pair of children using a suitable mating function.
   c) Apply a mutation operation to each of the newly created children.
   d) Evaluate the fitness function for each of the children.
   e) Based on the fitness of each of the children and the fitness of each of the solutions in the current pool, decide which solutions will be placed in the new solution pool. Copy the chosen solutions into the new solution pool.
   f) Replace the current solution pool with the new one. So, the new solution pool becomes the current one.

4. Choose the fittest solution of the final generation as the best solution.

Fig. 2. The Genetic Algorithm

III. A Parallel Genetic Algorithm Attack

We will attack the polyalphabetic substitution cipher using a number of genetic algorithms running in parallel, each solving a different part of the problem. Fig. 3 is a pictorial representation of this strategy with M GA's running in parallel and communicating every $k$iterations.

Let's consider a polyalphabetic substitution cipher consisting of M monoalphabetic or simple substitution ciphers. There will then be M genetic algorithms (call them GA1, GA2, ..., GAM) solving each of the M simple substitution ciphers. GAj ($1 < j < B$), which is attempting to find the key to the cipher of position j, in determining the cost of each of the solutions in its pool, GAj uses the current best key from each of its neighbours to find the bigram and trigram statistics.

Before implementing the parallel attack a number of design problems have to be solved.

A. Suitability Assessment

The technique used to compare candidate keys is to compare n-gram statistics of the decrypted message with those of the language (which are assumed known). Equation 1 is a general formula used to determine the suitability of a proposed key (k). Here, A denotes the language alphabet (i.e., for English, [A, ..., Z, _], where _ represents the space symbol), K and D denote known language statistics and decrypted message statistics, respectively, and the indices u, b and t denote the unigram, bigram and trigram statistics, respectively. The values of α, β and γ allow assigning of different weights to each of the three n-gram types.

$$C_k = \alpha \cdot \sum_{i \in A} \left| K^u_{(i)} - D^u_{(i)} \right| + \beta \cdot \sum_{i,j \in A} \left| K^b_{(i,j)} - D^b_{(i,j)} \right| + \gamma \cdot \sum_{i,j,k \in A} \left| K^t_{(i,j,k)} - D^t_{(i,j,k)} \right| \quad (1)$$

Spillman [4], in his attack on the simple substitution cipher use Eq. (2). This equation is based on unigram and bigram statistics.

$$C_k \approx \alpha \cdot \sum_{i \in A} \left| K^u_{(i)} - D^u_{(i)} \right| + \beta \cdot \sum_{i,j \in A} \left| K^b_{(i,j)} - D^b_{(i,j)} \right| \quad (2)$$

The only difference between these assessment functions is the inclusion of different statistics. In general, the larger the n-grams, the more accurate the assessment is likely to be. It is usually an expensive task to calculate the trigram statistics - this is, perhaps, why they are omitted in Eq. (2). The complexity of determining the fitness is $O(N^3)$ (where N is the alphabet size) when trigram statistics are being determined, compared with $O(N^2)$ when bigrams are the largest statistics being used.

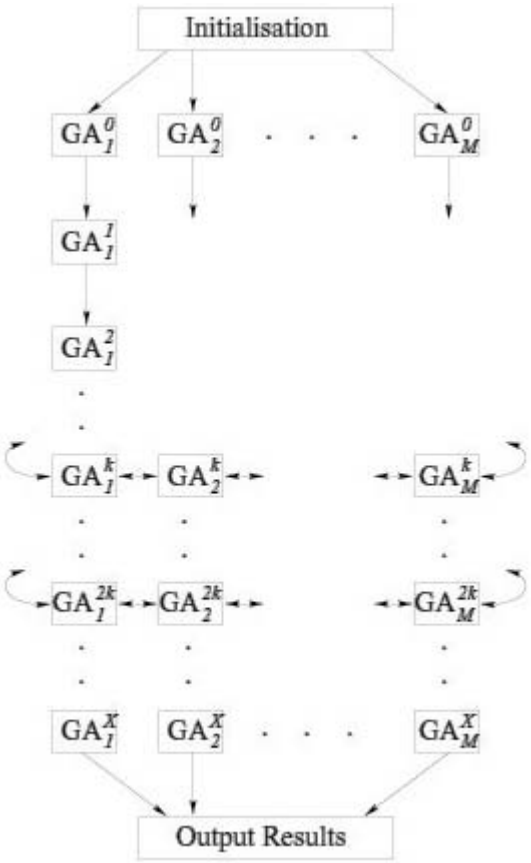


Fig. 3. A parallel genetic algorithm

In the case of the polyalphabetic substitution cipher, without knowledge of the keys for the two adjacent block positions it is impossible to determine bigram or trigram statistics. To overcome this problem the following strategy is used:

1. Initially only unigram statistics are used in determining the cost of the solutions in any pool.
2. Every k iterations of each GA, the most fit solution in the current pool is sent to each of the neighbouring GA's. Each GA has knowledge of the entire ciphertext message so it is able to determine a fitness based on unigram, bigram and trigram statistics using ciphertext characters in its position, the position to the left and the position to the right.

B. The Reproduction Process

The mating function utilised here is similar to the one proposed by Spillman [4], who use a special ordering of the key. The characters in the key string are ordered such that the most frequent character in the ciphertext is mapped to the first element of the key (upon decryption), the second most frequent character in the ciphertext is mapped to the second element of the key, and so on. The reason for this ordering will become apparent upon inspection of the mating function. For example, the key FGHIJKLMNOPQRSTUV_WXYZAB indicates that the most frequent character in the ciphertext represents a plaintext F; the second most frequent character in the ciphertext represents a plaintext G, etc.

Given two parents constructed in the manner just described, the first element of the first child is chosen to be the one of the first two elements in each of the parents, which is most frequent in the known language statistics. This process continues in a left to right direction along each of the parents to create the first child only. If, at any stage, a selection is made which already appears in the child being constructed, the second choice is used. If both of the characters in the parents for a given key position already appear in the child then a character is chosen at random from the set of characters that do not already appear in the newly constructed child.

The second child is formed in a similar manner, except that the direction of creation is from right to left and, in this case, the least frequent of the two parent elements is chosen.

Parent 1
`FGHIJKLMNOPQRSTUV_WXYZABCDE`
Parent 2
`CDEFGHIJKLMNOPQRSTUVWXY_ZAB`

Child 1
`CDEIGHLMNOPQRSTUV_WXYZABFJK`
Child 2
`CIEFGHLJKLMNOPQRSTUVWXY_ZDB`

Fig. 4. The mating process

C. Description of the Algorithm

The implementation of each genetic algorithm proceeds as follows:

1. Each GA is given language statistics for unigrams, bigrams and trigrams, the ciphertext, the block size (B) and this GA's position within the block, j ($1 \leq j \leq B$), the frequency of inter-GA communications (f), the maximum number of iterations for the GA (G) and the solution pool size (M).
2. Generate a random pool of M simple substitution cipher keys for position j and calculate the cost for each using unigram statistics only. Call this pool of solutions PCURR.
3. For iteration i (i= 1, ..., G) do:
   a) If i mod $k'0$ send the best key from PCURR to each of the neighbouring GA's (i.e., the GA's solving for positions j 1 and $j + 1$). Also receive the best keys from each of these GA's.
   b) Select M =2 pairs of solutions from PCURR to be the parents of the new generation. The selection should be biased towards the most fit of the current generation (i.e., the keys in PCURR).
   c) Mate using each pair of parents with the algorithm given above. This produces M children that become the new generation (i.e., the solutions of PNEW).

d) Mutate each of the children in PNEW using the same swapping procedure as described in the attack on the simple substitution cipher.

e) Calculate the cost of each of the children in PNEW using the neighbouring keys obtained in Step 3a and Equation 3.1.

f) Select the M best keys from the two pools PCURR and PNEW. Replace the current solutions in PCURR with these solutions.

4. Output the best key from PCURR.

Experimental results obtained from this algorithm are now given.

IV. Results

It is clear that the parallel implementation of the attack will perform much more efficiently than a serial version since the parallel attack is solving each key of the polyalphabetic cipher simultaneously. The overhead of communication between the parallel processors is minimal leading to an attack of the polyalphabetic substitution cipher which would be expected to complete in roughly the same time as a similar attack on a monoalphabetic substitution cipher.

In this section results based on the amount of ciphertext provided to the attack are given. The attack was implemented with a polyalphabetic substitution cipher with a block size of three. The attack was run 100 times for each of 200, 400, 600, 800 and 1000 known ciphertext characters per key. The average results for the polyalphabetic substitution cipher are given in Fig. 5.

These results indicate that the parallel genetic algorithm is an extremely powerful technique for attacks on polyalphabetic substitution ciphers. It could be surmised from the experimental results given above that the attack could be used on polyalphabetic ciphers with very large periods provided that sufficient ciphertext and a parallel machine with sufficient nodes to implement the attack are available to the cryptanalyst.

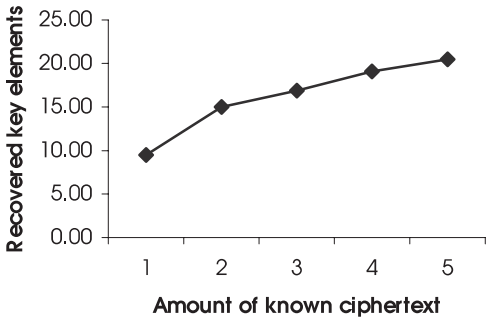


Fig. 5. Known ciphertext versus percent recovered (key and message).

References

[1] J.Holland. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, Michigan, 1975.

[2] D.E. Goldberg. Genetic Algorithms in Search. Optimization and Machine Learning. Addison Wesely, Reading, Massechusetts, 1989.

[3] K.A. DeJong. An Analysys of the Behavious of a Class of Genetic Adaptive Systems. University of Michigan Press, Ann Arbor, Michigan, 1975. Doctoral Disertation.

[4] R.Spillman, M.Janssen, B.Nelson, M.Kepner. Use of a genetic algorithm in the cryptanalysis of simple substitution ciphers. Cryptologia, January 1993.

[5] A. Clark and E. Dawson. A parallel genetic algorithm for cryptanalysis of the polyalphabetic substitution cipher. *Cryptologia*, April 1997.

Attacks on the Transposition Ciphers Using Optimization Heuristics

A. Dimovski[1], D. Gligoroski[2]

Abstract – **In this paper three optimization heuristics are presented which can be utilized in attacks on the transposition cipher. These heuristics are simulated annealing, genetic algorithm and tabu search. We will show that each of these heuristics provides effective automated techniques for the cryptanalysis of the ciphertext. The property which make this cipher vulnerable, is that it is not sophisticated enough to hide the inherent properties or statistics of the language of the plaintext.**

Keywords – **Transposition substitution cipher, Cryptanalysis, genetic algorithm, simulated annealing, tabu search**

I. Transposition Ciphers

First, we will describe a simple transposition cipher. A transposition or permutation cipher works by breaking a message into fixed size blocks, and then permuting the characters within each block according to a fixed permutation, say P. The key to the transposition cipher is simply the permutation P. So, the transposition cipher has the property that the encrypted message i.e. the ciphertext contains all the characters that were in the plaintext message. In other words, the unigram statistics for the message are unchanged by the encryption process.

The size of the permutation is known as the period. Let's consider an example of a transposition cipher with a period of six 6, and a key $P = \{4, 2, 1, 5, 6, 3\}$. In this case, the message is broken into blocks of six characters, and after encryption the fourth character in the block will be moved to position 1, the second remains in position 2, the first is moved to position 3, the fifth to position 4, the sixth to position 5 and the third to position 6.

Table 1. Example of the transposition cipher key and encryption process

KEY:
Plaintext: 123456
Ciphertext: 421563

ENCRYPTION:	
Position:	123456123456
Plaintext:	HOW_ARE_YOUX
Ciphertext	_OHARWO_RUXY

In Table 1 is shown the key and the encryption process of the previously described transposition cipher. It can be noticed that the random string "X" was appended to the end of the message to enforce a message length, which is a multiple of the block size. It is also clear that the decryption can be achieved by following the same process as encryption using the "inverse" of the encryption permutation. In this case the decryption key, P^{-1} is equal to $\{3, 2, 6, 1, 4, 5\}$.

[1]A. Dimovski, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University Arhimedova b.b., PO Box 162, 1000 Skopje, Macedonia adimovski@ii.edu.mk 2

[2]D. Gligoroski, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University Arhimedova b.b., PO Box 162, 1000 Skopje, Macedonia gligoroski@yahoo.com

II. Attacks on the Transposition Cipher

In this section, we will describe three optimization heuristics for attacks on the transposition cipher. Also, a method of assessing intermediate solutions, in the search for the optimum, is discussed.

A. Suitability assessment

The technique used to compare candidate keys is to compare n-gram statistics of the decrypted message with those of the language (which are assumed known). Equation 1 is a general formula used to determine the suitability of a proposed key (k). Here, A denotes the language alphabet (i.e., for English, [A,...,Z,], where represents the space symbol), K and D denote known language statistics and decrypted message statistics, respectively, and the indices b and t denote the bigram and trigram statistics, respectively. The values of β and γ allow assigning of different weights to each of the two n-gram types.

$$C_k = \beta \sum_{i,j \in A} |K^b_{(i,j)} - D^b_{(i,j)}| + \gamma \sum_{i,j,k \in A} |K^t_{(i,j,k)} - D^t_{(i,j,k)}|. \quad (1)$$

As I said above, the unigram frequencies for a message are unchanged during the encryption process of a transposition cipher and so, they are ignored when evaluating a key i.e. in Equation 1.

In attacks, which are proposed here, we will use assessment function based on bigram statistics only. The basic reason for this, it is an expensive task to calculate the trigram statistics. The complexity of determining the fitness is $O(N^3)$ (where N is the alphabet size) when trigram statistics are being determined, compared with $O(N^2)$ when bigrams are the largest statistics being used.

B. Simulated annealing attack

In this section an attack on the transposition cipher using simulated annealing is presented.

Simulated annealing is based on the concept of annealing. In physics, the term annealing describes the process of slowly

cooling a heated metal in order to attain a minimum energy state.

The idea of mimicking the annealing process to solve combinatorial optimization problems is attributed to Kirkpatrick et al [4]. The algorithm is (usually) initialized with a random solution to the problem being solved and a starting temperature. The choice of the initial temperature, T_0 is such that $T \gg \Delta E$. At each temperature a number of attempts are made to perturb the current solution. For each proposed perturbation is determined the change in the cost ΔE. And then, if $\Delta E < 0$ then the proposed perturbation is accepted, otherwise it is accepted with the probability indicated by Metropolis Equation 2 which makes a decision based on this cost difference and the current temperature.

$$\mathrm{Pr}obability(E_1 \Rightarrow E_2) = \exp\Big(-\frac{\Delta E}{T}\Big). \qquad (2)$$

If the proposed change is accepted ($\Delta E < 0$ or Probability $(E_1 \Rightarrow E_2) > 0.5$) then the current solution is updated. Generally, the temperature is reduced when either there a predefined limit in the number of updates to the current solution has been reached or after a fixed number of attempts have been made to update the current solution. The algorithm finishes either when no new solutions were accepted for a given temperature, or when the temperature has dropped below some predefined limit.

Here, in this attack, the Equation 1 is utilized when determining the cost of the solutions, and candidate solutions are generated from the current solution by swapping two randomly chosen positions. Description of this algorithm is given in Fig. 1.

This simulated annealing was implemented and the experimental results are given below in Section 3, which compare this technique with the genetic algorithm attack described in Section 2.3 and the tabu search attack, which is described in Section 2.4.

1. Inputs to the algorithm are the intercepted ciphertext, the key size (permutation size or period) P, and the bigram statistics of the plaintext language.
2. Initialize the algorithm parameters: the maximum number of iterations MAX, the initial temperature T_0, and the temperature reduction factor T_{FACT} .
3. Set T = T_0 and generate a random initial solution K_{CURR} and calculate the associated cost C_{CURR}.
4. For i = 1, . . . MAX, do:
   (a) Set N_{SUCC} = 0.
   (b) Repeat 100 · P times:
      i. Generate a new candidate key K_{NEW}:
         A. Choose n_1,$n_2 \in$ [1, P], $n_1 \neq$ n2.
         B. Swap n_1 and n_2 in K_{CURR} to create K_{NEW}.
      ii. Calculate the cost C_{NEW} of K_{NEW}. Find the cost difference $\Delta E = C_{NEW} - C_{CURR}$ and consult the Metropolis criterion, Equation 2 to determine whether the proposed transition should be accepted.
      iii. If the transition is accepted set $K_{CURR} = K_{NEW}$ and $C_{CURR} = C_{NEW}$ and increment N_{SUCC}.
   (c) If $N_{SUCC} > 10 \cdot P$ go to step 4d.
   (d) If $N_{SUCC} = 0$ go to Step 5.
   (e) Reduce T (T = T · T_{FACT}).
5. Output the current solution K_{CURR}.

Fig. 1. Simulated annealing attack on the transposition cipher

C. Genetic algorithm attack

The genetic algorithm is more complicated than the simulated annealing attack. This is because a pool of solutions is being maintained, rather than a single solution. An extra level of complexity is also present because of the need for a mating function and for a mutation function.

In this attack also, the Equation 1 is utilized when determining the cost of the solutions. The mating i.e. reproduction technique used here for creating the two children is now given:

1. Notation: p_1 and p_2 are the parents, c_1 and c_2 are the children, $p_i(j)$ denotes the element j in parent i, $c_i(j)$ denotes element j in child i, $\{C_i^{j,k}\}$ denotes the set of elements in child i from positions j to k with the limitation that if $k = 0$ or $j = P + 1$ then $\{C_i^{j,k}\} = \{\emptyset\}$,

2. Child 1:

(a) Choose a random number $r \in [1, P]$
(b) $c_1(j) = p_1(j)$ for $j = 1, ..., r$
(c) For $i = 1, ..., P - r$ and $k = 1, ..., P$
If $p_2(k) \notin \{C^{1,i+r-1}\}$
then
$- c_1(i + r) = p_2(k)$
else $k = k + 1$

3. Child 2:

(a) Choose a random number $r \in [1, P]$
(b) $c_2(j) = p_1(j)$ for $j = P, ..., r$
(c) For $i = 1, ..., r$ and $k = P, ..., 1$
If $p_2(k) \notin \{C^{r-i+1,P}\}$
then
$- c_2(r - i) = p_2(k)$
else $k = k - 1$

The mutation operation is identical to the solution perturbation technique used in the simulated annealing attack. That is, randomly select two elements in the child and swap those elements.

In Fig. 2 is an algorithmic description of the attack on a simple substitution cipher using a genetic algorithm.

The results of the genetic algorithm attack are given in Section 3.

D. Tabu search attack

The transposition cipher can also be attacked using a tabu search. This attack is similar to the simulated annealing one with the added constraints of the tabu list. The same perturbation mechanism i.e. swapping two randomly chosen key elements is used to generate candidate solutions. In each iteration the best new key found replaces the worst existing one in the tabu list. The overall algorithm is described in Fig. 3.

The experimental results obtained using all three attacks are presented in the next Section.

1. Inputs to the algorithm are the intercepted ciphertext, the key size (permutation size or period) P, and the bigram statistics of the plaintext language.
2. Initialize the algorithm parameters: the solution pool size M, and the maximum number of iterations MAX.
3. Generate an initial pool of solutions (randomly) P_{CURR}, and calculate the cost of each of the solutions in the pool using Equation 1.
4. For i = 1, . . . ,MAX, do:
   (a) Select M/2 pairs of keys from P_{CURR} to be the parents of the new generation.
   (b) Perform the mating operation described above on each of the pairs of parents to produce a new pool of solutions P_{NEW}.
   (c) For each of the M children perform a mutation operation described above.
   (d) Calculate the cost associated with each of the keys in the new solution pool P_{NEW}.
   (e) Merge the new pool P_{NEW} with the current pool P_{CURR}, and choose the best M keys to become the new current pool P_{CURR}.
5. Output the best solution from the current key pool P_{CURR}.

Fig. 2. Genetic algorithm attack on the transposition cipher

1. Inputs to the algorithm are the intercepted ciphertext, the key size (permutation size or period) P, and the bigram statistics of the plaintext language.
2. Initialize the algorithm parameters: the size of the tabu list S_TABU, the size of the list of possibilities considered in each iteration S_POSS, and the maximum number of iterations to perform MAX.
3. Initialize the tabu list with random and distinct keys and calculate the cost associated with each of the keys in the tabu list.
4. For i = 1, . . . ,MAX, do:
   (a) Find the best key i.e. the one with the lowest cost in the current tabu list, K_{BEST}.
   (b) For j = 1, . . . S_POSS do:
      i. Apply the perturbation mechanism described in the simulated annealing attack to produce a new key K_{NEW}.
      ii. Check if K_{NEW} is already in the list of possibilities generated for this iteration or the tabu list. If so, return to Step 3(b)i.
      iii. Add K_{NEW} to the list of possibilities for this iteration.
   (c) From the list of possibilities for this iteration find the key with the lowest cost, P_{BEST}.
   (d) From the tabu list find the key with the highest cost, T_{WORST}.
   (e) While the cost of P_{BEST} is less than the cost of T_{WORST}:
      i. Replace T_{WORST} with P_{BEST}.
      ii. Find the new P_{BEST}.
      iii. Find the new T_{WORST}.
5. Output the best solution from the tabu list K_{BEST}.

Fig. 3. Tabu search attack on the transposition cipher

III. Experimental Results

The three techniques were implemented in Java as described above and a number of results were obtained.

The first comparison is made upon the amount of ciphertext provided to the attack. These results are presented in Table 2. Here each algorithm was run on differing amounts of ciphertext – 50 times for each amount. The results in Table 2 represent the average number of key elements correctly placed for a key size of 15 in this case. Note that because a transposition cipher key, which is rotated by one place, will still properly decrypt a large amount of the message, a key element is said to be correctly placed if its neighbors are the same as the neighbors for the correct key (except for end positions). In that case, the message will almost certainly still be readable, especially if the period of the transposition cipher is large. It can be seen from the results that each of the three algorithms performed roughly equally when the comparison is made based upon the amount of known ciphertext available to the attack.

Table 2. The amount of key recovered versus available ciphertext, transposition size 15

Amount of ciphertext	SA	GA	TS
200	5	7	4.75
400	10.75	11.5	9.25
600	11.25	12.5	11.5
800	12.5	13	12.25
1000	12.75	13.25	13

Table 3 shows results for the transposition cipher based on the period. It should be noted that for period less than fifteen, with one thousand available ciphertext characters, each of the algorithms could successfully recover the key all the time. The table shows that the simulated annealing attack was the most powerful. For a transposition cipher of period 30 the simulated annealing attack was able to correctly place 25 of the key elements, on the average.

Table 3. The amount of key recovered versus transposition size, 1000 known ciphertext characters

Transposition size	SA	GA	TS
15	12.75	13.25	13
20	16.5	17	16.75
25	20.15	21.5	21
30	25	25.25	25.5

References

[1] Fred Glover, Eric Taillard, and Dominique de Werra. A users guide to tabu search. Annals of Operations Research, 41:328, 1993.

[2] D.E. Goldberg. Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, Reading, Massechusetts, 1989.

[3] Robert A. J. Matthews. The use of genetic algorithms in cryptanalysis. Cryptologia, 17(2):187201, April 1993.

[4] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi. Optimization by simulated annealing. Science, 220(4598):671680, 1983.

(4,2)-Formal Languages

Violeta Manevska[1], Donco Dimovski[2]

Abstract – **The aim of this paper is to define a (4,2)-semigroup automaton on free (4,2)-semigroup, with special attention on (4,2)-formal languages recognizable by them.**

Keywords – **(4,2)-semigroup, (4,2)-semigroup automaton, (4,2)-language**

I. Introduction

Our goal in writing this talk is to examine a (4,2)-formal language and to proof some properties about them. In that means, we are given an example.

II. (4,2)-Semigroups and (4,2)-Semigroup Automata

Here we recall the necessary definitions and known results. From now on, let B be a nonempty set and let $(B,\cdot)$ be a semigroup, where $\cdot$ is a binary operation.

A semigroup automaton is a triple $(S,(B,\cdot),f)$, where S is a set, $(B,\cdot)$ is a semigroup, and $f : S\times B\to S$ is a map satisfying

$$f(f(s,x),y)=f(s,x,y), \tag{1}$$

for every $s\in S$, $x,y\in B$.

The set S is called the set of **states** of $(S,(B,\cdot),f)$ and f is called the **transition function** of $(S,(B,\cdot),f)$.

A nonempty set B with the (4,2)-operation $\{\ \}: B^4\to B^2$ is called a **(4,2)-semigroup** iff the following equality

$$\{\{xyzt\}uv\}=\{xy\{ztuv\}\} \tag{2}$$

is an identity for every $x, y, z, t, u, v\in B$. It is denoted with the pair $(B,\{\})$.

Example 1: Let $B=\{a,b\}$. Then the (4,2)-semigroup $(B,\{\})$ is given by Table 1.

This example of (4,2)-semigroup is generated by an appropriate computer program.

A **(4,2)-semigroup automaton** is a triple $(S,(B,\{\}),f)$ where S is a set, $(B,\{\})$ is a (4,2)-semigroup, and $f : S\times B^2\to S$ is a map satisfying

$$f(f(s,x,y),z,t)=f(s,\{xyzt\}), \tag{3}$$

for every $s\in S$, $x,y,z,t\in B$.

The set S is called the set of **states** of $(S,(B,\{\}),f)$ and f is called the **transition function** of $(S,(B,\{\}),f)$.

Table 1. (4,2)-Semigroup

{ }	
a a a a	(a,a)
a a a b	(a,a)
a a b a	(a,a)
a a b b	(a,a)
a b a a	(a,a)
a b a b	(a,b)
a b b a	(a,a)
a b b b	(a,a)
b a a a	(b,b)
b a a b	(b,b)
b a b a	(b,a)
b a b b	(b,b)
b b a a	(b,b)
b b a b	(b,b)
b b b a	(b,b)
b b b b	(b,b)

2.1^0**.** Let $(S,(B,\cdot),\varphi)$ be a semigroup automaton. Then $(S,(B,\{\}),f)$ is a (4,2)-semigroup automaton with (4,2)-operation $\{\}: B^4\to B^2$ defined by $\{xyzt\}=(x\cdot y\cdot z,t)$ and the transition function $f : S\times B^2\to S$ defined by

$$f(s,x,y)=f(s,x\cdot y). \tag{[2]}$$

2.2^0**.** If $S,(B,\{\}),f)$ is a (4,2)-semigroup automaton, then:

i) $(B^2,*)$ is a semigroup, where the operation $*$ is defined by $(x,y)*(u,v)=\{xyuv\}$ for every $(x,y)(u,v)\in B^2$;

(ii) $(S,(B^2,*),\psi)$ is a semigroup automaton, where the transition function $\psi : S\times B^2\to S$ is defined by

$$\psi(s,(x,y))=f(s,x,y). \tag{[2]}$$

Example 2: Let $(B,\{\})$ be a (4,2)-semigroup given by Table 1 from Example 1 and $S=\{s_0,s_1,s_2\}$. A (4,2)-semigroup automaton $(S,(B,\{\}),f)$ is given by Table 2 and the graph in Fig. 1. This example of (4,2)-semigroup automaton is generated by computer.

III. Free (4,2)-Semigroups and (4,2)-Semigroup Automata on Them

Let B be a nonempty set. We define a sequence of sets $B_0,B_1,...,B_p,B_{p+1},...$ by induction as follows:

$B_0=B$. Let B_p be defined, and let A_p be the subset of B_p of all the elements u_1^{2+2s}, $u_\alpha\in B_p$, $s\geq 1$. Define B_{p+1} to be $B_{p+1}=B_p\bigcup A_p\times\{1,2\}$.

Let $\bar{B}=\bigcup_{p\geq 0}B_p$. Then $u\in\bar{B}$ iff $u\in B$ or $u=(u_1^{2+2s},i)$ for some $u_\alpha\in\bar{B}$, $s\geq 1$, $i\in\{1,2\}$.

[1]Violeta Manevska, University "St. Clement Ohridski"-Bitola, Faculty of Technical Sciences, Bitola, Ivo Lola Ribar b.b. 7000 Bitola, Macedonia, e-mail: violeta.manevska@uklo.edu.mk

[2]Donco Dimovski, Univesity "Sts. Cyril and Methodus"-Skopje, Faculty of Mathematics and Natural Sciences, Institute of Mathematics, e-mail: donco@iunona.pmf.ukim.edu.mk

Table 2. (4,2)-Semigroup

f	(a,a)	(a,b)	(b,a)	(b,b)
s_0	s_1	s_1	s_0	s_1
s_1	s_1	s_1	s_1	s_1
s_2	s_1	s_1	s_2	s_1

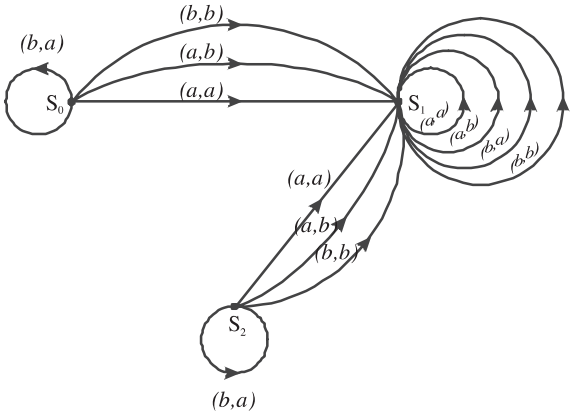


Fig. 1. (4,2)-semigroup automaton

Define a length for elements of $\bar{B}$, i.e. a map $|\ |: \bar{B} \to N$ (N is a set of positive integers) as follows:

1^0 If $u \in B$, then $|u| = 1$;

2^0 If $u = (u_1^{2+2s}, i)$, then $|u| = |u_1| + |u_2| + \cdots + |u_{2+2s}|$.

By induction on the length we are going to define a map $\varphi : \bar{B} \to \bar{B}$. For $b \in B$, let $\varphi(b) = b$. Let $u \in \bar{B}$and suppose that for each $v \in \bar{B}$ with $|v| < |u|$, $\varphi(v) \in \bar{B}$ and

(1) If $\varphi(v) \neq v$, then $|\varphi(v)| < |v|$;

(2) $\varphi(\varphi(v)) = \varphi(v)$.

Let $u = (u_1^{2+2s}, i)$. Then, for each α, $\varphi(u_\alpha) = v_\alpha \in \bar{B}$ is defined, $|\varphi(u_\alpha)| \leq |u_\alpha|$ and $\varphi(\varphi(u_\alpha)) = \varphi(u_\alpha)$. Let $v = (v_1^{2+2s}, i)$.

(i) If for some α, $u_\alpha \neq v_\alpha$, then $|v_\alpha| < |u_\alpha|$, and so, $|v| < |u|$. In this case let $\varphi(u) = \varphi(v)$.

Because $|v| < |u|$, it follows that $\varphi(v)$ is defined, and moreover, (1) and (2) imply that

$$|\varphi(u)| = |\varphi(\nu)| \leq |v| < |u|,\ \varphi(u) \neq u \text{ and}$$
$$\varphi(\varphi(u)) = \varphi(\varphi(v) = \varphi(v) = \varphi(u).$$

(ii) Let $u_\alpha = v_\alpha$ for each α. Then $u = v$. Suppose that there is $j \in \{0, 1, ..., 2s\}$ and $r \geq 1$, such that $u_{j+v} = (w_1^{2r+2}, i)$ for each $v \in \{1, 2\}$ and let t be the smallest such j. In this case, let

$$\varphi(u) = \varphi(u_1^t w_1^{2r+2} u_{t+3}^{2s+2}, i).$$

Because $|(u_1^t w_1^{2r+2} u_{t+3}^{2s+2}, i)| < |u|$ it follows that $\varphi(u)$ is well defined, and moreover, (1) and (2) imply that

$$\varphi(u) \neq u,\ |\varphi(u)| < |u| \quad \text{and} \quad \varphi(\varphi(u)) = \varphi(u).$$

(iii) If $\varphi(u)$ cannot be defined by (i) or (ii), let $\varphi(u) = u$. In this case,

$$\varphi(\varphi(u)) = \varphi(u) = u \quad \text{and} \quad |\varphi(u)| = |u|.$$

The above discussion and (i), (ii) and (iii) complete the inductive step, and so we have defined a map $\varphi : \bar{B} \to B$. Moreover, we have proved the following:

Lemma:

(a) For $b \in B$, $\varphi(b) = b$;

(b) For each $u \in \bar{B}$, $|\varphi(u)| \leq |u|$;

(c) For $u \in \bar{B}$, if $\varphi(u) \neq u$, then $|\varphi(u)| < |u|$;

(d) For each $u \in \bar{B}$, $\varphi(\varphi(u)) = \varphi(u)$. ■

Now, let $Q = \varphi(\bar{B})$. By Lemma (d),

$$Q = \{u | u \in \bar{B}, \varphi(u) = u\}.$$

Define a map [ ]: $Q^4 \to Q^2$, by $[u_1^4] = (v_1^2) \Leftrightarrow v_i = \varphi(u_1^4, i)$ for each $i \in \{1, 2\}$.

Because $u_j \in Q$, it follows that $(u_1^4, i) \in \bar{B}$, and so $\varphi(u_1^4, i) \in Q$ for each $i \in \{1, 2\}$. Hence [ ] is well defined.

Theorem: $(Q, [\ ])$ is a free (4,2)- semigroup with a basis B. ([1])

Let $S, (B, \{\}), f)$ be a (4,2)-semigroup automaton.

Now, we define a sequence of maps $\psi_0, \psi_1, ..., \psi_p, \psi_{p+1}, ...$ for a sequence of sets $B_0, B_1, ..., B_p, B_{p+1}, ...$ by induction as follows:

$\psi_0 : B_0 \to B_0$ with $\psi_0(b) = b$, for each $b \in B_0$;

$\psi_1 : B_1 \to B_0$ with $\psi_1(b_1^n, i) = \{b_1^n\}_i$;

$\psi_2 : B_2 \to B_0$ with $\psi_2(u_1^n, i) = \{\psi_1(u_1)...\psi_1(u_n)\}_i$;

$\vdots$

$\psi_p : B_p \to B_0$ with $\psi_p(u_1^n, i) = \{\psi_{p-1}(u_1)...\psi_{p-1}(u_n)\}_i$;

$\vdots$

Because $\bar{B} = \bigcup_{p \geq 0} B_p$, we define a map $\psi : \bar{B} \to B_0$ with $\psi(u) = \psi_p(u)$ for $u \in \bar{B}$ and $|u| \leq p$. Now we will prove that ψ is well defined. If

$$u = (u_1^r (w_1^{2+2s}, i_1)(w_1^{2+2s}, i_2) u_{r+3}^{2+2t}, i),$$
$$v = (u_1^r w_1^{2+2s} u_{r+3}^{2+2t}, i)$$

and $\varphi(u) = \varphi(v)$, we have to prove that $\psi(u) = \psi(v)$. We have

$$\begin{aligned}
\psi(u) = \psi_p(u) &= \\
&= \psi_p(u_1^r (w_1^{2+2s}, i_1)(w_1^{2+2s}, i_2) u_{r+3}^{2+2t}, i) = \\
&= \{\psi_{p-1}(u_1)...\psi_{p-1}(u_r)\psi_{p-1}(w_1^{2+2t}, i_1)\psi_{p-1}(w_1^{2+2s}, i_2) \\
&\psi_{p-1}(u_{r+3})...\psi_{p-1}(u_{2+2t})\}_i = \\
&= \{\psi_{p-1}(u_1)...\psi_{p-1}(u_r)\{\psi_{p-2}(w_1)...\psi_{p-2}(w_{2+2s})\}_{i_1}... \\
&\{\psi_{p-1}(w_1)...\psi_{p-2}(w_{2+2s})\}_{i_2}\psi_{p-1}(u_{r+3})...\psi_{p-1}(u_{2+2t})\}_i = \\
&= \{\psi_{p-1}(u_1)...\psi_{p-1}(u_r)\psi_{p-1}(w_1)...\psi_{p-1}(w_{2+2s}) \\
&\psi_{p-1}(u_{r+3})...\psi_{p-1}(u_{2+2t})\}_i,
\end{aligned}$$

Also,

$$\begin{aligned}
\psi(v) = \psi_p(v) = \psi_p(u_1^r w_1^{2+2s} u_{r+3}^{2+2t}, i) &= \\
= \{\psi_{p-1}(u_1)...\psi_{p-1}(u_r)\psi_{p-1}(w_1)...\psi_{p-1}(w_{2+2s}) \\
\psi_{p-1}(u_{r+3})...\psi_{p-1}(u_{2+2t})\}_i.
\end{aligned}$$

Hence $\psi(u) = \psi(v)$. On the other hand, $Q = \varphi(\bar{B})$, so it follows that the restriction of ψ on Q is well defined.

Now again, we define a sequence of maps $\tau_0, \tau_1, ..., \tau_p, \tau_{p+1}, ...$ for a sequence of sets $B_0, B_1, ..., B_p, B_{p+1}, ...$ by induction as follows:

$\tau_0 : S \times B_0^2 \to S$ with $\tau_0(s, u, v) = f(s, u, v)$;

$\tau_1 : S \times B_1^2 \to S$ with

$\tau_1(s, (u_1^n, i), (v_1^k, j)) = f(s, \psi_1(u_1^n, i), \psi_1(v_1^k, j))$;

$\tau_2 : S \times B_2^2 \to S$ with

$\tau_2(s,(u_1^n,i),(v_1^k,j)) = f(s,\psi_2(u_1^n,i),\psi_2(v_1^k,j));$

$\vdots$

$\tau_p : S \times B_p^2 \to S$ with

$\tau_p(s,(u_1^n,i),(v_1^k,j)) = f(s,\psi_p(u_1^n,i),\psi_p(v_1^k,j)).$

Now we define a map τ for the sequence of maps $\tau_0, \tau_1, ..., \tau_p, \tau_{p+1}, ...$ by $\tau : S \times \bar{B}^2 \to S$, so that $\tau|_{B_p} = \tau_p$ and

$$\tau(s,(u_1^n,i),(v_1^k,j)) = \tau_p(s,(u_1^n,i),(v_1^k,j)) =$$
$$= f(s,\psi_p(u)1^n,i),\psi_p(v_1^k,j)) = f(s,\psi(u_1^n,i),\psi(v_1^k,j)).$$

Because ψ is well defined, it follows that τ is well defined. On the other hand, $Q = \varphi(\bar{B})$ so $\bar{\varphi}$ denotes the map $\bar{\varphi} : S \times Q^2 \to S$ defined by

$$\bar{\varphi}(s,(u_1^n,i),(v_1^k,j)) = \tau(s,(u_1^n,i),(v_1^k,j)) =$$
$$= f(s,\psi(u_1^n,i),\psi(v_1^k,j)).$$

Moreover, $(S,(Q,[\,]),\bar{\varphi})$ is a (4,2)-semigroup automaton, where $(Q,[\,])$ is a free (4,2)-semigroup with a basis B.

IV. Recognizable (4,2)-Languages

Any subset $L^{(4,2)}$ of the universal language $Q^* = \bigcup_{p\geq 1} Q^p$, where Q is a free (4,2)-semigroup with a basis B, is called a **(4,2)-language (formal (4,2)-language)** on the alphabet B.

A (4,2)-language $L^{(4,2)} \subseteq Q^*$ is called **recognizable** if there exists:

(1) a (4,2)-semigroup automaton $(S,(B,\{\,\}),f)$, where the set S is finite;

(2) an initial state $s_0 \in S$;

(3) a subset $T \subseteq S$ such that

$$L^{(4,2)} = \{w \in Q^* | \bar{\varphi}(s_0,(w,1),(w,2)) \in T\},$$

where $(S,(Q,[\ ]),\bar{\varphi})$ is the (4,2)-semigroup automaton constructed above, for the (4,2)-semigroup automaton $(S,(Q,\{\,\}),f)$.

We also say that the (4,2)-semigroup automaton $(S,(Q,\{\ \}),f)$ **recognizes** $L^{(4,2)}$, or that $L^{(4,2)}$ is **recognized** by $(S,(Q,\{\,\}),f)$.

Example 3: Let $(S,(Q,\{\ \}),f)$ be a (4,2)-semigroup automaton given in Example 2. We construct the (4,2)-semigroup automaton $(S,(Q,[\,]),\bar{\varphi})$ for the (4,2)-semigroup automaton $(S,(Q,\{\,\}),f)$.

A (4,2)-language $L^{(4,2)}$, which is recognized by the (4,2)-semigroup automaton $(S,(Q,[\,]),\bar{\varphi})$, with initial state s_0 and terminal state s_1 is

$$L^{(4,2)} = \{w \in Q^* | w = w_1 w_2 ... w_{2q},$$

where

$$w_l = \begin{cases} (u_1^n,i),\ n \geq 4,\ u_\alpha \in Q \\ (a^*b^*)^* \end{cases}, l \in \{1,2,...,q\},\ q \geq 3$$

and:

a) If $i = 1$, then:

a1) $(u_1^n,1) = a$, where

$\psi_{p-1}(u_1)...\psi_{p-1}(u_n) = a(a \cup b)(a^t b^j)^*,$

a2) $(u_1^n,1) = b$, where

$\psi_{p-1}(u_1)...\psi_{p-1}(u_n) = b(a \cup b)(a^t b^j)^*;$

b) If $i = 2$, then

b1) $(u_1^n,2) = a$, where

$\psi_{p-1}(u_1)...\psi_{p-1}(u_n) = (ba)^+ \cup (a(a \cup b)(a^t b^j)^* \backslash (ab)^+),$

b2) $(u_1^n,2) = b$, where

$\psi_{p-1}(u_1)...\psi_{p-1}(u_n) = (ab)^+ \cup (b(a \cup b)(a^t b^j)^* \backslash (ba)^+),$

where $t + j = 2k$, $t, j \in \{0,1,2,...\}$, $k \geq 1$, and finally

$$\psi_p(w_1)...\psi_p(w_q) =$$
$$= (ba)^*(bb \cup a(a \cup b))((a \cup b)(a \cup b))^* =$$
$$(ba)^*(aa \cup ab \cup bb)(aa \cup ab \cup ba \cup bb)^*\}.$$

4.1^0. Let $L^{(4,2)}$ be a (4,2)-language on the set B recognized by (4,2)-semigroup automaton $(S,(Q[\ ]),\bar{\varphi})$. Let $(S,(Q,[\ ]),\bar{\varphi})$ be a (4,2)-semigroup automaton eith initial state s_0 and a set of terminal states $T \subseteq S$. Then $\tilde{L}^{(2,1)} \subseteq L^{(4,2)}$ for any language $L^{(4,2)}$, which is recognized by the semigroup automaton $(S,(Q^2,*),\psi)$ with the same initial state and the same set of terminal states, where $\psi : S \times Q^2 \to S$ is a transition function defined by $\psi(s,(u,v)) = \bar{\varphi}(s,u,v)$ and $\tilde{L}^{(2,1)} = \{\tilde{w} | w \in L^{(2,1)}\}$.

Proof: $L^{(4,2)}$ is a recognizable (4,2)-language on the set B by the (4,2)-semigroup automaton $(S,(Q,[\,]),\bar{\varphi})$ with initial state s_0 and a set of terminal states $T \subseteq S$, so

$$L^{(4,2)} = \{w \in Q^* | \bar{\varphi}(s_0,w) \in T\}.$$

By Proposition 2.2^0, $(S,(Q^2,*),\psi)$ is a semigroup automaton. It recognizes a language $L^{(2,1)}$ with a same initial state s_0 and a same terminal states $T \subseteq S$, so it is of the form

$$L^{(2,1)} = \{w \in (Q^2)^* | \psi(s_0,w) \in T\}.$$

Let $w \in L^{(2,1)}$. It follows that $w \in (Q^2)^*$ and $\psi(s_0,w) \in T$. That

$$\bar{\varphi}(s_0,(\tilde{w},2),(\tilde{w},2)) = \bar{\varphi}(s_0,w) = \psi(s_0,w) \in T.$$

Thus $\tilde{w} \in L^{(4,2)}$, i.e. $\tilde{L}^{(2,1)} \subseteq L^{(4,2)}$. ■

4.2^0. Let $L^{(2,1)}$ be a recognizable language on the set B by a semigroup automaton $(S,(B,||\,||),\xi)$ with an initial state s_0 and a set of terminal states $T \subseteq S$, and $(S,(B,\{\,\}),f)$ be a (4,2) semigroup automaton constructed by a semigroup automaton $(S,(B,||\,||),\xi)$. Let $f : S \times B^2 \to S$ be a transition function defined by $f(s,x,y) = \xi(s,x,y)$. Then $L^{(2,1)} \subseteq L^{(4,2)}$, where $L^{(4,2)}$ is a recognizable (4,2) language on the set B by the (4,2)-semigroup automaton $(S,(Q,[\,]),\bar{\varphi})$ with initial state $s_0 \in S$ and a set of terminal states $T \subseteq S$.

Proof: A language $L^{(2,1)}$ is recognizable by a semigroup automaton $(S,(B,||\,||),\xi)$ with an initial state $s_0 \in S$ and a set of terminal states $T \subseteq S$, so

$$L^{(2,1)} = \{w \in B^* | \xi(s_0,w) \in T\}.$$

By Proposition 2.1^0 $(S,(B,\{\ \}),f)$ is a (4,2)-semigroup automaton. We construct a (4,2) semigroup automaton

$(S,(Q,[\,]),\bar{\varphi})$, where $Q=\varphi(\bar{B})$ and $\bar{\varphi}: S\times Q^2\to S$ is a transition function defined by

$$\bar{\varphi}(s,(y_1^n,i),(v_1^k,j))=\varphi_p(s,(u_1^n,i),(v_1^k,j))= \\ =f(s,[\bar{u}_i^n]_i,[\bar{v}_1^k]_i),$$

where

$$\psi_p(u_1^n,i)=[\psi_{p-1}(u_1)...\psi_{p-1}(u_n)]_i=[\bar{u}_1^n]_i,$$
$$\psi_p(v_1^k,j)=[\psi_{p-1}(v_1)...\psi_{p-1}(v_k)]_j=[\bar{v}_1^k]_j$$

It follows that a recognizable (4,2)-language $L^{(4,2)}$ on the set B by (4,2) semigroup automaton $(S,(Q,[\,]),\bar{\varphi})$, with initial state $s_0\in S$ and a set of terminal states $T\subseteq S$ is of the form

$$L^{(4,2)}=\{w\in Q^*|\bar{\varphi}(s_0,w)\in T\}.$$

Let $w\in L^{(2,1)}$ and $|w|\geq 2$. Then

$$\bar{\varphi}(s_0,(w,1),(w,2))=\bar{\varphi}(s_0,w)=\xi(s_0,w)\in T.$$

Thus $w\in L^{(4,2)}$, i.e. $L^{(2,1)}\subseteq L^{(4,2)}$. ■

V. Conclusion

The results was given in this paper, are of the scientific interest, because there was defined a (4,2)-languages as a consequence of the generalization of the semigroup automata in case (4,2). Also, here was given the connection between (2,1)-languages and (4,2)-languages..

References

[1] D. Dimovski, "Free vector valued semigroups", *Proc. Conf. "Algebra and Logic", Cetinje*, (1986), 55-62.

[2] D. Dimovski, V. Manevska, "Vector valued (n+k)-formal languages", *Proc. 10th Congress of Yugoslav Mathematicians*, Belgrade, (2001), 153-159.

[3] V. Manevska, D. Dimovski, "Properties of the (3,2)-languages recognized by (3,2)-semigroup automata", MMSC, Borovets, (2002), 368-373

Using Decomposition to Produce High-Level System Organization of Software Source Code

Violeta T. Bojikova[1]

Abstract – **Software clustering techniques are useful for extracting architectural information about a system directly from its source code structure. In this paper is discussed the evaluation problem of clustering algorithms.**

Keywords – **software clustering algorithms, program decomposition, program restructuring**

I. Introduction

Software architecture is a critical asset to a project due to the ever increasing complexity and the demand to reduce maintenance cost for evolution. One of the areas in software architecture is architecture recovery through reverse engineering of existing implementations.

Clustering techniques have been used in many disciplines to support grouping of similar objects of a system. Clustering analysis is one of the most fundamental techniques adopted in science and engineering. The primary objective of clustering analysis is to facilitate better understanding of the observations and the subsequent construction of complex knowledge structure from features and object clusters. Examples include botanic species and mechanical parts. The key concept of clustering is to group similar things into clusters, such that intra-cluster similarity or cohesion is high, and inter-cluster similar or coupling is low. Coupling has great impact on many quality attributes, such as maintainability, verifiability, flexibility, portability, reusability, interoperability, and expandability. The main objective of clustering is similar to that of software partitioning.

Most existing clustering approaches often are limited to architecture recovery activity in the reverse engineering process only. But clustering techniques can be applied to software during various life-cycle phases. Clustering techniques can be used to effectively support both software architecture partitioning at the early phase in the forward engineering process and software architecture recovery of legacy systems in the reverse engineering process. Lung demonstrated that clustering techniques can also be used to effectively facilitate software architecture restructuring instead of simply being used for software architecture recovery of existing systems.

Because the structure of software systems are usually not documented accurately, researchers have expended a great deal of effort studying ways to recover design artifacts from source code. Since many software systems are large and complex, appropriate abstractions of their structure are needed to simplify program understanding and restructuring.

For small systems, source code analysis tools can easy extract the source level components (e.g., modules, classes, functions) and relations (e.g., method invocation, function calls, inheritance) of a software system. For large systems there is significant value in identifying the abstract (high-level) entities, and then modeling those using architectural components such as subsystems and their relations.

Subsystems generally consist of a collection of source code resources that collaborate with each other to implement a feature or provide a service to the rest of the system. Typical resources found in subsystems include modules, classes, and possibly, other subsystems. Subsystems facilitate program understanding by treating sets of source code resources as high-level entities.

The entities and relations needed to represent software architectures (high-level component) are not found in the source code. Thus, without external documentation, we seek other techniques to recover a reasonable approximation of the software architecture using source code information. Researchers in the reverse engineering community have applied a variety of software clustering approaches to address this problem.

Many of the clustering techniques published in the literature can be categorized by the way they create clusters. These techniques determine clusters (subsystems) using source code component similarity concept analysis, optimization [144,145], or information available from the system implementation such as module, directory, and/or package names.

II. Fundamental Questions Pertaining to the Software Clustering Problem

1. How can a software engineer determine – within a reasonable amount of time and computing resources – if the solution produced by a software clustering algorithm is good or not?

2. Can an algorithm be created that guarantees a solution – within a reasonable amount of time and computing resources – that is close to the ideal solution?

From a practical aspect, the answers to these questions are important because they provide increased confidence to software engineers who analyze systems. From a theoretical aspect, these answers are important because they provide an approximation algorithm to a known NP-Hard problem, in addition to a method for comparing any solution, even those produced by other algorithms that use the same clustering

[1]Violeta Bojikova is with the Department of Computer Science Varna Technical University, Bulgaria, e-mail: vbojikova@yahoo.com, bojikov@nat.bg

criterion we do (i.e., coupling-cohesion tradeoff), to the optimal solution.

III. Sub-Optimal Decomposition Algorithm – SOAD

In this paper is presented an evaluation of a clustering algorithm, which first version is presented in [5-7] and which uses heuristic search technique to determine the subsystems of a software system. The goal of the software clustering process is to partition the graph model – MDG of the system into a set of clusters such that the clusters represent subsystems.

Since graph partitioning is known to be NP-hard, obtaining a good solution by random selection or exhaustive exploration of the search space is unlikely. SOAD overcomes this problem by using heuristic-search techniques.

Formalization: The MDG $= (X, U)$ is a directed graph where the source code components are modeled as nodes, and the source code dependencies are modeled as edges:

– X is finite set of components (nodes), where $N = |X|$ is the number of components – classes, modules, files, packages, etc.;

– $U \subseteq X \times X$ is the set of ordered pairs $\langle x_1, x_2 \rangle$ that represent the source-level relationships between module x_1 and module x_2 (inherit, import, include, call, instantiate, etc.)

Once the MDG is created, SOAD produce an initially partition of the MDG and evaluates the "quality" of this partition using a fitness function that is called Modularization Quality (k) [doklad]. After producing the initially solution from the search space, SOAD improves it using iterative algorithm. Given that the fitness of an individual partition can be measured, heuristic search algorithms are used in the iterative clustering phase in an attempt to improve the MQ of the initially generated partition. SOAD implement a hill-climbing algorithm.

The Goal of SOAD is to "Find a good partition of the MDG."

A partition is the decomposition of a set of elements (i.e., all the nodes of the graph) into mutually disjoint clusters.

A "good partition" is a partition where:

– highly interdependent nodes are grouped in the same clusters;

– independent nodes are assigned to separate clusters The k function is designed to penalize excessive inter-cluster coupling. k increases as the inter-edges (i.e., external edges that cross cluster boundaries) increase.

Modularization Quality and restrictive condition. Modularization Quality (k) is a measurement of the "quality" of a particular MDG partition.

The assumption is: "Well designed software systems are organized into cohesive clusters that are loosely interconnected".

The weight – W_k of each cluster $g \in$ MDG with $x_i \in X$ components (nodes) corresponds to the restrictive condition W_0, where w_i – is the label of node x_i and present the number of node's elements (i.e. functions):

$$W_k = \sum_{x_i \in X_k} w_i \Longleftarrow W_0.$$

The value of "k", where k_{ij} is the number of inter-edges (i.e., external edges that cross cluster boundaries) between nodes x_i and x_j is calculated as follow:

$$k = \sum_{i=1,\ldots,M} \sum_{j=1,\ldots,M} k_{ij} = \min, \quad \forall i \neq j.$$

M represents the number of clusters in the current partition of the MDG.

IV. Comparing the Results Produced by Software Clustering Algorithms

Now that a plethora of approaches to software clustering exist, the validation of clustering results is starting to attract the interest of the Reverse Engineering research community. Numerous clustering approaches have been proposed in the reverse engineering literature, each one using a different algorithm to identify subsystems. Since different clustering techniques may not produce identical results when applied to the same system, mechanisms that can measure the extent of these differences are needed.

Many of the clustering techniques published in the literature present case studies, where the results are evaluated by the authors or by the developers of the systems being studied. This evaluation technique is very subjective. Recently, researchers have begun developing infrastructure to evaluate clustering techniques, in a semi-formal way, by proposing similarity measurements [1-3]. These measurements enable the results of clustering algorithms to be compared to each other, and preferably to be compared to an agreed upon "benchmark" standard. Note that the "benchmark" standard needn't be the optimal solution in a theoretical sense. Rather, it is a solution that is perceived as being "good enough".

Existing clustering techniques neither provide a guarantee on the quality of their solutions nor any indication of a solution's proximity to the optimum. Bunch [1,3], for example, uses several methods to find solutions, such as hill-climbing and genetic algorithms. Hill-climbing only guarantees local optimality, but makes no guarantees of global optimality.

Genetic algorithms are another type of search, like hill climbing, that does not guarantee the quality of its solution, not even with respect to local extreme. Neither method indicates how good a solution is with respect to the optimal solution. Not being able to meet either of these criteria is unsatisfactory.

Researchers have begun formulating ways to measure the differences between system decompositions.

For example, Anquetil et al. developed a similarity measurement based on Precision and Recall.

Much of the research on measuring the similarity between decompositions only considers the assignment of the system's modules to subsystems. Mancoridis and al. [1] argue that a more realistic indication of similarity should consider how much a module depends on other modules in its subsystem, as well how much it depends on the modules of other subsystems.

Mojo measures the distance between two decompositions of a software system by calculating the number of operations

needed to transform one decomposition into the other [4]. The transformation process is accomplished by executing a series of Move and Join operations. In MoJo, a Move operation involves relocating a single resource from one cluster to another, and a Join operation takes two clusters and merges them into a single cluster.

Tzerpos and Holt also introduce a quality measurement based on MoJo. The MoJo quality measurement normalizes MoJo with respect to the number of resources in the system. Given two decompositions, A and B, of a system with N resources, the MoJo quality measurement is degfined as:

$$\text{MoJoQuality}(A; B) = (1 - \text{MoJo}(A; B)/N) * 100\%$$

Koschke and Eisenbarth present in 2000 a framework for experimental evaluation of clustering techniques [2]. The goal this evaluation is to have an oracle to compare the results of automatic techniques.

Let software engineers detect modules $\longrightarrow$ references R

Let automatic techniques propose modules $\longrightarrow$ candidates C

Let compare candidates to references

– identify immediate corresponding candidates and references $\longrightarrow$ good match

– identify corresponding submodules; i.e., a module corresponds only to a part of another module $\longrightarrow$ O.k. match

– measure accuracy of correspondences $\longrightarrow$ detection quality

Types of matchs:

1. Good match $C \approx pR$:
Iff $|\text{elements}(C) \cap \text{elements}(R)|/|\text{elements}(C) \cup \text{elements}(R)| >= p$, where p is a tolerance parameter.

- if $p = 1$, C and R must be the same
- more pragmatically, $p = 0.7$
  – C and R overlap at 70%
  – $\{a, b, c, d\} approx 0.7\{b, c, d\}$, overlap is $3/4 = 0.75$
  – $\{a, b, c, d\}$ no $\approx 0.7\{b, c, d, e\}$, overlap is $3/5 = 0.6$

2. Part-of matches
Matching relation $S \subseteq pT$:
Iff $|\text{elements}(S) \cap \text{elements}(T)|/|\text{elements}(S)| >= p$, where p is tolerance parameter as above S is part of T

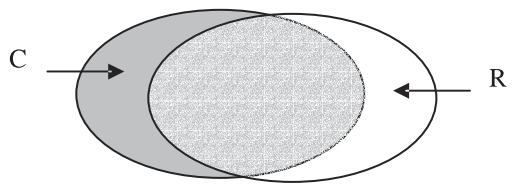


Fig. 1. Mutually part-of matches

3. Mutually part-of matches

- $C \approx pR \Rightarrow C \subseteq pR \wedge R \subseteq pC$
- but not: $C \subseteq pR \wedge R \subseteq pC \Rightarrow C \approx pR$
- yet, there is a distinct correspondence between C and R (fig. 1): if $C \subseteq pR \wedge R \subseteq pC$

$$\text{Overlap}(C, R) = (C \cap R)/(C \cup R) < 70\%$$

C and R are a mutually part-of match:

iff $C \subseteq pR \wedge R \subseteq pC$

– mutually part-of matches are part-of matches

– good matches are mutually part-of matches

Part-of and mutually part-of are O.k. matches

4. The accuracy of each match is:

$$T(C, R) = \text{overlap}(C, R),$$

where: $\text{overlap}(C, R) = (C \cap R)/(C \cup R)$.
Accuracy $T(M)$ for class of matches is:

$$T(M) = \frac{\sum_{C_i, R_i \in M} T(C_i, R_i)}{|M|}.$$

5. There are multiple aspects of detection quality [2]:
– Numbre of false positives and true negatives
– Granularity of matches = good matches/all matches
– Accuracy of each match and the class of matches

6. SOAD is evaluated using the Koschke similarity technique and the Mojo distance evaluation. The result, produced by SOAD are stable. Ideally, the results produced by SOAD could be compared to an optimal reference solution, but this option is not possible since the graph partitioning technique used by SOAD for software clustering is NP-Hard.

In the case is used the benchmark standard. SOAD has been tested for open-source software systems, which has

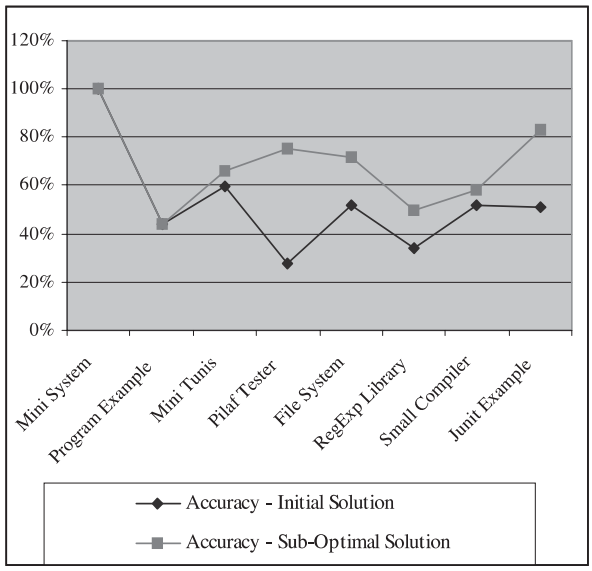


Fig. 2. Accuracy results

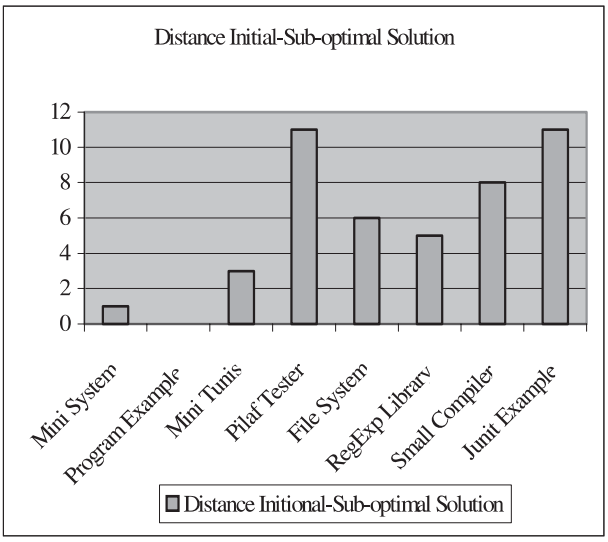


Fig. 3. Distance evaluation

been used by other clustering algorithms [3]: Apache Regular Expression Class Library (RegExp Package), Mini Tunis, Small Compiler, File System, PlafTester Program etc. All systems are small or middle size. We dispose with the MDG graph for these systems and with the reference partition (i.e. a partition presented by the experts, or by other techniques for clustering, or in the design documentation).

Figures 2,3 show the clustering results and the value of the accuracy of each class of matches (initial partition and sub-optimal partition). The results depend from the graph strength (number of nodes and edges).

V. Conclusion

Most interesting software systems are large and complex and, as a consequence, understanding their structure is difficult. Such software systems are composed of many resources. The structure of these systems can be represented as a graph where the nodes are the resources and edges are the relations between the resources.

Without automated assistance the software structure graph provides little value when being used to understand the design of practical systems because of its large number of nodes and edges.

Decomposing source code components and relations into subsystem clusters is an active area of research. The primary goal of clustering tools is to propose subsystems that expose abstractions of the software structure. However, the various clustering tools use different algorithms, and make different assumptions about how subsystems are formed.

Now that many clustering techniques exist, some researchers have turned their attention to evaluating their relative effectiveness. There are several reasons for this: Many of the papers on software clustering formulate conclusions based on case studies, or by soliciting opinions from the authors of the systems presented in the case studies.

Much of the research on measuring the similarity between decompositions only considers the assignment of the system's modules to subsystems. We argue that a more realistic indication of similarity should consider how much a module depends on other modules in its subsystem, as well how much it depends on the modules of other subsystems.

When decompositions are compared, all source code resources tend to be treated equally. Conclusions are often formulated based on the value of a similarity or distance measurement.

In the paper we examine two similarity measurements that have been used to compare decompositions.

We show the results of our study of similarity measurements. The conclusion is that SOAD shows good results referring to these measurements.

References

[1] Comparing the Decompositions Produced by Software Clustering Algorithms using Similarity Measurements, Spiros Mancoridis and Brian Mitchell IEEE Proceedings of the 2001 International Conference on Software Maintenance (ICSM'01). IEEE.

[2] Rainer Koschke and Thomas Eisenbath, "A Framework for experimental evaluation of clustering techniques", *International Workshop on Program Comprehension*, June, 2000.

[3] Search Based Reverse Engineering, by B. S. Mitchell, S. Mancoridis, M. Traverso. In the *ACM Proceedings of the 2002 International Conference on Software Engineering and Knowledge Engineering (SEKE'02)*, Ischia, Italy, July, 2002. pp. 431-438.

[4] Mojo: A distance metric for software clustering. V. Tzerpos and R. C. Holt. In Proc. Working Conf. on Reverse Engineering, Atlanta - 1999, pp.187-193.

[5] Software Architecture Decomposition, Violeta Bojikova, M. Mitev, Proceedings of the 14th Int'l Conference SAER'2000, Varna, 2000, pp. 173-177.

[6] An Approach to measure the Cost of Program restructuring, Violeta Bojikova, M. Karova, Proceedings of papers, Volume 2, pp 669-671, 2002, Nish, Yugoslavia

[7] Elementary Operations and Program Restructuring, V.Bojikova, M.Karova, Proceedings of the Int'l Conference Tehnonav 2002, Constanca, June-2002, pp.192-197.

An Approximation Algorithm for Scheduling Problem on a Finite Number of Processors with Communication Delays*

Vassil G. Guliashki[1]

Abstract – **A polynomial approximation algorithm for scheduling a finite number of tasks on m identical processors with nonzero communication times between the processors is presented. The created algorithm is compared with other approximation algorithms on the grounds of theoretical results for their worst-case performance. The possibility for anomalous behavior is commented.**

Keywords – **scheduling, makespan, communication delays**

I. Introduction

During the last two decades the interest in solving problems for scheduling a finite number of tasks on limited number of processors grows up rapidly. The real problems of this kind require consideration of communication delays between two consecutive tasks when they are not assigned performance to one and the same processor. For convenience we will assume that a precedence relation between two tasks i and j is available if task i needs data from task j before being started.

We will consider the problem for making a schedule to proceed n tasks on m processors, for which the task duplication is not allowed, the communication between any two processors is possible and the communication delays depend only on the corresponding tasks. The precedence constraints and the processing times are arbitrary. The objective is to find the schedule that minimizes the overall finishing time, or the "makespan". Let ρ denotes the ratio of the greatest communication delay to the smallest processing time. We will assume that the greatest communication time between any two different processors is smaller than the processing time, needed for the completion of the smallest task, i.e. $\rho \leq 1$. This problem is known as Small Communication Time problem (SCT problem).

There are surveys studying scheduling problems (see for example [1,7,13]), where some theoretical results about this problem are presented. As it is mentioned in [1] Picouleau has proven in 1992 that this problem is NP-hard. Jakoby and Reischuk have shown in [6] that the special case with unlimited number of processors and unit processing time is NP-hard even when the in-degree of each node is at most two. Using a similar reduction, they also proved that for a binary tree, unit processing times and arbitrary communication times the problem is NP-complete. For fixed $m \geq 3$, no algorithms which ensure optimal schedules are known yet. For this reason different kind of approximation algorithms have been developed (see for example [2,4,5,9,11,12]). The parallelism of multiprocessor problem in combination with the communication delays causes difficulties at the design of approximation algorithms, because the problem is combinatorial one. The worst-case performance of all of them is as bad as possible (see [4]), especially if a great number of processors is assumed. The performance ratio for the known approximation algorithms varies around 2 and tends to 3 when m – the number of processors – is fixed. The best known approximation algorithms for this problem are those presented in [9] and [4] with performance ratio 7/3, and $7/3 - 4/(3m)$, correspondingly. For the problem with unlimited number of processors Hanen and Munier (see [4]) have created an approximation algorithm with 4/3 performance ratio. The aim of this paper is to present the approximation algorithm AASCT, which could to a great extent avoid the anomalous behavior, arising when the number of processors increases, and from another side could improve the performance of approximation algorithm presented in [4]. The time complexity of AASCT is $O(\gamma n^2(n-m))$.

II. Preliminaries

First of all will be defined the SCT task system. With this aim we will introduce some symbols. The set of n tasks will be denoted by T and the corresponding processing times by $p_1, ..., p_n$. Let $G = (T, E)$ be a directed acyclic graph (DAG). An arc $(i, j) \in E$ corresponds to the data transfer from task i to task j, that occurs after i has been finished and before the start of j. The duration of this data transfer is a constant delay, equal to c_{ij} in case i and j are performed by different processors and 0 if i and j are performed by one and the same processor. The task system $\Im(T, p, G, c)$ is called SCT task system, if the following constraint on the communication delays holds:

$$\rho = \frac{\max_{(ij)\in E} c_{ij}}{\min_{i=1,...,n} p_i} \leq 1. \qquad (1)$$

In some cases (see [1,9]) the SCT system is defined by weaker conditions, but the algorithm presented in section 3 is based on condition (1).

Here we consider the problem of scheduling n tasks of the SCT task system on m processors under condition (1), where n and m are finite numbers.

A schedule $S = (t, \pi)$ assigns a starting time t_i and a processor π_i to each task i, so that

1) for any pair of tasks (i, j) if $\pi_i = \pi_j$, then $t_i + p_i \leq t_j$ or $t_j + p_j \leq t_i$;

*This study is partly supported by the Ministry of Education and Science, National Science Fund, contract No. I-1203/2002, Sofia, Bulgaria.

[1]Vassil G. Guliashki is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Bulgaria, Sofia 1113, "Acad. G. Bontchev" Street, Bl. 29 A, E-mail: vgoul@iinf.bas.bg, vggul@yahoo.com .

2) for any arc (i,j) of G, if $\pi_i = \pi_j$, then $t_j \geq t_i + p_i$; else $t_j \geq t_i + pi + c_{ij}$;
3) if m processors are available: $\forall i \in T$, $\pi_i \in \{1,...,m\}$.
The makespan of the schedule is denoted by ω:

$$\omega = \max_{i \in T}(t_i + p_i). \quad (2)$$

We will denote the optimal (minimal) makespan by ω_{opt}.

It will be assumed that the task i precedes task j if there is a path in G from i to j. The task i is called predecessor of j and the task j is called successor of i. This relation will be denoted by $i \rightarrow j$. A task i is said to be an immediate successor (resp. predecessor) of a task j if there is an arc (j,i) (resp. (i,j)) in G. For any task i we denote by $\Gamma^+(i)$ (resp. by $\Gamma^-(i)$) the set of immediate successors, (resp. predecessors) of i. In case one of immediate successors of a task j satisfies the following condition:

$$t_j < t_i + p_i + c_{ij}, \quad (3)$$

j is called the favorite successor of i. It follows from (1) that there is only one favorite successor j of i. Similarly i is called a favorite predecessor of j.

The usual approximation algorithms used for scheduling tasks on m processors, called list scheduling (LS) algorithms, build a schedule by means of a greedy process, that schedules a new task at each iteration. Assuming a partial schedule is already built for the time period $[0, t_{k-1}]$, the greedy algorithm scans each processor to find a task that is ready for it at the moment t_k and if any, to assign to it the first ready task in the list at this moment. Graham (see [2]) has proposed such algorithm for the problem without communication delays. For this case he obtained the performance ratio $\omega/\omega_{\text{opt}} = 2 - 1/m$. Rayward-Smith has shown in [12] that any list scheduling algorithm with unit execution times and unit communication times (UET-UCT) satisfies $\omega < (3 - 2/m)\omega_{\text{opt}} - (1 - 1/m)$.

When general communication delays are considered (not necessarily SCT), an extension of the usual schema has been proposed [5], called ETF (i.e. earliest task first) that can be outlined as follows:

While there remains an unscheduled task, the set of ready tasks R (the predecessors of which have been already scheduled) is determined. Then for each couple (i,π), $i \in R$, $\pi \in 1,...,m$, the earliest starting time of task i on processor π, denoted by $e(i,\pi)$ is computed. Then the earliest starting time $e = \min_{(i,\pi)} e(i,\pi)$ is determined and a task i, for which there is a couple (i,π) with $e(i,\pi) = e$ is chosen and scheduled at time e. Finally a processor, for which $e(i,\pi) = e$ is assigned to i.

The ETF algorithm is analyzed in [5] and its performance ratio has the following bound: $\omega/\omega_{\text{opt}} \leq 2 - 2/m + \rho$.

As commented in [4] and [5] the relative performance of ETF can be decomposed in two parts. One of them is the Graham's bound $2 - 1/m$ and the other is the contribution of communication delays along a path of the graph, i.e. the ratio ρ. The time complexity of ETF (see [5]) is $O(mn^2)$.

Hanen and Munier [4] proposed an approximation algorithm called FS, based on an algorithm for unlimited number of processors and on a modification of ETF algorithm. They have proved that the performance ratio of their algorithm has the following worst-case bound:

$$\omega/\omega_{\text{opt}} \leq \frac{4+3\rho}{2+\rho} - \frac{2+2\rho}{m(2+\rho)}.$$

Möhring, Schäffter and Schulz [9] proposed another approximation algorithm, that is simpler than the algorithm in [4]. They first compute a schedule that regards all constraints except for the processor restrictions. This schedule is then used to construct a provable good feasible schedule for a given number of processors and as a tool in the analysis of the algorithm. The performance ratio of this algorithm is: $\omega/\omega_{\text{opt}} \leq 7/3$. In the next section is presented an approximation algorithm that in contrast to the above mentioned algorithms not is not based on a greedy procedure.

III. An Approximation Algorithm for the Small Communication Times Problem (AASCT)

The algorithm AASCT is based on the idea, that the arcs (i,j) of G having great c_{ij}-values should connect tasks performed by one and the same processor. In this way the tasks become favorite successors and the great delays are eliminated, which leads to reducing the greatest processing time and minimizing the makespan.

At the first step of AASCT is constructed a consequence of tasks (chain), beginning with the root of the spanning tree of G, so that to the current task i is added the task j for which the c_{ij}-value is maximal. In case there are many arcs having one and the same c_{ij}-value, then task j, for which p_j is maximal, is chosen as a next in the chain under composition. If there are many tasks, having one and the same processing time, then the task with smallest index is chosen. In case the current chain is composed (i.e. no more tasks can be added to it), the chain is assigned to the next processor in the list of idle processors. If there is not available idle processor, then assign the composed chain to the first processor which becomes idle. In case the starting task of the current chain needs data transfer from a task assigned to another processor, the corresponding communication delay should be added. A new graph G' is created by removing all tasks in this chain from G. Then graph G is replaced by G' and this step is repeated until no more tasks are available for composing new chains.

At the second step the task consequence on the processor π_k with greatest finishing time (makespan) is considered. This consequence is investigated, checking for each of tasks in it whether this task could be changed by one of the tasks assigned to the other processors π_i, $i = 1,...,m-1$; $i \neq k$; in such a way that the makespan would be reduced. In case the makespan has been reduced by such a change, than this step is repeated. In case on all other processors π_i, $i = 1,...,m-1$; $i \neq k$; there does not exist a task, which could change one of the tasks on π_k, which leads to reducing the makespan, than the algorithm passes to the third step. At this step $O(n-m)$ comparisons are performed, but the number of improvements may be arbitrary large because of combinatorial nature of the problem. For this reason we re-

strict the number of repetitions of this step to the small positive integer γ. Hence there are necessary $O(\gamma n^2(n-m))$ mathematical operations for the performance of this step.

At the third step a check-up is done, whether some one of the tasks on the processors π_i, $i = 1,,m-1$; $i \neq k$; could be started at an earlier moment on the same processor, so that some communication delay c_{ij} could be moved to an earlier moment and the makespan would be reduced. If the makespan is reduced in this way, repeat this step. Because there are n tasks available, the maximal number of check-ups is equal to $n-1$ and the corresponding number comparisons are $O(n^2)$.

Description of AASCT:

Step 0. Initialize $G' \equiv G$, $T' = T$, $n' = n$, s' is an empty chain.

Step 1. Chose an initial task i from T' (the root of G') and add it to the current chain s'.

For $j = 1,...,n'$; $(j \neq i, i \in s')$ find $c_{ij} = \max_{ij \in E'}\{c_{ij}\}$. and add the corresponding task j to s'. If more then one such arc has the same c_{ij}-value add the task j, having greatest p_j, to s'.

Repeat the **For** cycle until there are not available successors of the last task in the chain.

If there are idle processors available, assign the chain s' to the next processor π in the list of idle processors, otherwise assign s' to the first processor, which will become idle.

Remove all tasks in s' and their connecting arcs from G'. Initialize s' as an empty chain.

Repeat Step 1. until there are no more unassigned tasks.

Step 2. Set $icount = 0$. Find the processor π_k having the greatest finishing time t_k after processing all tasks assigned to it.

For each task $j' \in \pi_j$, $(j = 1,...,m-1; j \neq k)$ check whether it is possible one of the tasks $j \in \pi_k$, to be changed by j', reducing the makespan. If it is possible, do it and set $icount = icount + 1$, otherwise proceed Step 3 without repetitions and if reducing the makespan is achieved set $icount = icount + 1$, otherwise $icount = icount$.

If $icount < \gamma$ repeat Step 2, otherwise go to Step 3.

Step 3. For each processor π_j, $(j = 1,...,m-1; j \neq k)$, for each task j' assigned to π_j, check whether it is possible some to start it on the place of a preceding task on the same processor, so that the makespan is reduced. If it is possible, do it and repeat Step 3, otherwise stop the computations (End of AASCT).

Theorem: The time complexity of AASCT is $O(\gamma n^2(n-m))$, where m and n are the number of processors and the number of tasks correspondingly.

Proof: The most time-consuming step is Step 2, which is executed γ times in the worst case. This step requires $O(n-m)$ comparisons. Taking into account the operations of Step 3, which will be performed each time Step 2 is repeated, the running time of the AASCT algorithm is $O(\gamma n^2(n-m))$. □

IV. Basic Features of AASCT

As commented in [2] and [12] the increase the number of processors sometimes degrades the performance of the approximation algorithm ("anomalous behavior"). This feature has it reason in the essence of greedy procedures used. At Step 1. of AASCT algorithm the composed chains are dispatched uniformly to all processors, so that the anomalous behavior is reduced to a great extent. Step 2 and Step 3 reduce the makespan by means of changes of tasks on different processors and changing the starting time of a task on the same processor correspondingly. I this way they also contribute anomalous behavior to be avoided. The experiments performed by means of AASCT confirm this good characteristic of the proposed algorithm.

Another important feature of AASCT algorithm is that it does not use artificial delays, waiting for a more important task (in contrast to the algorithm presented in [4]) and it is reasonable to expect that the generated best schedule would have smaller makespan (finishing time) than the algorithms, which use artificial delays. The illustrative example presented in the next section confirms this presumption. The mentioned features lead to better performance of AASCT than that one of some other approximation algorithms as it is demonstrated in the next section.

V. An Illustrative Example

We will consider the illustrative example used in [4]:

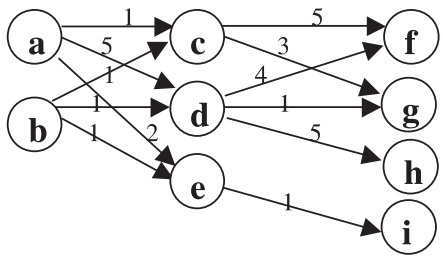


Fig. 1. Graph G with communication delays

The corresponding SCT task system for the example on Fig. 1. is presented in Table 1:

Table 1. SCT task system for the graph G from Fig. 1

a	b	c	d	e	f	g	h	i
6	7	9	8	10	6	6	10	6

On Fig. 2 and Fig 3. are presented two schedules as shown in [4].

Obviously the ETF algorithm creates a schedule with makespan (maximal finishing time) equal to 34. For the same example FS algorithm (see [4]) creates schedule with makespan equal to 29.

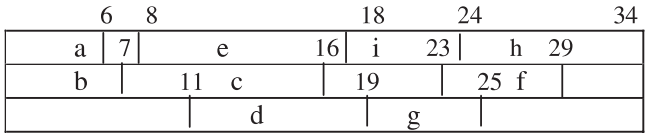


Fig. 2. An ETF schedule (3 processors)

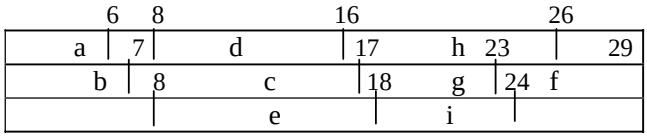


Fig. 3. A FS schedule (3 processors)

On Fig. 4 is presented the result found by means of AASCT algorithm for the same example (presented first in [4]). After Step 1 AASCT schedules on first processor tasks a, d and h; on second processor – tasks b, c, f and g; and on the third processor – tasks e and i, starting e at moment $t = 8$. The makespan is equal to 28. After Step 2 AASCT obtains the result on Fig. 4 with makespan equal to 27.

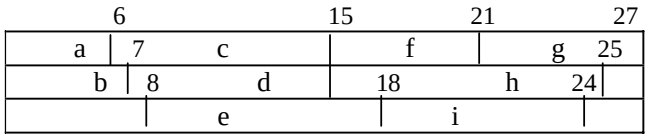


Fig. 4. An AASCT schedule (3 processors)

On Fig. 5 is presented the result obtained by AASCT algorithm for the same example but on two processors. After Step 1 the task schedule for the first processor is a, d, h, g; and for the second processor: b, c, f, e, i; The makespan is 38. After Step 2 task a is replaced by task b and the schedule is: for the first processor is b, d, h, g; and for the second processor: a, c, f, e, i; The makespan is reduced to 37. Step 2 is repeated and task g is replaced by task e. After applying Step 3 the result on Fig. 5. is obtained. The makespan is reduced to 35.

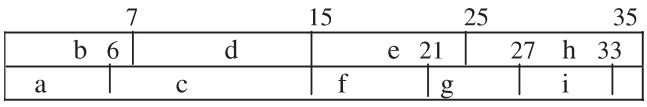


Fig. 5. An AASCT schedule (2 processors)

VI. Conclusions

An approximation algorithm called AASCT is presented in this paper. Its time complexity is $O(\gamma n^2(n-m))$. Since all approximate algorithms in the literature have polynomial computational complexity, the main criterion for comparison of their performance is the makespan (finishing time) of the generated schedule. The smaller the makespan is, the better the corresponding performance is. In this connection the relevant important features of AASCT algorithm are the avoiding anomalous behavior and artificial delays, which lead to its better performance in comparison to that one of ETF and FS algorithms, as well as (most likely) of some other approximation algorithms, based on greedy procedures.

References

[1] Chrétienne P., C. Picouleau (1995) *Scheduling Theory and its Applications*, P. Chrétienne, E. G. Coffman, J. K. Lenstra and Z. Liu (Eds.), 1995, John Wiley & Sons Ltd, pp. 65-90.

[2] Graham R. L. (1969) "Bounds on multiprocessing timing anomalies", *SIAM J. Appl. Math.*, Vol. 17, No. 2, pp. 416-429.

[3] Hanen C., A. Munier (1994) "Performance of Coffman-Graham schedules in presence of unit communication delays", http://citeseer.nj.nec.com/hanen94performance.html

[4] Hanen C., A. Munier (1995) "An approximation algorithm for scheduling dependent tasks on m processors with small communication delays", Preprint, Laboratoire Informatique Theoretique et Programmation, Institute Blaise Pascal, Universite Pierre et Marie, Curie.

[5] Hwang J. J., Y. C. Chow, F. D. Anger, and C. Y. Lee (1989) "Scheduling precedence graphs in systems with interprocessor communication times", *SIAM J. Comput.*, Vol. 18, No.2, pp.244-257.

[6] Jakoby A., R. Reischuk (1992) "The Complexity of Scheduling Problems with Communication Delays for Trees", *Lecture Notes in Computer Sciences*, No. 621, Vol. 3, pp. 165-177 Springer, Berlin.

[7] Liu Z. (1995) "Worst-case analysis of scheduling heuristics of parallel systems", No 2710, Institut National de Recherche en Informatique et en Automatique, http://citeseer.nj.nec.com/cache/papers/cs/1573/ftp:zSzzSzftp.inria.frzSzINRIAzSzpublicationzSzpubli-ps-zzSzRRzSzRR-2710.pdf/liu95worstcase.pdf

[8] Moukrim A., A. Quilliot (1998) "Scheduling with communication delays and data routing in message passing architectures", http://ipdps.eece.unm.edu/1998/scoop/moukrim.pdf

[9] Möhring R., M. Schäffter, A. Schulz (1996) "Scheduling jobs with communication delays: using infeasible solutions for approximation", http://citeseer.nj.nec.com/cache/papers/cs/5233/http:zSzzSzwww.math.tu-berlin.dezSzcogazSzpeoplezSzformer_members_pageszSzschaeffterzSzPaperszSzExtendedAbstract517.pdf/schedu

[10] Munier A., C. Hanen (1995) "Using duplication for scheduling unitary tasks on m processors with unit communication delays", http://citeseer.nj.nec.com/munier95using.html

[11] Munier A., J-C. König, (1997) "A Heuristic for a scheduling problem with communication delays", *Operations Research*, 45 (1), pp. 145-148.

[12] Rayward-Smith V. J. (1987) "UET Scheduling with unit interprocessor communication delays", *Discrete Applied Mathematics* 18, 1987, pp. 55-71.

[13] Veltman B., B. J. Lageweg and J. K. Lenstra (1990) "Multiprocessor scheduling with communication delays", *Parallel Computing* 16, pp. 173-182.

Capabilities of R-, R^+- and R^*-Tree Indexing Spatial Data in Network Space

Vladan Mihajlović[1] and Slobodanka Djordjević-Kajan[2]

Abstract – **Because of the complexity and large amount of data in spatial databases, SDBMS demands additional structures, i.e. indexes, for speeding up query processing. Indexes can be organized in three mayor groups of techniques depending on the way that the structure divides data space. This paper consider three hierarchical indexes, R-tree, R^+-tree and R^*-tree. An experiment is performed on data extracted from a telecommunication network in order to examine performance of named indexes in managing network spaces. The results of experiment indicate that R^*-tree is the most efficient and the most robust indexing structure among these three structures.**

Keywords – **R-tree, spatial data, network space**

I. Introduction

Capabilities of computers in nowadays, larger hard disks and faster processors, make possible using large scale of complex data. Spatial data complexity depends on a method applied for modeling real data. The amount of spatial data, growing with new demands, decreases application efficiency. Standard DBMS do not preserve the information about dimensionality of space and do not satisfied new trends. This disadvantage invokes creating Spatial DBMS (SDBMS), which make connection between scalar and spatial data.

Deference between standard and spatial DBMS is in part that defines spatial semantics. This part is additional layer above the DBMS, which provide interface for application. The core of this layer consists of spatial taxonomy, spatial data models, spatial query languages, query processing algorithms and spatial access methods [1].

Spatial taxonomy depends on real problem domain. It introduces restrictions in modeling relationship between the objects in real world. The most popular taxonomies are set based space, topological space, Euclidean space, metric space and network space.

Selected spatial taxonomy affects the selection of spatial data model. There are two groups of spatial data model: the models based on mathematical fields and the object model [2]. Nowadays, object models are used by most applications.

Execution of spatial query is much slower than execution of standard query. The reasons are complexity of spatial data and variety of relationship between them. Comparison of spatial data approximation, instead of real data, speeds up query processing. Nevertheless, this does not satisfy terms set by user due to great amount of data. Organize the data in groups following some spatial criterion will help to solve this problem. Index is a structure that arranges data in groups according to some group property.

An index is a data structure, which create records with property that each record contains data with specific key or that satisfied specific criterion. Thus query processing is reduced to examine particular record or smaller group of records and isolating data in it. In this way, indexes make spatial query processing efficient by decreasing the amount of data, which is needed to be retrieved from hard disk.

Nowadays, index is part of every DBMS, which manage database with large amount of records and objects of complex data types. Development of indexes for complex data types began in the field of spatial data. The indexes in spatial application are necessary due to complexity of spatial data objects. Indexes are usually built of simpler objects, which are approximations of real data. Accordingly they spare checking time and accessed disk pages. Indexes, as par of SDBMS, are used in Geographic Information Systems (GIS), CAD Systems, Computer Vision, Multimedia Information Systems and Data Warehousing Systems.

The rest of this paper considers application of spatial indexes in GIS. The second section gives an overview of different approaches to indexing spatial data. The third section outline definitions of R-tree, R^+-tree and R^*-tree, which are three very efficient and relatively simple indexing structures. Next section describes experiment, which use real data to compare performance of the three structures. Data used in experiment are extracted from telecommunication network. Results of comparison are presented in fifth section. The end of paper emphasizes accomplished research and gives some recommendation for selecting some of tested indexes in our applications.

II. Different Approaches to Spatial Data Indexing

Indexes designed for scalar data types are pretty simple since the key value can be determined strictly. A basic feature of spatial data is multidimensionality. This invokes problems since memory has only one dimension. This difference diminishes performance of indexes designed for scalar data types. This also provokes creation of fully new indexes designed for multidimensional data. Two major features of an index are how it partitions the data space and how it associates data with subspaces. According to these two features, indexing structures for spatial data can be classified in the three following approaches [3]:

[1]Vladan Mihajlović, Faculty of Electronic Engineering, University of Nis, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: wlada@elfak.ni.ac.yu

[2]Slobodanka Djordjević-Kajan, Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Nis, Serbia and Montenegro, E-mail: sdjordjevic@elfak.ni.ac.yu

1. The transformation approach. This approach comes in two flavors:

– *Parameter space indexing*. Object described by n vertices in a k-dimensional space are mapped in an nk-dimensional space. Points generated by such mapping can be stored directly in indexing structure designed for point data. An advantage of this approach is that is does not require creation of new index or modification of existing one. The main drawback is that the spatial proximity is not preserved in most cases.

– *Mapping to a single attribute space*. The k-dimensional data space is partitioned into grid where cells have the same size. The cells are labelled according to some curve-filling methods. So, spatial object is represented by set of numbers depend on cells that intersect with it. Therefore, objects can be indexed using conventional indexes for scalar data, like B$^+$-tree [4]. Shortcoming of this approach is multiplication of index entries.

2. The non-overlapping native space indexing approach. There exist two classes of techniques:

– *Object duplication*. A k-dimensional data space is partitioned into pairwise disjoint subspaces. These subspaces are then indexed. An object identifier is duplicated and stored in all subspaces it intersects.

– *Object clipping*. This is very similar to previous approach. Instead of duplicating the identifier, an object is decomposed into several disjoint smaller objects so that smaller sub objects are totally included in a subspace.

The most important property of this approach is that data structures used are straightforward extensions of the underlying point structures. This kind of indexes can store both points and objects with extensions together without having to modify the basic structure. Huge drawback of this approach is duplication of objects, which requires extra storage space and, more expensive insertion and deletion operations.

3. The overlapping native space indexing approach. The leading idea of this approach is to partition the data space into manageable number of smaller subspaces, which are hierarchically organized. Point object is included in one subspace, but nonzero sized objects may extend over more than one subspace. To assign nonzero sized object to exactly one subspace, subspaces are allowed to overlap. Spatial objects are indexed in their native space using this approach and that is its main advantage beside hierarchical organization. Main drawback is higher costs of insertion and deletion operations.

The overlapping native space indexing is the newest one. The basic advantage of this approach is preserving proximity relationship between objects that represent real data. Proximity is the basic relationship in every metric space. This indexing approach furthermore preserves proximity relationship between the entries on each level of hierarchical structure. A major design criterion for this approach is to minimize the overlap between subspaces and the coverage of subspace. The R-tree uses this kind of indexing approach.

III. Three Basic Types of R-Tree

An R-tree [5] is a high-balanced, hierarchically organized structure. A leaf of R-tree contains identifiers of database tuples, i.e. pointers to data objects. This tree is based on B-tree. Size of node corresponds to memory page size. R-tree is completely dynamic and no periodic reorganization is required.

R-tree is formed over data objects approximation. In two-dimensional space objects are approximate with minimum bounding rectangle (MBR). In the following we will describe two-dimensional R-tree. A leaf of tree consists of set of (I, id) entries, where id is unique database tuple identifier and I is MBR of data object identified by id. An internal node of tree contains entries of the form (I, ptr) where ptr is address of child node in the tree, and I is MBR that covers all rectangles in the lower node's entries.

Two values are typical for every R-tree. These are maximum number of entries that will fit in one node (M) and minimum number of entries in node $(m < M/2)$. Maximum number of entries is determined according to size of disk page. Minimum can be changed to improve structure performance. Only root node can have fewer entries than minimum. During the tree making node overflow or underflow will appear. The underflow is settled by reinsertion of entries that are rest in node. The overflow is resolved by splitting node in two parts. Two split algorithms are offered, linear and quadratic. Linear algorithm assigns elements to one of two new nodes one by one. Quadratic algorithm form two groups around two furthest elements using the criterion of minimum area covered.

Next variant of R-tree is R$^+$-tree [6]. Consider the advantage of the non-overlapping native space indexing approach author of R$^+$-tree modified R-tree so that overlapping between entries in internal node is equal to zero. Since there is no overlapping query, and there is no need to traverse multiple paths and the queries will execute faster. To achieve disjunction property between entries of level immediately above the leaf level same data object can be element of more than one leaf node. R$^+$-tree has not lower bound in number of node's entry (m) and if deleting is frequent performance of the tree can be deteriorate. This tree is not dynamic structure and requires periodic reorganization.

Objective of authors of R*-tree [7] is to prove that overlapping between internal node's entries does not lead necessary to index which has low efficiency. R*-tree has the same hierarchical structure as R-tree. The difference between them is in insertion and deletion algorithms. The authors carried out several criteria, which were believed to have the greatest effect on index performance. These criteria are: minimization of the area covered by rectangles in internal nodes, minimization of the overlap between rectangles in internal nodes, minimization of the margin of rectangles in internal nodes and optimization of storage utilization. The authors of R*-tree analyzed interdependencies among different parameters and optimization criteria and define insertion and deletion operation. Thus they proposed a split algorithm for R*-tree that firstly determines the axis perpendicular to which the split will be performed and then determines the best distribution of elements in two groups along that axis. One of following three criteria can be used in both parts of algorithm, unused area minimization, overlapping minimization and margin minimization. The new parameter introduced in

R*-tree is p. R*-tree treats overflow differently if it appears for the first time on particular level. In this case reinsertion is forced. Parameter p defines the number of entries that need to be reinserted. If an overflow appears for the second time in the same level of the tree, the split is performed. This forced reinsertion reorganizes the tree structure and improves its performance.

IV. Experiment Layout

Purpose of the experiment was to compare performance of three types of R-tree tested on real data. For this experiment all three R-trees are implemented in C^{++} language. Implemented trees manage two-dimensional data. Special software was designed in Visual C^{++} 6.0 programming environment. This software is used for performing large scale of tests on real and semantic data space.

In this paper will be presented tests performed on real data space. Data in this space is network of telecommunication canals and cables. Three tables of network data extracted from the spatial database. In the first (P1), the telecommunication cables represent spatial data. This space consists of 192 data and MBRs cover 10% of space area. The second table (P2) contains linear segments of the cables. This table has 848 tuples and data approximation occupy 1% of space area. Overlapping of data MBRs in previous two spaces is minimal. The third set of data (P3) consists of start and end points of cable segments and some key points in the telecommunication network. Total number of points is 5929.

Query represents search operation, which returns a set of data that satisfy requested criterion. For testing index structures, the three basic queries are used: point query and two types of rectangle query. The point query returns data, which MBRs contain given point. The overlap rectangle query returns data, which MBRs overlaps with given rectangle. The second type of rectangle query returns data, which MBRs enclose given rectangle. The first group of queries is point query (U1). The second are enclose query with query rectangle which area is 20% and 50% of elementary data (U2 and U3). Elementary data is square which area is equal to space area divided with number of data in space. Size of query rectangle in overlap query is equal to elementary data (U4), five times bigger (U5) and ten times bigger (U6) then elementary data.

Within each indexing structure the variant with the best performance is chosen for mutual comparison. To determine the best variant of R-tree was selected by changing parameter m from 10% to 50% with step of 5%. R-tree with linear and quadratic split algorithm was observed separately. Considered variants of R$^+$-tree have nodes filled from 75% to 95% of their capacity with step of 5%. To select best variant of R*-tree parameters m and p had values from 10% to 50% with step 10%, and with all possible combination of three criterion for choosing axes and distributions. Trees with different maximum number of elements in node (M) were compared separately because this parameter was imposed by memory page size. This parameter took four values, 13, 28, 56 and 113, according to page sizes of 0.5 kB, 1 kB, 2 kB and 4 kB.

Purpose of index is to minimize the number of disk access. Accordingly performance of three types of R-trees measured by number of memory pages accesses. Because all three structures have property that nodes exactly fit to one page, the efficiency is measured by number of nodes accesses during the query execution.

V. Interpretation of The Results

Because of small number of data in space P1 it is difficult to make general conclusions, but some trends can be isolated. R-tree that use quadratic split algorithm shows better results than linear counterpart. Property of R$^+$-tree that internal nodes must be disjunctive, even in this space with small number of nonzero sized data, demonstrate its main drawback significant increase in number of nodes in tree. All examined variants of R*-tree, with different criteria used in split algorithm, have nearly same performance (difference is below 10%). Variants of R*-tree with greater freedom in selecting minimum number of elements in node (m) have better results than the more restrictive variants.

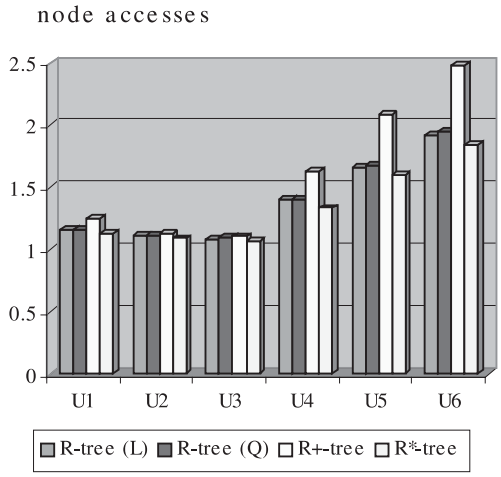


Fig. 1. Performance of indexes tested on linear segments as data objects of real network space

The results of query processing of P2 space demonstrate performance of tested index structures in real network spaces. Both split algorithms for R-tree, linear and quadratic, show almost equal performance. The variants of R-tree with greater freedom in number of elements in node appear to be superior. This advantage, which goes with greater freedom, is result of small MBRs of line segments and that each segment is a part of multi-line string that represent one cable. Same feature show R*-tree. This invokes rising of minimum node elements threshold for improving storage utilization. Criteria that are most efficient for R*-tree split algorithm are minimizing of the margin for choosing the axis and minimizing the overlap for selecting appropriate distribution. Parameter p has minimal influence on R*-tree performance. R$^+$-trees in P2 space show two features. The first is more elements in R$^+$-tree than in R-tree or R*-tree constant value for parameter M. The second is that efficiency of tree does not depend on storage utilization, as it is determined for point data. Fig-

ure 1 present performance of linear R-tree, quadratic R-tree, R$^+$-tree and R*-tree for different types of queries and parameter M set to 56. R$^+$-tree demonstrates poor performance, especially for overlap queries. This shortcoming is affected by traversing multiple paths since query rectangle intersect more then one entry in node. This is a consequence of poor algorithm for tree creation. R*-tree has minimum node accesses per query due to good minimization of node area and node overlapping. The other values for maximum number of entries in a node do not make qualitative changes on results.

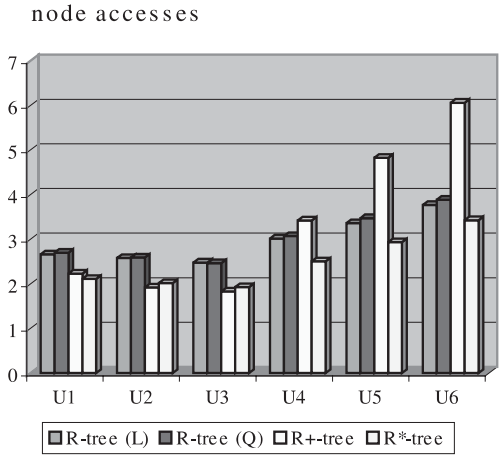


Fig. 2. Performance of indexes tested on points as data objects of real network space

Experiment on P3 space of points shows some distinction from expectation. Previous experiment performed by different authors concludes that quadratic split algorithm has better results than linear. The test on point data extracted from telecommunication network indicates to different conclusion. Linear variant of R-tree has better results than quadratic no matter what value is choused for M. Reason for this is probably the particularity of data space. Tests on synthetic data show that the best R*-tree variant uses minimization of margin and overlap as criteria for splitting node. In this space, minimization of node area results with most efficient tree. Although, the performances of variant that use other criteria are only 1% worse than the best variant. Also, storage utilization for R*-trees is much lesser than in synthetic data. If it is allowed to deteriorate efficiency about 1%, the variant of R*-tree with far better storage utilization can be found. Figure 2 show number of nodes access per query for different types of queries and M set to 56. R$^+$-tree has minimal node access when point and enclosing rectangle query are performed because the overlap between node's entries is equal to zero. But this advantage disappears with overlap rectangle query. That fact can be explained by large volume of unused space in internal node entries. Consider all six groups of queries R*-tree is the best choice overall. More data about the experiments can be found in [8].

VI. Conclusions

In previous papers that consider performance of R-tree authors agreed that a quadratic split algorithm creates better structure than a linear split algorithm. Linear and quadratic R-tree formed over telecommunication network space presented in this paper show nearly same efficiency. Larger overlapping between entries in internal nodes causes degradation in search performance. Basic property of R$^+$-tree, that overlapping between entries in internal nodes is zero, results in very efficient performing of rectangle enclosure query in point space. Disadvantage of this structure is unsatisfactory insert algorithm, which form entries with large unused space and make R$^+$-tree useless for nonzero sized data. Main drawback of this index is that it is not dynamic structure.

Algorithms of R*-tree are designed to enable greater influence of used criteria in creating nodes of index, which improve structure quality, i.e. minimize overlap and unused node area. Number of node access during query processing over R*-tree has logarithmic proportion with the number of data in space, i.e. depends directly on number of levels in tree. Quality of R*-tree has slightly affected by size and shape of data. Involved experiments with network data, which form variety of shapes, strongly affirm this fact. R*-tree is more robust and more efficient structure than the other two.

Acknowledgement

The research was partially supported by the project "Geographical Information System Designed to Improve the Local Municipal Function based on Internet/WWW Technologies", funded by Ministry of Science, Technology and Development, Republic of Serbia and Municipality of Niš, Contract No. IT.1.23.0249A.

References

[1] S. Shekhar, S. Chawia, S. Ravada, A. Fetterer, X. Liu, C. T. Lu, "Spatial Databases - Accomplishments and Research Needs", *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 1, Jan./Feb. 1999.

[2] M. Worboys, *GIS: A Computing Perspective*, Taylor & Francis, 1998.

[3] E. Bertino, B. C. Ooi, "The Indispensability of Dispensable Indexes", *IEEE Trans. on Knowledge and Data Engineering*, vol. 11, no.1, pp. 17-27, Jan./Feb. 1999.

[4] R. Bayer, E. McCreight, "Organization and Maintenance of Large Order Indices", *Proc. 1970 ACM-SIGFIDET Workshop on Data Description and Access*, pp. 107-141, Houston, Texas, Nov. 1970.

[5] A. Guttman, "R-Trees: A dynamic Index Structure for Spatial Searching", *Proc. ACM SIGMOD Intl. Conference on Management of Data*, pp. 47-57, 1984.

[6] T. Sellis, N. Roussopoulos, C. Faloutsos, "The R$^+$-Tree: A Dynamic Index for Multidimensional Objects", *Proc. of the 13th Intl. Conference on Very Large Databases Conference*, pp. 507-518, Brighton, 1987.

[7] N. Beckmann, H. P. Kriegel, R. Schneider, B. Seeger, "The R*-tree: An Efficient and Robust Access Method for Points and Rectangles", *Proc. ACM SIGMOD*, pp. 322-331, June 1990.

[8] V. Mihajlović, *Spatial data indexing*, Diploma thesis, Faculty of Electrical Engineering, Niš, May 200

Julia Multitudes and Theirs Computer Design

Slava Milanova Yordanova[1], Mariana Tsvetanova Stoeva[2]

Abstract – **In the recent issue are shown the Gaston Julia's multitudes and their computer presentation.**

Keywords – **Gaston Julia, multitude, segment**

I. Introduction

Many physical systems that are deterministic, it means that their future behavior is defined completely of the past condition of the objects, are so sensitive to the beginning conditions that theirs behavior is difficult to predict. The first difficulties in the deterministic approach in defining the condition of more complex system appear when the mathematicians as Cantor, fon Kox, Peano and Julia show geometric curves that are different from the previous. They are characterized as "selfsimilarity" it means that the shape of every little segment from the curve has the shape of the bigger segment. The length could not be easily defined and their size differ from that of the line and is situated between the size of straight line and plane.

The mathematician B. Mandelbrot concludes that many simple mathematical expressions could lead to "Chaotic" nonperiodical functions that although have stable behavior defined from the beginning conditions. Mandelbrot for the first time uses the concept "fractal"-curve, which size by Hausdorf-Bezicovich is higher than the size of the Euclid's space. The term "fractal" comes from the Latin "fractus", that means "irregular or fragmented". Different algorithms are made for generation of graphic computer fractal images. The mathematician M. F. Barnsly by examining the "Julia" multitudes searches different ways for generation of real images. He invented the method of "iteration functional systems" (IFS) that is complex of "iterative affined transformations" that define the connections between the parts of the image.

In the recent issue are shown the Gaston Julia's multitudes (1893-1978) and their computer presentation. [2,3]

II. Mathematical Presentation of the Julia's Multitudes

Let with C mark the Gaus's plane of complex numbers, and with $\bar{C}$ – the Riemann's sphere $C \bigcup \{\infty\}$ [2]. Let R is a rational function:

$$R(x) = P(x)/Q(x), \quad x \in \bar{C}, \tag{1}$$

where P and Q are polynomials that have not common divisors.

We presume that the function degree of R, $\deg R = \max\{\deg P, \deg Q\}$ is greater than 1. This degree is equal of the number of the proimages of the point x,

$$R^{-1}(x) = \{y \in \bar{C} : R(y) = x\}. \tag{2}$$

The Julia multitude J_R is commonly a multitude of exceptional points of the function $R : R^n(x) = R(...(R(R(x)))...)$, $n = 1, 2, 3, ...$ The addition to the multitude J_R is called Fatu multitude $F_R = \bar{C} \setminus J_R$. The classic definition for the Julia multitude is very comfortable for intuitive learning. That's why we take another definition, more accessible for understanding; it means the *periodic trajectory*. In every $x_0 \in \bar{C}$, the correlation $x_{n+1} = R(x_n)$, $n = 1, 2, ...$, defines certain sequence of points. This sequence is called *positive semitrajectory* of the point x_0 and is marked with $Or^+(x_0)$ [2,3]. When defining the negative semitrajectory could spring up difficulties because of the noncomplex of the reverse image R^{-1}. Although, taking all proimages we put:

$$Or^-(x_0) = \{x \in \bar{C} : R^k(x) = x_0 \text{ for } k = 0, 1, 2, ...\} \tag{3}$$

If $x_n = x_0$ in $Or^-(x_0)$ in known n, could be said that x_0 is periodic point. In this case $Or^-(x_0)$ is called periodic trajectory or circle that is marked with $\gamma = \{x_0, R(x_0), ..., R^{n-1}(x_0)\}$. If n is the smallest natural number having the pointed attribute, than n is called trajectory period.

In the case $n = 1$ there is the equality $R(x_0) = x_0$, it means that x_0 is immovable point of the function R. It is obvious that if x_0 is periodic point of the period n, than x_0 is unmovable point of the function R^n. (There have not to be mixed up the iterations of R and the rank of R, it means $R^n(x) = R \circ R \circ \cdots \circ R(x)$ and $(R(x))^n$.)

For the characterization of the stability of the periodic point x_0 with period n, have to be calculated the derivative. The complex number $\lambda = (R^n)'(x_0)$ $\left(' = \frac{d}{dx}\right)$ is called *self meaning* of the point x_0. Using the rule for differencing the complicated function we see that that number is the same for every point from the cycle. The periodic point x_0 is called:

- *supergravitation* $\Leftrightarrow \lambda = 0$,
- *neutral* $\Leftrightarrow |\lambda| = 1$,
- *gravitation* $\Leftrightarrow 0 < |\lambda| < 1$,
- *repulsion* $\Leftrightarrow |\lambda| > 1$.

The Julia multitude J_R could be described with the rational function R. Let P is multitude of all the repulsion periodic points of the function R. If x_0 is arbitrary gravitation unmovable point, than we examine its gravitation zone

$$A(x_0) = \{x \in \bar{C} : R^k(x) \to x_0, \text{ when } k \to \infty\}; \tag{4}$$

[1]Slava Yordanova is with the Technical University of Varna, Bulgaria, E-mail: slavov@ms3.tu-varna.acad.bg, sl@windmail.net

[2]Mariana Stoeva is with the Technical University of Varna, Bulgaria, E-mail: mariana_stoeva@hotmail.com

$A(x_0)$ consists of these points x, which positive semitrajectories $Or^+(x)$ agree in point. This multitude consists of the negative semitrajectory of the point x_0, $Or^-(x_0)$. If $\gamma = \{x_0, R(x_0), ..., R^{n-1}(x_0)\}$ is gravitation cycle of the period n, then each of the unmovable points $R^i(x_0)$, $i = 0, 1..., n-1$ of the function R^n has its own gravity zone and $A(\gamma)$ is just union of these zones [2].

III. Fundamental Features of the Julia Multitude

1. J_R consists of more than numbered multitude of points.
2. The Julia multitudes of Julia functions R and R^k, $k = 2, 3, ...$ coincide.
3. $R(J_R) = J_R = R^{-1}(J_R)$.
4. For each point $x \in J_R$ its negative semitrajectory $Or^-(x_0)$ is continuous in J_R.
5. If γ is gravity cycle of the function R, than $A(\gamma) \subset F_R = \bar{C} \backslash J_R$ and $\partial A(\gamma) = J_R$.

(Here $\partial A(\gamma)$ means the limit of the multitude $A(\gamma)$, it means that $x \in \partial A(\gamma)$, if $x \notin A(\gamma)$ and exists a sequence of points from $A(\gamma)$, agree in point x.)

On fig. 1 and fig. 2 are shown examples for Julia multitudes restricting two, even four different zones of gravity unmovable points.

6. If the Julia multitude has internal points (points $x \in J_R$, such that for known $\varepsilon > 0$, $\{x : |x - \bar{x}| < \varepsilon\} \subset J_R$) then $J_R = \bar{C}$.
7. Such a situation obviously is met rarely and although, one of the examples gives the function $R(x) = ((x - 2)/x)^2$
8. If $\bar{x} \in J_R$ and $\varepsilon < 0$, then exists whole n, in which $R^n(J^n) = J_n$.

From the features results of that each rational image has big reserve of repulsion points. That's why the Julia multitude is not changing when the image R is acting, but the dynamic of J_R is chaotic. The fifth feature shows the calculation algorithm for receiving images of the multitude J_R. Unfortunately the negative semitrajectory of the point $\bar{x} \in J_R$ usually is not distributed equally in the Julia multitude. That's why we need more complex algorithms for solving each time which of the branches of the tee structure $Or^-(\bar{x})$ have to be taken for most effective building of the image. Such algorithms are made and used for creation of our images. The sixth feature shows that in many of the cases the multitude J_R has to be fractal. For example, if R has more than 2 gravity unmovable points $a, b, c, ...$, then

$$\partial A(a) = J)R = \partial A(b) = J_R = \partial A(c) = \cdots , \quad (5)$$

it means that the limits of all gravity zones coincide. So, if R has 3 or 4 gravity unmovable points, then J_R consists of three sided or four sided points according to the gravity zones [2,3].

IV. Methods for Receiving of Computer Images of the Julia Multitudes

There are two different methods for receiving computer images of the Julia multitudes. One of them is based on the fifth feature and the other on the sixth feature. None of the methods has special advantages. In some cases the first method works better, in others the second. There are many cases when the two methods work perfect. But there are whole class of Julia multitudes for which is very difficult to be received satisfying images (if it is possible to receive any images). This class consists of Julia multitudes that limit parabolic regions, it means that correspond to images with parabolic periodic point.

A. Method of reverse iteration

If a rational image R is given and is known one periodic repulsion point $\bar{x} \in J_R$, then the feature (5) permits to be calculated

$$J_R^n = \{x \in \mathbf{C} : R^k(x) = \bar{x} \text{ for given } k \leq n\}. \quad (6)$$

B. Modified method for reverse iteration

The strategy is the following: over J_R is put rectangular grid with small size β. After this, for each cell B from the grid, have to be stopped the use of points from it for reverse iteration if certain number of $N_{\max}$ points in B are already have been used. It is usually that the optimal choice of β and $N_{\max}$ depends of μ_R and of the parameters of the computer image as the resolution of the used system. Therefore an iteractive and adaptive algorithms are necessary [2].

On fig. 1 and fig. 2 are shown two figures of images received from the Julia multitudes by the modified method for reverse iteration.

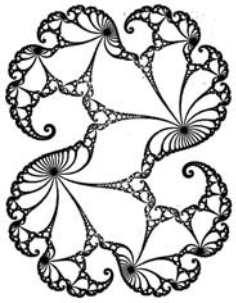


Fig. 1.

On fig. 1 is shown:

$$x \to (1 + \varepsilon)\lambda x + x^2 \quad (7)$$
$$\lambda = e^{\frac{2\pi i}{20}}, \quad x_0 = 0, \quad \varepsilon = 0.001$$

On fig. 2 is shown image of the Julia multitude with parabolic unmovable point for: [1,2,4]

$$x \to \lambda x + x^2 \quad (8)$$
$$\lambda = e^{\frac{2\pi i}{20}}, \qquad x_0 = 0.$$

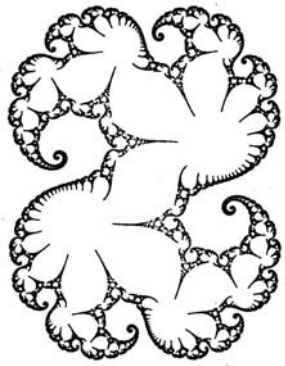


Fig. 2.

V. Additional Remarks

1. It is possible to be received interest pictures and with the help of two different colors – black and white. It is possible the use of color till $K = 200$;
2. The picture that is received is symmetric according the beginning of the coordinate system;
3. The time for calculation could be reduced two times because of the symmetry of the calculation process;
4. All the points that not incline to infinity after K steps, will be colorized in black, including the points that lay in regions of repulsion of other attractors, if exist such.

VI. Conclusion

The fractal image is infinite series of iteration and in the reverse iteration, the resolution of the display is not important, as independent from the level of zooming of the image, the level of the details is not changing.

A conclusion could be made that the method for signal and image compression through their fractal presentation is one of the most perspective and interesting in this field. It may be attention for additional scientist researches with purpose of creating new real program instrumentation for mass programs in effective and cheap computer multimedia systems. For example for the purposes of the television with high density – HDTV, such results will be necessary because of its future global use.

References

[1] Sl. Yordanova, B. Rachev, V. Naumov, *Multimedia and image compression*, Sofia 1999.
[2] H. Payten,, P. H. Rihter, *The Fractal Beauty*, 1993.
[3] Davis P.J. Hersh R., *The Mathematical Experience*, Birkhauser, Boston.
[4] G. Shuster, *Determinated Chaos*, 1999.

Business Aspects of Mobile Payments

Ljupco Antovski[1], Marjan Gusev[1]

Abstract – **M-Commerce like E-Commerce can be B2B (business to business), P2P (person to person) or B2C (business to customer) oriented. No successful mobile payment system has yet lived up the different requirements from the market and thereby not been a success. A brief research on the state of the market is given to present a framework for possible solutions. The purpose of this paper is to describe the factors that affect the introduction of a successful M-Payment system.**

Keywords – **M-Commerce, J2ME, HTTPS, cryptology, FSP**

I. Introduction

Mobile phones are already approaching penetration rates higher than 80 per cent in some parts of the world. Penetration is considerably lower but growth rates are high. High market penetration and a number of technical features make mobile phones very interesting commerce devices.

With the growing momentum of wireless revolution and M-Commerce explosion, it is evident that mobile devices are becoming a critical component of the new digital economy.

The transactions are rapidly transitioning from fixed locations, to anytime, anywhere and anyone. New forms of mobile technologies are rapidly transforming the marketplace. Optimists are of the opinion that the new world economy will witness the transition of mobile devices from a simple communication device to a payments mechanism. [14]

There have been different definitions of M-Commerce. Lehman defines M-Commerce as "*the use of mobile handheld devices to communicate, inform, transact and entertain using text and data via connection to public and private networks*" [25]. Their reason for using such a broad definition is because the borders between messaging and commerce have become too blurred to separate these categories. Another definition is "finance transaction especially buying and selling: trading"[26]. Durlacher research's use a fairly broad definition as they as more distinct and is as follows: "any transaction with a monetary value that is conducted via a mobile telecommunication network" [25] M-Commerce contributes the potential to deliver most of what the internet can offer, plus the advantage of mobility. M-Commerce gives mobile communication devices as mobile phones and personal digital assistants (PDA) the ability to pay for goods and services.

II. Services

While most of existing eCommerce application can be modified to run a wireless environment, M-Commerce also involves many more new applications that become possible only due to the wireless infrastructure.

These applications include mobile financial services, user and location specific mobile advertising, mobile inventory management, wireless business re-engineering, and mobile interactive games. In addition to device and wireless constraints, M-Commerce would also be impacted by the dependability of wireless infrastructure.

M-Commerce existing and futures possible application include:

- Mobile banking service (check account information, money transfer)
- Mobile trade service (stock quotes, selling/buying)
- Credit card information (account balance)
- Life insurance account information (account information, money transfer)
- Airline (online reservation, mileage account check)
- Travel (online reservation, timetables)
- Concert ticket reservation (online or telephone booking)
- Sales (online books, CDs)
- Entertainment (games)
- News/information (headline, sports, weather, horse racing information, business,
- technology, regional)
- Database, application (yellow pages, dictionary, restaurant guide)
- Location based application (area information and guides)

III. M-Commerce Segments

M-Commerce like E-Commerce can be B2B (business to business), P2P (person to person) or B2C (business to customer) oriented. The scope of this paper is on the B2C model.

In the B2C area, M-Commerce is still in its infancy. This is due to the limitations of present, intermediate technologies such as WAP, and to the relative lack of compelling contents and services. Certain B2C services (e.g. online banking) may charge a small monthly free, but it is similar to that of comparable offline service (e.g., maintenance fee for checking accounts) and are waived under certain circumstances (e.g., if a minimum balance criterion is met), hence monetary cost is not a constraint on B2C E-Commerce acceptance [27].

The M-Commerce framework divides into couple sub areas based on user's distribution criterion. Mobile E-Commerce addresses electronic commerce via mobile devices, where the consumer is not in physical or eye contact

[1]Both authors are with the Institute of Informatics, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University, Arhimedova b.b., PO Box 162, 1000 Skopje, Macedonia, Email: anto@ii.edu.mk, marjan@ii.edu.mk

with the goods that are being purchased. On the contrary in M-Trade the consumer has eye contact with offered products and services. In both cases the payment procedure is executed via the mobile network [1,5].

M-Commerce involves procedures of M-Payments (Mobile Payments) defined as payments carried out via mobile devices. The highest state of security has to be implemented in these procedures in order to ensure full reliability and trust from the customers in the system [1].

Principally, M-Payments can be used for M-Commerce, E-Commerce and in the real world. In the real world, it is the number of mobile phones that makes them a promising payment device. In 2000, trade via handy, pager and handheld has created revenues of EUR 1.3 billion in Europe and is expected to rise to EUR 3.8 billion in 2003 (BITKOM). The corresponding estimate for global M-Commerce in 2003 is USD 13 billion (Barnett/Hodges/Wilshire). By this estimates by 2005, data traffic is expected to be more important than voice traffic [12]. Similar research by Andersen [13] estimates that the European mobile content market size could range between EUR 7.8 billion to EUR 27.4 billion in 2006, with a median forecast of EUR 18.9 billion.

Many mobile operators have started offering M-Payment services. These services are in early stage and still in beta state. Several operators team up with banks while others manage M-Payment on their own [10].

There is a wide range of solutions concerning mobile payments services. The security implementation spreads from SMS messaging, PIN confirmation to financial message signing, encryption, use of tamper-resistant devices and digital certificates. Main characteristic of all this solutions is that they could only be used by limited number of users that fulfill the required technical specification.

IV. Current Protocols and Technologies

No new special network standard is needed to carry out M-Payment transactions. M-Payments are therefore carried out through existing networks, which could be Cellular networks (GSM/2,5G/3G), Wireless LAN (IEEE 802.11 protocol), Bluetooth and Infrared (irDa)

The most important technologies for M-Payment connectivity are: SIM Application Toolkit (SAT), WAP/WTLS/WIM, Voice and Manufacturer specific Applications SAT is a technology that allows configuring and programming the SIM card [15]. The SIM card contains simple application logic that is able to exchange data with the SMSC, to carry out M-Payment transactions. The specific mobile operator provides the application logic and is responsible of providing the SIM card.

Phones equipped with a WAP-browser are able to exchange data with a web server. Data is transmitted via wireless application protocol and the networks are GSM, 2.5G or 3G. WTLS is a layer in the WAP stack and is the wireless edition of the SSL 3.0 in a reduced scale. WTLS can provide secure connections for transferring confidential data [16]. WIM is a module for storing data in the mobile device and is usually used in relation to WAP transactions. WIM is used with WTLS transaction to protect permanent, typically certified, private keys. The WIM stores these keys and performs operation using these keys [17].

The end-user can via a normal phone call state his credit card number to the merchant that transfers the funds via interface provided by a PSP. A voice response system at the payment service provider can also call the end-user and guide him through a payment procedure. Voice recognition can also be used as an authentication tool for payment settlement.

The mobile phone manufacturers can chose to install native applications, which in interaction with one of the above technologies enables M-Payment opportunities.

V. Critical Success Factors

There are six main actors involved in a Mobile Payment System(MPS) [ShSw98] [Pay01]: Financial service providers (FSP), Payment service providers (PSP), Merchants, End-users, Network service Providers (NSP) and Device Manufacturers. These are further divided in users and system providers. There are different critical success factors and requirements considering the involvement of different actors.

Table 1. Critical success factors

Factor	Features
Ease of use	few clicks, intuitive, flexibility, performance, installing/download
Security	privacy, confidentiality, integrity, authentication, verification / non repudiation
Comprehensiveness	transferability, divisibility, standardization.
Expenses	set up fees, transaction fees, subscription fees
Technical Acceptability	integration effort, interoperability, scalability, remote access, performance

An important means of getting a successful MPS is obtaining acceptance from all the participants in the network and thereby achieving a critical mass. By comprehensive study from several authors [18, 19] success factors are identified: Ease of use, Security, Comprehensiveness, Expenses and Technical Acceptability. The Table 2 is an overview of the main factors features.

VI. New M-Payment Method

The foundation and ideology Java 2 Micro Edition (J2ME) brings itself a reasonable set of potentials of being a part in a MPS. There are several concrete arguments that indicate why J2ME should be considered as an interesting supplement for

M-Payments as: Broad customers experience, Comprehensiveness, Lower network and server load, Internet Enabled, Constant storage, Sun Microsystems have added an unofficial support for HTTPS (kSSL) as a part of the MIDP 1.0.3 reference implementation and the J2ME Wireless Toolkit version 1.0.3 [23]. HTTPS is not required by the MIDP 1.0 specification but if device manufactures releases devices supporting HTTPS, they will in theory be able to carry out secure transactions. In order to overcome the cryptographic gap a concrete initiative called Bouncy Castle has released a lightweight API (BC-API) with cryptology and certificate facilities, designed for J2ME. The BC-API provides a security toolbox obtained from the original Java Cryptography Architecture (JCA) and the JAVA Cryptography Extension (JCE) and has been boiled down to support the CDC and CLDC devices [24].

Considering the above exposed features of J2ME we propose a new M-Payment protocol that has the HTTP protocol as bearer. Due to the fact that SSL is still not supported in MIDP specification, the encryption, signing and certificate verification is managed at application level using the BC-API third party classes.

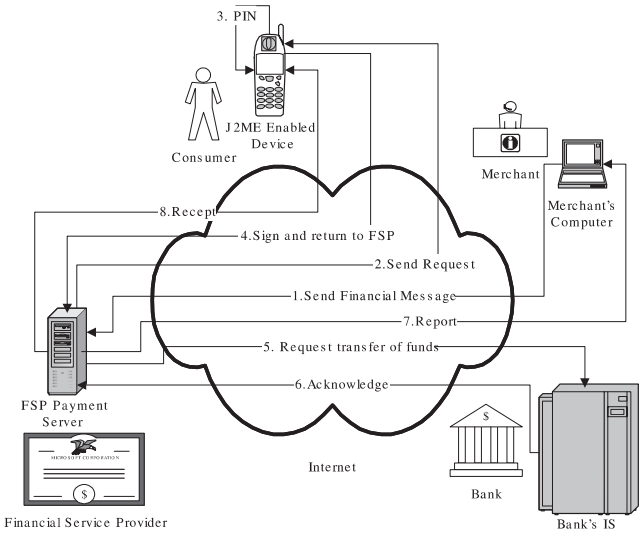


Fig. 1. M-Payment Protocol

The protocol (Fig.1) is executed in the following manner:

1. The merchant's computer issues a financial message that is encrypted and signed. Over secure Internet connection, (over SSL) the FSP receives the message.

2. The FSP verifies the source, signs, encrypts and redirects the message to the designated mobile user.

3. The user receives the message and verifies the source. If the source is the FSP gateway, the procedure continues otherwise it terminates. Afterwards the user enters PIN (or password) which is used to decrypt the encrypted private key stored in the persistent record store. Then the message is encrypted by asymmetric algorithm with session secret and sent to the FSP.

4. The encrypted message is send to the FSP. It validates the message source.

5. The FSP validates the signature. Then a request is send to the bank's information server to begin transaction from customers to merchant's account. In other scenarios the transfer of funds is from one account to another in the mobile operator's network. These accounts could be prepaid or postpaid, that involves additional procedures for validation and clearing.

6. The FSP is acknowledged after successful transfer of funds.

7. The merchant receives notification.

8. The user receives receipt in digital manner.

The procedure emphasized above addresses the M-Trade scenario. In the mobile E-Commerce scenario the procedure differs in the first steps when the user chooses the products and services and in the last steps when the merchant receives the report of successful payment and initiates shipment.

In order to lower the network load a new message system is introduced. The message transferred by the Interactive Message System (iMS) is predefined and contains financial and address data. The message represents a virtual envelope with enclosed letter. The message is divided in three sections [1]. The Extendable Markup Language (XML) is used to define the structure of the message [8].

VII. Conclusion and Future Work

It is evident that M-Payment methods are here to stay with M-Commerce gaining momentum. Lack of standards and security within devices as well as network may be pertinent issues for the future of M-Payments. A range of solutions involving financial institutions and mobile service providers seem to be in progress, and perhaps is the key to addressing these issues. The lack of standards across economies may be addressed through various consortiums, involving many economic forums, mobile operators and also financial institutions, if M-Commerce has to be diffused into the mass-market. Security has been an issue of M-Commerce development right from the start of this effort. Current infrastructures considering the limitations and enhancements, offer a comfortable environment for secure mobile payment transactions. Many challenges are involved in building an M-Commerce solution, and just as many ”solutions” available on the market. The comprehensive M-Payment suite combines strategy and analysis with rapid, fully customized technical solution development and implementation, resulting in a high return on the investments. The above proposed models of mobile payments are easy to implement considering the available technology infrastructure. The models are simple, secure and scalable.

References

[1] M. Gusev, Lj. Antovski, G. Armenski; “Models of mobile payments”, Proceedings of the 2nd WSEAS International Conference on Multimedia, Internet and Video Technologies (ICOMIV 2002), ISBN 960-8052-68-8, 25-28 September 2002, Skiathos, pp. 3581-3586

[2] O. Pfaff, Identifying how WAP can be used for secure m-business, Proc. of 3RD Wireless m-business Security Forum, 29-30 January 2002, Barcelona

[3] D. Amor, The E-business Revolution, New Jersey: Hewlett Packard Books, 2002

[4] Lj. Antovski, M. Gusev, Ebanking-developing future with advanced technologies. Proc. of 2nd Conf. on Informatics and IT, December 2001, Skopje, pp. 154-164

[5] D. Bulbrook, WAP: A Beginner's Guide, New York: Osborne/McGraw-Hill, 2001

[6] M. Gusev, E-Commerce, a big step towards e-business. Proc. of 2nd SEETI Conf. On Trade Initiative and Commerce, November 2000, Skopje

[7] R. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM, Feb.1978 Vol.21, pp.120-126

[8] W3C: http://www.w3.org (accessed 20.10.2002)

[9] WAP-forum: http://www.wapforum.org (accessed 15.10.2002)

[10] H. Knospe, S. Schwiderski - Grosche, Online payment for access to heterogeneous mobile networks, Proc. of IST Mobile & Wireless Telecommunications Summit 2002, June 2002, pp.745-752

[11] S. Pantis, N. Morphis, E. Felt, B. Reufenheuser, A. Bohm, Service Scenarios and business models for mobile commerce, Proc. of IST Mobile & Wireless Telecommunications Summit 2002, June 2002, pp 551-561

[12] Niko Mykkanen, Mobile Payments - A report into the state of the market, Commerce Net, Scandinavia, October 2001

[13] European Commission DGIS, Digital content for global mobile services final report, Andersen, Europe, February 2002

[14] M.Ding, and C. Unnithan, Mobile Payments (mPayments) – An Exploratory Study of Emerging Issues and Future Trends, School Working Papers Series 2002, Deakin University

[15] Guthery, Scott B. & Cronin, Mary j, Mobile Application Development with SMS and the SIM Toolkit, McGraw-Hill 2002

[16] WMLScript Crypto API Library Specification, WAP-161-WML Script Crypto - 20010620-a, Version 20-Jun-2001.

[17] Wireless Application protocol Forum Ltd, "Wireless Identity Module Specification, WAP-260-WIM-20010412-1", Version 12-july-2001.

[18] Shon T.W. and Swatman P.M.C., "Effectiveness Criteria for Internet Payment Systems", Internet Research: Electronic Networking Applications and Policy, Vol. 8, No. 3, 202-218, 1998

[19] Heijden, Hans van der, "Factors affecting the successful introduction of mobile payment systems", Vrije Universiteit Amsterdam, 2002

[20] Sun Microsystems, "Designing Wireless Enterprise Applications Using java. Technology, http://java.sun.com/blueprints/, Jan 2002.

[21] Mobile Information Device Profile (MIDP) Specification ("Specification"), Version: 1.0, Release: September 15, 2000, Copyright 2000 Sun Microsystems, Inc.

[22] Sun Microsystems, Inc. "Connected, Limited Device Configuration (CLDC) Specification, version 1.0a", Sun Microsystems, Inc., may 19, 2000.

[23] Mahmoud, Qusay H. "Secure Java MIDP programming using HTTPS with MIDP", http://www.wireless.java.sun, june 2002

[24] Bouncy Castle, the Specification, http://www.bouncycastle.org, v. 1.1.4, 2002

[25] Lehman Brothers Moving in mobile media Mode (1995, p.8)

[26] Haddon, Communication on the move: the experience of mobile technology in the 1990s. COST 248, European Commission, Sweden, Telia AB, 1997.

[27] Bhattacherjee- Anol, Acceptance of E-Commerce Services: The case of Electronic Brokerage, Man and Cybernetics, 2000

User-Level DMA Extension for NOW/Cluster Applications

Alexander P. Kemalov[1]

Abstract – **Direct Memory Access /DMA/ is previously used to transfer data between the main memory of host computer /PC/ and the network → to another one. This method is used to free the processor from the burden of transfer operations. DMA procedures commonly are initiated by the operating system kernel to separate one application and its data with another.**

A Network of Workstations /NOW/ architecture suggest that interconnections get faster and overhead and latency in networks go down while operating system operations get slower. In NOW or clusters these factors are very important because an intensive data transfers between hosts. These trends imply that DMA operation becomes slower /using operating system kernel/, compared to interconnection network.

This paper proposes several algorithms that allow applications to start DMA operation without OS kernel. The algorithms allow user-level applications to have direct access to the DMA engine. This approach is achieved without requiring changes to the OS kernel. Using our algorithms, DMA operation can be initiated faster /in comparison to OS kernel/.

Keywords – **DMA operation, memory allocation, networks, operating system kernel**

I. Introduction

Direct Memory Access /DMA/ is a common method for routing data directly between memory of host computer /PC/ and an input/output device /network controller/ without requiring enervation by the CPU. DMA management has been traditionally done by Operating System kernel, which provides protection, memory, buffer management, DMA registers and address translations. The overhead of this kernel initiated DMA transaction is hundreds of CPU instructions. There are two reasons for the necessity of the OS involvement in starting a DMA operation:

1. Atomicity – DMA operation start with transfer to DMA engine a source address, destination address and size of DMA packet. The process invokes OS to start an and schedule DMA operation and when finished, start another.
2. Protection from program errors – DMA engine works only with physical addresses, not allowed to access to user programs.

The user program use virtual address and must be translated to physical one. Virtual-to-physical address translation is performed by OS kernel. The physical memory pages used for DMA must be pinned to prevent the virtual memory system form paging them out while DMA data transfers are in progress.

[1]Al. Kemalov is with ICCS institute of BAS, Akad. G. Bonchev str. Bl.2 1113 Sofia Bulgaria, sasho@hsi.iccs.bas.bg

In the last years, high-speed LANs offers great performance and communication throughput and overheads of the OS involvement in DMA operation are still sensitive. For this reason, several researches have started to address the problem of letting user applications initiate DMA operation. Projects SHRIMP [1] and FLASH [2] have pinpointed the importance of user-level DMA. A disadvantage of these approaches is needs of modification of OS kernel.

A DMA procedure has three arguments: *source address, destination address* and *size* of packet. A DMA engine is responsible to perform above sources.

The OS translate the virtual source and destination addresses to their corresponding physical addresses and size to the DMA engine registers and start a DMA transfer. One of the common used techniques to secure virtual-to-physical address translation is the notion of *shadow addressing* [3]. For each virtual address **vaddr** correspond physical – **paddr** and **shadow(paddr)**. The shadow address is concatenating the physical one. The difference is shadow bit in address / for example 0x0FFFF → regular address; 0x1FFFF → its shadow address; range is within the physical address range; it is made in initialization time/.

An access to shadow address is interpreted by the DMA engine – virtual address **vaddr** is mapped to physical address **paddr** and virtual shadow address **shadow(vaddr)** - to **shadow(paddr)**. A transformation virtual-to-physical addresses use TLB (page-table) and is performed by memory controller. When a user application tray to pass the DMA engine, it will be treated as an argument passing operation and reject access to regular physical address. Thus it makes an access to virtual **shadow(vaddr)**. The DMA engine recognizes the shadow address and takes the physical address **paddr** by applying function shadow to physical address **shadow(paddr)**.

The mechanism of shadow addressing is fast and reliable, to pass physical addresses to a DMA engine from user memory space.

Another problem is to guarantied atomicity of a DMA operation. If there were a way to execute two instructions uninterrupted, then the problem will be solved. But from security point of view, it is dangerous because malicious users may be monopolizing an execution of programs – a decision is OS control.

II. First User-Level DMA Algorithm

The DMA engine is equipped with /4 to 8/ register contexts. Each context has a source, destination and size registers with their meaning. Each context is mapped into memory address space so that the processor can access it. Distinct context are

mapped into distinct memory pages so that each process gets access rights for only a single context. Each process can start user-level DMA operation /to write into single group context registers/. Thus if a process gets interrupted while starting a DMA, its arguments can't be mixed with another process's arguments. Each process has its own space in the DMA engine.

Unfortunately, user-level application can't use regular load/store operations to access these registers and load them with arguments of a DMA operation. Thus a process that would like to pass a physical address to a register context will pass context identification as a data argument of store operation, since the address argument of store has already been reserved to pass the shadow address:

STORE context_id **TO** shadow(vaddr)

The DMA engine extracts the **paddr** from **shadow(paddr)** and put it in register context **context id**. To start a DMA, a process makes a sequence of above store operations. Unfortunately any process will be allowed to write an address argument into any register context. To prohibit this, we introduce a key that implies the user process is allowed to register context. Thus a physical address is passed to a DMA engine:

STORE key#context_id **TO** shadow(vaddr),

i.e. to proof key in OS and in an instruction is permitted to store a physical address as an argument in the register context. Using the above instruction the address arguments are securely passed to the DMA engine.

The last argument that must be passing is size of DMA packet. In this case we used regular store operation to the address that corresponds to the register context /size register/.

A user-level DMA operation is performed in fig. 1.

The last argument that must be passing is size of DMA packet. In this case we used regular store operation to the address that corresponds to the register context /size register/.

A user-level DMA operation is performed in fig.1.

We used store /not load/ instructions to load address arguments because a process that have both read and write access to the source address will be able to start user-level DMA operation from it. Most parallel and distributed applications use DMA procedures that have both read and write accesses to these data.

```
/* the KEY allows the process to write arguments into CONTEXT_ID */
global  KEY, CONTEXT_ID;
 /* The register context is mapped into address REGISTER_CONTEXT */
global  address REGISTER_CONTEXT ;
DMA(vsouce, vdestination, size)
      /* pass the destinatio n argument */
      STORE KEY#CONTEXT_ID TO shadow(vdestination);
      /* pass the source argument */
      STORE KEY#CONTEXT_ID TO shadow(vsource);
      /* pass the size argument */
      STORE size TO REGISTER_CONTEXT;
      /* did it succeed? */
      LOAD return_status FROM REGISTER_CONTEXT
```

Fig. 1.

III. Second DMA Algorithm

The algorithm proposed above, achieves user-level DMA operation without OS kernel modification. But theoretically may be broke from a lucky user, who manages to guess another user's key. To avoid this one, we make the identification of the process part of the shadow address.

We introduce some bits of the physical address that will be passed as an argument to the DMA engine corresponds to the process identification. These bits are set by the OS when it creates the mappings from shadow virtual addresses → to shadow physical addresses. Part of the shadow physical address is now the context_id /2 bits/, i.e. 4 processes will be able to start user-level DMA operation from the same processor /fig. 2/.

```
DMA(vsource, vdestination, size)
   /* pass physical address shadow(vdestin.) and size to the DMA engine */
   STORE size TO shadow(vdestination);
   /* pas s physical addr shadow(psouce) to the DMA engine and read if successful */
   LOAD return_status FROM shadow(vsouce)
```

Fig. 2.

By checking the context_id , the DMA engine knows which process the shadow address belongs to. The DMA engine has several register contexts to save these addresses, receives in the appropriate contexts and start the DMA operation when all arguments are available. If there are no register contexts and DMA engine receives pairs of STORE and LOAD instructions, it checks for the context_id value of the two physical addresses. If they are different, DMA is not started and an error is returned by the last LOAD instruction.

IV. Third Algorithm

In the last algorithm we tray to start user-level DMA operation without the need extra bits in the physical address /context_id /. If a process passes at least one shadow address more than once, the DMA engine may be able to determine if the user process was interrupted. The proof is checking the two successive accesses to the same shadow addresses. The DMA engine initiates a DMA operation only if it sees a sequence of the form LOAD, STORE, LOAD and arguments of the two load instructions are the same. If the process is interrupted while trying to start a DMA, then the DMA engine will receive a non valid sequence of shadow addresses, and DMA is not start.

The above sequence may lead to error data transfer, if abused by malicious user – a possibility of interleave of shadow address.

We introduce additional instruction to protect this situation:

```
DMA(vsouce, vdestination,size)
      STORE size TO shadow(vdestination)
      LOAD return_stat.1 FROM shadow(vsouce)
      STORE size TO shadow(vdestination)
      LOAD return_stat.2 FROM shadow(vsource)
      LOAD return_stat.2 FROM shadow(vsource)
```

If a malicious user does not have access to addreses vsouce, vdestination, above sequence will work correctly. To provide this and avoid interleaving, we include additional instruction: /Fig. 3/

```
1: STORE size TO shadow(vdestination)
2:LOAD return_stat.1FROM shadow(vsouce)
   If (return_stat.1==FAILURE) go to 1:
3:STORE size TO shadow(vdestination)
4:LOAD return_stat.2 FROM shadow(vsource)
   If (return_stat.2==FAILURE) go to 1:
5: LOAD return_stat.3 FROM shadow(vdestin.)
   If (return_stat.2==FAILURE) go to 1:
```

Fig. 3.

The shadow(vsource) address pass twice to the DMA engine, while shadow(vdestination) address – three times. The DMA engine is prepare to receive 5 instruction sequence to shadow address space and a DMA operation start only if there are sequence STORE, LOAD, STORE, LOAD, LOAD and the address arguments in instructions 1,2,5 and 2,4 are the same.

V. Proof of Correctness

We proof above algorithms with a testbed including a pair Pentium III workstations, running MSC.Linux OS rev. nov.02 /special version for cluster and distributed applications/. The test applications consists a server and client /ping-pong/ messages with acknowledgments and different size of packets.

A DMA operation would be initiated incorrectly if a user process attempt to start a DMA, are interrupted and interleave their address arguments. Suppose that process P1 want to start DMA from A1 memory location → to A2. Suppose that there are several other processes P2...Pn interleave their instructions with P1. Although other processes may have read only access to A1, they do not have access to A2. Assume that all P2Pn execute subroutine in fig. 3 and want to write/read the same physical address. If processes P2...Pn belong to different applications, then they should not be able to write-share the same physical memory location, since different applications do not write-share physical memory. Thus such an interleaving can't happen.

If P2...Pn belong to the same application, then there should be some synchronization operations be include before they all attempt to write/read the same memory location. This synchronization should serialize DMA operations.

If all access to A1 were issued to P1, that process has also issued two interleavening LOAD instructions to A2 as well. Thus all trying to access to A2 is reached from DMA engine. If a DMA started all five instructions must have been issued by the same process /P1-successfully started DMA/.

VI. Conclusion

UDMA allows user process to initiate DMA procedure to or from I/O nodes at a cost of only two user-level memory reference and additional instructions. These extremely low overheads with using of UDMA.

The UDMA procedure does not require much additional hardware because it takes advantage of both hardware and software in the existing virtual memory system. This is very important in a process of an implementation of cluster architectures in practice.

In the future we tray to implement these procedures in I/O operations in a cluster architecture and GRID middleware. This task transform DMA procedure in remote I/O paradigm in which applications use familiar parallel I/O interfaces to access remote file systems.

References

[1] M. Blumrich, R. Alpert, Y. Chen, D. Clark, C. Dubnicki, Design Choices in the SHRIMP system: An Empirical study. In *Proc. Of 25th Intern'l Symp. On Computer Architecture*, 1998

[2] J. Heinlein, K. Gharachorloo, S. Dresser, A. Gupta, Integration of Message Passing and Shared Memory in the Stanford FLASH Multiprocessor. In *Proc. of 6th Intern'l Conf. on Architectural Support for Progr. Languages and OS*, 1994

[3] C. Dubnicki, A. Bilas, Y. Chen, K. Li, VMMC-2: Efficient Support for Realible Connection Orienteted Communication. In *Proc. of Hot Interconnects*, 1997

[4] R. Dimitrov, A. Skjellum, An Efficient MPI Implementation for Virtual Interface (VI) Architecture - Enabled Cluster Computing. http://www.mpi-softtech.com

[5] Intel Corp. Intel Virtual Interface Architecture - Developer's Guide http://developer.intel.com/design/servers/vi/developer

[6] M. Buchanan, A. Chien, Coordinated Thread Scheduling for Workstation Clusters under Windows NT. In *Proc. of USENIX Windows NT Workshop*, 1997

[7] NERSC PC Cluster Project at Lawrence Berkeley Nat'l Laboratory M - VIA: A High Performance Module VIA for Linux; http://www.nersc.gov /research/FTG/via

[8] St. Muir, J. Swift, Functional divisions in the Piglet multiprocessor operating system, In *ACM SIGOPS European Workshop*, 1998

[9] K. Schwan, R. West, M. Rosu, A Network Co-processor based Approach to Scalable Media Streaming in Servers. In *Intern'l Conf. on Parallel Processing*, 2000

[10] G. Banga, J. Mogul, Scalable kernel performance for Internet servers under realistic loads. In *USENIX Technical Conference*, 1998

Software System for Multicriteria Decision Making

Filip B. Andonov[1], Krassimira B. Genova[1], Boris A. Staykov[1], Mariyana V. Vassileva[1]

Abstract – **The paper discusses a multicriteria decision support system, called MOLIP, designed to model and solve linear and linear integer problems of multicriteria optimization with the help of an innovative classification-oriented algorithm. The structure, the functions and the user's interface of the system are described.**

Keywords – **multicriteria decision support systems, multicriteria optimization, interactive algorithms**

I. Introduction

The multicriteria decision support systems (MDSS) are interactive computer-based systems, designed to aid the decision maker (DM) in solving multicriteria problems for optimization and analysis [4]. Some well-known MDSS, which solve problems of multicriteria optimization, are the systems VIG [2], NIMBUS [4], DIDAS [3], LBS [1], DINAS [5], MOLP-16 [7], MONP-16 [7], MOIP [8]. In each one of these software systems a well-known interactive algorithm of multicriteria optimization is implemented. The quality of any of these algorithms defines to a great extent the quality of the system as a whole.

An experimental MDSS, called MOLIP, is described in this paper. It is designed to solve linear (continuous and integer) problems of multicriteria optimization. The system operates under MS Windows operating system. The optimization modules of the system realize two new interactive classification-oriented multicriteria algorithm [9] and two single-objective algorithms. The first single-criterion algorithm [6] is designed to solve linear problems, while the second one – linear integer problems [10]. The two single criterion algorithms are realized in LINDO Callable library [www.LINDO.com]. The interactive multicriteria algorithms allows the DM define at each iteration not only the aspiration level, as it is usual in most of the interactive algorithms known up to now, but also set the desired or acceptable intervals and directions of change in the values of the separate criteria. In this way the DM can describe his/her wishes and preferences with greater precision, flexibility and reliability.

II. Function and Structure of the System

MOLIP system is designed to solve linear and linear integer problems of multicriteria optimization (MO).

[1]The authors are with the Institute of Information Technologies, "Acad. G. Bonchev", bl. 29A, BAS, 1113, Sofia, Bulgaria; Filip Andonov, E-mail: vonodna@yahoo.com, Krassimira Genova, E-mail: krasi@iinf.bas.bg Boris Staykov, E-mail: bstaykov@iinf.bas.bg Mariyana Vassilev, E-mail: mari@iinf.bas.bg

The linear integer problems of MO and the linear problems of MO have not a mathematically well-defined optimal solution. That is why it is necessary to choose one of the (weak) non-dominated solutions, which is most appropriate with reference to DM's global preferences. This choice is subjective and it depends entirely on the DM.

The software realizations of two innovative classification-oriented interactive algorithms [9] are built in with the purpose to solve these multicriteria problems. The two interactive algorithms are oriented towards learning, which means that the existence of an implicit utility function of the DM is not presumed. These algorithms give the DM wide capacities to describe his/her local preferences with the help of desired or acceptable levels, directions and intervals of change in the values of a part or of all the criteria.

A software system for multicriteria decision making MOLIP is realized in MS Visual Basic. It consists of the following three main modules: a control program, optimization modules and interface modules.

The control program is an integrated software environment for creating, processing and saving of files associated with MOLIP system (ending by ".mlp" extension) and also for linking and executing different types of software modules. The basic functional possibilities of the control program can be divided in three groups. The first group includes possibilities to use the standard for MS Windows applications menus and system functions – "File", "Edit", "View", "Window", "Help" and others in MDSS own environment. The second group of control program facilities includes the control of the interaction between the modules realizing:

– creating, modification and saving of ".mlp" files associated with MOLIP system, which contain input data and data concerning the process and the results from solving MO linear and linear integer problems;

– interactive solution of the linear and linear integer MO problems which have been entered;

– localization and identification of errors occurring during MDSS operation.

The third group of control program functional features consists of possibilities for visualization of important information concerning the DM and the system operation as a whole.

The control program is developed on the principle of Multiple Document Interface (MDI) in MS Visual Basic software environment. In its main form it has a menu containing the standard for MS Windows applications drop-down menus for control of files, editing, windows control and Help. The main functions of the system are realized with the help of several daughter forms and context menus.

The optimization module realizes the two classification-

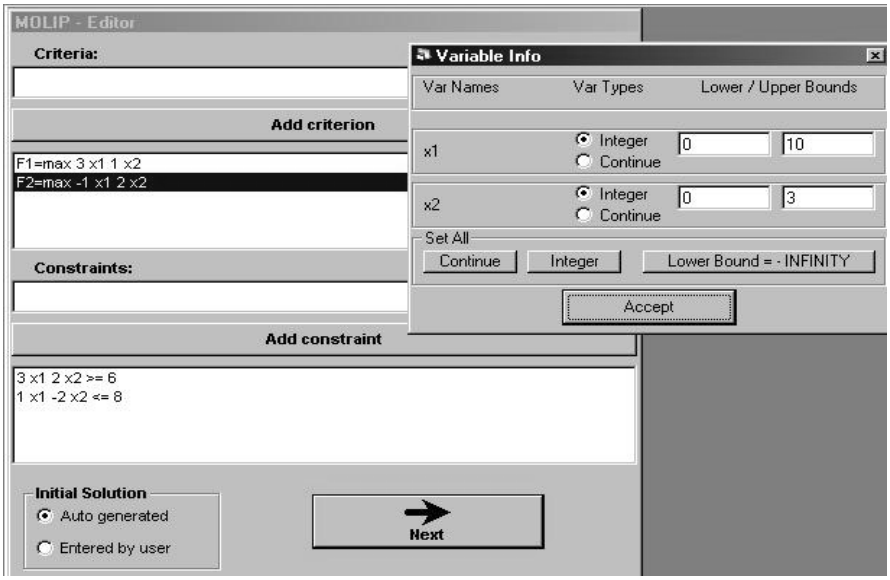


Fig. 1. The "MOLIP Editor" window

oriented interactive algorithms and two single criterion algorithms of linear [6] and mixed integer [10] programming. The single criterion algorithms are generated by LINDO Callable library.

The interface modules accomplish the dialogue between the DM and MDSS during the entry and correction of the input data of the multicriteria problems being solved, during the interactive process of these problems solution, and also in dynamic digital and graphic visualization of the process main parameters. The editing module enables the entry, alteration and storing of the descriptions of the criteria, the constraints, and also the variables type and limits of alteration. Two types of graphic representation of the information about the criteria values at different steps and some possibilities for comparison are provided by another interface module. Dynamic Help is also available, which shows information about the purpose and way of use of each one of the GUI elements.

III. Operation with MOLIP System

MOLIP system is working under MS Windows. It can be added to *Programs* group and/or with a *Desktop* icon, from where it is started. The system registers the ".mlp" extension and associates it. Thus at double clicking on a valid ".mlp" file, the system will be started and this file will be loaded. There is a menu in the main window with the standard for MS Windows drop-down menus and commands.

With their help the operation of a new file is started or an existing ".mlp" file is loaded and the operation may continue with the information stored in it.

The entry and correction of the problem criteria and constraints is realized in "MOLIP Editor" window. Every criterion and every constraint is entered separately in the respective text field for edition. Syntax check is accomplished when they are added to the data already entered. The syntax accepted is similar to the mathematic record of this class of optimization problems. The type of the optimum looked for is entered first – "min" or "max". After that the digital coefficient with its sign is entered, followed by the variable name it refers to. The variables names can be an arbitrary set of letters and numbers. Each one of these elements is separated by a space. The constraints have similar syntax – digital coefficients and variables names are successively entered. The type of the constraints is defined by some of the symbols "<=", ">=" or "=". By double clicking on the constraint or criterion already entered, they are transferred to the editing field again, if subsequent corrections are necessary.

Variable Info form can be opened in this window, where information concerning variables type and limits of alteration is entered. All the variables are by default of "Integer" type, with "Lower Bound"=0 and "Upper Bound"=1E+30, which is considered as ∞. The information about all the variables can be automatically altered with the help of two buttons – Continue and Lower Bound = INFINITY. The closing of "Variable Info" window and the corrections made are saved pressing Accept button.

Fig. 1 shows the windows of "MOLIP Editor" and "Variable Info" with the following illustrative example entered:

$$\mathrm{F}_1 = \max(3\mathrm{x}_1 + \mathrm{x}_2);\ \mathrm{F}_2 = \max(-\mathrm{x}_1 + 2\mathrm{x}_2);$$
$$\text{subject to}: 3\mathrm{x}_1 + 2\mathrm{x}2) \le 6;\ \mathrm{x}_1 - 2\mathrm{x}_2) \le 8;$$
$$\mathrm{x}_1 \le 10;\ \mathrm{x}_2) \le 3.$$

With the help of "Accept" button "MOLIP Solving" window is called, where the generated initial solution is output.

The "MOLIP Solving" window is divided into several zones. Its upper part contains a band with buttons that realize the main functions of the process for interactive solution of MO linear and linear integer problems. These are the buttons:

Solve – for starting the optimization module in order to find a new current solution of MOLIP, solving the scalarizing problem generated at this iteration;

Info – for visualization of the variables values at the current solution in a separate window;

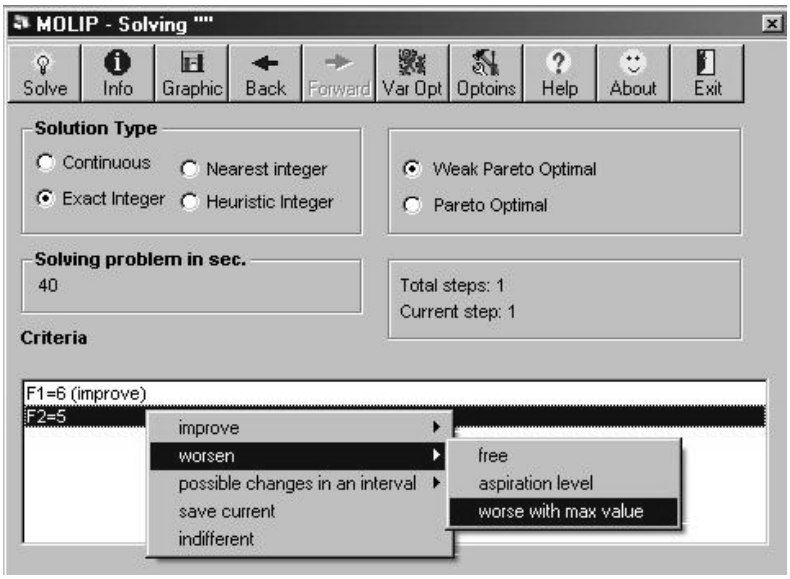


Fig. 2. The "MOLIP Solving" window

Graphic – for opening the window for graphic comparison of the results obtained at the separate steps. The upper bar-graphics gives the possibility for visual comparison of the solutions found at two iterations, selected by the rotating fields below it. The low graphics can trace the alteration of the values of the different criteria at different steps of the interactive process for better solution search. The buttons for rotation enable the selection of an initial and final step of the interval, in which the values of all the criteria are observed;

Back and Forward – buttons for navigation. They allow the DM go back to preceding steps and reconsider the solutions found. In case the DM wishes, he/she can change his own preferences concerning the criteria alteration at any of the previous steps and start the process for better solution search from there on;

Options – for opening different system setups: of the data file, which is active at the moment – it can be associated with ".mlp" extension; changing the names of the system variables if "alfa" and "beta" have another user's meaning in the problem being solved; changing the values of the default parameters of the scalarizing problems solved;

Help – for output of help information with basic directions about entry, editing and solving of MO linear and linear integer problems in MDSS environment;

About – for providing information about the team and system information about the computer system used;

Exit – for MOLIP system exiting with or without storing of the data and the results from the recent work in a file.

The next field of MOLIP Solver window contains radio buttons for setup of the MOLIP solution looked for: continuous, integer, approximate integer, the closest integer, as well as Weak Pareto optimal or Pareto optimal.

Below them information is found about the time of MDSS operation for the current problem in seconds, the number of the step being currently considered and the total number of the executed steps.

Two text fields follow. The first one outputs successively the values of the criteria obtained at the current step. It is an operating field where DM's preferences relating to the search of the next solution are set. After marking each one of the criteria, a context field is opened with the help of the mouse right button, where the DM sets the desired alteration in the value of this criterion at a following iteration. In case the selection is connected with the necessity to enter a particular value, MOLIP system opens an additional dialogue window and waits for the entry of the corresponding digital information.

The solution of the illustrative example found after the second iteration is shown in Fig. 2. For the next iteration the DM sets his/her preferences for improvement of the first criterion F_1 and worsen by a maximally feasible value 1 for criterion F_2. After that the new nondominated solution $F_1 = 9; F_2 = 4$ is output. The graphic presentation in Fig. 3 enables the DM to consider the alteration of the criteria values in the process of search for the best compromise solution.

When interactive algorithms are used for MO problems solving, it is an advantage to present information not only about the last solution found, but also about the process of search, about all the previous steps. Given that some significant solutions are made on the basis of these results. It is important for the DM to be able to "testify" how he has reached this solution. That is why the information about the interactive process of MO problem considered, which consists of the problem input data, the solutions obtained at each step, the preferences set by the DM for a new search and the constructed scalarizing problems, saved in *.mlp files associated with MOLIP system serves not only for restarting an interrupted solution process, but also for documentation. "*Print*" command from the main menu can be used for selective print of the type of information chosen by the DM.

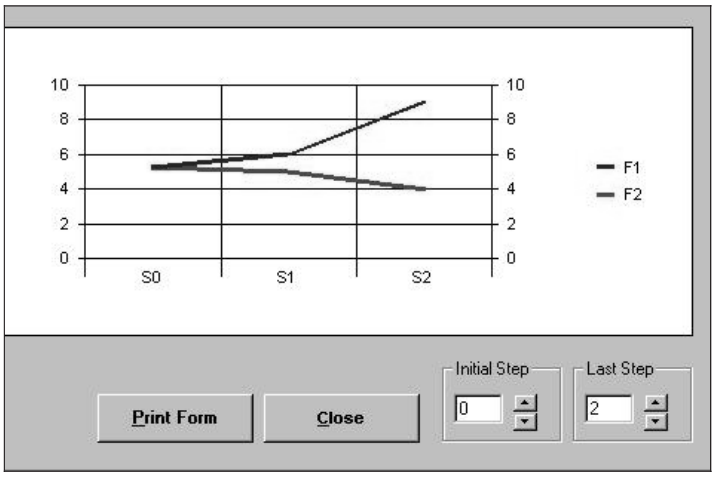


Fig. 3. The "MOLIP Graphic of results" window

IV. Conclusion

A software system for multicriteria decision making MOLIP is developed on the basis of two new classification-oriented interactive algorithms. These algorithms enabled the design of a system with a very user-friendly interface. The experiments recently accomplished prove that MOLIP is a convenient and reliable software tool supporting the solution of linear multicriteria problems. The applied software tools enable its future development for operation in a network.

References

[1] A. Jaszkiewicz and R. Slowinski, "The Light Beam Search Over a Non-Dominated Surface of a Multiple-Objective Programming Problem". In: *Mutiple Criteria Decision Making* (G.H. Tzeng, H.F. Wang, U.P. Wen and P.L. Yu, Eds.), pp.87-99, Berlin, Spinger-Verlag, 1994.

[2] P. Korhonen, "VIG - A Visual Interactive Support System for Multiple Criteria Decision Making", *Belgian Journal of Operations Research, Statistics and Computer Science* 27(1), pp. 3-15, 1987.

[3] A. Lewandowski and A. Wierzbicki, eds: "Aspiration Based Decision Support Systems", *Lecture Notes in Economics and Mathematical Systems*, 331, Berlin, Spinger Verlag, 1989.

[4] K. Miettinen, Nonlinear Multiobjective Optimization. Boston, Kluwer Academic Publishers, 1999.

[5] W. Ogryczak, K. Stuchinski, K. Zorychta, "DINAS: A Computer-Assisted Analysis System for Multiobjective Transhipment Problems with Facility Location", *Computers and Operations Research*, 19, pp. 637-648, 1992.

[6] M. Padberg, Linear Optimization and Extensions. (*Algorithms and Combinatorics*, vol. 12), Berlin, Springer Verlag, 2000.

[7] V. Vassilev, A. Atanassov, V. Sgurev, M. Kichovitch, A. Deianov, L. Kirilov, "Software Tools for Multi-Criteria Programming". In: *User-Oriented Methodology and Techniques of Decision Analysis and Support* (J. Wessels and A. Wierzbicki, Eds.), Berlin, Spinger Verlag, pp. 247-257, 1993.

[8] V. Vassilev, S. Narula, P. Vladimirov, V. Djambov, "MOIP: A DSS for Multiple Objective Integer Programming Problems", In: *Multicriteria Analysis* (J .Climaco, Ed.), Berlin, Springer Verlag, pp. 259-268, 1997.

[9] M. Vassileva, K. Genova, V. Vassilev, "A Classification based Interactive Algorithm of Multicriteria Linear Integer Programming", *Cybernetics and Information Technologies*, Vol. 1, No 1, pp. 5-20, Sofia, BAS, 2001.

[10] L. Wolsey, Integer Programming, John Wiley & Sons Interscience, 1998.

Three-Dimensional Geographical Information Systems

Dejan D. Rančić[1] and Vladan T. Mihajlović[2]

Abstract – **The paper describes three-dimensional data models for relief modeling which are used in modern three-dimensional geographical information systems. The paper, also, describes a three-dimensional information system Ginis-3D which is developed in Computer Graphics and GIS Laboratory at Faculty of Electronic Engineering in Niš. Special attention is paid to the three-dimensional relief modeling and visualization. Algorithm for this purpose is described in detail, starting from faces generation for three-dimensional surface to the texture adding and other methods needed for the realistic relief visualization.**

Keywords – **Three-dimensional GIS, Terrain visualization**

I. Introduction

Visual representation of Earth's surface has a long history. Whenever, people tried to represent on drawing the landscape, which surround them. The oldest drawings have been found before 4000 in Mesopotamia. These findings were found on pottery and the drawings depict mountains and rivers in two dimensions [1]. Further evolution of cartography and relief representation was three-dimensional (3D) representation of Earth's surface on two-dimensional medium (paper or screen). During the history, models of relief representation varied from different symbols that depict mountains throw the special technique for represent inclination of terrain using different type of lines, to the wide used abstract symbolization that use contour lines [2]. Computers, and specially Geographic Information Systems (GIS), improve the possibilities for 3D representation of Earth's surface. Early versions of GIS have enabled founding same additional information (attributes) from two-dimensional map. The one of the possibilities is retrieving the additional information about third dimension (altitude) of arbitrary point on the map. Computer's and software's progress enable new opportunities to make the representation of landscape more realistic. The new trend in realistic modeling of real world is Virtual Reality (VR) [3]. The VR is not mature enough, but from the beginning the results are impressive. We witness the amassing progress of new technologies, which provide 3D data representation (such as 3D representation of real world). Naturally, development of GIS has pursued this progress. This was result in development the new generation of 3D GIS, which provides three-dimensional interactive topological maps without spatial, temporal and thematic constraints.

[1]Dejan D. Rančić is with Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: ranca@elfak.ni.ac.yu

[2]Vladan T. Mihajlović is with Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: wlada@elfak.ni.ac.yu

II. Terrain Modeling

The basic information in three-dimensional relief representation is data of terrain altitude. This data can be acquired in different ways, from digitalization of two-dimensional maps using different software for semi or fully automated detection of contour lines [4] to the usage of Global Positioning Systems (GPS) and conventional gathering of attitude extracted from aero-photos taken from multiple different positions. No matter which method is used, the goal is creating of digital space model as bases for generating 3D representation. During the development of GIS two models for representing 3D spatial data are standardized:

1. Digital Elevation Model (DEM), and
2. Triangulated Irregular Network (TIN).

DEM [5] is regular mesh of points in space and their altitudes (see Fig 1).

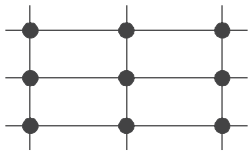


Fig. 1. DEM data model

Using the regular DEM mesh, the triangle and tetragon surfaces are constructed. 3D representation consists of these surfaces. Resolution of points in mesh depends on the generalization level. Usually the resolution of commercial DEM data is 25 meters. Of course, it is always possible to get higher resolution applying different types of interpolations.

The second type of digital representation model is TIN model [6]. TIN uses irregular point mesh with the information about altitude gathered using space triangulation process, i.e. tessellation of the convex hull of some points into triangles (see Fig 2).

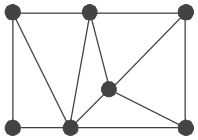


Fig. 2. TIN data model

The second model space is defined by altitude of arbitrary set of points. Usually these points represent the points where the terrain slope is changed. Thereby, the amount of data, which specify particular space, is radically reduced (if it is

compared with DEM). This set of points is the bases for creating the triangle and tetragon surfaces as a part of 3D representation.

Frequently, the hybrid model is used. The base of this hybrid model is DEM, and TIN is assigned to it as an extension for generating more accurate terrain representation.

The next step, after creating the triangular and tetragon surfaces, is 3D scene creation. This includes standard procedures from the computer graphics domain (computation of normal on the surface in defined point, light placement and defining the position of camera). In purpose to get a more realistic view, it becomes normal to place two-dimension texture on the surface. The scanned raster maps, orthophotos or satellite images are used as 2D texture for designing the 3D view. The 3D view reality is improved by using one of known shading models, one of color interpolation methods applied on triangular and tetragon surfaces during 3D scene rendering or different kind of texture filtering.

III. Ginis-3D

Ginis-3D is three-dimensional GIS developed in Computer Graphics and GIS Laboratory on Electronic Faculty in Niš. This system is hybrid GIS (georeferenced raster map + vector layers) with capability of terrain 3D view generation, finding information about altitude of arbitrary location and some three-dimensional analysis (3D profile of terrain).

DEM is used as 3D model. The resolution is 100 meter created with altitude digitalization from paper maps. The georeferenced raster map is used for texture generation [7]. Microsoft DirectX [8] is used as 3D graphics engine.

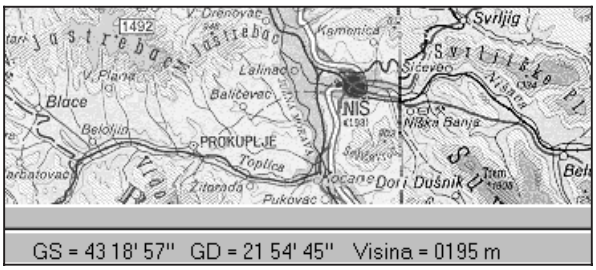


Fig. 3. Altitude information in Ginis-3D application

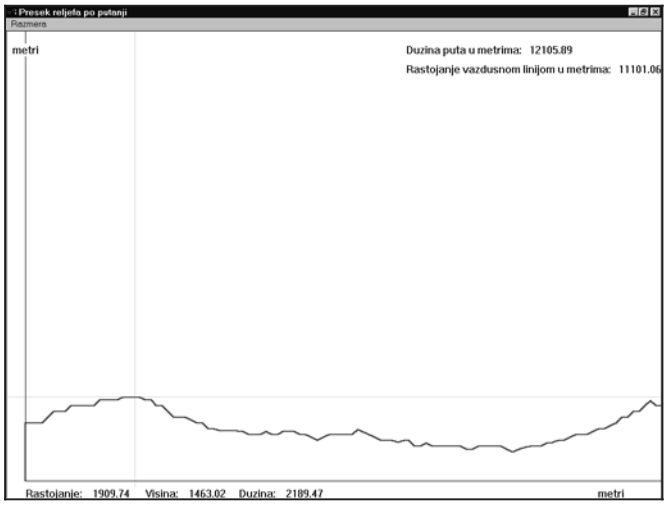


Fig. 4. 3D terrain profile in Ginis-3D application

In two-dimensional mode Ginis-3D provides information about altitude of arbitrary location on the raster map. That value computes continually and display depending on current mouse position (see Fig 3).

2D mode provides a possibility to generate 3D terrain profile. The mouse is used to select point on raster map that define the path on which user want to analyze 3D profile. When the points are selected 3D profile is displayed (see Fig 4).

3D profile view enables user to measure distance from the start point of the path to an arbitrary position on the profile. Two types of measurement are provided: measure the strait distance or the real distance (include information of terrain altitude). The information of an arbitrary position altitude is, also, displayed in the profile view.

Three-dimensional mode in Ginis-3D provides a lot of information. In this mode 3D terrain model is displayed. To create 3D view, these steps must be followed:

1. In 2D mode user defines the rectangle that contains the area for 3D model that will be formed.
2. Software retrieves the DEM data about selected area and creates regular grid.
3. The part of raster map, which will be used in 3D view, is insulated, based on the defined rectangular area.

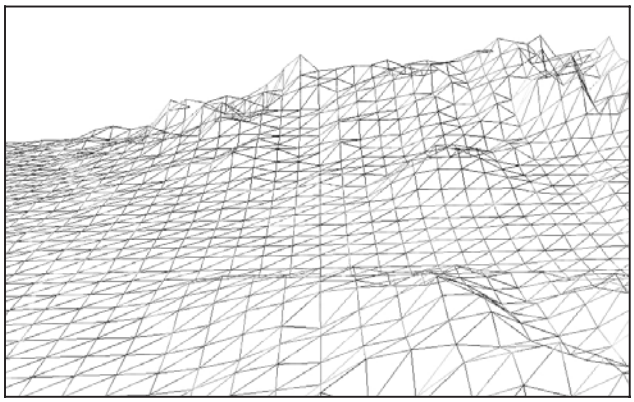


Fig. 5. Creating of 3D surface using space triangulation in Ginis-3D application

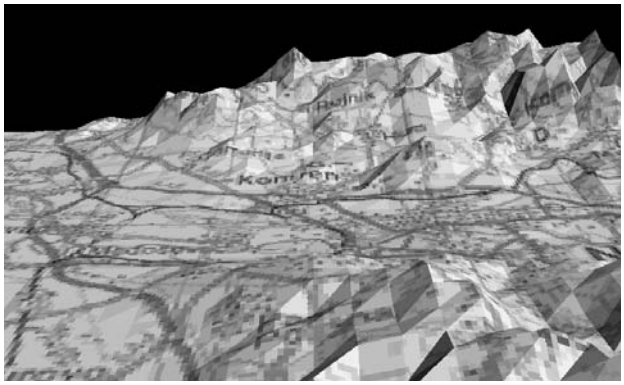


Fig. 6. Dropping the texture over 3D surface

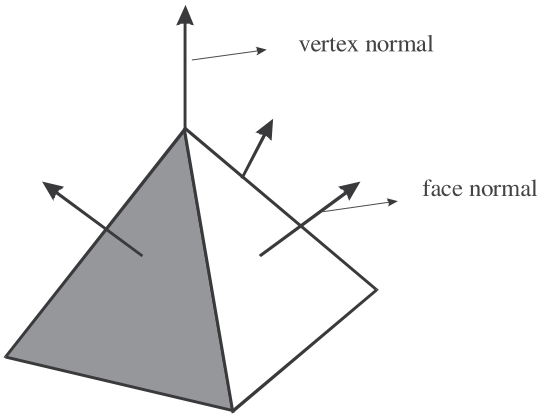


Fig. 7. Normal calculation

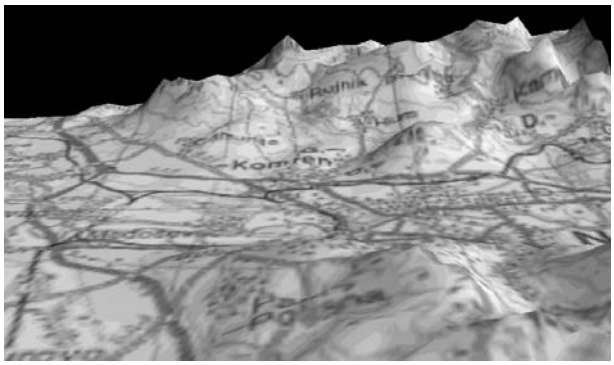


Fig. 8. 3D view made using Gouraud shading and texture filtering

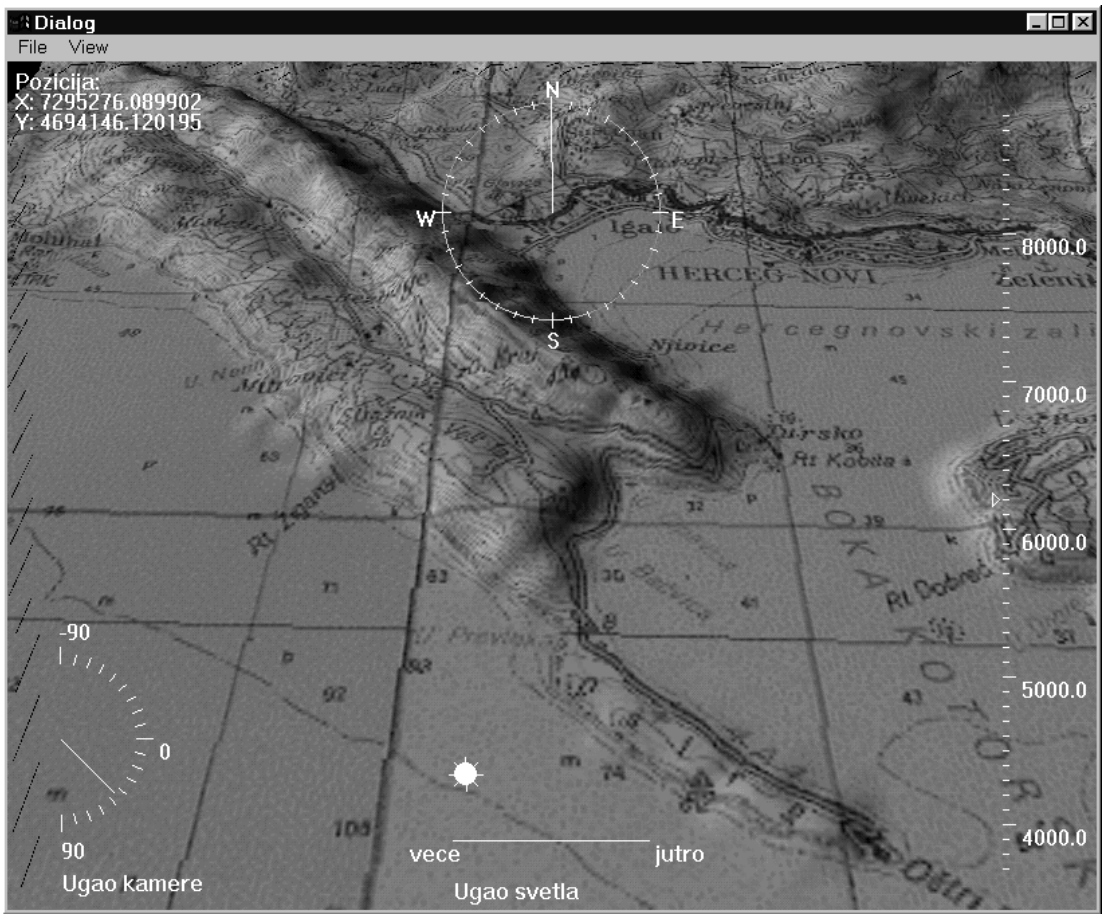


Fig. 9. 3D terrain view in Ginis-3D application

4. The space triangulation is performed and 3D surface is produced from the triangular faces (see Fig 5).
5. For each triangular face of surface the patch of the raster map is computed and dropped over it (see Fig 6).
6. The normal on each vertex on surface and each face of surface is computed (see Fig. 7). These normal are used later for shading.
7. The type of shading is defined. Ginis-3D uses Gouraud type of shading. The normal are used in purpose of generating the smooth surfaces (see Fig. 8). The color and intensity of adjacent vertices are interpolated along the space between them and generate seamless view.
8. The type of color interpolation is determined. Ginis-3D use linear interpolation of color among three vertices on each face.
9. The type of texture filtering is determined. Ginis-3D use linear interpolation among four adjacent pixels that significantly improve the quality of representation.

10. The type and location of the light is determined. Two types of light are used: direct and ambient light [8].
11. The position of the camera is determined (the perspective view of 3D terrain).
12. Using previously defined parameters 3D terrain rendering is performed.

Result of algorithm is 3D scene of terrain. The user has possibility to change the light position and camera position in real time. The effect is flying over the terrain. The system provides information about current location, azimuth and altitude of camera (see Fig 9).

IV. Conclusions

The old human desire, to represent the surrounding world in a realistic way, finally begins to realize. The progress of computer hardware and software with discovering the new powerful technologies for spatial data acquisition (scanning satellites with 1m resolution) provide appearance of three-dimensional geographic systems. These systems can generate 3D terrain model with great efficiency and ability to perform different types of 2D and 3D analysis. The new generation of GIS has improved quality and expanded the application domain (specially in military domain, domain of hydrometeorology and telecommunication domain). Ginis-3D, three-dimensional information system is developed following the actual trends in Laboratory of Computer graphics and GIS on Electronic Engineering Faculty in Nis. This system is natural extension of existing Ginis system that was standard two- dimensional hybrid GIS. Ginis-3D has all functionalities of standard 3D geographic systems (3D terrain generation and 3D analysis) and represent the base on which can be made systems applied in different domains (telecommunications, military service etc.).

References

[1] E. Imhoff, *Cartographic Relief Presentation*, De Gruyter, 1982.

[2] J. Wood, "Visualization of Scale Dependencies in Surface Model", *Proceedings of the ICA/ACI Cartographic Conference*, Ottawa, 1999.

[3] G. Buziek, J. Dolner, "Concept and Implementation of an Interactive, Cartographic Virtual Reality System", *Proceedings of the 19th ICA/ACI International Cartographic Conference*, Ottawa, 1999.

[4] M. Huber, "Contour-to-DEM, a new algorithm for contour line interpolation", *Proceedings of the JEC '95 on GIS*, http://lbdsun.epfl.ch/huber/ctdjec.html

[5] J. Wood, *The geomorphologic characterization of digital elevation model*, PhD Thesis, University of Leicester, U. K. 1996.

[6] M. Van Kreveld, "Digital Elevation Models and TIN Algorithms ", In: M. v. Kreveld, J. Nievergelt, T. Roos, and P. Widmayer (Eds.), *Algorithmic Foundations of Geographic Information Systems*, Lecture Notes Computer Science, Springer-Verlag, 1997, pp.1340:37-38.

[7] D. Rančić, S. Djordjevic-Kajan, "Mapedit: Solution to Continuous Raster Map Creation", *Computer and Geosciences*, Vol. 29, No. 2, Elsevier Science, 2003, pp.115-122.

[8] *DirectX 5.0*, Microsoft Press, 1997.

[9] J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, *Computer Graphics Principles And Practice*, Second Edition in C, Addison-Wesley, 1996.

Interactive Computer System for Solving Problems with Multiple Criteria*

Ivo Marinchev[1] and Leoneed Kirilov[2]

Abstract – **A Computer System for solving Multiple Attribute Decision Making Problems (MADMP) is presented in the paper. It is assumed that the set of alternatives is explicitly known and finite one. The attributes are assumed to be given as numerical values. The system incorporates some of the well-known and classical methods for solving MADMP. These are methods ELECTRE, PROMETHEE. It also includes the RDM (Reference Direction Method).**

Keywords – **decision making, multiple attributes, multiple criteria, JAVA**

I. Decision Making with Multiple Objectives and Decision Support

Multiple Criteria Decision Making (MCDM) is a choice among a set of decisions/variants/alternatives made by an expert on given problem according to multiple criteria/objectives/goals. The expert is called Decision Maker (DM). The set of alternatives is generated from multicriteria model according to given rule. Usually it is an optimization procedure. The multiple criteria models are natural generalization of single criteria ones. The set of objectives is optimized in total. For more details, different models, and approaches for solving them the reader can refer to [1,6,7].

Decision Support System (DSS) is every interactive computer system designed to support the process of Decision Making (DM) [8]. Its basic purpose is to support but not to substitute the DM in the process of the decision making (DM). The DSS has three basic components:

1. Model;
2. Optimization module(s) (solver);
3. Man-machine interface or shortly interface.

The model which is usually a mathematically one, is hidden for the user. The analyst makes the choice of the model. The analyst also chooses a solving method, constructs a suitable model for given problem with the help of the DM. He makes the conception for the DSS.

The optimization module implements one or more methods for solving the model. For general purposes a sufficiently general model is selected/constructed. The latter is solved by appropriate method(s). But when trying to solve real problems for real users a modification of a model is done and possibly a modification of a method to solve it. All mentioned is a responsibility of an analyst in cooperation with the DM (an expert of a given problem).

An interface is an important element of a DSS. At first, it is viewed from the DM. Therefore it has to be sufficiently attractive. Sometimes one DSS could be chosen on the base of its interface. The interface has to be full of matter and convenient (user-friendly).

Usually an input/output/editing module is also available to the DSS. The following natural requirement follows from the said above.

If one DSS is designed for solving a very specific problem, then it is not easy to use it for another problem without modification. On the other hand, if it solves a general model, then sometimes it would be necessary to customize it for solving certain real problems.

Usually a sufficiently general model is realized that can solve a number of problems, for example – linear model, or nonlinear, or integer, etc. Such DSS we name *universal systems*. The other class DSS we name *specialized systems*. The user uses such DSS to solve his/her problem. The analyst constructs a model and an appropriate method. Maintenance in the future is provided.

Multi-Criteria DSS (MCDSS) is a DSS with multi-objective model(s). Multiple objective models are generalization of a single objective ones. But are they better alternative? What are their disadvantages?

One such disadvantage is related to multi-objectivity. This leads to non-uniqueness of the produced optimal solutions in the objective space. Indecision arises about the ”best” solution. The question is, what is better to the DM? To use single objective model with one optimal solution which could be very close to real solution or to use multi-objective model with a set of solutions. The DM has to choose one of them on the base of compromises.

The other question is about convergence of multi-objective methods. This subject is not investigated completely. This is compensated by the fact that one DM could intuitively find satisfactory solution for a small number of iterations. Most multi-objective DSS provide tools for avoiding cycling. Single-objective methods are well studied for convergence as a rule. But when trying to solve real problem the question is - to use complicated model without guarantee for finding optimal solution or to use simple model with optimal but not real solution.

*The work report in this paper has been partially supported by Project IIT-010051 “Advanced methods and tools for knowledge representation and processing” and Project IIT-010049 “MCDM methods”.

[1]Ivo Marinchev is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 29A, 1113 Sofia, Bulgaria. E-mail: ivo_m@iinf.bas.bg

[2]Leoneed Kirilov is with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 29A, 1113 Sofia, Bulgaria. E-mail: lkirilov@iinf.bas.bg

Further, most real DMs are indecisive to use MCDSS in their activities. The available MCDSS are experimental. They can be used to prove the efficiency of a given method or they can be used for educational purposes. Some of them are called commercial according to their authors but not by the software companies [3].

The conclusion from the above is that MCDM approach still has not taken its place among other optimization approaches for solving real problems.

II. Multiple Attribute Decision Making

Multiple Attribute Decision Making (MADM) is one of the two major fields of Multiple Criteria Decision Making. The other one is Multiple Objective Mathematical Programming. In MADM it is assumed that the feasible set consists of a finite number explicitly known alternatives. A best one has to be chosen according to the set of k ($k \geq 2$) objectives [1].

Thus, the problem has simply the matrix formulation of dimension (n, k) – decision matrix, where the elements a_{ij}, $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$ are the values of i-th alternative according to the j-th attribute (objective). We assume that the rows of a matrix form the alternatives and the columns form the objectives.

A number of approaches are available for solving this problem. They can be classified according to the type of information required by the DM and its basic features, and according to the strategy used to find solution(s). Some basic approaches can be found in [2].

What is the best method for solving MADM is a "nonsense". The choosing of a method depends on the nature of the problem solved, on the preferences of the Decision Maker (DM), and other factors.

A number of computer systems for solving MADM problems (MADMP) are available. Note that the users better know these problems and respective computer systems than the MCDM problems. Also commercial systems for this class dominate. But the accent is mainly on the method implementation without taking into account the user-friendliness of the interface. Example of such well-known systems are ELECTRE I-IV [9,10], PROMCALC and GAIA [11, 12], TRIMAP [13]. Also EXPERT CHOICE of Saaty [14] – an implementation of Analytic Hierarchy Process, proposed by the same author.

In this paper we present a Decision Support System (DSS) for solving basic problem of MADM. It incorporates a number of well-known, classical and effective methods for their solving. These are ELECTRE and PROMETHEE methods [3,4]. It also includes Reference Direction Method (RDM) [5]. As it is known, RDM is basically designed for solving Multiple Objective Mathematical Programming Problems. Here a version for MADM problems is realized.

The reason for choosing such method is that most methods for solving MADM are non-interactive. Also the input information, requested by the DM, has not clear interpretation for him/her. The proposed result (alternative(s)) has not explicit connection with the input information. On the other hand, most Interactive Methods (IMs) have clear and simple interpretable dialog. After that a series of solutions are generated. If the compromise solution is not among them, the process repeats.

The ELECTRE method for example uses the following strategy for searching best alternative. On the base of a set of weights (input parameters) for each attribute ELECTRE constructs "outranking relationship". This means that for two nondominated alternatives A and B (that are incomparable in general) the DM could accept for example A to be more acceptable than B. The result is that:

1. The dominated alternatives are simply eliminated;
2. The nondominated alternatives are outranked, i.e. a subset of nondominated alternatives is presented to the DM. In the ideal case this is one alternative.

ELECTRE with its simple logic, full using of information in the decision matrix is one of the best methods [2].

III. Motivation for Selecting Java Technologies for Implementation of Our System

We have implemented our system using Java programming language. The reason of making this decision is that Java is not only general purpose programming language but complete development platform that has enormous diversity of APIs (Application Programming Interfaces) supporting almost all contemporary programming technologies. Most of them are embedded in the standard programming libraries that are part of any standard compliant Java 2 virtual machine.

One technology that we consider very suitable for our system is *Java Web Start*. It allows launching applications through the network using recently introduced Java network launching protocol. Java Web Start applications are extensions of the Java applets technology that have some very useful advantages:

1. Applications can be launched through a web browser but they no longer depend on the Java virtual machine built in it, hence they are not restricted to Java 1.1 which is default JVM embedded in the Internet Explorer browser (i.e. when no Java 2 plug-in is used).
2. Applications are cached locally and on any subsequent activation they are started from the local hard drive. At the same time the system launcher checks their web site whether new version is available. If so it downloads the newer version and replaces the old one with it. This solves the issue with versioning (supporting many different versions of a certain application or its components) in a very graceful manner.
3. Like Java applets, Java web start applications are executed in the secure sandbox that isolates them from the local system resources. This restriction protects the user from executing malicious code. But Java web start technology allows restricted (user confirmation is required) access to local file system to store any persistent data on it. This solves the big issue with privacy because users

often prefer keeping their data locally and not giving access to it to anyone else.

Indeed, Java Web Start technology was one of the main motives in selecting Java technologies for the implementation of our system. Although the current version of it is developed as a stand-alone application in the future it will be very easy to separate it in two components according to the client-server computational model. The thin client component implementing user interface and user interactions code and server component performing all numerical computations. The client component will be Java web start application that will be executed on the user's local machine. The server component will solve optimization tasks sent by the client components and will return the final results.

And last but not least is Java's cross platform compatibility. The language completely justifies Sun's "Write once, run anywhere" paradigm. We develop our system mainly on the Windows workstations, but the final system can be deployed and used on any Java 2 enabled platform – Windows, Linux, FreeBSD, Solaris, Mac OS, etc. At the same time client part and server part can be executed on different operating system allowing any available system to be used as a client or server.

IV. System Architecture

Current version of the system is implemented as a stand-alone Java application. It uses Swing library for graphic user interface (GUI). We have selected this library because it supports (and enforces) the use of the model-view programming model in user interface programming. The later allows clean separation between the internal data structures (data model) and processing and their visual representation and user interactions with them.

Internally our implementation is organized according MV (model-view) architecture, which is supported by the GUI library. The basic idea behind this architecture is that every entity that the program processes is represented with some sort of internal model – a collection of interrelated data structures. This internal model has one or more views associated with it. Every view is a user interface component that renders the model on some sort of external device (usually monitor screen). The separation between data processing and data visualization has many advantages. The most important of them are:

1. One data model can have many views that represent its different aspects. For example in our program every table (from the Tabbed Pane control) is a different view to the same data model. Every view renders only the data needed for its corresponding optimization method and ignores all irrelevant information (parameters).
2. If some data in the data model is changed by the user (through any of the views associated with it) or by some processing algorithm, it becomes immediately visible in all associated views. In practice it is implemented with the subscription-notification mechanism (listeners in Java terms) built in the Swing library. By means of it every view that is registered to receive certain events (changes in the data model data in our case) is notified for the changes. Receiving this event the corresponding view repaints its canvas to reflect the new data model state.
3. Data model and view can be changed independently. The data exchange between the model and the view is channeled through a well-defined interface (TableModel interface) that is part of the Swing library. This architecture allows the model and the view internals (data structures and algorithms used) to be changed independently as far as the interface is properly implemented. So the system is easily extensible and different peoples can work on the different part of it simultaneously without the complications of implementation synchronization.

ELECTRE | PROMETEE I | PROMETHEE II

	F1	F2
Minimize	max	max
Weights	0.5	0.5
A1	0.0	20.0
A2	1.0	18.0
A3	3.0	15.0
A4	5.0	12.0
A5	7.0	10.0
A6	9.0	5.0
A7	10.0	0.0

ELECTRE | PROMETEE I | PROMETHEE II

	F1	F2
Minimize	max	max
Weights	0.5	0.5
q	0.5	0.5
p	1.0	1.0
s	0.5	0.5
TYPE (I,II,III,IV,V,VI)	TYPE I	TYPE I
A1	0.0	20.0
A2	1.0	18.0
A3	3.0	15.0
A4	5.0	12.0
A5	7.0	10.0
A6	9.0	5.0
A7	10.0	0.0

Fig. 1. Different views of a single data model

V. User Interface

We have created the user interface of our system that conforms to the following preliminary defined criteria:

1. The interface must organize a lot of information (decision matrixes) in the least possible space.

2. The interface must be easily extensible in order to be possible to add new optimization algorithms with little efforts.
3. All optimization algorithms (methods) must have uniform look and feel.
4. The interface must be easy to understand and use.

In order to comply with these criteria we have selected to organize the user interface with the use of the Tabbed Pane (Tab Strip) control. This control is broadly used when a lot of categorized information has to be confined to single screen (usually in options and/or preferences dialogs) and it is in perfect accordance with all specified criteria. Every tab pane is associated with a given optimization method and displays its related data in the form of generalized matrix (containing not only the alternatives and criteria but also all constraints or parameters used in the corresponding optimization algorithm.

VI. An Illustrative Example

We shall demonstrate the work of the system on the following example, described in [2]. A country has to buy a fleet of jet fighters from the U.S. The Pentagon officials offered the characteristic information about four models of fighters. The Air Force analyst team of the country agreed that six characteristics should be considered: F1 – maximum speed; F2 – ferry range, F3 – maximum payload, F4 – purchasing cost, F5 – reliability, F6 – maneuverability. The values of each attribute for each alternative are given in table 1.

Table 1. A problem for selection fighter aircraft

	F1 Maximum speed (Mach)	F2 Ferry range (NM)	F3 Maximum payload (pounds)	F4 Acquisition cost ($*10^6)	F5 Reliability High - low	F6 Maneuverability High – low
A1	2.0	1500	20000	5.5	Average	Very high
A2	2.5	2700	18000	6.5	Low	Average
A3	1.8	2000	21000	4.5	High	High
A4	2.2	1800	20000	5.0	Average	Average

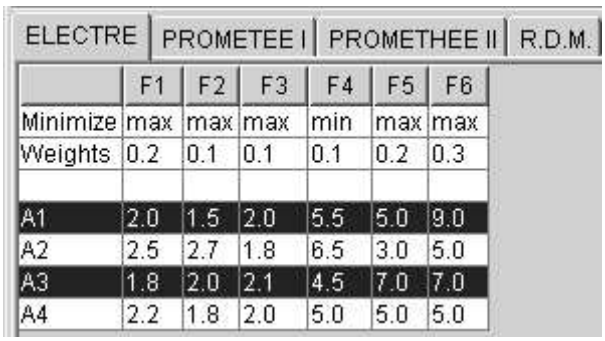


Fig. 2. The ELECTRE solution of the problem

As it is seen the 5th and 6th attributes are qualitative. They are converted to quantitative ones by using bipolar scale (interval scale). Using 10-point scale and setting 0 points to the minimum attribute value and 10 points to the maximum attribute value we receive the following relations (see for more details [2]) – Very low (1), Low (2), Average (5), High (7), Very high (9). Fig. 2 shows the ELECTRE solution of the above problem.

VII. Conclusion

In the present paper we present decision support system for solving multiple attribute analysis problems. The main features of the system are:

1. It implements several multiple criteria methods (ELECTRE, PROMETHEE, and RDM), allowing the decision-maker to compare the solutions obtained with the different of them.
2. It is implemented in JAVA hence it is completely portable and can be used on any Java enabled platform (including web browsers).
3. The unified interface allows the user to work with the different methods with ease.

References

[1] R. Steuer, *Multiple Criteria Optimization: Theory, Computation and Application*, John Wiley & Sons, New York, 1986.

[2] Ch. Hwang, K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, Berlin, 1981.

[3] Ph. Vincke, *Multicriteria Decision Aid*, John Wiley & Sons, New York, 1992.

[4] C. Bana e Costa (Ed.), *Readings in Multiple Criteria Decision Aid*, Springer-Verlag, Berlin, 1990.

[5] S.C. Narula, L. Kirilov, V. Vassilev, "Reference Direction Approach for Solving Multiple Objective Nonlinear Programming Problems", *IEEE Transactions on Systems, Man, and Cybernetics*, vol.24, No5, pp.804-806, 1994.

[6] Sawaragi Y., Nakayama H., Tanino T., *Theory of multiobjective optimization*, Acad. Press, Inc., Orlando, Florida, 1985.

[7] Miettinen K., *Nonlinear multiobjective optimization*, Kluwer, Norwell. USA, 1999.

[8] Eom H., "The Current State of Multiple Criteria Decision Support Systems", *Human Systems Management* 8, 113-119, 1989.

[9] Roy B., Skalka J., "Electre IS – Aspects methodologiques et guide d'utilization". Document du Lamsade 30, Univ. Paris Dauphin, 1984.

[10] Skalka J., Bouyssou D., Bernabeu Y., "ELECTRE III et IV: aspects methodologiques et guide d'utilization". Document du Lamsade 25, Univ. Paris Dauphin, 1984.

[11] Mareschal B., "Weight Stability Intervals in Multicriteria Decision Aid", *European J. of Operational Research* 33, 54-64, 1988.

[12] Mareschal B., Brans J., "Geometrical Represenation for MCDA", *European J. of Operational Research* 34, 69-77, 1988.

[13] Climaco J., Henggeler Antunes C., "TRIMAP: an Interactive Tricriteria Linear Programming Package", *Foundations of Control Engineering* 12(3), 101-119, 1987.

[14] Saaty T., *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.

Sliding Mode Control of Third-Order Objects with Stable Finite Zero[1]

Čedomir Milosavljević[2], Milić M. Pejović[3], Goran Milosavljević[4]

Abstract – **In this paper we consider a new approach for sliding mode control of third-order objects with stable finite zero. The topology of proposed approach contains variable structure controller, object and model-observer intended to obtaining differentials of controlled variable. This traditional control scheme for objects without finite zero is completed by conventional PI part in observer's control channel, and with integral part in the controlled object control channel. Parameters of PI part are chosen in such way that the two control channels become identical for object's nominal parameters. For stability and disturbance rejection improvement a new control action is introduced to the controlled object input from detector of observation error. The described topology is very robust to the internal and external disturbances.**

Keywords – **Variable structure control systems, sliding modes, minimum phase systems**

I. Introduction

As it is well known, in the variable structure control systems (VSS) the control signal is formed as a discontinuous function of system state coordinates. The sign of control signal is determined by the system state with respect to some hyperplane, which intersects the phase space origin. The control is determined in such way to provide, with no respect to the limited parameter or external disturbances, driving of the system state from any initial state toward the above mentioned hyper-plane, and that the further motion takes place on the hyper-plane in the sliding mode. After the sliding mode has been reached, the system motion is not function of control signal, object parameters and disturbance. It is only determined by parameters of sliding hyper-plane. If those parameters are chosen in such way to provide desired, stable system motion dynamics, then the process of system design is successfully achieved.

With respect to the fact that the control signal is discontinuous high frequency signal, the control of the object with the zeros in transfer function in this class of VSS is facing certain problems, caused by differential action upon the control signal. This problem was noticed, and has been studied at the beginning of VSS theory development [1,2].

Four methods are known. The first one is limited to second order systems [1]. It is based on appropriate state coordination transformation. The second method [2] is based on introduction of cascade of low pass commutation filters in the object control loop. The third method [4] is based on sliding mode organization in the subsystem of lower order. Above mentioned methods are not immune to unwonted motion phenomenon (chattering), caused by presence of nonmodelled object's dynamics and high frequency switching control signal. Beside that, it was shown [4] that the conventional asymptotically stable sliding modes couldn't be organized in the systems with the right zeros, so the sliding modes based on generalized minimal variance [5] must be implemented in this case. That problem won't be explained in this paper, i.e. it will be considered only the problem of control of minimal phase objects.

In this paper it's made an attempt on eliminating the shortages of conventional methods as well as providing the quality control over one class of third order objects with the stable finite zero. The proposed way of control algorithm synthesis is based on combination of well-known VSS control laws with the observer and conventional linear control laws of PI type. The proposed way in fact represents combination of three above-mentioned methods [1,2,4], because it includes state coordinate transformation [1], filter implementation [2] and organization of sliding mode in the subsystem of lower order. Only single input single output systems have been considered.

In the second section it will be more preciously defined the structure of proposed solution for VSS, it will be stated some preliminary considerations with respect to synthesis of sliding control for systems without finite zero, with application of linear PI element in the control loop. The third section contains quality verification of proposed solution by simulation of concrete example.

II. Preliminary Considerations

Let us consider fully controllable and observable single input single output object, with bounded parameter perturbations, described by following transfer function

$$W_{ob}(s) = \frac{C(s)}{Us} = \frac{k_o(Ps+1)}{A(s)}\,, \qquad (1)$$

where $z = 1/P$ is finite zero ($z_{\min} \leq z \leq z_{\max}$), k_o is object's gain $k_{o\min} \leq k_o \leq k_{o\max}$, and $A(s)$ is n order polynomial de-

[1]This research is supported by Ministry of Scientific, Development and Technologies of Serbian Government under Contract No 0125.

[2]Čedomir Milosavljević is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia, E-mail: milosavljevic@elfak.ni.ac.yu

[3]Milić M. Pejović is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia.

[4]Goran Milosavljević is with The Advanced Technical School, Beogradska 20, 18000 Niš, Serbia.

scribed as

$$s^n + \sum_{i=1}^{n} a_i s^{i-1};\ a_{i\min} \le a_i \le a_{i\max}\ . \tag{2}$$

It is assumed that the object parameters are non-stationary, but the rate of their change is much slower then dynamics of control system process.

In the input of control object we will introduce integral part, what leads to transfer function of extended control object in the form of

$$W_{ab}^{p}(s) = \frac{C(s)}{U_{kl}(s)} = \frac{(Ps+1)}{s}\frac{k_o}{A(s)} = \\ = \left(P + \frac{1}{S}\right)\frac{k_o}{A(s)}\ . \tag{3}$$

Extended control object can be considered as object without finite zero with additional PI part (PI regulator) with determined parameters. Beside full object model let us introduce reduced object model with nominal parameters without finite zero. Relation describing such model is

$$W_{ob}^{s} = \frac{k_o}{A(s)}\ . \tag{4}$$

The reduced object model can be implemented by computer. From that model, all canonical state coordinates (x_i) are available, and its mathematical model is

$$\dot{x}_i = x_{i+1}\ , \\ \dot{x}_n = -\sum_{i=1}^{n} a_i x_i + k_o u\ . \tag{5}$$

For such object-model it can be synthesized variable structure control law, for example in the following form,

$$u = \sum_{i=1}^{n-1} \Omega_i x_i\ , \tag{6}$$

$$\Omega_i = \begin{cases} \omega_{i1} & for \quad gx_i > 0\ , \\ \omega_{i2} & for \quad gx_i < 0\ , \end{cases} \tag{7}$$

$$g = \sum_{i=1}^{n} c_i x_i;\ c_i = \text{const} > 0,\ c_n = 1\ , \tag{8}$$

which will provide realization of continuous sliding mode on the hipper-plane $g = 0$ for any variation of parameters a_i and k_o in predefined boundary.

In the monograph [1] it was shown that choice of commutation function parameters (7) in the form of

$$\omega_{i1} > \sup_{a_i, k_o} \frac{-a_i + c_{i-1} + c_i a_n - c_i c_{n-1}}{k_o}\ , \\ \omega_{i2} < \min_{a_i, k_o} \frac{-a_i + c_{i-1} + c_i a_n - c_i c_{n-1}}{k_o}\ , \\ i = 1, 2, \ldots, n-1\ , \tag{9}$$

provides stable sliding mode in the system (5) on $g = 0$, after short reaching period.

In such systems, if object is of zero type, sliding mode could be lost in the small vicinity of equilibrium point, when static error occurred, and some unwonted self-oscillations are possible. Introducing proportional-integral (PI) action in control law can successfully solve that problem. In that case instead of (6), the following control law should be introduced

$$u = \sum_{i=1}^{n-1} \Omega_i x_i + \int_0^t \left(\sum_{i=1}^{n-1} \Omega_i x_i\right) \mathrm{d}t\ . \tag{10}$$

It was shown [6] that by such control sliding conditions (9) are more easier to fulfill. System has zero error and control becomes smooth in the steady state. Chattering exists only in time interval from the reaching moment until the moment when system enters steady state. The system (5) with control (10) can be threaten as a system with finite zero, for which sliding mode is organized in subsystem of lower order [4], because system output signal and its differentials are available for measurement, what is the basic assumption for third above mentioned method.

In the theory and practice of VSS beside control algorithm (6), often is used algorithm in the form

$$u = U_o \text{sgn}(g)\ , \tag{11}$$

which is characterized with powerful switching signal in the reaching mode of sliding hipper-plane as well as in sliding mode itself. From one point of view, it makes system more robust, especially to external disturbances. In the meantime, such signal leads to strong excitation of non-modeled system dynamics. Besides that, for systems with discrete-time data processing, static error might occur. That error could be eliminated by implementation of control algorithm in the form

$$u = U_o(\text{sgn}(g) + \int_0^t \text{sgn}(g)\mathrm{d}t)\ , \tag{12}$$

i.e. by introducing PI variable structure control.

So, for the object-model, with added PI action (10) or (12), sliding mode on hipper-plane (8) can be organized in model space (5) of n-order. Under conditions that object parameters are known and unchangeable, and that the model parameters are the same and that there are no external disturbances affecting the object, obtained control will be at the same time also valid for the extended object. Because on the input of extended object is simple integrator, switching control signal in the input of real object will be continuous and chattering will be eliminate.

As the object parameters, according to assumption, are changing within the known boundary and the object is exposed to external disturbance, in order to provide stability and desired accuracy it is necessary to introduce closed loop from error signal between outputs of object and model.

Above stated considerations lead to the structure block diagram of control system for regulation of considered class of objects, Fig. 1. SMC is sliding mode controller, designed in above described, or some other known methods for reduced object. I-term stands for integral action, which is added to the object input. PI stands for proportional-integral action on the input of objects model (observer).

At this point a remark should be made. We are considering object-model. However, it is *de facto* observer in its basic

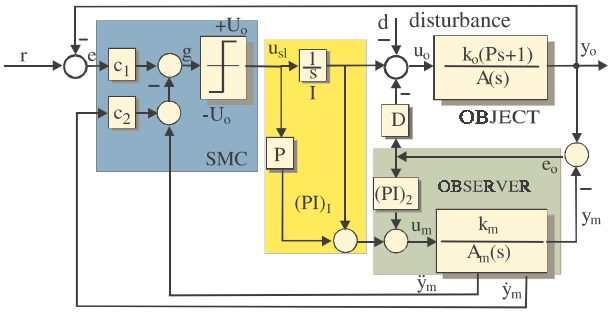


Fig. 1. Proposed control scheme.

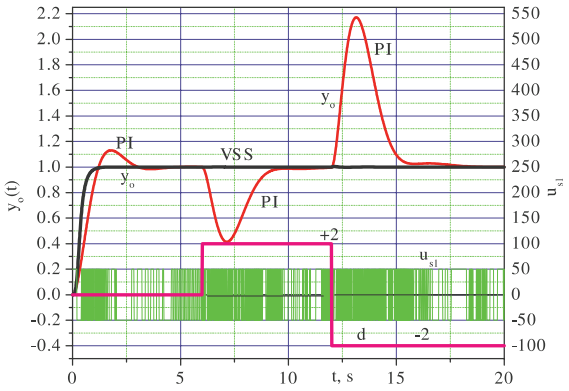


Fig. 2. Step response of the controlled variable y_o for the object with nominal parameters, sliding control signal u_{sl} and disturbance d.

sense, which is formed by feed backing the object model by $(PI)_2$ feedback. Integral action in the observation error loop is introduced in order to increase the observer ability to detect slowly varying disturbances $d(t)$. At the object input it could be introduced additional action from the observation error signal over the element D. That action could be P or PI type for second order objects, while for third order objects it must be of the D type. Without such action, system very slowly rejects external disturbances or oscillations might occur.

III. Verification of Proposed Approach by Digital Simulation

In order to verify efficiency of proposed approach computer simulation was performed for randomly chosen example of third order object, which nominal parameters are

$$a_1 = 10,\ a_2 = 17,\ a_3 = 8,\ k_o = 10,\ P = 0.25;$$
$$d(t) = 2h(t-6) - 4h(t-12)\ .$$

The sliding control was chosen in the following form

$$u_{sl} = 50\text{sgn}(g);\ g = 50x_1 + 15x_2 + 2\ . \qquad (13)$$

$(\text{PI})_2$ is $200(1+1/s)$ while the differential part is $D = 50$ s. A real differentiator with time constant of 1 ms was used in the simulation.

Fig. 2 shows: step response of controlled variable y_o for nominal parameters when conventional PI and proposed VSS controller are used, sliding mode control signal and load disturbance. The PI controller is approximately optimally tuned

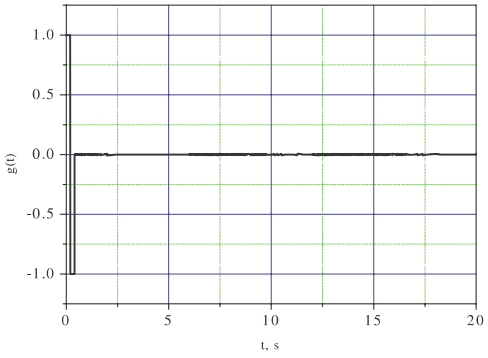


Fig. 3. Switching function $g(t)$.

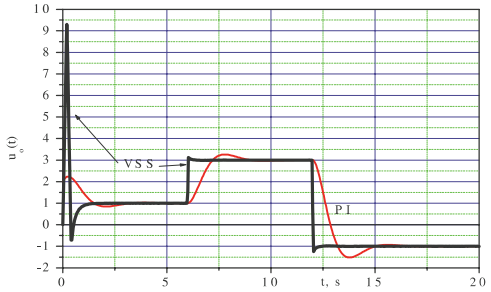


Fig. 4. Control signals u_o for nominal parameters.

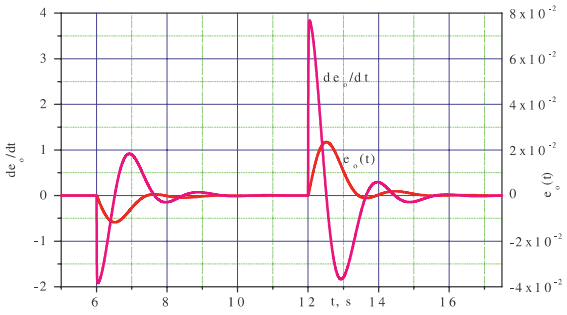


Fig. 5. The observation error e_o and its differential.

by the technical (magnitude) optimum method ($W_{\text{PI}}(s) = 2(1+1/s)$). As it can be seen, the system with the proposed VSS controller is invariant to the given disturbance, while the system with the PI controller doesn't have that ability.

Fig. 3 represents switching function $g(t)$ which (as well as the control signals u_{sl} in Fig. 2) indicates existence of sliding mode in the proposed system ($g(t)$=0) after short reaching time.

Fig. 5 shows observer error signal $e_o(t)$ and its derivative ($50de_o/\text{d}t$), obtained by real differentiator (D) with time constant of 1 ms .

Fig. 6 represents simultaneously process of load rejection for the proposed VSS and the conventional PI controller, while the process for VSS is amplified 100 times.

Fig. 7 shows control process for the real zero changes of $\pm 50\%$ with respect to the nominal value of 0.25.

Fig. 8 gives regulation processes when object gain is changed $\pm 50\%$ with respect to the nominal value (k_o=10).

Fig. 9 presents control signals for the case when object have non modeled inertial mode with time constant of 0.02 s,

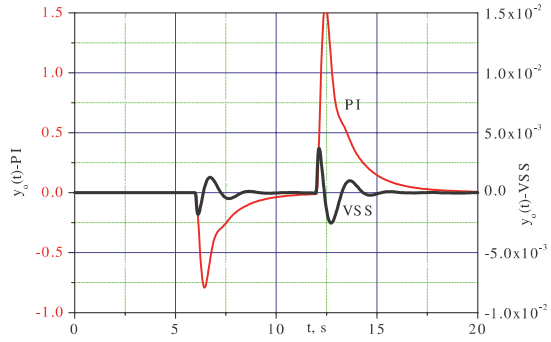


Fig. 6. Load rejection.

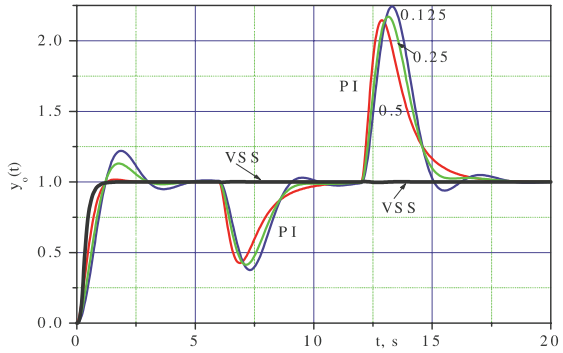


Fig. 7. Step response of the controlled variable y_o for different values of finite zero (D).

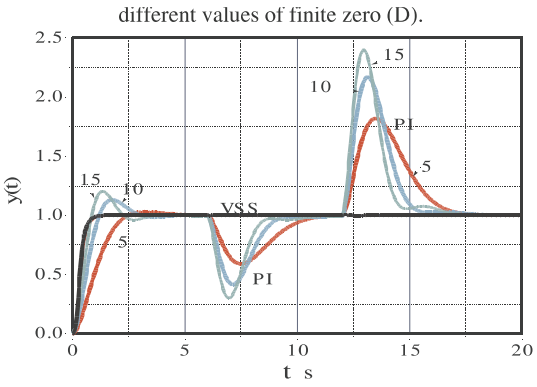


Fig. 8. Step response of the controlled variable y_o for different gain k_o.

(ten times lower then the smallest time constant of object model). In this case, temporary control oscillations occurs for VSS controller. Inertial mod is placed on the object input. In the steady state VSS control signal is smooth and therefore chattering is eliminated.

All computer simulations were performed with integration time of $5{\cdot}10^{-4}$ s with Euler integration method. The proposed system preserves his characteristics with sampling data processing by sampling time lower or equal 1 ms.

IV. Conclusions

The basic properties of proposed solution is considerable robustness to the variations of internal and external disturbances, i.e. practical invariance, chattering elimination, because the control signal is free from high frequency discontinuities. System is also robust to the non-modelled inertial dynamics if the non-modelled time constant is for the order of magnitude smaller then the smallest modelled time constant.

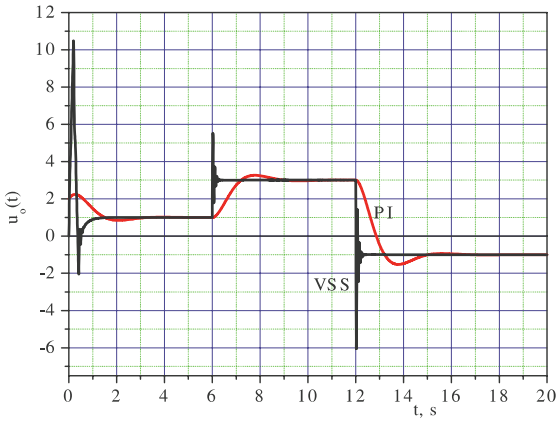


Fig. 9. Control signals with unmodelled inertial dynamics T_i=0.02 s.

On the basic of this research it can be concluded that the proposed solution for the mentioned class of objects is worth to investigate in detail in order to establish the sliding mode existence conditions, stability and robustness for systems of general case, for single input single output, multiple input multiple output systems, as well as possibility of using this approach to the systems without finite zeros.

For the final verification of proposed control algorithm, it should be implemented on the adequate real object.

References

[1] Емельянов С. В., *Теория систем с переменной структурой*, "Наука", Москва, 1970.

[2] Костылева Н. Е., "Об управлении возмущенным движением объектов описываемых дифференциальными уравнениями с оператором дифференцирования в правой части", *Системы с пере-менной структурой и их применение в задачах автоматизации полета*, под ред. Петрова Б. Н. и Емельянова С. В., "Наука", Москва, 1968.

[3] Уткин В. И., *Скользящие режимы и их применения в системах с переменной структурой*, "Наука", Москва, 1974.

[4] Уткин В. И., *Скользящие режимы в задачах оптимизации и управления*, "Наука", Москва, 1980.

[5] Furuta, K., (1993) "VSS type self-tuning control", *IEEE Trans. IE-40*, No 1, Feb. 1993, pp. 37-44.

[6] Milosavljevic, Č., "Variable structure systems of quasi relay type with proportional - integral action", Facta Universitatis (1997) **2**, *Mechanics, Automatic Control and Robotics*, No 7, pp. 301-314 (University of Niš).

A Brief Review of Model-Based Fault Diagnosis

Gordana Janevska[1]

Abstract – **This paper outlines the basic concept of fault detection and isolation (FDI), i.e. fault diagnosis in dynamic systems based on analytical process models. It gives a brief review of the most important approaches in literature.**

Keywords – **fault diagnosis, analytical redundancy**

I. Introduction

Within the last two decades there has been increasing interest in the field of fault diagnosis both within the academic community and in industry. The increasing complexity of automatic control systems and their use in safety critical areas such as flight control and chemical and power plants has helped to fuel this interest. An undetected fault in such a system could often result in dire consequences and the vast array of data which complex system generate can make in difficult, if not impossible, for operators to assess the location and nature of a fault. Automated fault diagnosis has an important role both in systems such as flight control, where automatic reconfiguration of sensors and control systems can be carried out, and with process management, where the role of a fault diagnosis system is more one of information processing and filtering with the final decision and choice of action being performed by a human operator.

Away from the safety critical areas, fault diagnosis is also attracting interest from those wishing to improve productivity and reduce plant downtime. The knowledge of the state of a plant can be used to schedule maintenance and allow reconfiguration or operating point changes to be carried out more effectively to increase the efficiency of a plant's operation. Such a scheme is often termed condition monitoring and can result in significant decreases in the running costs of plants.

The increase in the need for fault diagnosis systems has been matched with advanced in computer technology which facilitates the analysis of large amounts of data. Methods that would have been computationally infeasible only a decade age can now be applied in real time.

II. The Fault Diagnosis Problem

Diagnosis is a procedure to detect and locate faulty components in a dynamic process. Faults and failures in complex automated control systems are, in general, unavoidable facts and they require quick detection, location and identification. A diagnosis scheme is of importance in, for example, nuclear plants, aeroplanes, automotive engines. This is due to increasing demand for higher performance, higher safety and reliability. Different fault detection and isolation techniques have been developed over the recent years.

A general diagnosis procedure for a dynamic system consists of several tasks. In literature the following steps are suggested.

- *Fault detection*: Detect when a fault has occurred. That is often done with a suitable comparison, for example in parameter estimation, the estimated physical parameters are compared to their nominal values;
- *Fault isolation*: Isolate the fault. Primarily to determine the faults origin but also the fault's type, size and time.

These two tasks are commonly referred to as FDI (*fault detection and isolation*), which sometimes is referred to as *diagnosis* and the other way around.

The system to be diagnosed often includes a control loop, which further complicates the problem. A control loop tends to hide or mask a faulty component or sensor making it even more important, in a controlled system, to detect faults. The control loop can also damp the system's signals making it necessary to excite the signals from the system.

We speak of *faults* and *failures* in diagnosis. In diagnosis literature there is a distinction between the two and the definition can be written as:

Definition 1. *A failure suggests a complete breakdown of a process component while a fault is thought of as an unexpected component change that might be serious or tolerable.*

Fault diagnosis and fault detection is not a new problem and before model based fault diagnosis, they were accomplished e.g. by introducing hardware redundancy in the process. A critical component was then duplicated, triplicate (TMR) or even quadrupled and a majority decision rule was then used. Hardware redundancy methods are fast and easy to implement but they have several drawbacks

- Extra hardware can be very expensive
- It introduces more complexity in the system
- The extra hardware is space consuming which can be of great importance, e.g. in a space shuttle. Also the components weight sometimes has to be considered.

Instead of using hardware redundancy, analytical redundancy can be utilized to reduce, or even avoid, the need for hardware redundancy. Analytical redundancy is in principle the relationships that exist between process variables and measured output signals. If an output signal is measured, there is information about all variables that influences the output signal in the measurement. If the relationships are known, by quantitative or qualitative knowledge, this

[1]Gordana Janevska is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, R. Macedonia E-mail: gordana.janevska@uklo.edu.mk

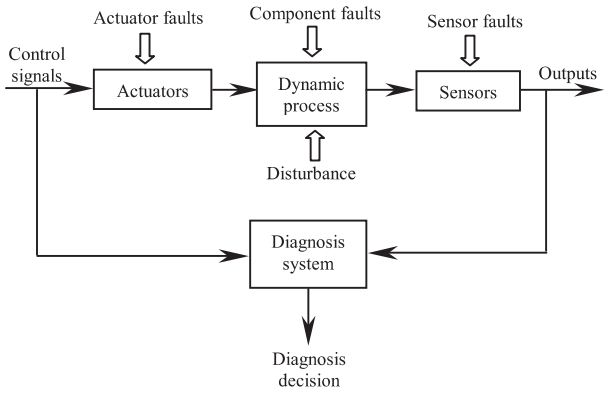


Fig. 1. Structure of a diagnosis system

information can be extracted and the extracted information from different measurements can be checked for consistency against each other.

There are different types of analytical redundancy. Instead of measuring several outputs, the different output measurements at different times can be compared. If the relationship between time series of outputs and inputs are known, from this relationship fault information can be extracted. This kind of analytical redundancy is called temporal redundancy.

The faults acting upon a system can be divided into three types of faults.

- *Sensor (Instrument) faults*: Faults acting on the sensors
- *Actuator faults*: Faults acting on the actuators
- *Component (System) faults*: fault acting on the system or the process we wish to diagnose.

A general FDI scheme based on analytical redundancy can be illustrated as in Fig. 1, an algorithm with measurements and control signals as inputs and fault detection as output.

It is unrealistic to assume that all signals acting on the process can be measured; therefore an important property of an algorithm is how it reacts on these unknown inputs. It is also unrealistic to assume a perfect model; the modelling errors can be seen as unknown inputs. An algorithm that continues to work satisfactory even when unknown inputs vary is called *robust*. Some approaches give the possibility to achieve disturbance decoupling, i.e. make the isolation decision independent of unmeasured disturbances.

III. Advantages of Model Based Diagnosis

This paper outlines the model based diagnosis, i.e. the procedure of diagnosis based on the mathematical model of the system. Why is there need for a mathematical model to achieve diagnosis? It is easy to imagine a scheme where important entities of the dynamic process is measured and tested against predefined limits. The model based approach instead performs consistency checks of the process against a model of the process. There are several important advantages with the model based approach.

- Outputs are compared to their expected value on the basis of process state, therefore the thresholds can be set much tighter and the probability to identify faults in an early stage is increased dramatically.
- A single fault in the process often propagates to several outputs and therefore causes more than one limit check to fire. This makes it hard to isolate faults without a mathematical model.
- With a mathematical model of the process the FDI scheme can be made insensitive to unmeasured disturbances, and also feasible in a much wider operating range.
- It might be possible to perform the diagnostic task without installing extra sensors, i.e. the sensors available for e.g. control might suffice.

There is of course a price to pay for these advantages in increased complexity in the diagnosis scheme and a need for a mathematical model.

IV. Quantitative Approaches to Diagnosis

In quantitative approaches the diagnosis procedure is explicitly parted into two stages, the residual *generation* stage and the residual *evaluation* stage, as illustrated in Fig. 2.

The residual evaluation can in its simplest form be a threshold test on the residual, i.e. a test if $|r(t)| > Threshold$. More generally the residual evaluation stage consists of a change detection test and a logic inference system to decide what caused change. A change here represents a change in normal behavior of the residual.

The residual generation approaches can be divided into three subgroups, *limit & trend checking*, *signal analysis* and *process model based*.

- **Limit & trend checking** – This approach is the simplest imaginable, testing sensor outputs against predefined limits and/or trends. This approach needs no mathematical model and therefore it is simple to use, but it is hard to achieve high performance diagnosis.
- **Signal analysis** – These approaches analyze signals, i.e. sensor outputs, to achieve diagnosis. The analysis can be made in the frequency domain, or by using a signal

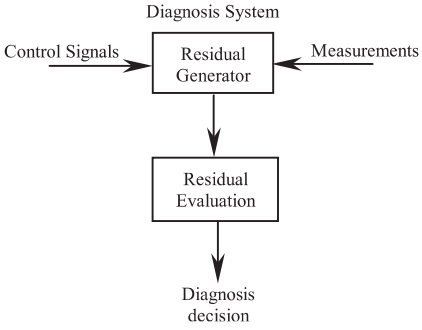


Fig. 2. Two stage diagnosis system

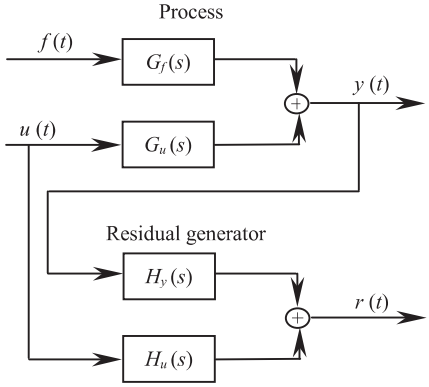


Fig. 3. General structure of a linear residual generator

model in the time domain. If fault influence is known to be greater than the input influence in well known frequency bands, a time-frequency distribution method can be used.

- **Process model based residual generation** – These methods are based on a process model. The process model based approaches are further parted into two groups, *parameter estimation* and *geometric approaches*.

The approaches mention here generate residuals which can be defined as:

Definition 2. *A residual (or parity vector) $r(t)$ is a scalar or vector that is 0 or small in the fault free case and ≠0 when a fault occurs.*

The residual is a vector in the parity space. This definition implies that the residual $r(t)$ has to be independent of, or at least insensitive to, system states and unmeasured disturbances.

A general structure of a linear residual generator can be described as in Fig. 3. The transfer function from the fault $f(t)$ to the residual $r(t)$ then becomes

$$r(s) = H_y(s)G_f(s)f(s) = G_{rf}(s)f(s) \qquad (1)$$

To be able to detect the i:th fault the i:th column of the response matrix $[G_{rf}(s)]_i$ has to be nonzero, i.e.

Definition 3. ***Detectability*** *The i:th fault is detectable in the residual if $[G_{rf}(s)]_i \neq 0$*

This condition is however not enough in some practical situations. This leads to another definition,

Definition 4. ***Strong detectability*** *The i:th fault is said to be strongly detectable if and only if $[G_{rf}(s)]_i \neq 0$*

Note that in Definition 4, the frequency $\omega = 0$ is made particularly important. Which frequency that is particularly important depends on which type of faults that are interesting. There are three different types of temporal fault behavior:

- Abrupt, step faults
- Incipient (developing) faults
- Intermittent faults

V. Isolation Strategies

In the case of the strongly detectable residuals, the literature describes two general methods for isolation,

- Structured residuals
- Fixed direction residuals

The idea behind **structured residuals** is that a vector valued of residuals is designed making each element in the residual insensitive to different faults or subset of faults whilst remaining sensitive to the remaining faults, i.e. if three faults should be isolate then a three dimension residual should be designed with components $r_1(t)$, $r_2(t)$ and $r_3(t)$ insensitive to one fault each. Then if component $r_1(t)$ and $r_3(t)$ fire it can be assumed that fault 2 has occurred.

The idea with fixed direction residuals is the basis of the f*ault detection filter* (FDF) where the residual vector get a specific direction depending on the fault that is acting upon the system.

VI. Robustness

One problem, as was noted earlier, is that unmeasurable signals often act upon the system plus the influence by modelling errors. This makes it hard to keep the false alarm rate at an appropriate level. This problem is called the robustness problem and a diagnostic algorithm that continues to work satisfactory, even when subjected to modelling errors and disturbances, is called robust.

Since the ideal situation never occurs in a real application, the robustness aspect is one of the most important issues when designing a diagnosis system. The methods to tackle the robustness problem can be divided into two categories

- Robust residual generation, active robustness
- Robust residual evaluation, passive robustness

Robust residual generation methods strive to make the residuals insensitive or even invariant to model uncertainty and disturbances, and still retain the sensitivity towards faults. There are two different types of disturbances, structured and unstructured disturbances. If it is ”known” exactly how a disturbance signal influences the process it is called structured uncertainty and this high degree of disturbance knowledge is enough to actively reduce or even eliminate the disturbance influence on the residual. However if no knowledge of the disturbance is known, no active robustness can be achieved.

However, it is possible to increase robustness in the fault evaluation stage, i.e. in the threshold selection step, for example by using adaptive threshold levels or statistical decoupling. This is called passive robustness. It is not likely that one method can solve the entire robustness problem; a likely solution is one where disturbance decoupling is used side by side with passive robustness.

VII. Model Structure

To proceed in the analysis of residual generation approaches, an analytical model is needed. A state representation of the model is given with the following equation:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)) \end{aligned} \tag{2}$$

The linear (time-continuous) state representation

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{3}$$

As was noted earlier, there are three general types of faults: **sensor (instrument) faults**, **actuator faults** and **component (system) faults**. There are also uncertainties about the model or unmeasured inputs to the process. If these uncertainties are structured, i.e. it is known how they enter the system dynamics, this information can be incorporated into the model.

In the linear case and model uncertainties are supposed structured, the complete model becomes

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B(u(t) + f_a(t)) + Hf_c(t) + Ed(t) \\ y(t) &= Cx(t) + Du(t) + f_s(t) \end{aligned} \tag{4}$$

where $f_a(t)$ denotes actuator faults, $f_c(t)$ component faults, $f_s(t)$ sensor faults and $d(t)$ disturbances acting on the system. H and E is called the distribution matrices for $f_c(t)$ and $d(t)$.

VIII. Parameter Estimation

Process model based residual generators could be parted into two approaches: parameter estimation and geometric approaches.

A parameter estimation method is based on estimating important parameters in a process, e.g. frictional coefficients, volumes or masses, and compares them with nominal values. The typical parameter estimation diagnosis method can be outlined with three steps

- **Data processing**, with the help of the model and measured output data, model parameters can be estimated
- **Fault detection**, which includes a comparison between the estimated parameters and the nominal values
- **Fault classification**, in the case of fault presence, isolation of the fault source is the final stage in a parameter estimation method.

IX. Parity Space Approaches

Geometric approaches to residual generation are called parity space approaches because they generate residuals that are vectors in the parity space. The methods can be divided into open- and closed-loop approaches. In an open-loop approach there are, as the name suggests, no feedback from previously calculated residuals.

The idea behind closed-loop approaches, i.e. observer base approaches, is to use a state estimator as a residual generator. There are a number of approaches suggested in literature like

- State observers
- Fault detection filter
- Unknown Input Observers
  - By parity equations
  - By Kronecker canonical form
  - By eigenstructure assignment of observer

Note that these are methods to design the residual generator. Several of these designs may result in the same residual generator in the end.

X. Summary of Approaches in Literature

To summarize the relationships between the different diagnosis methods a tree-structured is presented in Fig. 4. The different residual generation methods are related as in Fig. 5. All these methods have their advantages and disadvantages and it is likely that in a complete diagnosis application several of these methods will be used.

The presentation done here is in no way complete as there exist numerous of approaches, e.g. the neural network approach.

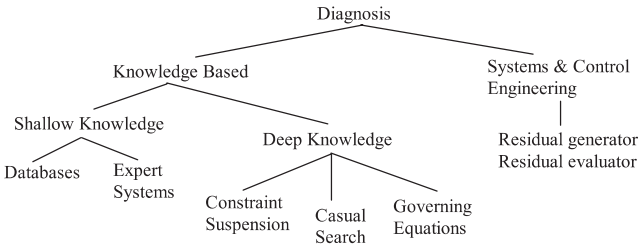


Fig. 4. Categorization of FDI methods

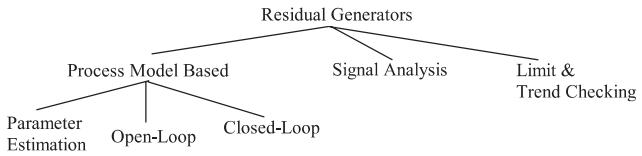


Fig. 5. Categorization of residual generation methods

References

[1] P.M. Frank, "Fault Diagnosis in dynaic system using analytical and knowledge based redundancy – A survey and some new results", *Automatica*, Vol.26, (3), pp.459-474, 1990.

[2] P.M. Frank, "Application of fuzzy logic to process supervision and fault diagnosis",*IFAC Symp. SAFEPROCESS '94*, pp.531-538, Espo, Finland, 1994.

[3] P.M. Frank, "Analytical and qualitative model-based fault diagnosis – a survey and some new results",*European Journal of Control*, no. 2, pp.6-28, 1996.

[4] E. Frisk, *Model-based fault diagnosis applied to a SI-engine*, Master's thesis, Reg.nr: LiTH-ISY-EX-1679, Linköping University, 1996.

[5] M. Nyberg, *Model Based Fault Diagnosis: Methods, Theory and Automotive Engine Applications*, Phd. thesis 591, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 1999.

[6] R.J. Patton, "Fault-Tolerant Control Systems: the 1997 situation", *IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes*, Vol. 3, pp.1033-1054, Kingston Upon Hull, UK, 1997.

[7] R.J. Patton, F. J. Uppal & C. J. Lopez-Toribio, "Soft Computing Approaches to Fault Diagnosis for Dynamic Systems: A Survey", *4th IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes*, Vol. 1, pp.298-311, Budapest, 2000.

An Algorithm for Synchronized Control of Multi-Motor Drive Systems

Mikho R. Mikhov[1]

Abstract – **A general-mode algorithm for synchronized control of multi-motor drive systems is presented in this paper. Application of the offered control algorithm to speed synchronization has been investigated and discussed. The developed computer simulation models and the results obtained can be used in optimization and final tuning of such types of drive systems.**

Keywords – **fault diagnosis, analytical redundancy**

I. Introduction

Automation of a number of production lines requires control of multi-motor electric drive systems. By technological reasons it is often necessary to maintain exact synchronization of the main controlled variables such as position, speed and acceleration.

The required synchronization can be achieved applying a common reference signal and individual stabilization of the regulated variables for all the electric drives.

In order to realize more precise synchronization control, some corrections of the slave electric drive reference signals may be introduced, in accordance with the respective differences between the master and slave drives controlled variable values [1].

A method and relevant device for control of dual-motor electric drives aiming at synchronized maintenance of the regulated variables have been proposed in [2]. The performance of some dual-motor electromechanical systems with synchronized speed and position control applying this method has been presented in [3].

This paper describes and discusses a general-mode control algorithm for multi-motor drive systems, which provides the possibility to maintain reference synchronization accuracy of the respective controlled variables. Results from the investigation of such a drive system where the offered algorithm has been applied for synchronized speed control, are also represented.

II. Control Algorithm

The simplified block diagram of the multi-motor drive system under consideration is shown in Fig. 1, where the following notations have been used: LC – logical control block; ED1, ED2, ..., EDn – electric drives; S1, S2, ..., Sn – sensors for the controlled variables; L1, L2,..., Ln – loads of the electric drives; $x(t)$ – the input common reference signal for the main controlled variable; $x_1(t)$, $x_2(t)$, ..., $x_i(t)$, ...,$x_n(t)$ – the reference signals for the respective electric drives; $y_1(t)$, $y_2(t)$, ..., $y_i(t)$, ...,$y_n(t)$ – the feedback signals; $z_1(t)$, $z_2(t)$, ..., $z_i(t)$, ...,$z_n(t)$ – the disturbances applied to the electric drives; $\triangle y_r$ the reference dead zone determining the synchronization accuracy.

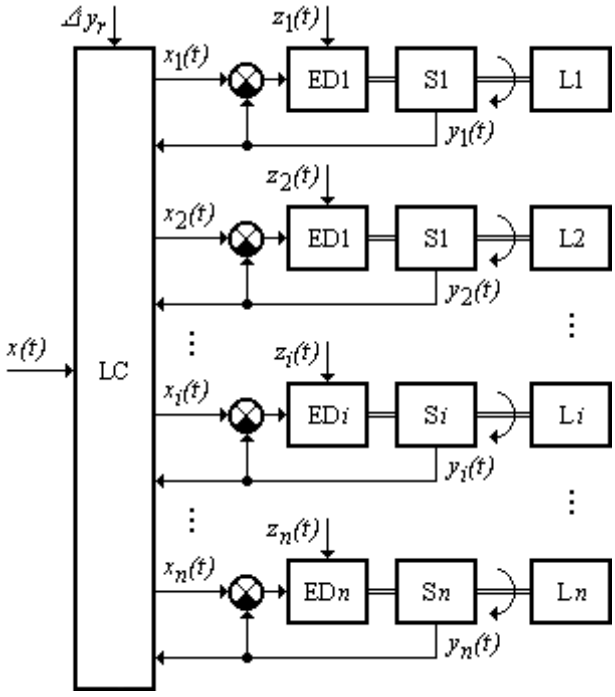


Fig. 1. Simplified block diagram of a multi-motor drive system

In this system the respective feedback signals are being used for both individual stabilization of the controlled variables and realization of synchronized control of the electric drives.

Fig. 2 shows a simplified flowchart of the proposed control algorithm in its general-mode. The algorithm provides for synchronization of the principal controlled variables in multi-motor drive systems such as speed or position.

According to the adopted control strategy, at the beginning of the starting process, as well as during operation in steady-state regimes, the reference signals for the electric drives are as follows:

$$x_1(t) = x_2(t) = ... = x_i(t) = ... = x_n(t) = x(t) \qquad (1)$$

The control continues in relation to Eq. 1 while the maximum error is within the limits of the reference error

$$\Delta y_{max} \leq \Delta y_r \qquad (2)$$

[1]Mikho R. Mikhov is with the Faculty of Automatics, Technical University of Sofia, 8 Kliment Ohridski Str., 1797 Sofia, Bulgaria, Email: mikhov@tu-sofia.acad.bg

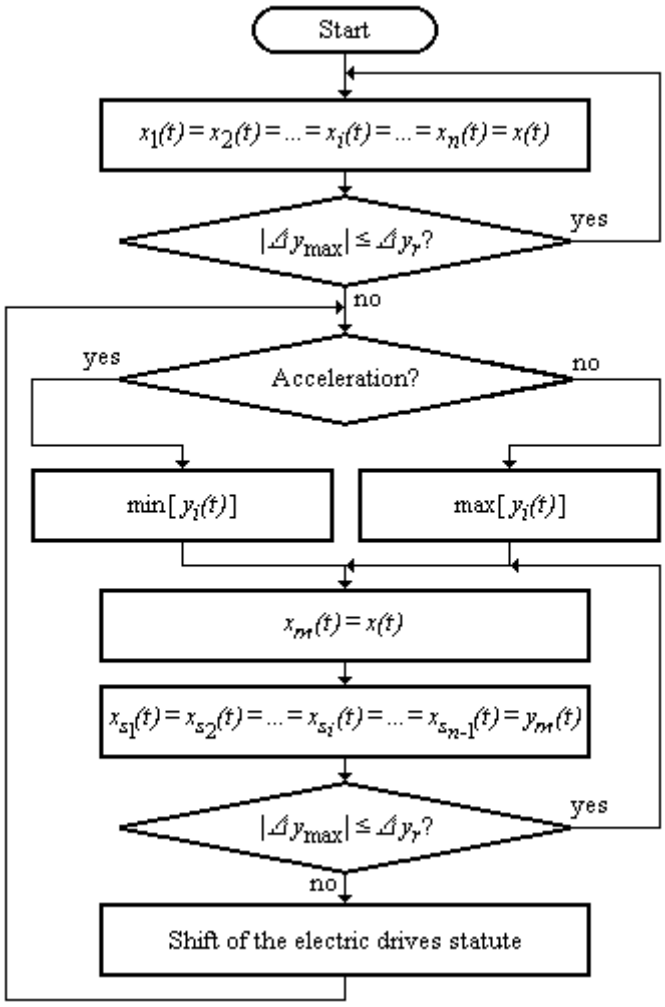


Fig. 2. Simplified flowchart of the proposed control algorithm

The respective feedback signals $y_i(t)$ are compared to determine the smallest value signal. The electric drive that for the given moment appears as most lagging assumes the statute of a master one, while the rest of $n-1$ electric drives become slaves. Therefore, the master drive is the one limiting the performance of the multi-motor system.

The master electric drive continues to be synchronized by the input reference signal

$$x_m(t) = x(t) \tag{3}$$

while the reference signals for the slaves are determined by the feedback signal of the master electric drive:

$$x_{s\,1}(t) = x_{s\,2}(t) = ... = x_{s\,i}(t) = ... = x_{s\,n-1}(t) = y_m(t) \tag{4}$$

The subscripts used are as follows: m – for the master drive and s – for the all slave drives.

When a new lagging occurs, the slowest electric drive automatically becomes the master and the control continues in compliance with Eq. 3 and Eq. 4 respectively.

During acceleration the electric drive that has the lowest level of the controlled variable is assumed as a master - $\min[y_i(t)]$, but at deceleration the master drive will be the one that has the highest controlled variable value - $\max[y_i(t)]$. Therefore, the shifting conditions for the electric drives statute are as follows:

$$y_m(t) = \min[y_i(t)] \;-\; \text{for acceleration} \tag{5}$$
$$y_m(t) = \max[y_i(t)] \;-\; \text{for deceleration} \tag{6}$$

This way the required accuracy is achieved as in the transient regimes (accelerating and decelerating) with different loads, so in cases of disturbances applied to the respective electric drives.

To verify the offered control algorithm functionality a number of computer simulation models have been developed using the MATLAB/SIMULINK software package. Various electric drives have been modelled including both DC and AC motor types. The respective models allow study of a wide range of multi-motor electromechanical systems with different controlled variables.

III. Investigation of a Sample Drive System

The block diagram of a sample multi-motor drive system of the investigated type is shown in Fig. 3. The used notations are as follows: LC – logical control block; SC1, SC2 and SC3 – speed controllers; CC1, CC2 and CC3 – current controllers; PC1, PC2 and PC3 – power electronic converters; M1, M2 and M3 - electrical motors; SS1, SS2 and SS3 – speed sensors; SF1, SF2 and SF3 – speed feedback blocks; CF1, CF2 and CF3 -current feedback blocks; L1, L2 and L3 – loads of the motors; ω_2, ω_2 and ω_3 – angular speeds; T_1, T_2 and T_3 – motor torques; T_{l_1}, T_{l_2} and T_{l_3} – load torques applied to the motors; J_1, J_2 and J_3 – summary inertias referred to the motor shafts.

The corresponding signals and variables for this tri-motor drive system are as follows:

$$\begin{aligned} &x(t) \to V_{sr}\,; \\ &x_i(t) \to V_{sr_i}\,; \\ &y(t) \to V_{sf_i}\,; \\ &z(t) \to \triangle T_{l_i},\, \triangle J_i\,; \\ &\triangle y_r \to \triangle \omega_r\,; \\ &i = 1,\, 2,\, 3 \end{aligned} \tag{7}$$

The system input includes the common reference signal for the main controlled variable V_{sr} as well as the reference dead zone determining the synchronization accuracy $\triangle \omega_r$. Both load torque and inertia changes in the drive system can be considered as disturbances.

In the electromechanical system under consideration all the electric drives are of the same type and act as dual-loop cascade control structures.

Fig. 4 represents some simulation results illustrating the performance of one of the electric drives. The transient start and stop processes are shown, as well as the drive response to the disturbances, expressed in changes of the load torque acting upon the motor shaft. The starting current is limited to the maximum admissible value $I_{a\max}$, which provides a maxi-mum starting motor torque. The load torque is equal

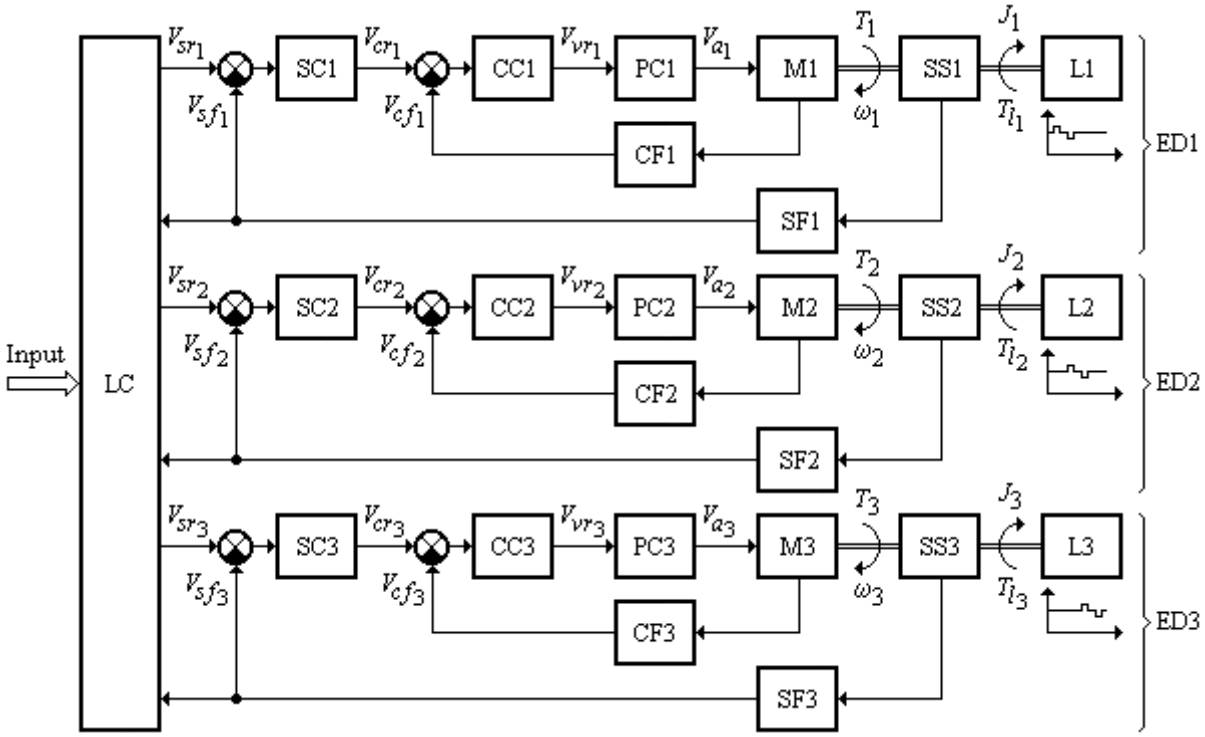


Fig. 3. Block diagram of the investigated multi-motor drive system with synchronized speed control

to the nominal value, while the disturbances applied sequentially are $\triangle T_l$=+25% and $\triangle T_l$=-25%, respectively. The motor cur-rent value at the steady-state regimes is $I_a = I_{a\text{nom}}$

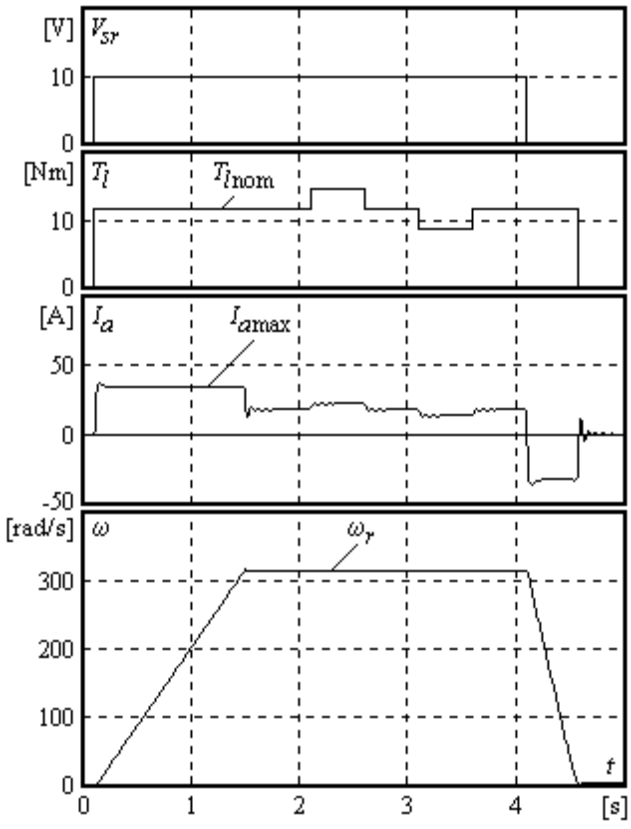


Fig. 4. Simulation results for one of the controlled electric drives

and it conforms to the nominal load torque $T_{l\text{nom}}$. The setting electric drive parameters in this case are as follows: reference speed $\omega_r = 314$ rad/s; maximum current $I_{a\max}$=35.2 A.

Fig. 5 shows the time diagrams of speeds and relative speed errors obtained as simulation results for non-synchronized control. The speed errors $\triangle\omega_1(t)$, $\triangle\omega_2(t)$ and $\triangle\omega_3(t)$ are calculated with respect to the slowest electric drive (in this system it is ED1). Because of that the respective error is $\triangle\omega_1(t)$. The rather large speed discrepancies are due to the differences between the load inertias J_1, J_2 and J_3 of the controlled electric drives. The reference speeds are as follows: $\triangle\omega_{r_1}(t)$=157 rad/s; $\triangle\omega_{r_2}(t)$=314 rad/s.

The respective time diagrams of speeds and relative speed errors at synchronized speed control are represented in Fig. 6. The working conditions at which this investigation has been carried out are exactly the same as those of non-synchronized control but the resulting speed errors $\triangle\omega_i(t)$ have been represented in a different scale. As evident, the errors of synchronized control are considerably reduced and the accuracy achieved in this case is $|\triangle\omega_{\max}| = 0.5\%$.

IV. Conclusion

A general-mode algorithm for synchronized control of multi-motor drive systems has been represented. Relevant models for computer simulation of such systems with control utilizing the described algorithm have been developed. The possibilities for application of this algorithm to speed control have been investigated and discussed.

The study has been carried out for electromechanical systems with controlled rectifier DC motor drives. The basic parameters of the used motors are as follows: P_{nom}=

3.4 kW; V_{nom}= 222 V; $I_{a\,\text{nom}}$= 17.6 A; ω_{nom}= 314 rad/s; $T_{l\text{nom}}$= 11.76 Nm.

The detailed analysis of the drives performance at the respective transient and steady state regimes bring us to the following basic conclusions:

- the control algorithm presented here provides for a good electric drives synchronization at different loads and disturbances;
- except for speed, this algorithm for synchronization control is suitable for other basic regulated variables of multi-motor electromechanical systems, such as position and acceleration;
- this type of synchronization control can be applied also in maintaining reference ratios of the respective regulated variables.

The developed computer simulation models and the results obtained can be used in optimization and final tuning of such types of multi-motor drive systems.

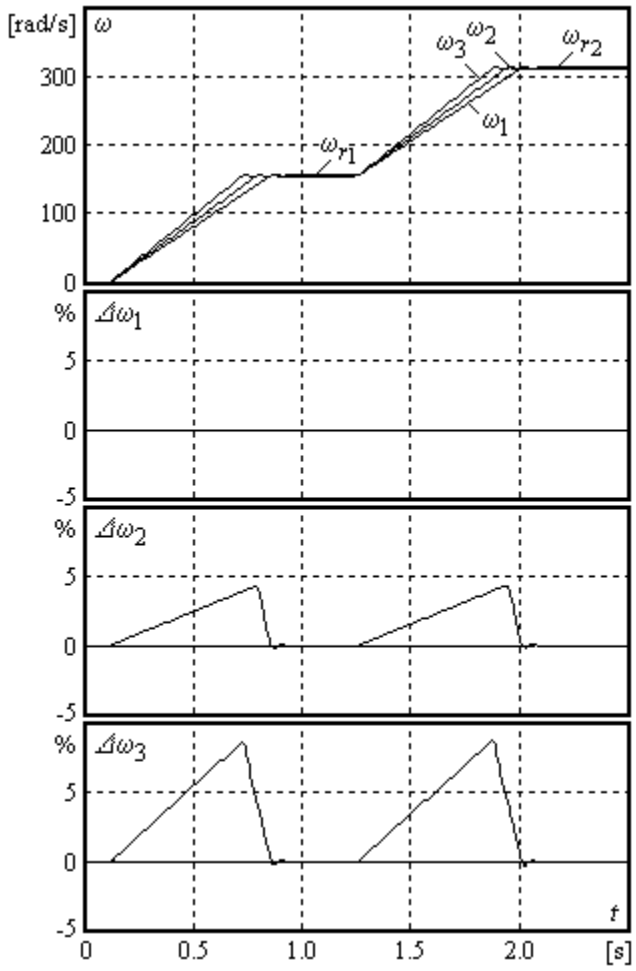


Fig. 5. Time diagrams at non-synchronized speed control

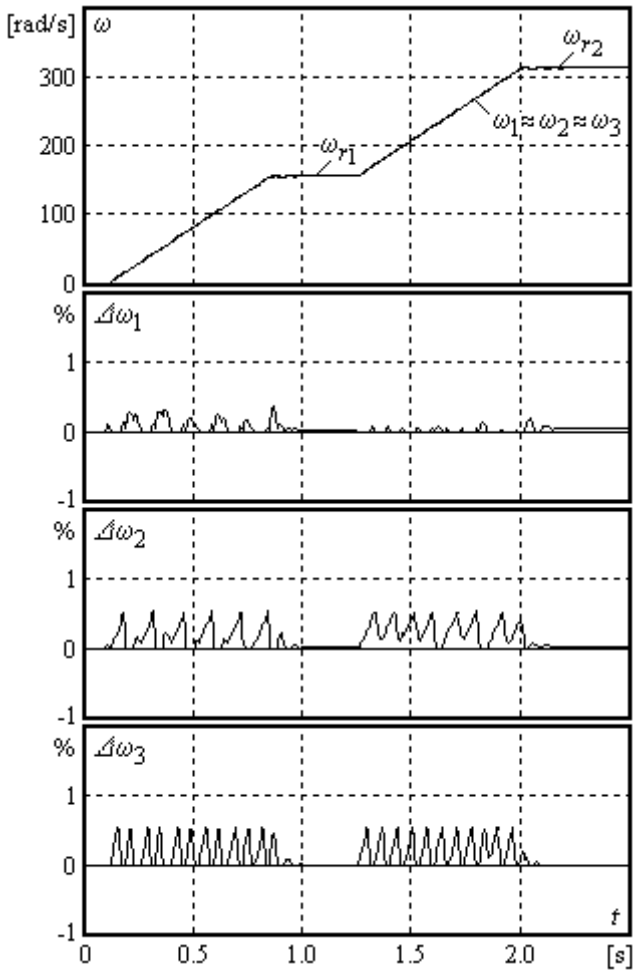


Fig. 6. Time diagrams at synchronized speed control

References

[1] R. P. Litchev, L. T. Dombalova and M. R. Mikhov, "A System for Automation of Extruded Asbestos-Cement Panels Production", *Proceedings of the International Conference AEDTP' 87*, pp. 240-246, Plovdiv, Bulgaria, 1987.

[2] M. R. Mikhov, "Method and Device for Controlling Double Motor Electric Drive", *Patent Abstracts of Bulgaria*, vol. 10, no. 10, pp. 44-45, 1998.

[3] M. R. Mikhov, "Control of dual-motor drive systems", *Proceedings of the International Conference on Mechatronics and Information Technology (ICMIT'01)*, pp. 198-203, Yamaguchi, Ja-pan, 2001.

[4] U. Keuchel and R. M. Stephan, *Microcomputer-Based Adaptive Control Applied to Thyristor-Driven DC Motors*, London, Springer-Verlag, 1994.

[5] K. Ogata, *Solving Control Engineering Problems with MATLAB*, New Jersey, Prentice-Hall, 1994.

Comparative Analysis of Some Discrete-Time Sliding Mode Chattering-Free Control Algorithms

Bojan T. Milosavljević[1]and Čedomir Milosavljević[2]

Abstract – **This paper performs a comparative analysis of three discrete-time sliding mode control algorithms that tend to reduce chattering in the controlled variables. First, a brief review of each algorithm is given, based on the original papers. In order to compare presented algorithms, computer simulations have been carried out prior to verify their robustness features. The robustness tests are not limited to parameter uncertainties and external disturbances only, but robustness to the existence of unmodelled dynamics is also considered. Simulation results reveal the best performance algorithm.**

Keywords – **discrete-time sliding mode, chattering-free**

I. Introduction

Variable structure systems with sliding mode theoretically possess very desirable features, such as robustness to controlled plant parameter variations and to external disturbances in very wide range, simple definition of requested motion dynamics, being described by differential equations of lower order than one of the controlled object, high compatibility of modern electronic components and devices to the requests of such a control etc. However, in a real system, there exists parasitic high frequency motion around the sliding surface, the so-called chattering. This phenomenon exists due to discontinuities, high gains, sampling effects and finite switching speed in the system. It can cause damage to actuators or the plant. There are essentially two ways to overcome this problem. One way is to use higher order sliding mode, and the other way is to add a boundary layer around the switching surface and use continuous control inside the boundary. The problem with the first method is that the derivative of the certain state variable is not available for measurement, and therefore methods have to be used to observe that variable. A modification of this method was presented in [1], where the control algorithm is still based on state- and control input derivatives, but combining the equivalent control method and Lyapunov theory, direct use of those variables was avoided. Therefore, it was possible to achieve a continuous control input and thus reduce chattering without observing any variable. In the second method, it is important that the trajectories inside the boundary layer do not try to come outside the boundary after entering the boundary layer.

A number of algorithms can be found in papers based on one of those techniques. Most of them attempt to ensure robustness of the system to parameter uncertainties and external disturbances only. Incomplete knowledge of the system dynamics is very common in engineering practice. Therefore it is very important to provide robust algorithms to the existence of unmodelled dynamics. Then it would be possible to simplify the control design being applied to lower-order system. This paper provides a comparative analysis of three control algorithms regarding robustness properties.

II. Algorithm 1

The following discrete-time system is considered in [3]:

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \triangle\mathbf{A}\mathbf{x}(k) + \mathbf{b}u(k) + \mathbf{f}(k) \\ y(k) &= \mathbf{h}^T\mathbf{x}(k) \end{aligned} \tag{1}$$

$\mathbf{x}$ is the $n\times 1$ state vector, $\mathbf{A}$ is an $n\times n$ matrix, $\mathbf{b}$ and $\mathbf{h}$ are $n\times 1$ vectors, u is the system input and y is the system output. In this equation, the $n\times n$ matrix $\triangle\mathbf{A}$ represents parameter uncertainties and the $n\times 1$ vector $\mathbf{f}$ denotes external disturbances, satisfying matching conditions. The time-varying switching surface is defined as follows:

$$s(k) = \mathbf{c}^T\mathbf{x}(k)\ . \tag{2}$$

Disturbances and parameter uncertainties are bounded so that the following relation holds:

$$d_l \le d(k) = \mathbf{c}^T\triangle\mathbf{A}\mathbf{x}(k) + \mathbf{c}^T\mathbf{f}(k) \le d_u\ . \tag{3}$$

The lower and upper bounds d_l and d_u are known constants. The average value of $d(k)$ (d_0) and its maximum admissible deviation (δ_d) are introduced as follows:

$$d_0 = \frac{d_l + d_u}{2}, \quad \delta_d = \frac{d_u - d_l}{2}\ . \tag{4}$$

First, the required evolution of the time-varying switching surface $s(k)$ is specified:

$$s(k+1) = d(k) - d_0 + s_d(k+1) - \sum_{i=0}^{k}\left[s(i) - s_d(i)\right]\ . \tag{5}$$

The evolution of the time-varying hyperplane is

$$s_d(k) = \begin{cases} \dfrac{k^* - k}{k^*}s(0) & k = 0, 1, .., k^*, k^* < \dfrac{s(0)}{2\delta_d} \\ 0 & k > k^* \end{cases} \tag{6}$$

[1]Bojan T. Milosavljević is a postgraduate student at the Faculty of Electronic Engineering, University of Niš, Beogradska 14, Serbia, Serbia and Montenegro, E-mail: mbojanks@ptt.yu

[2]Čedomir Milosavljević is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, Serbia, Serbia and Montenegro, E-mail: milosavljevic@elfak.ni.ac.yu

[3]Dimitar Stoyanov is with the Institute of Electronics at the Bulgarian Academy of Sciences, 72, Tzarigradsko chausee, 1784-Sofia, Bulgaria, e-mail: dvstoyan@ie.bas.bg

The constant k^* is a positive integer chosen by the designer in order to achieve good tradeoff between the fast convergence rate of the system and the magnitude of the control u required to achieve this convergence rate.

Then a control law $u(k)$ is proposed which drives the system in such a way that the variable $s(k)$ actually changes according to the specification:

$$u(k) = -(\mathbf{c}^T\mathbf{b})^{-1}\Big\{\mathbf{c}^T\mathbf{A}\mathbf{x}(k) + d_0 - s_d(k+1) + \sum_{i=0}^{k}[s(i) - s_d(i)]\Big\} \quad (7)$$

This control design procedure is referred to as the reaching law approach.[2]

As it is shown in [3], the following holds:

$$|s(k) - s_d(k)| = |d(k) - d(k-1)| \le \triangle_d \overset{\text{Eq. (6)}}{\Longrightarrow} |s(k)| \le \triangle_d,\ k > k^* \quad (8)$$

$\triangle_d$ denotes the disturbance-rate and the parameter- change-rate limit.

III. Algorithm 2

In [1] the sliding mode motion design is proposed generating a continuous control input, thus eliminating chattering. Neither the explicit calculation of the equivalent control nor high gain inside the boundary layer is used. The algorithm is performed by means of the Lyapunov theory and is applied to a nonlinear system shown in the regular form:

$$\frac{\mathrm{d}\mathbf{x}_1}{\mathrm{d}t} = \mathbf{f}_1(\mathbf{x}_1, \mathbf{x}_2),\ \frac{\mathrm{d}\mathbf{x}_2}{\mathrm{d}t} = \mathbf{f}_2(\mathbf{x}_1, \mathbf{x}_2) + \mathbf{B}_2(\mathbf{x})\mathbf{u} + \mathbf{B}_2(\mathbf{x})\mathbf{d}(t)\ ;$$
$$\mathbf{x}_1 \in \Re^{n-m}, \mathbf{x}_2 \in \Re^m, \mathbf{u} \in \Re^m, \mathbf{f} \in \mathbf{F}^m, \quad (9)$$
$$rang\,[\mathbf{B}_2(\mathbf{x})] = m\ .$$

The components of control input and of the vector $(\mathrm{d}\mathbf{x}_2/\mathrm{d}t)$ are assumed bounded:

$$u_i \in [u_{i_{min}} u_{i_{max}}]\ ;$$
$$(\mathrm{d}\mathbf{x}_2/\mathrm{d}t) \in [\alpha_{min}, \alpha_{max}],\ (i = 1, ..., m)\ . \quad (10)$$

The motion of the system is restricted to belong to the manifold S

$$S = \{\mathbf{x} : \varphi(t) - \sigma_a(\mathbf{x}) = \sigma(\mathbf{x}, t) = 0\}\ ;$$
$$\sigma_a^T = [\sigma_{a1}, \sigma_{a2}, ..., \sigma_{am}] \in \mathbf{F}^m\ , \quad (11)$$
$$\varphi_a^T = [\varphi_{a1}, \varphi_{a2}, ..., \varphi_{am}] \in \mathbf{F}^m\ .$$

$\sigma_{ai}(t)$ and $\varphi_{ai}(t)$, $(i = 1, ..., m)$ are continuous functions. $\varphi_{ai}(t)$ and their first time-derivatives are bounded. These functions can be interpreted as the references to be traced by selected combinations $\varphi_{ai}(\mathbf{x})$ of the system's states.

For the system described by Eqs. (9), (10), and (11), the following design procedure is adopted:

- select a Lyapunov function candidate $\nu(\sigma)$, such that, if the Lyapunov stability criteria are satisfied, the solution $\varphi(t) - \sigma_a(\mathbf{x}) = 0$ is stable on the trajectories of the system described by Eqs. (9), (10), and (11);
- select a form which the time-derivative of the Lyapunov function should satisfy, and find control u such that selected form is achieved on the trajectories of the system described by Eqs. (9), (10), and (11);
- find the equations of motion on the selected manifold with designed control.

The selection of Lyapunov function should be as simple as possible; hence, the first choice is a quadratic form

$$\nu = \frac{\sigma^T\sigma}{2}\ . \quad (12)$$

According to the Lyapunov theory, the solution $\sigma(\mathbf{x}, t) = 0$ will be stable if the time-derivative of the Lyapunov function can be expressed as

$$\frac{\mathrm{d}\nu}{\mathrm{d}t} = -\sigma^T\mathbf{D}\sigma,\ \mathbf{D} > 0\ . \quad (13)$$

It is shown in [1] that control can be calculated as

$$u = sat\left[u_{eq} + (\mathbf{B}_2)^{-1}\mathbf{D}\sigma\right] \quad (14)$$

and, using equality $\mathbf{B}_2 u_{eq} = \mathbf{B}_2 u + \mathrm{d}\sigma/\mathrm{d}t$, control u is finally

$$u(t) = sat\left[u(t^-) + (\mathbf{B}_2)^{-1}\left(\mathbf{D}\sigma + \frac{\mathrm{d}\sigma}{\mathrm{d}t}\right)\right]\ , \quad (15)$$
$$t = t^- + \triangle, \triangle \to 0\ .$$

Here $\triangle$ denotes the time-delay necessary for the calculations.

For the system given in the regular form, the following model holds during the sliding mode:

$$\frac{\mathrm{d}\mathbf{x}_1}{\mathrm{d}t} = \mathbf{f}_1(\mathbf{x}_1, \mathbf{x}_2)\ \wedge\ \frac{\mathrm{d}\sigma}{\mathrm{d}t} + \mathbf{D}\sigma = 0 \quad (16)$$

Discrete-time versions of Eqs. (15) and (16) can be written as

$$u(kT) = sat\Big[u(kT - T) + (\mathbf{B}_2)^{-1}\Big(\mathbf{D}\sigma(kT) + \frac{\mathrm{d}\sigma(kT)}{\mathrm{d}t}\Big)\Big] \quad (17)$$

$$\sigma(kT) = (\mathbf{I} - T\mathbf{D})\sigma(kT - T) \quad (18)$$

T is the sampling interval, $\mathbf{I}$ is the identity matrix. If matrix $\mathbf{D}$ is selected diagonal with $d_{ii} = 1/T$ then Eq. (18) equals zero and sliding mode will occur after finite number of sampling intervals. Further simplifications can be introduced by substituting $\mathrm{d}\sigma(kT)/\mathrm{d}t$ by its first order approximation

$$u(kT) = sat\Big[u(kT - T) + (\mathbf{B}_2 T)^{-1}\Big((\mathbf{I} + T\mathbf{D})\sigma(kT) - \sigma(kT - T)\Big)\Big]\ . \quad (19)$$

IV. Algorithm 3

In [4] a linear time-invariant system is considered:

$$\dot{\mathbf{x}} = \mathbf{A}_c\mathbf{x}(t) + \mathbf{b}_c u(t) \quad (20)$$

with scalar sample & hold control

$$u(t) = u(kT),\ kT \le t < (k+1)T,$$
$$k \in N^0 = \{0, 1, 2, ...\},\ T > 0\ . \quad (21)$$

An equivalent discrete-time representation is then for the perturbed system

$$\delta\mathbf{x}(kT) = A_\delta(T)\mathbf{x}(kT) + A_\delta(T)\mathbf{x}(kT) + \mathbf{b}_\delta(T)u(kT) + \mathbf{d}_\delta(T)\mathbf{f}(kT) \quad (22)$$

$$\dot{\mathbf{x}} \approx \delta\mathbf{x}(kT) = \left[\mathbf{x}((k+1)T) - x(kT)\right]/T \quad (23)$$

and $\triangle\mathbf{A}_\delta(T) \in \Re^{n\times n}$ is a matrix of uncertainties, $\mathbf{d}_\delta(T) \in \Re^{n\times 1}$, $\mathbf{f}(kT)$ is a bounded external disturbance with

$$|\mathbf{f}(kT)| \le \mu,\ \forall k \in N^0\ . \quad (24)$$

The matching conditions are assumed, and therefore

$$\mathbf{d}_\delta(T) = \mathbf{b}_\delta(T)\ . \quad (25)$$

The goal is to impose $s = 0$ as the sliding mode hyperplane

$$s = \mathbf{c}_\delta(T)\mathbf{x},\ \mathbf{c}_\delta(T) \in \Re^{1\times n}\ . \quad (26)$$

The following assumption ensures that the relative degree of variable s, seen as an output, with the respect to the control signal u is one

$$\mathbf{c}_\delta(T)\mathbf{b}_\delta(T) = 1\ . \quad (27)$$

The reaching law is defined as follows

$$\delta s(k) = -\Phi(s(k), \mathbf{X}(k))\ ,$$
$$\delta s = \frac{s(k+1) - s(k)}{T} = \mathbf{c}_\delta(T)\delta\mathbf{x}(k)\ , \quad (28)$$

$$\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}(k) \\ \hat{\mathbf{x}}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{x}(k-1) \end{bmatrix}\ ,$$
$$\text{by definition } \hat{\mathbf{x}}(0) = \mathbf{x}(0)\ , \quad (29)$$

The control u is then

$$u(k) = -\mathbf{c}_\delta(T)\mathbf{A}_\delta(T)\mathbf{x}(k) - \Phi(s(k), \mathbf{X}(k))\ . \quad (30)$$

According to the theorem from [4], it is sufficient that the following conditions for Φ are met:

$$\Phi(s, \mathbf{X}) = s/T,\ \mathbf{X} \in \mathbf{S}(T)\ ,$$
$$\gamma T(d_m||\mathbf{X}||_1 + \mu)/|s| < T\Phi(s, \mathbf{X})/s < 1\ , \quad (31)$$
$$\mathbf{X} \notin \mathbf{S}(T),\ \gamma > 1,\ \varepsilon > \mu,\ \eta_2 > d_m\ .$$

The following function satisfying these conditions is proposed in [4]

$$\Phi(s, \mathbf{X}) = \min\left(\frac{|s|}{T}, \sigma + q|s| + r||\mathbf{X}||_1\right)\mathrm{sgn}(s)\ ,$$
$$0 \le qT < 1,\ r \ge d_m\gamma,\ \sigma > \gamma\mu\ . \quad (32)$$

The vicinity of the hyperplane is defined by

$$\mathbf{S}(T) = \Big\{\mathbf{X} \in \Re^{2n}:$$
$$s = |\mathbf{c}_\delta(T)\mathbf{x}| < \frac{\sigma T + rT||\mathbf{x}||_1 + rT||\hat{\mathbf{x}}||_1}{1 - qT}\Big\}\ , \quad (33)$$

$$\sigma = 9,\ q = 0,\ r = 0.011,\ (\gamma = 1.1).$$

V. Comparison of Algorithms

In order to compare the proposed algorithms, each of them is applied to the model of a DC motor with neglected electric-time constant, also used in [4] for the verification purposes

$$\dot{x_1} = x_2,\ \dot{x_2} = -16x_2 - 680u\ , \quad (34)$$

where $x_1 = \theta_d - \theta$ (θ is the angular position of the rotor shaft), $x_2 = -\omega$ (ω is the rotor velocity), and u is the control signal. Corresponding to the matrix representation of the Eq. (20), it can be written

$$\mathbf{A}_c = \begin{bmatrix} 0 & 1 \\ 0 & -16 \end{bmatrix},\ \mathbf{b}_c = \begin{bmatrix} 0 \\ -b_c \end{bmatrix} = \begin{bmatrix} 0 \\ -680 \end{bmatrix}\ . \quad (35)$$

The Eq. (34) with external disturbance included, according to Eqs. (22), (25), satisfying Eq. (24), is as follows

$$\dot{x_1} = x_2,\ \dot{x_2} = -16x_2 - 680(u - f)\ , \quad (36)$$
$$f = \mu\big(|2 - t| - 1\big)\ . \quad (37)$$

The adopted external disturbance waveform f of Eq. (37) with $\mu = 0.7$ (the same as in [4]) is shown in Fig. 1.

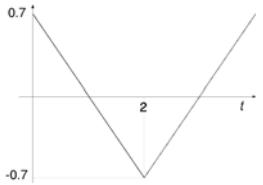


Fig. 1. The external disturbance waveform f

A relation between two discrete representations (given by the Eqs. (22) and (1)) should be established to evaluate the external disturbance in the Eq. (1) in the form $\mathbf{f}(kT) = [0\ \ f_1]^T$. It can be performed easily after rearranging the terms in the Eq. (22) and dividing by T(sampling period) both sides of the equation; hence, the following holds

$$\mathbf{A}_\delta = \frac{\mathbf{A} - \mathbf{I}}{T},\ \mathbf{b}_\delta = \frac{\mathbf{b}}{T},\ \triangle\mathbf{A}_\delta = \frac{\triangle\mathbf{A}}{T}\ , \quad (38)$$

$$b_\delta f = \frac{f_1}{T}\ . \quad (39)$$

The discrete-time representation parameters expressed in the form of Eq. (1) for the system described by Eq. (36), are [5]

$$\mathbf{A} = \exp(A_cT) = \begin{bmatrix} 1 & \frac{1}{16}(1 - \exp(-16T)) \\ 0 & \exp(-16T) \end{bmatrix},$$
$$\mathbf{b} = \int_0^T \exp(\mathbf{A}_c\tau)\mathbf{b}_c\mathrm{d}\tau = \quad (40)$$
$$= \begin{bmatrix} \frac{85}{32}(1 - 16T - \exp(-16T)) \\ \frac{85}{2}(1 - \exp(-16T)) \end{bmatrix}.$$

The reference value θ_d is 100 rad. The desired system response is related to the degree of exponential stability, being $\exp(-\alpha T)$, α=15 s^{-1}. According to the procedure described

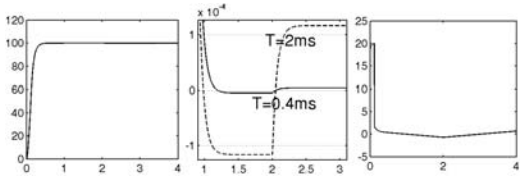


Fig. 2. Algorithm 1: time response, steady-state error and control signal respectively

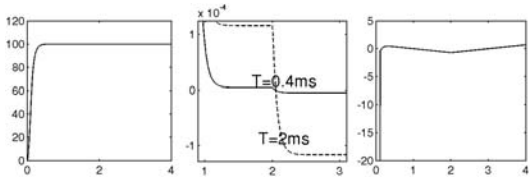


Fig. 3. Algorithm 2: time response, steady-state error and control signal respectively

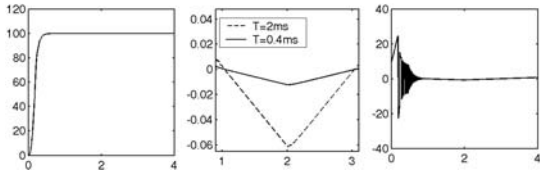


Fig. 4. Algorithm 3: time response, steady-state error and control signal respectively

and applied in [4], the sliding hyperplane parameters $\mathbf{c}_\delta$ are as follows

$$\begin{aligned} T=0.4ms:\ & \mathbf{c}_\delta(T)=[-0.0220632-0.00147088],\\ & \mathbf{c}_\delta(T)\mathbf{A}_\delta(T)=[0,0.00146618]\,;\\ T=2ms:\ & \mathbf{c}_\delta(T)=[-0.0220808-0.00147204],\\ & \mathbf{c}_\delta(T)\mathbf{A}_\delta(T)=[0,0.00144863]\,;\\ T=10ms:\ & \mathbf{c}_\delta(T)=[-0.0221655-0.00147758],\\ & \mathbf{c}_\delta(T)\mathbf{A}_\delta(T)=[0,0.00136288]\,. \end{aligned} \tag{41}$$

To apply the Algorithm 1, Eq. (40) is used, along with

$$\mathbf{b}_\delta=\frac{\mathbf{b}}{T}\Rightarrow(\mathbf{c}_\delta\mathbf{b}_\delta T)^{-1}=(\mathbf{1}\cdot T)^{-1}=\frac{1}{T} \tag{42}$$

and a certain approximation, arising from Eqs. (38) and (41)

$$\mathbf{c}_\delta\mathbf{A}_\delta T=\mathbf{c}_\delta(\mathbf{A}-\mathbf{I})\Rightarrow\mathbf{c}_\delta\mathbf{A}=(\mathbf{c}_\delta\mathbf{A}_\delta)T+\mathbf{c}_\delta\approx\mathbf{c}_\delta\,. \tag{43}$$

Finally, the control signal is generated by

$$u(k)=-\frac{1}{T}[s(k)-s_d(k+1)+Z(k)]\,, \tag{44}$$

$$Z(k)=Z(k-1)+s(k)-s_d(k)\,. \tag{45}$$

The choice of k^* is governed by the desired sliding-surface reaching time, being $t_r=120$ ms. Hence

$$k^*=\frac{t_r}{T}\,. \tag{46}$$

The simulation results for the Algorithm 1, applied to the perturbed system of Eq. (36), are given in the Fig. 2. The system response waveform is in the Fig. 2a, and steady-state error in Fig. 2b, but for three different sampling periods, the same as in Eq. (41). The control signal is presented in the Fig. 2c.

The steady-state error of the controlled variable x_1 can be evaluated by means of Eq. (8). $\triangle_d$ can be expressed as (Eqs. (3), (39) and (37))

$$\begin{aligned} \triangle_d &\overset{\text{Eq. (3)}}{=} c_2\triangle f_1 \overset{\text{Eq. (39)}}{=} c_2 b_\delta T\triangle f\\ &= c_2 b_\delta T\big(f(t)|_{t=kT}-f(t)|_{t=(k-1)T}\big) \end{aligned} \tag{47}$$

$$\triangle_d \overset{\text{Eq. (37)}}{=} c_2 b_\delta T\cdot\mu T=c_2 b_\delta\mu T^2\,.$$

Since in steady state $x_2\to 0$ and therefore $|s(k)|\approx c_1 x_1$

$$x_1\approx\frac{c_2}{c_1}b_\delta\mu T^2\,. \tag{48}$$

The simulation values from Fig. 2b match the values from Eq. (48). Thus the simulation scheme and the Algorithm 1 itself are verified.

Control design according to the Algorithm 2 is not based on the discrete-time representation like the other two algorithms. The continuous-time system (36) is already in the regular form [6]. That is why c_2 is chosen to be $c_2=1$, and to keep the same system response features, c_1 takes the value of c_2/c_1 from the Eq. (41). It is also chosen that $d_{ii}=1/T\Rightarrow T\mathbf{D}=\mathbf{I}$, and according to the Eq. (19)

$$u(kT)=sat[u(kT-T)+(b_\delta T)^{-1}(2s(kT)-s(kT-T))]. \tag{49}$$

The simulation results shown in Fig. 3 are as expected.

The Algorithm 3 is designed according to the Eqs. (30), (32) and (33) and simulation results shown in Fig. 4 verify the simulation scheme and the algorithm itself.

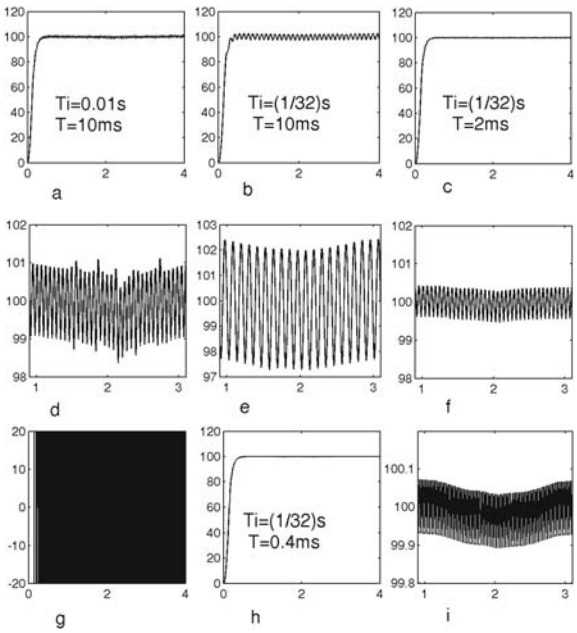


Fig. 5. The effects of unmodeled dynamics in the Algorithm 2:
– First two rows: time responses for respectively Ti=0.01s, T=10ms; Ti=(1/32)s, T=10ms and Ti=(1/32)s, T=2ms
– The third row: control signal u and time response for Ti=(1/32)s and T=0.4ms

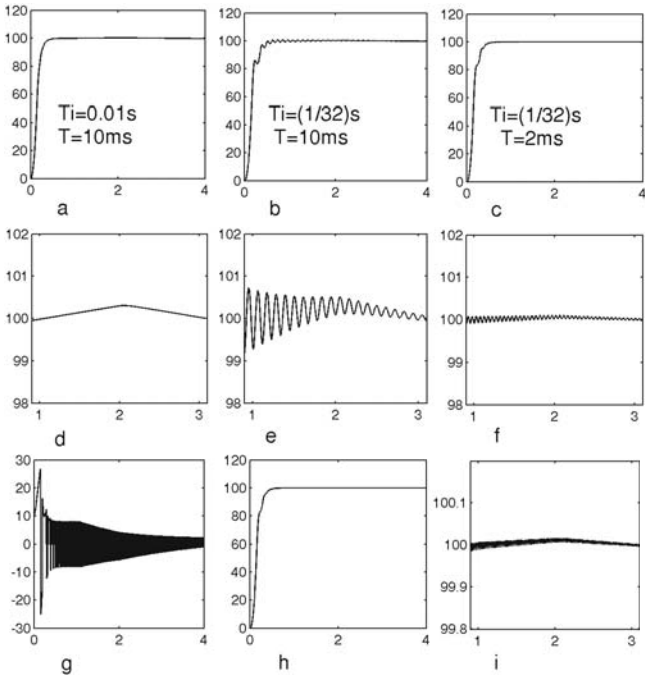


Fig. 6. . The effects of unmodeled dynamics to Algorithm 3:
– First two rows: time responses for respectively Ti=0.01s, T=10ms; Ti=(1/32)s, T=10ms and Ti=(1/32)s, T=2ms
– The third row: control signal u and time response for Ti=(1/32) s and T=0.4ms

Unmodelled dynamics is introduced in the system as an additional pole with time-constant T_i and the Coulomb friction. The simulation of the corresponding system with Algorithm 1 could not finish because going to infinity. If the Algorithm 2 is applied, the simulation results are those shown in Fig. 5.

The simulation results for Algorithm 3 are shown in Fig. 6.

Obviously, it can be seen from Figs. 5 and 6 that Algorithm 3 is better solution if chattering amplitude reduction is the most important task.

The simpler version of Algorithm 1 (without sums in the Eqs. (5) and (7)) provides the sliding surface steady-state error of $|s(k)| \leq \delta_d$. It should be emphasized that the system controlled by this version of the Algorithm 1 is stable with the chattering amplitude greater than in two other cases.

VI. Conclusion

Three possible control algorithms are compared regarding chattering reduction. Their robustness to parameter uncertainties and external disturbance presented in the original papers are verified. Their different behaviors regarding chattering reduction at presence of unmodelled dynamics are treated in this paper and the best performance algorithm is proposed.

References

[1] Asif Šabanović, Karel Jezernik and Kenzo Wada, *Chattering-free sliding modes in robotic manipulators control*, Robotica (1996), vol. 14, pp 17-29 1996 Cambridge University Press

[2] Weibing Gao, Yufu Wang and Abdollah Homaifa, *Discrete-Time Variable Structure Control Systems*, IEEE Trans. on Industrial Electronics, vol. 42, No. 2, April 1995

[3] Andrzej Bartoszewicz, *Discrete-Time Quasi-Sliding-Mode Control Strategies*, IEEE Trans. on Industrial Electronics, vol. 45, No. 4, August 1998

[4] Goran Golo, Čedomir Milosavljević, *Robust discrete-time chattering free sliding mode control*, Systems & Control Letters 41 (2000) 19-28, Elsevier

[5] Milić Stojić, *Discrete-Time Control Systems*, Nauka (Science), Belgrade, 1998 (in Serbian)

[6] C. Edwards and S. Spurgeon, *Sliding Mode Control: Theory and Applications, universiteits bibliotheek*, Twente and Taylor & Francis, 1998

Two Stages Piece-Wise Linearization Method for Inteligent Transducers

Dragan B. Živanović[1], Miodrag Z. Arsić[2], Jelena R. Dordević[3]and Ilija S. Mladenović[4]

Abstract – **This paper presents the intelligent transducers suitable linearization methods, considering practical implementation. New method is proposed. By one look-up table, it transforms x-axis first, and then by other look-up table performs classic piece-wise linearization. Based on the example of inverse NTC characteristic, new method is compared with polynomial approximation and standard piece-wise linearization.**

Keywords – **Piece-wise, sensors, approximation, linearization.**

I. Introduction

Non-linear transfer characteristic of the sensor is possible to compensate with many software methods in intelligent transducers: segment linearization ("piece-wise") [1] and [2], polynomial approximation on the one or more segments, or approximation by the rational functions [3]. This enables the use of strong non-linear sensor with stable characteristic. Speed and simplicity of the implementation are very important during the obtaining of the response of the linearization methods, till the linearization tables and needed coefficients are determined by the help of PC, with use of much more resources.

In the intelligent transducers processor power and memory for placing program, linearization table and temporal variables are limited [4]. This gives advantage to the methods which are appropriate to implement in mathematics with fixed point, i.e. piece-wise linear approximation versus polynomial.

In the estimation of the quality of the linearization method the following parameters should be considered:

- achieved accuracy of the linearization, like least mean squares deviation, or maximal deviation of the whole linearization range.
- time needed for response calculation in the device which does the linearization.
- used memory space as in EPROM for the linearization table or coefficients, so as in RAM for the saving of the temporal calculation results.
- the use of the program memory for the implementation of the linearization procedures.

II. Standard Piece-Wise Linearization

All segmental linearizations and all graphic generation with appropriate error, presented in this paper, have done with specific purpose developed program for PC. Standard Piece-wise linearization is carried out in such a way that the range of input variable is divided into desired number of segments and the optimal line found for each segment by the method of least mean squares deviation. In order to obtain the continual transmition function some of the linear segments are connected in such a way that for the ordinate value on the border of segments the mean value of the ordinates adjacent lines are taken. It can be noticed that for the line determination the criteria of least mean squares deviation on segment is chosen, and the methods of the linearization are compared on the bases of maximal deviation. As for the all considered examples that are gained the maximal error is a little bit bigger than that which is optimally possible. On the other side, when the points of calibration are gained by real measuring, which itself considers the existence of less or bigger uncertain of the measuring results, the method of the minimal square is very appropriate in practice.

Piece-wise linearization gives very good results for almost all transferable characteristics and it is also used when the output values depend on two variables, for example for the temperature compensation as influential value [2].

As the example of the transferable function with distinctively non-linearity, this paper considers inverse characteristic of the NTC (negative temperature coefficient) resistant sensor. One way of temperature measuring by NTC is termistor voltage measuring with constant current (Fig. 1). Result of the A/D conversion is proportioned to resistance which is

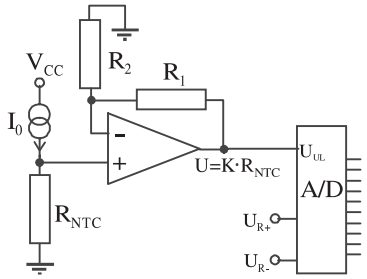


Fig. 1. NTC temperature sensor measuring circuit

[1]Dragan B. Živanović is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia, E-mail: dzile@elfak.ni.ac.yu

[2]Miodrag Z. Arsić is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia, E-mail: marsic@elfak.ni.ac.yu

[3]Jelena R. Djordjević is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia, E-mail: jelenadj@elfak.ni.ac.yu

[4]Ilija S. Mladenović is with the Faculty of Technology, University of Niš, Bul. Oslobodjenja 124, 16 000 Leskovac, Serbia

the function of temperature. In order to calculate the temperature for the bases of the conversion results it is needed to carry out the linearization function, which is inverted termistor NTC characteristic, i.e. function $T = f_T(R)$. In the paper temperature range from 0° C up to 100° C is considered. In order to use universal procedures in mathematics with fixed point, the input and output range of all functions are normalized to lie from 0 up to 65535. The R_N variable is proportioned to the A/D conversion result, Eq. 1:

$$R_N = k_2 \cdot U_{A/D} = k_2 \cdot k_1 \cdot R_{NTC} \,. \qquad (1)$$

The selection of the amplification, the range of the A/D converter, and conversion results averaging, in order to reject the noise, it can be adjusted that the R_N belongs to that range without additional scaling. After the linearization, final result of the temperature measurement is obtained by scaling according to formula $T_m[^\circ C] = aT + b$, which enables two points recalibration.

In the part of small resistances inverse characteristic of the NTC has very large slope and alteration of the slope. Classic linearization by 16 segments gives maximum approximation error about 10% (Fig. 2). It can be seen that for the linearization of the distinctively non-linear characteristics, the small error of linearization (for example ¡ 1%) demands classic segment linearization with lot of segments. Strong characteristic non-linearity is often present in the small part of the input range, and it can be reduced if the input range is divided on unequal segments. The input data for the application of the Piece-wise linearization is the current number of segments, i.e. the ordinal number of the point in the table and the rest of the input value in the segment itself. For equal segment size these data are easy and quickly found by integer divide of the input variable with the segmental size. But, the difference of the segment size causes slow linearization method response, i.e. the defining of the current segment must be done in the loop.

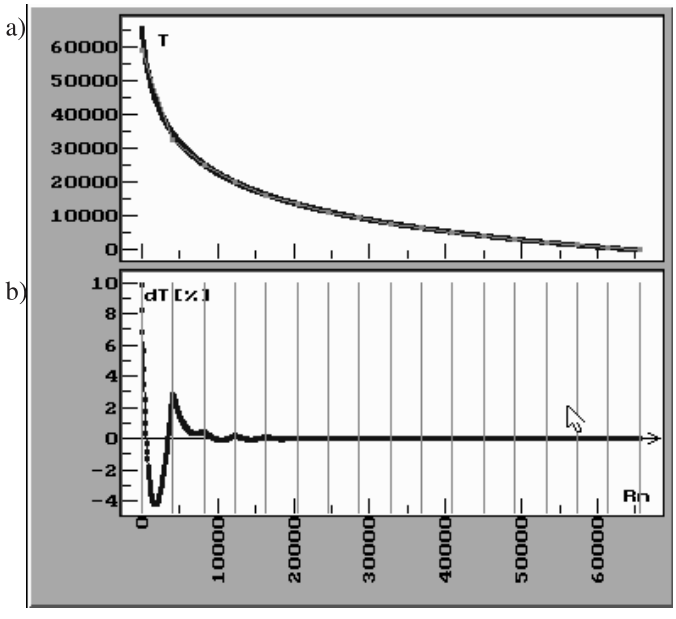


Fig. 2. Standard segment linearization

III. The Two Stage Piece-Wise Linearization Method

The method presented in this paper keeps simplicity and universality of the equal linear segment methods, and decreases total needed number of segments by linearization in two steps.

On the assumption that polynomial approximation is also implemented in the mathematics with fixed point, the price of this approach is the doubling of the calculation time. It can be compared with the second degree polynomial approximation on several segments. Concerning the memory space occupation, if we need 16 segments for the universal segment approximation, the accuracy will be compared with two-stage method of 8+8 segments.

The idea of the method is that the transformation of the x-axis is done before standard linear segment method application, so that the parts of the input range with strong nonlinearity are stretched, on account of the rest. The linear segment table also does this transformation. New characteristic in the function of transformed input variable is obtained, thus linear segment approximation can be performed with minor error than starting characteristic.

On the Figs. 3, 4 and 5 the example of the method by two times 8 segments is given. On the Fig. 3a function $f_2(z)$ is shown, i.e. temperature dependence of resistant with transformed abscissa, so that the starting part of the curve is stretched. After that, this function is approximated by 8 linear segments. The maximal error of this approximation of the whole range is 2,5% (Fig. 3b).

Input value is R. After the first transformation by linear segment table we get $z = f_1(R)$ (Fig. 4), and after the second transformation of z value, temperature is obtained as Eq. 2:

$$T = f_2(z) = f_2(f_1(R_N)) = f_T(R_N) \qquad (2)$$

Let's define f_1, first. The input range of f_T is divided to desired number of segments (n=8), and for each segment

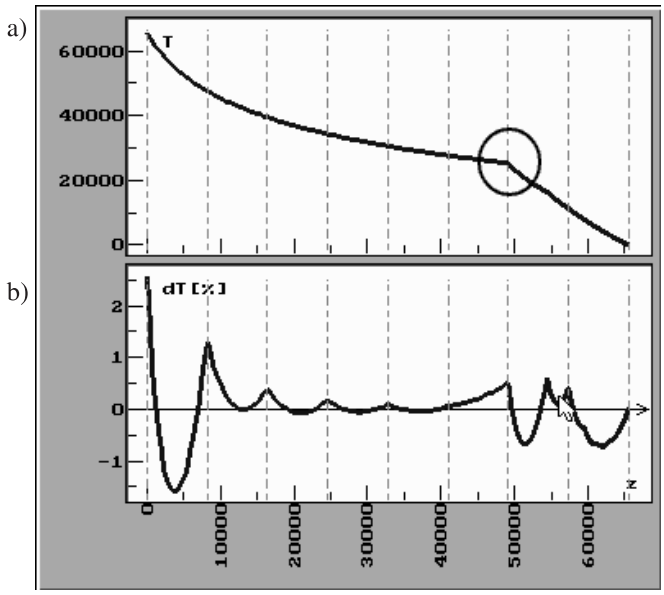


Fig. 3. Error of transformed curve piece-wise approximation

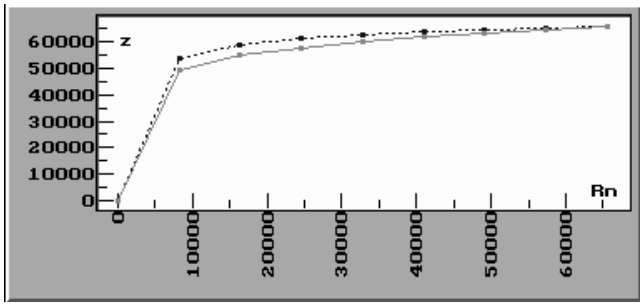


Fig. 4. The function which transforms R – axis

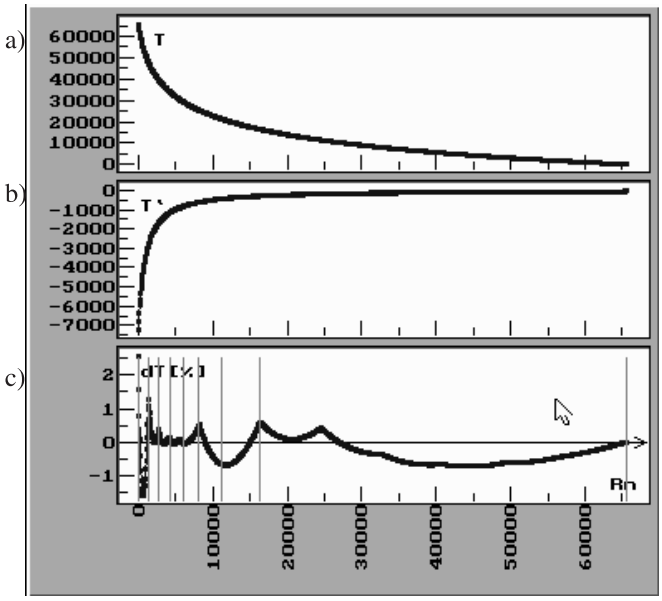


Fig. 5. Obtained error of presented method

apart defines maximal and minimal value of the derivation, and their difference, i.e. $\triangle T_i' = |T'_{i\max} - T'_{i\min}|$. Numerical obtained derivation of the $f_T(R_N)$ can be seen on Fig. 5b. Hence, the value F is determined by $F = \sum_{i=1}^{N} \triangle T_i'$. Knots of the first function $z = f_1(R)$ has ordinate according to Eq. 3:

$$z_0 = 0,\ z_i = z_{i-1} + \left[\frac{\triangle T_i'}{F} \cdot 65536\right] \tag{3}$$

The values on the R-axis are equidistant arranged as on Eq. 4 and Fig. 4 - a doted line.

$$R_i = i \cdot \left[\frac{65536}{n}\right] \tag{4}$$

Function f_1 is monotonous increased, which enables that one-valued of $f_2(z)$ is kept. It is achieved that the parts of the characteristic with the large second degree derivate stretch on the count of the rest of the part, which has been the aim of this transformation. The first Piece-wise linearization is stretched 6 times. The transformation of the input variable from the range of 0 up to 65535 to the same range is also obtained, which enables the use of the universal linearization procedure in intelligent transducer.

The Fig. 5a shows the function which is linearized (it's equal with the function on the Fig. 2a, the Fig. 5b represents the derivation of the function, and the Fig. 5c shows the linearization error. The error is obtained by the linearization function shown on the Fig. 3, but the same error is shown with linear change of the input value R_N. On the Fig. 3b the equidistant vertical lines represent the segment boundaries which respond to those on the Fig. 5c. The interspace between those segmental boundaries has been changed caused by transformation of the R_N - axis.

On the Fig. 3a it can be seen that after the first transformation smooth function $T = f_T(R)$ gets ridge points (circled area) so on them, caused by sudden change of the slope, valuable error of linearization can be expected. As to avoid this, the following procedure is applied: the ordinate values of the first table, which have large transformation in comparison with previous one, are rounded to values which correspond to abscissas of the second linearization table, to amount of $i \cdot \left[\frac{65536}{n}\right]$. In this way ridge points of the function $f_2(z)$ will coincide with the knots of the second linearization function, so they will not take in addicted error. Function $z = f_1(R)$ transformed by exposed method is shown in Fig. 4 – full line.

IV. Comparison to Classical Methods

Independent from the linearization methods, it is needed to consider the way of obtaining the input data, i.e. the way of the transfer function defining:

- by measuring of the pair of points in the calibration process, which is a long-lasting procedure for the great number of points and high accuracy.
- on the bases of the previous knowledge of the physics of sensor, known functional dependence is used, thus for the each concrete sensor several constants (at least the offset and the slope) has to be defined by several measurements.

In order to increase the accuracy in the real conditions, disregarding whether the transfer function is defined by the set of points or known mathematical formula, it is needed to carry out more measurements than the minimal needed for the coefficients defining. If the polynomial approximation or the approximation according to beforehand defined curve is continued, by the use of the mathematical programs of the general purpose (MCAD, ORIGIN...) the needed coefficients can be defined by the iterative methods, concerning as a criteria the minimization of the square error on the whole range of the input variable.

For the given instance of the NTC temperature sensor, direct dependency of the resistance from the temperature is known as Eq. 5:

$$R_{NTC} = R_{25} \cdot e^{\beta\left(\frac{1}{273.16+T} - \frac{1}{298.16}\right)} \tag{5}$$

On concrete sample of the sensor in the temperature chamber, by four wired measurement of the resistance by the use of multimeter HP3478A, the 20 pairs of calibration points have been obtained, on which base the determined coefficients R_{25} and β have been defined by the use of the ORIGIN

program. Afterwards, the inversion function is mathematically defined, and the input and the output ranges are linearly scaled to the range from 0 up to 65535. Based on such defined dependence, by the use of the ORIGIN program, the 1000 points of the function $T = f_3(R_N)$ has been generated. Those points are the input data for all linear and polynomial approximations which have been carried out, and whose accuracy is presented in this paper.

Standard polynomial approximation on the whole range of the input variable, even of the 9^{th} degree, gives the error larger than 2%, for the given example of the inverse NTC characteristic. In the Table 1 maximal error of the piece-wise polynomial approximation of the second and the third degree is given. As distinctively non-linearity is on the small part of the range, even second and the third degree polynomial, on the small number of the segments, does not give significantly better results.

Table 1. The maximal errors of the polynomial approximation

Number of segments	4	8	16	32
Second degree	16%	8%	3,5%	1,1%
Third degree	9%	4%	1,3%	0,3%

In the Table 2 results achieved by the standard segmental linearization are given. The results achieved by the suggested method (by the two tables with equal number of the segments) are also given in Table 2. By analyzing the required memory for the tables, or coefficients of polynomial approximation, classical segment linearization with 16 segments, should be compared with two-stage method with 8+8 segments, or approximation by the second-degree polynomial on 8 segments.

Table 2. The maximal errors of the standard segment approximation and suggested method

Number of segments	8	16	32	64
Classic segment	17,5%	10%	4,5%	1,75%
Number of segments	4+4	8+8	16+16	32+32
Two-stage method	12,5%	2,5%	0,7%	0,12%

For the given example of the distinctively non-linear function, the suggested method gives significantly better results than standard segment linearization, especially if the higher accuracy of linearization is required. It also gives better results than polynomial approximation, even the third degree, with easier implementation. The response of the linearization is obtain by doubled application of the same procedures, thus negligible larger program memory is required.

V. Conclusion

It is on disposal the whole range of software methods for the linearization of the transferable characteristics in intelligent transducers. However, for the linearization of the extremely non-linear characteristics, the accuracy in order of 1% is hardly achieved by the polynomial approximations and by the standard piece-wise linearization also. We should have in mind that the linearization has to be carried out by integer mathematics, or in fixed point, and that in the intelligent transducers the program memory, as well as the memory for the linearization tables saving, is very often limited.

Suggested two-stage method is easy to realize even with the limited processor recourses, and it gives less maximal linearization error in relation to the polynomial approximation and standard Piece-wise linearization. Program memory occupancy needed for the implementation of the linearization procedures is only insignificantly bigger than standard Piece-wise linearization.

References

[1] G. Horn, J. Huijising, *Integrated smart sensors, design and calibration, doctoral diss.*, Delft, Kluwer Academic Publisher, Netherlands, 1998.

[2] A. Flammini., D. Marioli, & A. Taroni, Application of an optimal look-up table to sensor data processing, *IEEE Trans. Instrum. Meas.*, 48 (4), 1999, 813-816.

[3] W. T. Bolk, A general digital linearising method for transducers, *J.Phys.E: Sci. Instrum.*, 18, 1985, 61-64.

[4] L. Barford, & Q. Li, Choosing designs of calibration transducer electronic data sheets, *HP Labs Technical Reports*, 1998, HPL-98-166, http://www.hpl.hp.com/techreports/98/HPL-98-166.pdf

[5] D. Živanović, M. Arsić "Universal Intelligent Sensor Module", *ETRAN, Conference Proceedings*, Vol. I, pp.121-123., Zlatibor, Serbia, 1995.

[6] D. Živanović, G. Djordjević, M. Arsić, "A Method Of Measurement Results Processing Definition For Intelligent Sensor Module", *IT'96, Conference Proceedings*, pp.468-471, Žabljak, Serbia,1996

Computer Based Remote Measuring and Acquisition of Dynamic Data

Aleksandar S. Peulić[1] and Sinisa S. Randjić[2], *Member, IEEE*

Abstract – **Target of this work is realization the PC based remote measuring and acquisition dynamic data of various processes, for example, car systems, processes in medical systems. The whole system can be operated as stand-alone as well as PC controlled. The simulations of dynamic data will be various sine waves in scale of frequency from 1 Hz to 20 Hz, from signal generator. The measured data will be collected by processor MSP430 TI family and transmitted by RF wireless line. All shapes of collected data will be showed.**

Keywords – **Dynamic data, car systems, transceiver, wireless line**

I. Introduction

The network in between the measuring modules and the base station is realized as a bi-directional multi-point, single master RF-link, operating in the LPD-frequency range (868 MHz) on a single channel. The structure of the network is fully dynamic and in operation re-configurable. The TRF6900 device integrates radio frequency (RF) with digital and analog technologies to form a frequency-agile transceiver for bi-directional RF data links. The TRF6900 device operates as an integrated-transceiver circuit for both the European (868-870 MHz), and the North American (902-928 MHz) ISM bands. This device is expressly designed for low-power applications over an operating voltage range of 2.2 V to 3.6 V, and is well suited for battery-powered operation. A key feature of the TRF6900 transceiver is the use of a direct digital synthesizer (DDS) to allow agile frequency setting with fine-frequency resolution.

II. Direct Digital Sinthesizser

The receiver uses single conversion, for use with either 10.7-MHz or 21.4 MHz IF filters. The TRF6900 supports frequency-shift keying (FSK)-modulated transmission or reception with bit rates up to 115.2 Kbps. The frequency reference, f_{ref} (which determines the frequency accuracy), is divided to set the step size. This step size is in turn multiplied up to a final output frequency. Characteristics such as operating-frequency range, step size, frequency accuracy, phase noise, switching time, and spurious-signal level are parameters that are balanced to yield a final design. The DDS-based synthesizer simplifies these design issues while maintaining the various performance requirements. The basic principle of operation of the DDS is to generate a signal in the digital domain and to reconstruct the waveform in the analog domain by D/A conversion. Adders and D-type flip-flops accomplish generation of the signal in the digital domain. The D-type flip-flops act as storage devices that change their state when clocked. All arithmetic operations are done using a modulo 2^N, where the N bits determine the output-frequency resolution (f_{ad}) at the phase detector as shown in equation (1).

$$f_{pd} = \frac{f_{ref}}{2^{24}} \tag{1}$$

Where f_{pd} is the minimum phase-detector input-frequency. This is the bit weight of the 2^0 bit of the DDS for the clock frequency f_{ref} used. The power in 2^{24} represents the number of registers of the DDS accumulator, which is 24 for the TRF6900. The value of f_{pd} is multiplied by N, the prescaler value (user-selectable values of 256 or 512), which yields a minimum frequency-step size $\triangle f$ as shown in equation (2) and (3).

$$\triangle f = N \cdot f_{pd} \tag{2}$$

$$\triangle f = N \cdot \frac{f_{ref}}{2^{24}} \tag{3}$$

As previously mentioned, generation of the signal in the digital domain begins with an accumulator whose output serves as a phase generator. Control inputs to the accumulator are a user-defined frequency word, and a reference clock used to clock the accumulator and other circuits (D/A, etc.). The accumulator output is a series of pulsed digital samples, spaced at the clock rate, in the form of a linear ramp. The slope of this ramped signal represents a phase, based on the user-defined inputs.

III. Communications Methods

In time division multiple access (TDMA), time slots differentiate users. The available radio spectrum is divided into time slots. Users can either transmit or receive in their dedicated time slot. Time slots for N number of users are collected into a periodic frame, with N time slots per frame. Because TDMA data is transmitted in bursts, transmission for a given user is not continuous. Temporal synchronization between a TDMA transmitter and receiver using time gating permits reception of a specific user's time-slot data, essentially turning the receiver on and off at the appropriate times. The users essentially "take turns" using the same radio channel. The composite signal indicates that the channel is in use

[1]Aleksandar S. Peulić is PhD student at Technical Faculty Cacak, Yugoslavia (e-mail: peulic@ptt.yu)

[2]Sinisa S. Randjić is with Technical Faculty, Cacak, 32000, Sv. Save 65, Yugoslavia (phone: +381-32-302-721, fax: +381-32-342-101, e-mail: rasin@tfc.kg.ac.yu)

for the complete length of time. TDMA is sometimes combined with time division duplex (TDD) or frequency division duplex (FDD). With TDD, half of the frame's time slots are used for transmission and half are used for reception. With FDD, different carrier frequencies are used for TDMA transmission and reception, even though similar time slots may be used for both functions.

TDMA systems using FDD usually add several time slots of delay to separate a particular user's transmission and reception time slots and to synchronize transmitters and receivers. The guard time between transmissions can be designed as small as the system's synchronization permits. (Guard times of 30 to 50 ms between time slots are common in practical TDMA systems). All mobile units must be synchronized to the base station to within a fraction of the guard time. Users of FDMA systems divide the frequency spectrum into narrow bandwidth channels where each user is assigned a specific frequency or frequencies for transmission and reception. The receiver demodulates information from the desired channel and rejects other signals nearby. Code-division multiple access (CDMA) systems are either direct-sequence spread spectrum (DSSS), which use orthogonal or un-correlated pseudorandom-noise (PN) codes to differentiate signals that overlap in both frequency and time, or frequency-hopping spread spectrum (FHSS), in which signals are randomly hopped about different portions of an available spectrum. In DSSS CDMA, a narrowband message signal is multiplied by a very large-bandwidth PN spreading signal with a chip rate that is orders of magnitude larger than the data rate of the message signal. (The chip period is the inverse of the spreading rate.) A large spreading rate can minimize the effects of multi path distortion, such as channel signal fading. Each user has a unique pseudorandom code. Due to the orthogonal of the codes, all other codes appear as noise to a given user. A matched filter extracts a specific user's signal, with little or no output signals resulting from other users. Armed with the proper code for that user, a CDMA receiver can effectively extract a user's signals from a channel with multiple signals at the same relative amplitude level. For CDMA systems to work effectively, the power levels of mobile units seen by the base-station receiver must be approximately equal. CDMA achieves this balance using dynamic power control techniques. The total power of multiple users at a receiver determines a CDMA system's noise floor after decorrelation. Unless the power of each user within a cell is controlled so that they are approximately equal at the base station receiver, the strongest received mobile signal can dominate the base-station's demodulator. Signals rising above the level of the other signals increase the noise level at the base station receiver. Higher-level signals decrease the probability that a desired signal will be received and decorrelated. As a result, CDMA systems employ power control at each base station to control the signal level received from each mobile unit, making them all approximately equal. In this way, a mobile unit close to the base station will not overpower the base station for a user much further away. Power control is implemented by rapidly sampling the received signal strength indicator (RSSI) level from each mobile unit and then sending a power change command over the forward radio link to each unit. Unfortunately, out-of-cell mobile units can still provide interference for the base station

IV. Experimental Results

The RF measuring network consists of two measuring modules and of one master module. Each measuring module has a TF6900 and microcontroller capable to make analog to digital conversions. The master acts in direct conjunction with PC by RS232, collecting information from RF link and sending to database by PC application. User action by PC application master translates to slaves by RF link. At the next pictures are showed simple application and transmitted sine wave from sine generator. Frequency band is form 1 Hz to 20 Hz and it mean that it is possible to transmit wide spectrum of measured signal, a temperature, a pressure, a EKG etc. The application is designed in Visual Basic environment and using demo version of graph module for graph representing of sine waves. The slave modules making digital scanning, for example two switches and sending information by RF link to master if switch is on or off.

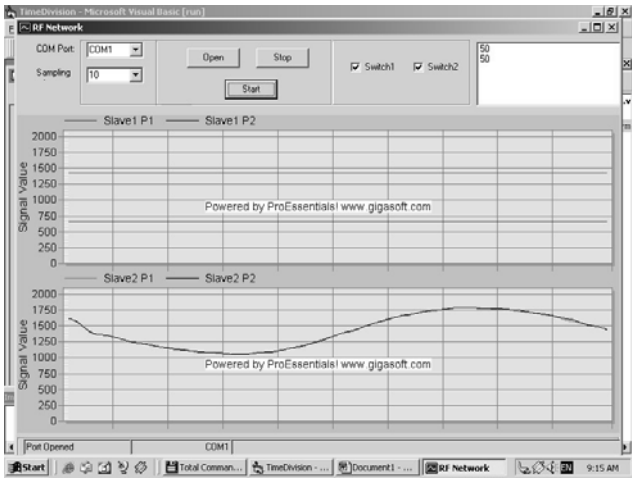


Fig. 1. The sine wave of 1 Hz

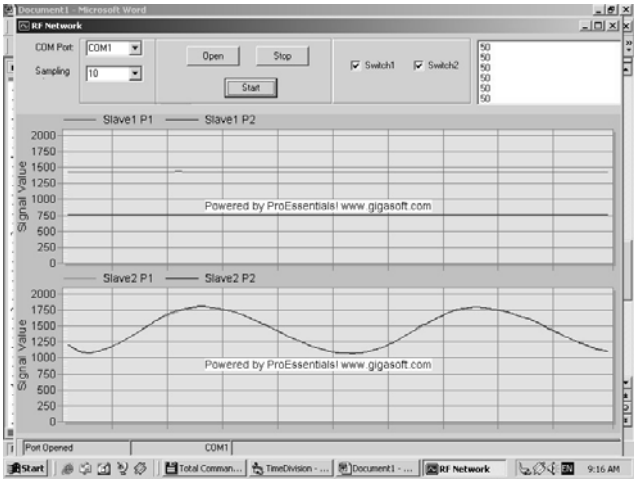


Fig. 2. The sine wave of 2 Hz

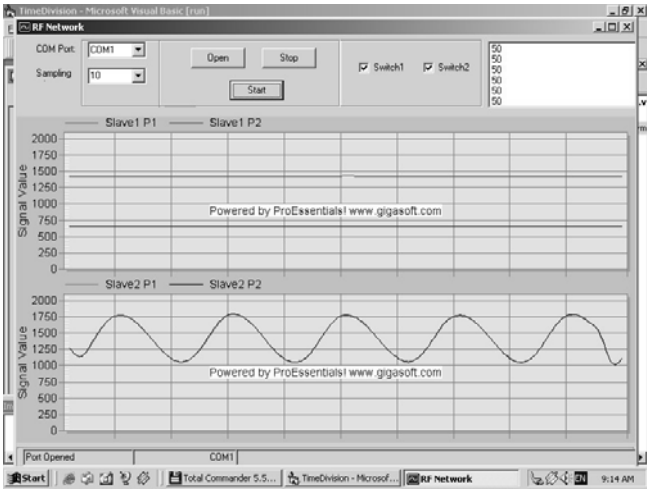


Fig. 3. The sine wave of 5 Hz

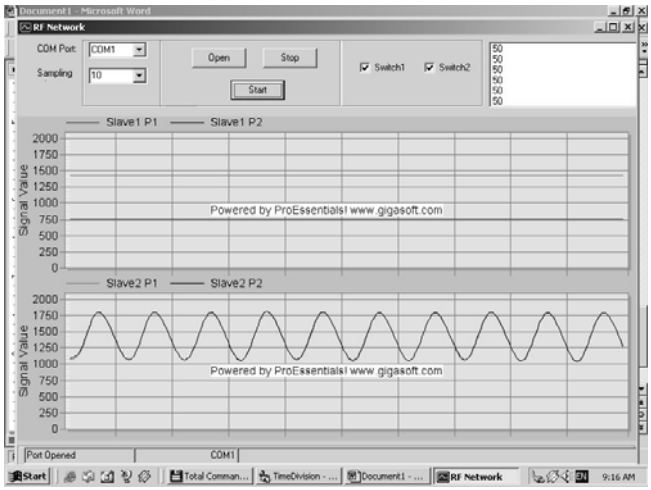


Fig. 4. The sine wave of 10 Hz

V. Conclusion

This application using new frequency band from 868 MHZ to 870 MHz and this band is getting flooded with applications of all kinds. The Rf link and communication between a computer and RF modules are realized. In praxis exist many kind of processes, which should be measuring. The final target is realization a stand alone base station with local data storage will be connected by internet to one big Internet based control, monitoring and operation system.

References

[1] Mr A.Peulić, Dr S.Randjić, *Radio Frequency Measured Data Transmit*, ICEST 01.-04. October 2002, Nis, Yugoslavia, CD zbornik, Measurement Technique

[2] P.J.Gawthrop, *Symbolic Algebra and Phisicakl Model Based Controll*, Report CCSC 96-00,june 1996

[3] P.Spevak, *Impelmenting a bi-directional FSK link*,july 2000

[4] H.J.Vaughan-Nichols, *Web Services: Beyind the Hype*,Computer, february 2002

[5] Hai Lin, *Internet Based Monitoring and Controls for HVAC*, Jan/Feb 2002.

High Resolution Virtual Absolute Encoders

Dragan B. Denić[1] and Goran S. Miljković[2]

Abstract – **Absolute encoders are well known electromechanical components of all control systems for positioning. This paper considers virtual absolute encoders as a new type of absolute encoders that is very up to date. The possibility of applying developed methods for serial code reading of chain codes to high resolution virtual absolute encoders is considered. The solution that eliminates problems regarding the serial code reading with two detectors is proposed.**

Keywords – **Position measurement, pseudorandom code, optical encoders.**

I. Introduction

The absolute encoders are well - known electromechanical components. As a main part of all control systems for positioning, they provide measured information about sensor head position (detector) related to the measurement scale. Considering that each position is coded, the momentary one is defined apart from the previous position. This is the basic quality of the absolute encoders and hence proceeds their main feature that after the power is turned on, an information about the current position of the movable system is instantly obtained. There is no need for any initial moving. For the purpose of angle movement detection, the measurement scale is realized using a disc with concentric tracks, which provides n - bit code word for each discreet angle position. Reading of these circular code tracks is done using a sensor array, where each single sensor serves for reading of one code track and it provides an output signal that represents one bit. Thus, n - bit output code, which represents momentary position of the movable system, is obtained at the output of this sensor array. Code tracks often consist of segments which can be optically detected using transmission or reflection methods. Also, code tracks could consist of segments which can be detected using magnet capacitive or inductive methods. Thus, depending on the applied method of code tracks bit detection, the encoders are divided into optical, inductive, capacitive and magnet encoders. In any case, high resolutions of position measurement are achieved by increasing the number of code tracks, and that way providing a higher number of output code bits.

Virtual absolute encoders represent a new type of absolute encoders that are a result of tendency of avoiding a use of large number of code tracks which is typical in case of high-resolution absolute encoders. This is achieved by using cyclic or serial codes, which possess a feature that two n - bit code words, which correspond to two consecutive positions, comprehend an identical sequence of (n-1) bits. In other words, the last (n-1) bits of the current code word, (meaning, all bits besides the first one), are equivalent to the first (n-1) bits of the subsequent code word. A possibility of overlapping of the records of all 2^n code words on one code word is evident, [2]. To begin with, such encoder has an enormous advantage and it does not only solve a problem of increasing the number of code tracks with increasing of the resolution, yet it always has only one code track regardless of the resolution. Since nothing is ideal neither are the absolute encoders; we would still need n sensors for the instant reading of n - output code bits, one for each output code bit. Much bigger problem is that distance between sensor heads changes with the change of resolution. A technical problem of allocation of n sensor heads within that small physical area could also occur in case of high measurement resolutions.

Fortunately, cyclic code features provide a new way of reading of the code bits using only one detector, [3]. This method of serial code reading implies collecting of code bits into a shift register used for code forming. Only one bit is being read for each new position of the movable system and entered into the mentioned shift register. After the initial movement that corresponds to space width of n bits, forming of the code word which corresponds to the current position of the movable system will be executed. For each of the following positions a new bit is being read, and along with (n-1) bits of the previous code word, an output code word of a new position is obtained. This new type of absolute encoders possesses all the features of conventional ones, except one. That is the necessity of initial moving after the first plugging in/out. In those cases, it is necessary that the movable system (MS) crosses a distance equivalent to space width of n code bits, so that the first valid output n - bit code could be formed. This is the reason that these absolute encoders are called virtual absolute encoders. In the case of high resolution encoders, mentioned distance of initial movement is very small. However, this is still a virtual absolute encoders disadvantage. This disadvantage is rather attenuated and becomes almost negligible in relation to a new quality that is provided by virtual absolute encoder. It is evident that in case of high resolution virtual absolute encoders one of the most interesting moments is code reading. In this paper, methods for serial code reading are considered and a new approach in realization of virtual absolute encoders is suggested.

[1]Dragan B. Denić is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: ddenic@elfak.ni.ac.yu

[2]Goran S. Miljković is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: goranm@elfak.ni.ac.yu

II. Serial Code Reading in Case of Virtual Absolute Encoders

In order to explain the method of serial code reading, a concrete simple example of virtual absolute encoder will be considered. A rotary disc consists of two tracks, Fig. 1. Let us consider that these two tracks consist of transparent and non - transparent segments. Also, let us consider that appropriate optical methods for detection are applied. Interior track is identical to the one of incremental encoder, [4], and it is used in this example for generating of two bits in the output code with the smallest weight. Its main role is providing synchronized code reading and it is often called synchronization or "tact" or time track, [5]. In this simple example which considers 5 - bit binary encoder, it is adopted that the space - time width of one incremental cycle is equivalent to the space width of one code track bit. Otherwise, that ratio can change. External track, a code track, is coded in a way to provide residual important bits needed for forming of the complete absolute output code word. Applied cyclic code, named a shift register code [1], provides a unique code word for each new position of the encoder, which alludes reading of a new bit from the code track.

For the purpose of obtaining the output position code, three detectors are being used. Serial bit reading from the code track is done by detector x(0). Obtaining of the signal from the synchronized track is realized using two detectors, as in case of conventional incremental encoder, [4].

In this example, classic quadrature signals are required (two sine signals dislocated by 90°), because two additional bits are planned which would magnify the position measurement resolution four times. These two signals are also used for determining of the encoder disc rotation direction. These signals are then shaped into rectangular signals, and whenever a transition of signal A (with signal B on logical "0") is detected, reading of a new code bit is being performed. In order to entirely explain a principle of serial code reading, an example of realization of electronic block of this virtual absolute encoder is shown in Fig. 2.

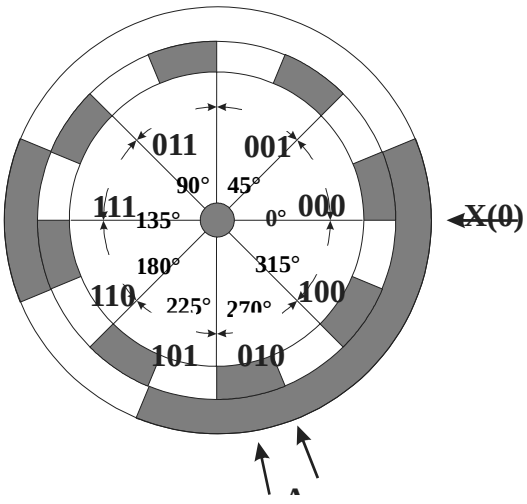


Fig. 1. Virtual absolute encoder disc

Light source, LED diode, for synchronization track is always actuated and it illuminates two detectors forming quadrature signals A and B at the comparator outputs C_1 and C_2. As said before, code reading is done whenever a transition of the signal A (with signal B on logical "0") takes place. Because of this, signal A goes to the input of a signal edge detection circuit, and then, the output signal of this circuit along with the signal B complement to the input of the AND circuit I_1. Whenever an impulse at the logical AND circuit output appears, a new bit reading is done. A simple realization of the signal edge detector is presented here. Rectangular signal A from the comparator output C_1 is brought to both inputs of the same EXOR circuit E_1, but with small delay at one of the inputs. In this case, the delay is generated using integrator in the form of RC circuit.

Whenever an impulse appears at the transistor T base, it leads, whereby the LED diode which illuminates code track is excited. Considering that at that moment the code track detector is located at the middle of the sector that defines current code track bit, reliable reading of that bit can be done. A logical value of the read bit is located at the comparator output C_3. That bit is brought to the appropriate shift register input depending on the disc rotation direction. Considering that impulses at the signal edge detector output always appear at the moment immediately after the detected transition on the synchronization track, then, based on the logical value of the quadrature signal A, encoder rotation direction is being determined, Fig. 1. If A equals 1 when the impulse appears, then the rotation direction is clockwise (CW). Then, an impulse appears at the output of the logical AND circuit I_4, shift register shifts to the left and newly read bit is accepted at the appropriate shift register input. In case of reverse encoder disc rotation, an impulse occurs at the output of the logical AND circuit I_5. After the initial movement of (n-2) bits at same direction, a correct code word is formed and there is a valid information about position at the output. It is obvious, that it is necessary to preconvert a cyclic code at the shift register output into a desired output code, usually into the natural binary code. There are few known methods for code conversion, [6, 7], of which the one named parallel conversion method that uses table memory located in PROM, is applied here. At the end, two bits of the smallest weight are obtained at the quadrature detector output, which consists of one EXOR circuit and one logical NOT circuit.

Basic reason for using of the impulse stimulus of the LED diode, is that there is one gap for bit reading, in contrast to conventional incremental optical encoders where a number of gaps is used for fine tracks observing (multiple-line slits). Impulse stimulus allows greater pick values for current, whereby greater momentary illumination is achieved and thus, a probability of amplitude loss due to usage of only one gap (single line slit), is reduced. Only in case of low resolution measurements when gap is width enough to provide enough signal amplitude from photodetector, DC activating of LED diode is possible.

Illustrated example in a simple way presents the manner in which the virtual absolute encoder functions, using only one detector x(0) for serial code reading. However, this method

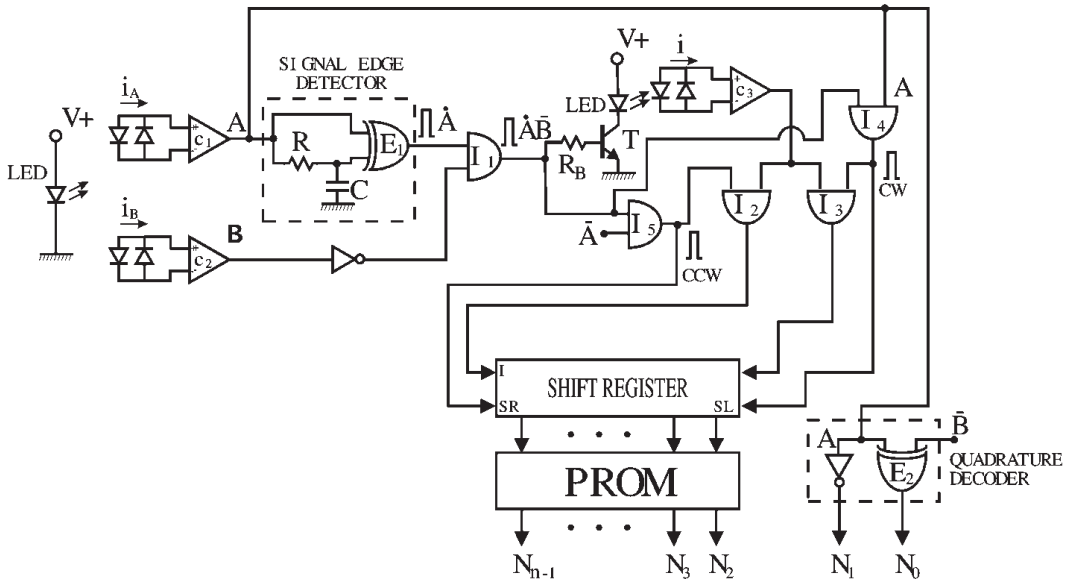


Fig. 2. An example of electronic block realization of virtual absolute encoder

for code reading meets a problem of loosing position information after each change of encoder disc rotation direction. After each direction change, initial movement is needed for obtaining valid information at the output of the encoder electronic block. This disadvantage could be resolved by various methods of additional operations after each change of direction, such as additional movement of shift register content, etc. If they do not cause significant performance violation of the encoder system, such solutions would certainly make it much more complex. A new method for code reading, adopted to high resolution encoders, is suggested here [8].

III. A New Method for Code Reading

The simplest way of solving the problem of loosing position information after each change of rotation direction is by introducing one more detector at distance of (n-2)q, where q is code track quantization step [8]. Using simple logic, consisted of two AND and one OR circuit [8], selection of one, out of two available heads for code reading, is done.. When moving to the left, code track bits read by head x(n-2) are accepted, and when moving to the right the ones read by sensor head x(0). In that way, code words formed after direction change correspond to current positions of the movable system. A circuit for error protection which can occur when MS changes the direction of movement, described in reference [3], is no longer needed. Most importantly, a continuity of code word forming is now achieved even in cases of possible MS oscillating in the direction of movement. In case of systems where there are potential oscillations, use of suggested method for code reading is fairly reasonable. Suggested arrangement of sensor heads additionally annihilates need for a correction element in form of parallel adder, because of the elimination of systematical errors made during code reading.

Besides it simply solves problems of serial code reading, this method, at the same time, provides continuous review of accuracy of formed code words [8,9]. Such method certainly detects real possible errors in code reading, [9]. Thus, using this high - quality error detector, virtual absolute encoder provides an output position information which is incomparably more reliable then the one obtained by any conventional absolute encoder. Because of its importance, this new method is cited and commented in the VI chapter of the newest edition of ”Measurement, instrumentation and sensors handbook” [10]. It was noticed that this new method is not appropriate for use in case of high resolution rotary encoders, since this solution was developed for positioning of flexible movable systems with large range of moving. Application of this solution in the field of micropositioning is critical because two sensors are located at small distance on the same armature. Because of that, different influences (like temperature or vibrations) can cause variation of space distance between these two sensors, which could cause errors in code reading in case of very high resolutions. Thus, a modified method is suggested for use in high resolution measurements. That could be achieved by introducing one additional code track, that would include the same code as the previous one, but it would be dislocated, revolved, for n-2 bits. This way, each of two code tracks would have their own code bit detector and they would be locating in the same line as in case of conventional absolute encoders. A module for code reading from the disc would be same for different encoder discs with different resolutions. Namely, all problems tied to the method of serial code reading using two detectors (two sensor heads) are this way eliminated. At the same time, all good features of this method are retained. Of course, price is introducing one more code track, which is negligible comparing to what is obtained using this method. Continuity in code word forming in case of multiple direction changes, extremely simple realization of encoder electronic block, a possibility of realizing new encoder functions, magnifying of system redundancy because of the possibility of working even if one of detectors breaks down, and simple realization

of high - quality error detector, are obtained. The last possibility would represent special, original quality of virtual absolute encoders and is itself more than enough of excuse for adding one more code track and using the suggested method for serial code reading.

IV. Conclusion

The virtual absolute encoders are momentary the greatest hit, as something new with entirely new quality. They are especially interesting because they own great number of possibilities for further upgrading of their performance. Their price is less than the one of conventional absolute encoders, in return of great new quality. That is magnifying of the system reliability and possibility of providing additional information to user about validity of output measured information. In contrast to pseudorandom encoders [2], virtual absolute encoders obligatory include one of methods for serial code reading using cyclic code. Modification of already developed new method for serial code reading [8] for its fully application in high resolution virtual absolute encoders, is proposed here.. The only price we need to pay for this new approach is introducing one additional code track.. Fortunately, that is not that big of a problem considering that the number of code tracks in virtual absolute encoder is fixed, apart from used measurement resolution. Two code tracks compared to 16 or more in case of conventional absolute encoders are entirely acceptable variant. This way, suggested method for code reading would significantly increase quality of virtual absolute encoders. It is quite enough to mention that high - quality error detector could be then directly applied [9] and that it would certainly detect each possible error during reading and forming of the code word.

References

[1] MacWilliams, F.J., Slone, N.J.A., ”Pseudo-random sequences and arrays”, *Proceeding of IEEE*, Vol. 64, No. 12, pp. 1715-1728, December 1976.

[2] E.M. Petriu, ”Absolute-type position transducers using a pseudorandom encoding”, *IEEE Trans. Instrum. and Meas.*, Vol. IM-36, No. 4, pp. 950-955, December 1987.

[3] E.M.Petriu, J.S. Basran, ”On the position measurement of automated guided vehicles using pseudorandom encoding”, *IEEE Trans. Instrum. and Meas.*, Vol. 38, No. 3, pp. 799-803, June 1989.

[4] T. Wigmore, ”Optical shaft encoder from SHARP”, *Elektor Electronics*, pp. 60-62, July/August, 1989.

[5] E.M. Petriu, J.S. Basran, F.C.A. Groen, ”Automated guided vehicle position recovery”, *IEEE Trans. Instrum. and Meas.*, Vol. 39, No. 1, pp. 254-258, February 1990.

[6] E.M. Petriu, ”New pseudorandom/natural code conversion method”, *Electronics Letters*, Vol. 24, No. 22, pp. 1358-1359, 1988.

[7] M. Arsić, D. Denić, ”Konvertor koda pseudosluajni/prirodni primenjen kod pozicionih enkodera”, *ETRAN, Ser. Elektronika*, str. 164-167, Jun 1995.

[8] M. Arsić, D. Denić, ”New pseudorandom code reading method applied to position encoders”, *Electronics letters*, Vol. 29, No. 10, pp. 893 - 894, 1993.

[9] D. Denić, M. Arsić, ”Checking of pseudorandom code reading correctness”, *Electronics letters*, Vol. 29, No. 21, pp. 1843 - 1844, 1993.

[10] John G. Webster: ”The measurement, instrumentation and sensors handbook”, CRC Press and IEEE Press, 1999.

Absolute Position Measurement Using the Method of Pseudorandom Code Parallel Reading

Dragan B. Denić[1], Ivana S. Randjelović[2] and Miodrag Z. Arsić[3]

Abstract – **Possibilities for digital measurement of absolute position using only one code track are considered in this paper. It is provided using an especial code technique based on pseudorandom binary sequences. Problems in the field of micropositioning are considered and a concrete solution is proposed. A method for parallel reading of pseudorandom code using linear photodetector array is applied.**

Keywords – **Position measurement, pseudorandom encoder, pseudorandom sequence.**

I. Introduction

One good alternative to the classical absolute encoder is the absolute encoder that uses the longitudinal code technique. The absolute position is coded by applying one code track with a pseudorandom binary sequence so that each group of n successive bits represents a unique code word. Two successive code words are overlapping and they differ only in one bit. Summary of basic problems in realization of pseudorandom encoders is given in [5]. Basic dilemma is about the application of the parallel code reading method or the method of serial pseudorandom code reading, Fig. 1. In the reference [6], a solution of parallel code reading is given in case of an application of the pseudorandom encoder with large measurement scale and with a relatively low measurement resolution. Application of n sensor heads for reading of n bit code words is problematic when dealing with high measurement resolution. In that case we use a linear integrated photodetector array [1] or CCD cameras. Disadvantage of such an approach is a significant time of signal processing at the output of these complex detection circuits and that is why these solutions are unacceptable for commercial encoders of high resolutions. Serial pseudorandom code reading [8,9] simplifies pseudorandom encoder and enables greater measurement resolution, but at the same time owns a disadvantage. Pseudorandom encoder with a serial code reading requires small initial movement for the first determination of the position after the power is turned on. For many applications and also for the commercial optical encoders this is not a problem, but this solution still does not represent a real absolute encoder. It owns all good features of an absolute encoder, it is almost as simple as an incremental encoder, very reliable and it has recently appeared on the market and represents a real hit. It is called a virtual absolute encoder because of the induced disadvantage above. However, there are applications where requiring of an initial movement after power is turned on is not desirable for determining the first absolute position. There are some special applications where it is impossible to demand such initial movement, such as systems for level measurement by using digital position converters. Thus it is very interesting to consider the possibilities of realization of the real high-resolution absolute pseudorandom encoder using a method of parallel code reading. Because of a great need for such a solution, in this paper, a possibility of algorithm solution acceleration using linear integrated photodetector array for parallel pseudorandom code reading, will be considered.

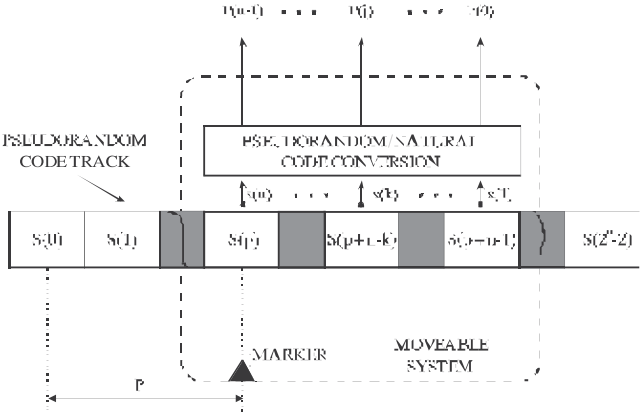


Fig. 1. Absolute position p is entirely identified by the pseudorandom code word $x(k)/k = n, ..., 1$

II. Pseudorandom Code and Its Usage in Encoders

The method of pseudorandom encoding, which for absolute position determining requires only one code track, represents an attractive alternative to the classic measurement method. Its advantages are significant in case of high-resolution position encoders and linear position encoders with very long code tracks. A coding is based on the "window property" of $PRBS_s$ $\{S(p)/p = 0, 1, ..., 2^n - 2\}$. According to this, any n - bit code word $\{S(p + n - k)/k = n, ..., 1\}$ provided by a window $\{x(k)/k = n, ..., 1\}$, of width n scanning the $PSBS_s$, is unique and may fully identify windows absolute

[1]Dragan B. Denić is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: ddenic@elfak.ni.ac.yu

[2]Ivana S. Randjelović is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: rivana@elfak.ni.ac.yu

[3]Miodrag Z. Arsić is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: marsic@elfak.ni.ac.yu

position p relative to the beginning of sequence, Fig. 1.

As shown on Fig. 1, code words are now arraying linearly or longitudinally (but not transversely like in case of classic coding), and they overlap. The first $(n-1)$ bits of such code word are identical to the last $(n-1)$ ones of the previous code word. Therefore, in contrast to transverse coding technique, which requires writing of a definite digital code in the transverse direction for each sector of code device, this method enables coding with one bit per sector.

III. Solutions Based of the Linear Photodetector Array

Commercial integrated photodetector arrays are available on the market with different intervals between photodetectors. Those intervals are 13 μm, 10 μm, 7 μm and smaller. Number of photodetectors can be few thousands, thus few tenths of sensors are used for one bit reading. Because of a exactly defined interval between two detectors, it is possible to use large number of photodetectors in order to increase a precision of absolute position measurement. A solution presenting basic principles of so far known solutions for parallel reading using linear photodetectors array is shown in references [1], [11]. A code track of transparent type is applied, so that there is a light source at one side and at other side is an integrated circuit, which consists of a linear photodetector array. This integrated circuit often offers a possibility of serial reading of photodetectors output signal. In case of using circuits with k photodetectors, after k tact pulses a read state is put into a shift register. Let us read $(n+c)$ bits and let nominal value of photodetectors number per one code bit be m. n is a number of bits needed for detecting the absolute position. Usually, at least one additional bit is read, and generally c additional bits. This way system redundancy is increased and use of some methods for code reading error detection is enabled. A condition $k \geq (n+c)m$ is always accomplished. Read code word is in the following form $\{00000111...110000111...1100....\}$.

A transition is detected at the border of two elements. In the ideal case, a number of consecutive ”1” or ”0” per one bit equals m. But, a deviation may occur due to not ideal drift of code elements on code track. After the reading of total output code, its conversion into natural code is done. A generator of used PSBS_s starts from the code word that corresponds to initial ”0” position, on the code track. Generator core is a shift register with appropriate feedback. With each tact that conduces generating of next PSBS byte, m tact from the shift register are performed. A code identical to the one that would be read in case of continuous MS movement from the position ”0”, is obtained that way. In an ideal case, after a certain number of register shifts, a code word identical to the read one would be obtained. That correspondence could be simply detected by digital comparator circuit. A number of steps counted by a counter until the moment of correspondence represents output position information in the natural code. Unfortunately, as said before, an error would often occur in practice, and it is enough that one bit is read from $(m+1)$ detectors and the PSBS_s generator will not generate such code word, thus, a correspondence will not be detected. This is why a digital correlator is used, although it is a much complex solution, but it solves the problem. Accuracy of the detected code is increased introducing a greater value for c. Accuracy of correlator output does not depend much on the accuracy of defining boundary locations of code elements.

In the reference [11], a possible realization is discussed. A classic pseudorandom/natural code conversion is done using a simple digital comparator. Additional fine position could be defined based on the detected, defined transitions. Using this procedure, measurement time would be reduced, which is still not good enough for general-purpose encoders. Further in the paper, this idea will be elaborated and with some additional modifications, we will attempt to realize a solution where rough position for each new code reading is not defined.

IV. Complete Solution Algorithm without Using Permanent Code Reading During the MS Movement

A realization of the electronic encoder block using discrete electronic circuits is shown in Fig. 2. Of course, the same function can be done using microprocessors and the appropriate software. Also, a realization using one of modern programmable logical circuits is possible. An example of a possible realization algorithm of a new solution proposed here is shown in Fig. 3. The basic idea is to exclude the digital correlator, separately from all the elements connected for the purpose of achieving any correlation function. Although digital correlators are well known and commercially available, in this case, they lead to a more complex system. In reference [11] it is pointed to a fact that with software realiza-

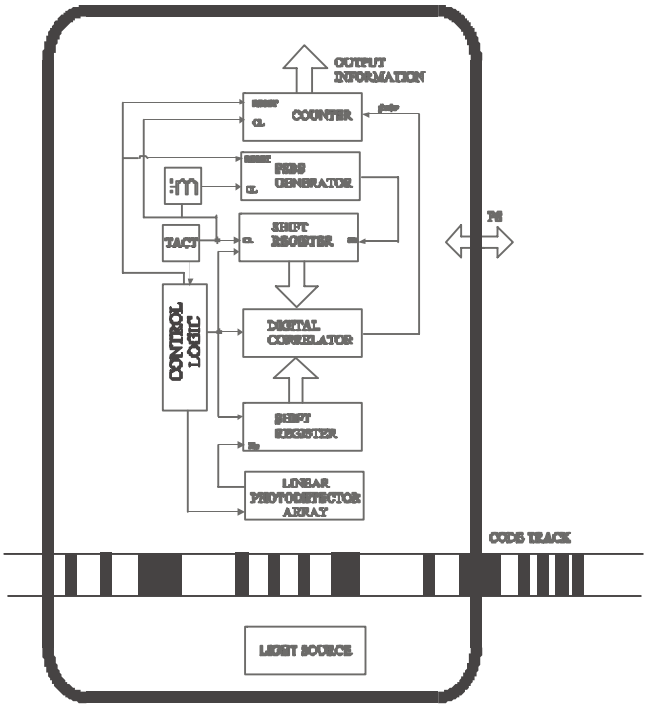


Fig. 2. High-resolution pseudorandom encoder with applied method of parallel code reading

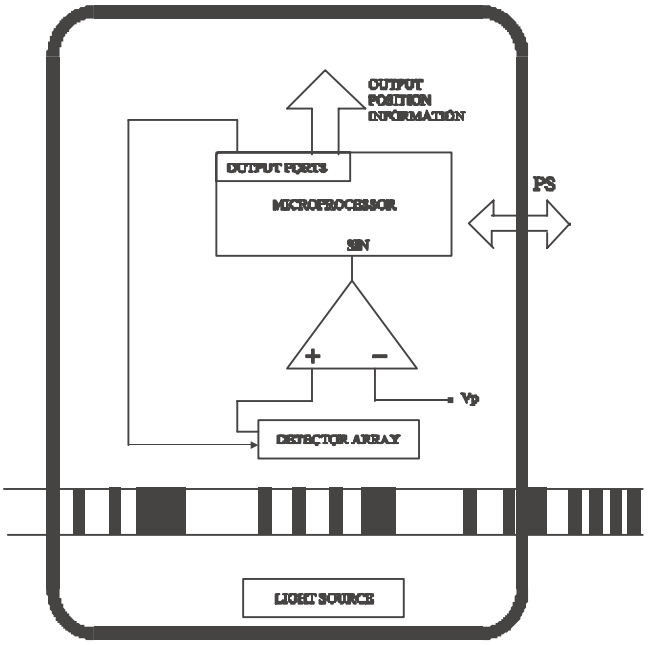


Fig. 3. An example of an encoder realization using a microprocessor

tion, simpler techniques can be used rather than the accurate and precise mathematical correlative ones. Our goal is to achieve shorter time for PSBS reconstruction read from the track, and then use of the well-known algorithm of pseudorandom/natural code conversion. The other idea is not to determine the rough position constantly, but only for certain values of the fine position. The fine position is defined by a number of rightmost photodetectors that read a bit that is not entirely included by the used circuit of photodetectors. This way, without using any arithmetical operations, a momentary position of the movable system is directly obtained with a fine position resolution defined by the distance of two consecutive photodetectors in the photodetector array circuit. For example, for the used 10 - bit PSBS and for m=64 (64 photodetectors per one bit), 16 - bit output position information is obtained by means of direct linking of 10 - bit code conversion result and 6 bits of smaller weight that define a number of photodetectors which detect next, not quite visible bit.

As it is shown in Fig. 3, a microprocessor defines the operation of photodetector array circuit using control signals. A reading of recorded PSBS sequence is done in a way that series of voltage levels, which indicate a light intensity on the corresponding elements of photodetector array, is lead to the input of a comparator. That way, with a defined voltage level Vp, this series of voltage levels is converted to a series of logical ”0” and ”1” at the comparator output, that is then lead to a serial microprocessor port. Afterwards, the microprocessor does the functions according to the software whose algorithm is shown in Fig. 4.

This way obtained bits are memorized and then analyzed. Memorized binary code word is in the form of $\{...000111...1111000...\}$, considering that m consecutive detectors are used for reading of one bit. However, as it is previously mentioned, it is possible that some deviations occur - $\{...0001011...1101000...\}$. Detector ordinals are then determined. Since nominal period (a number of detectors per one bit) is known, a difference of two consecutive detector ordinals is approximately dm, $d = \{1, 2, ..., n\}$. All deviations greater then this one can be updated, but the periods between the transitions will be different afterwards. There are other reasons why they will be different from nominal period, but those deviations must not be great.

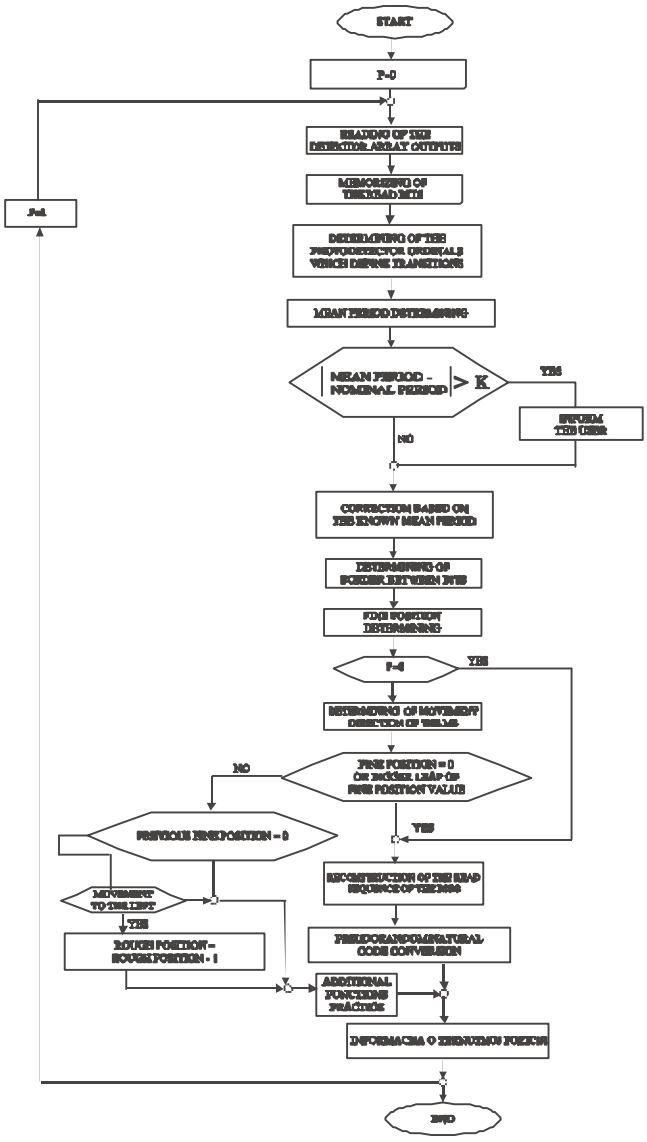


Fig. 4. Algorithm of the proposed solution

In any case, a calculation of period mean value is proposed here. In the ideal case, obtained mean period (meaning the integer part of the mean value) will be equal to the nominal period. In reality, there will be deviations, and if they are greater than the proposed value k, the user should be informed. The system can track down those deviations and perform diagnostics in part of additional functions. Based on the estimated mean period, certain corrections are eventually performed and the detector ordinals are determined. Fine position is now simply evaluated as a number of detectors that are outside of the last determined transition. The MS move-

ment direction is determined based of the previous value of the fine position. When the boundary between bits is known, it is easy to perform the reconstruction of the read part of PSBS and then the pseudorandom/natural code conversion. The value of the rough position is that way determined. Since the rough position has the same values for a great number of simultaneous code readings, there is no need for permanent performing of this, according to the necessary time, the most important part of the algorithm. It is being performed only when fine position equals ”0”, or when higher fine position value leap is detected. The latter ensures working in case of leap of the output position whose fine position equals 0. According to a well - know problem of necessary position correction in case of parallel pseudorandom code reading [6], rough position value is decreased by one in case of moving to the left.

All the additional functions, such as error detection, error diagnosing, code track status recording (location and size of dirt), are performed permanently, except in case when new rough position value is determined. Also, with the first passing through the algorithm after the turn on (parameter p equals ”0”), the rough position is immediately determined regardless to the fine position value. In the case of already considered example n=17, m=10 and the overall resolution of 1 μm, required time for execution is now at least 10 times less. In case of m taking considerably higher values respected to the pseudorandom code longitude (for instance, n=10, m=64), that acceleration is substantially higher.

V. Conclusion

The presented solution considerably decreases the basic problem, which is a substantial time needed for the execution of the position measuring. The process of determining the fine and rough position is being divided based on the recorded pseudorandom code, and now, all known pseudorandom/natural code conversions can be directly applied, which results in 10 times higher acceleration in relation to the already known solutions and methods in this field. What is proposed here is occasional determining of the rough position, which gives space for realization of many additional functions without increasing the maximal time for the execution of the algorithm.

References

[1] J.T.M Stevenson and J.R. Jordan, *Absolute position measurement using optical detection of code patterns*, J. Phys E. Sci. Instrum. 21, 1140-1145, 1988.

[2] Whitwell A.L., ”More techniques ensure unerring positional control”, *Design Engng November* 45-8, 1973.

[3] Jones B.E. and K. Zia, ”Digital Displacement transducer Using Pseudorandom Binary Sequences and a Microprocessor”, *IMEKO/IRAC*, Symposium Proceedings, London, Nov. 1980, pp 368-379.

[4] Arsić M., Denić D., ”Position measurements based upon the application of pseudorandom encoding”, *Facta Universitatis, Ser. Electronics and Energetics*, No. 6, pp. 13-23, 1993.

[5] John G. Webster, ”The measurement, instrumentation and sensors handbook”, *CRC Press and IEEEb Press*, 1999.

[6] Petriu E.M., Basran J.S., ”On the position measurement of automated guided vehicles using pseudorandom encoding”, *IEEE Trans. Instrum. and Meas.*, Vol. 38, No. 3, pp. 799-803, June 1989.

[7] Khalfallah H., Petriu E.M., Groen F.C.A., ”Visual position recovery for an automated guided vehicle”, *IEEE Trans. Instrum. and Meas.*, Vol. 41, No. 6, pp. 906-910, December 1992.

[8] J. N. Ross and P.A. Taylor, ”Incremental digital position encoder with error detection and correction”, *Electronics letters*, Vol. 25, 1436-1437, 1989.

[9] Arsić M., Denić D., ”New pseudorandom code reading method applied to position encoders”, *Electronics letters*, Vol. 29, No. 10, pp. 893 - 894, 1993.

[10] Denić D., Arsić M., ”Checking of pseudorandom code reading correctness”, *Electronics letters*, Vol. 29, No. 21, pp. 1843 - 1844, 1993.

[11] Johnston J. S., ”Position measuring apparatus”, UK Patent Application no GB 2126 444A.

Managing Calibration Confidence in Calibration Process

Vladan S. Djurić[1], Božidar R. Dimitrijević[2] and Ivana S. Randjelović[3]

Abstract – **Voltage calibrators as test standards have their own probability distribution producing uncertainty in the determination of an in-tolerance or out-of-tolerance condition. In the accredited metrology laboratory of the Faculty of Electronic Engineering in Niš, two calibrators FLUKE 5100B and METRAtop 53 are compared. In calibration process guardbanding strategy is proposed to equalizing the cost of faulty test decision between above two calibrators. Performed results of evaluation and comparison their uncertainties are presented in this paper.**

Keywords – **Voltage and current calibrator, Test uncertainty ratio, Confidence limits.**

I. Introduction

Accurate measurements are essential in test and measurement systems. However, if the measurement hardware is not calibrated, then there can be no certainty in the acquired measurements results.

With the increased acceptance of ISO standards [1,2], many users now find necessity to prove the accuracy of implemented measurements. They must produce some sort of traceable verification of their instruments in order to prove measuring correctness and specifications. In calibration process of particular concern is adequacy of standards which are used to calibrate units under test (UUT). Measurement uncertainty of used calibration standard is directly contributed the quality of calibration process. Accredited metrology laboratory must provide that uncertainty of measurement standards not exceed acceptable tolerance (manufacturer's specification).

The calibration support of the most accurate measuring instruments has always been a complex task. As technical advances make it easier for manufacturers to offer products with high-performance, the metrologist must find practical ways to calibrate measuring instruments that often need higher capabilities of the available standards.

High reliable calibration required that standards are at least ten times better than the instruments being compared to them, that is, a test uncertainty ratio (TUR) would be equal 10:1 [3]. Increased performance in the instrument being calibrated has resulted in a reduction of acceptable TURs to 4:1.

In this paper guardbanding strategy in calibration process is proposed to equalizing the cost of faulty test decision between two calibrators FLUKE 5100B and METRAtop 53. Evaluation of uncertainties is done using "Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results" [5]. New confidence limits are proposed to assure that calibration confidence is maintained.

[1]Vladan S. Djurić is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: djvladan@elfak.ni.ac.yu

[2]Božidar R. Dimitrijević is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: dimitrijevic@elfak.ni.ac.yu

[3]Ivana S. Randjelović is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: rivana@elfak.ni.ac.yu

II. Evaluation of Uncertainty

All measurements are estimates of the true value of the measured parameter and are subject to errors, described as uncertainty. The uncertainty of measurement is evaluated according to either a Type A or a Type B method of evaluation [5,6]. The evaluation of standard uncertainty Type A is the method of evaluating the uncertainty by the statistical analysis of a series of measurements. In this case the standard uncertainty is the experimental standard deviation of the mean that follows from an averaging procedure or an appropriate regression analysis. The evaluation of standard uncertainty Type B is the method of evaluating the uncertainty by means other than the statistical analysis of a series of observations. In this case the evaluation of the standard uncertainty is based on technical documentation provided by manufacturers.

Technical specifications of used calibrators [7,8] for dc voltage range 2 mV, 200 mV, 2 V and 20 V are presented in Table 1.

Table 1.

Ranges	Fluke 5100B	Metratop 53
20mV	0.005%·U+5,2μV	0.02%·U+50μV
200mV	0.005%·U+7μV	0.02%·U+50μV
2V	0.005%·U+25μV	0.02%·U+500μV
20V	0.005%·U+205μV	0.02%·U+5mV

Values in columns 2 and 3 of Table 1 represent specification limits, where value U is set referent dc voltage on calibrator output. The worst case is value for the highest voltage value in each range, and that is shown in Table 2.

Table 2.

Ranges	Fluke 5100B	Metratop 53	TUR
20mV	6,2μV	54μV	8,7
200mV	17μV	90μV	5,3
2V	125μV	900μV	7,2
20V	1,205mV	9mV	7,5

Type A method evaluation is done using HP3290A voltmeter. Measured difference between set output dc voltage values of two calibrators for ranges 2 mV, 200 mV, 2 V, 20 V are shown in Figs. 1 to 4.

For four ranges, max value, mean value of difference and standard deviation of mean value are shown in Table 3.

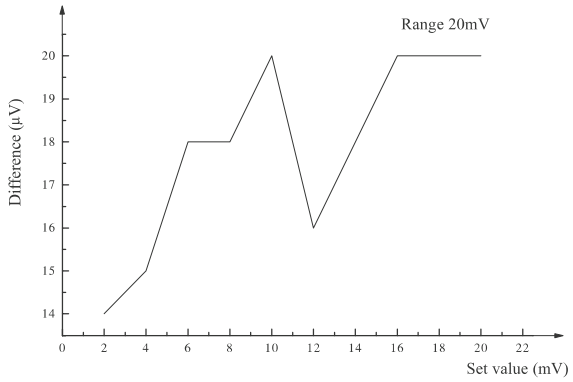


Fig. 1. Difference between set values of two calibrators

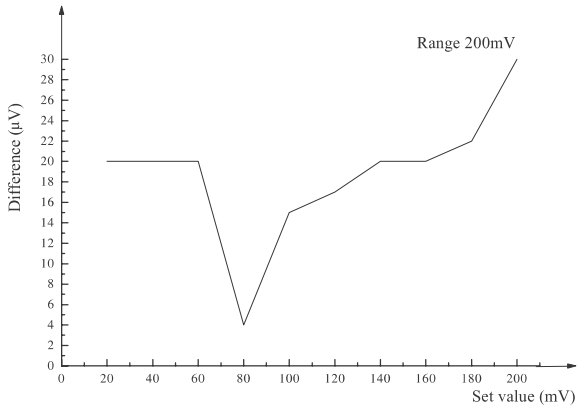


Fig. 2. Difference between set values of two calibrators

Test uncertainty ratios between two calibrators are calculated based on values in Table 2 and Table 3. Proposed method is based on correction of test uncertainty ratios provided by technical documentation of manufacturer, Table 2. Correction is done using relation between maximum value of difference, Table 3, and difference of uncertainties from Table 2. New test uncertainty ratios are shown in Table 4. For example, for calculating test uncertainty ratio on range 20 mV is shown in Eq. (1)

$$\text{TUR} = (8.7 \times 20\mu\text{V})/(54\mu\text{V} - 6.2\mu\text{V}) = 3.6 \qquad (1)$$

Table 3.

Ranges	Max value	Mean value	Standard deviation of mean value
20mV	20μV	18μV	2μV
200mV	30μV	19μV	6μV
2V	350μV	175μV	77μV
20V	1400μV	865μV	330μV

Table 4.

Ranges	Test uncertainty ratio (TUR)
20mV	3,6
200mV	2,2
2V	3,2
20V	1,3

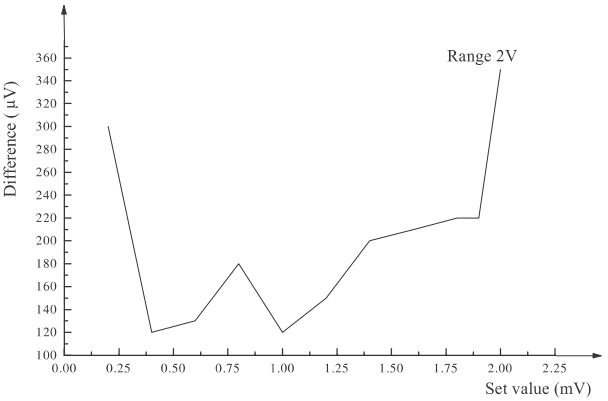


Fig. 3. Difference between set values of two calibrators

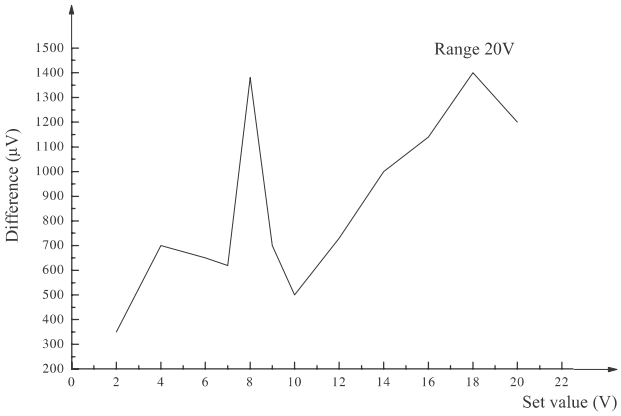


Fig. 4. Difference between set values of two calibrators

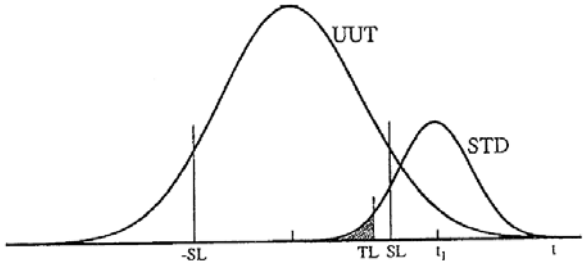


Fig. 5. Out-of tolerance unit reported as confirming

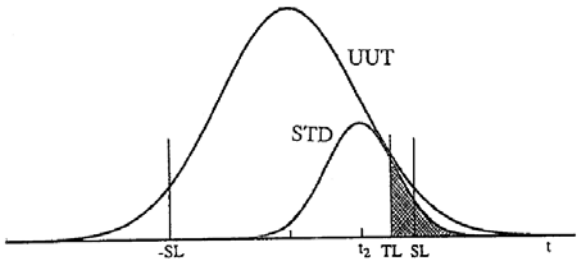


Fig. 6. In-tolerance unit reported non-confirming

III. Guardbanding Method for Estimating New Test Limits

Guardbanding is a statistical method for setting in-tolerance and out-of-tolerance limits so that calibration is done with

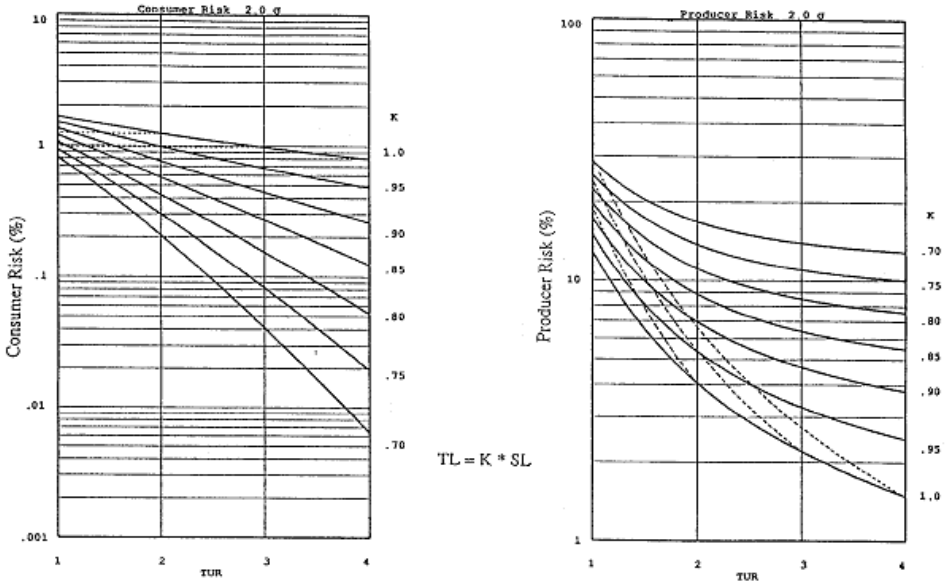


Fig. 7. Risks for different guardband limits and for $\pm 2\sigma$ (P=95%)

adequate confidence when test uncertainty ratios are small. New in-tolerance limits are calculated by comparing the uncertainty of the calibration standard with the specifications of the UUT. Consumer risk (CR) represents out-of tolerance unit reported as confirming. Producer risk (PR) represents in-tolerance unit reported non-confirming.

Though the probability of making faulty test decisions (consumer risk) increases with decreasing TURs, the test limits can be placed to set the desired level of consumer risk or producer risk. For example, it is possible, with a 2:1 TUR, to keep the same risk of accepting defective units as a 4:1 TUR by setting the test limits TL=K×SL, K<1, inside the specification limits SL. By factor K the specification limit is reduced to obtain the new test limit.

ISO Guide 25, drafts 5 and 6, released as ISO/IEC 17025 [2] is not explicit about confidence interval. Most of the literature, laboratory and industry practices, assumes $\pm 2\sigma$ confidence interval. New test limits are set to give a 95% probability of being within the UUTs specification limits.

Fig. 5 shows the effects of having a TL inside the SL for symmetrical limits, where are:

UUT: The distribution of possible values for the unit under test.

STD: The distribution of possible values for the Standard.

t: local variable for the UUT distribution.

t_1: a possible value of the UUT.

The shaded area to the left of t_1 in Fig. 5 illustrates the probability that a unit outside the SL will be accepted with new test limits. Out of tolerance probability is calculated by the double integral of Eq. (2). Test limit is obtained by reducing specification limits by factor K.

$$CR = \frac{1}{\pi} \int_{SL}^{\infty} \int_{-TUR\cdot(t+TL)}^{-TUR\cdot(t-TL)} \exp\left[-\frac{(s^2+t^2)}{2}\right] \mathrm{d}s\mathrm{d}t \qquad (2)$$

Shaded area shows the reduced probability of false accepts since units measuring inside the SL but greater than the TL will be rejected.

Similarly, the in tolerance with guardband is shown in Fig. 6.

$$PR = \frac{1}{\pi} \int_{-SL}^{SL} \int_{TUR\cdot(TL-t)}^{\infty} \exp\left[-\frac{(s^2+t^2)}{2}\right] \mathrm{d}s\mathrm{d}t \qquad (3)$$

ISO Guide 25 Draft 5 proposal is to use TL=SL when the TURs are sufficiently high, 10:1 TURs are recommended. 4:1 TURs might be tolerated.

For setting new test limits ISO Guide 25 Draft 5 proposal is to use Eq. (4).

$$TL = K \times SL = \left(1 - \frac{1}{TUR}\right) \times SL \qquad (4)$$

In Fig. 7 Consumer and producer risk for different guardband limits and for $\pm 2\sigma$ (P=95%) is presented.

In order to maintain calibration confidence for both calibrators, new test limits are calculated. Calculation of new test limits for FLUKE 5100 B is done using guardbanding method based on Eq. (4) and TURs in Table 4.

New test limits are shown in Table 5.

Table 5.

Ranges	Old Test Limits	New Test Limits
20mV	6,2µV	4,5µV
200mV	17µV	10µV
2V	125µV	86µV
20V	1,205mV	300µV

IV. Conclusion

Guardbanding is a method for setting in-tolerance and out-of-tolerance limits so that calibration is done with adequate confidence when test uncertainty ratios are small.

In this paper, guardbanding strategy in calibration process is proposed to equalizing the cost of faulty test decision between two calibrators. New confidence limits are proposed for equalizing the cost of faulty test decisions between the two calibrators and to assure that calibration confidence is maintained.

With proposed guardbanding method calibrator with lower test uncertainty ratio can be successfully used in the calibration process. This is particularly significant for portable calibrator Metratop 53, which is used for calibration outside laboratory.

References

[1] ISO/IEC Guide 25, International Standards Organization, August 1996.

[2] ISO/DIS 17025, International Standards Organization, March 1998.

[3] David K. Deaver, ”How to Maintain Your Confidence”, NCSL Workshop & Symposium, pp. 133-153, 1993.

[4] David K. Deaver, ”Managing Calibration Confidence in the Real World”, NCSL Workshop & Symposium, 1995.

[5] B. Taylor, C. Kuyatt, ”Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results”, 1994.

[6] Dave Abell ”Accreditation for Complex Electronic Instruments”, Agilent Technologies, 2001.

[7] W. T. Bolk,” A general digital linearising method for transducers”, Phys. E: Sci. Instrum., vol.18, pp. 61-64, 1985.

[8] ADAM 4000 Series Data Acquisition Modules , User’s Manual, Advantech Co., Ltd,

High Resolution Time-to-Digital Converter

Goran S. Jovanović[1] and Mile K. Stojčev[2]

Abstract – **This paper describes the architecture and performance of a high-resolution time-to-digital converter (TDC) based on a Vernier delay line. The TDC is used as a basic building block for time interval measurement in an ultrasonic liquid flowmeter. Operation of the TDC with 10 ps LSB resolution and 1 ms input range has been simulated using library models for 1.2 m double-metal double-poly CMOS technology. The TDC operates at clock frequency of 200 MHz, and is composed of 500 delay-latch elements. The difference in delay between two chains, one for the start and the other for stop pulse, is controlled by the delay locked loop (DLL).**

Keywords – **Time to digital conversion, Vernier delay line, DLL.**

I. Introduction

The precise measurement of the time interval between two events with very fine timing resolution is common challenge in the test and measurement instrumentation (logic analyzer, ATE system, nuclear instrumentation), industrial control (multichannel DAS, ultrasonic liquid flowmeters), electronic embedded control system (automotive controllers, medical devices avionics), etc. [1-3]. A time-to-digital converter (TDC) is one of the crucial building blocks installed into this type of equipment. High-resolution TDC is primarily used in application areas that require a resolution better than 10 ps, low dead-time (minimum time between two measurement, less than several microseconds) and large dynamic range (maximum time interval that can be measured, can be in the range of hundred seconds) at operating frequency from the minimum of around 1 MHz to the maximum of 500 MHz. Therefore the time intervals of interest for measurements in our case range from 10^{-11} s to 10^{2} s [4].

In principle, time interval measurement performed by the TDC can be decomposed into the following two steps. The first one is called short-time interval measurement, and is characterized with very fine timing resolution in the range from 10 ps up to 500 ps. The second referred as long-time-interval characterizes a coarse time resolution in the range from 3 μs up to 100 s. For evaluating long-time interval measurements standard counter based methods are used. The principle methods of digitizing short-time intervals have been reviewed in [1]. They include utilization of fast counters [5], analog methods based on generating voltage ramp [6] and dual-slope conversion [7], and various CMOS tapped delay line configurations [3,4,8].

[1]Goran S. Jovanović is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: joga@elfak.ni.ac.yu

[2]Mile K. Stojčev is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Niš, Serbia and Montenegro, E-mail: stojcev@elfak.ni.ac.yu

The goal of this works is to develop a TDC, which would enable the realization of an ultrasonic based liquid flowmeter in pipe under pressure with 1% time-interval measurement accuracy, and 10 ps time-resolution at operating frequency of 200 MHz.

II. TDC Principle of Operation

TDC's can traditionally be divided into the following two groups: analog and digital. In general, analog TDC's are based on current integration, while the digital on some counting methods. Recently alternative TDC architectures have emerged which are somewhere in between these two extremes.

For realization of the TDC that would be implemented into our ultrasonic liquid flowmeter several different candidate architectures have been considered. In the sequel, our experiences about the properties and limitations of such TDC's, mainly acquired involving simulation methods, are presented. The starting point for our comparative study was: a) available, transistor models for 1.2 μm double-metal double-poly CMOS technology, mainly used for synthesis of custom circuits; and b) commercially available FPGA/CPLD circuits (Xilinx, Spartan series) that we used for integration of the digital circuits needed to create a complete interpolating time counter on a single CMOS FPGA/CPLD chip. PSpice 9.2 and Xilinix Projekt Navigator V. 4.2WP2.X software were used as simulation and synthesis tools, respectively.

A. Analog TDC-Based on Current Integration

Analog high resolution TDC consists of a constant current source used to charge a capacitor, whose voltage is sampled when a trigger pulse occurs. An ADC then converts the analog voltage to a digital value. Analog time-to-voltage converters can provide about 10 ps resolution over a dynamic range of 1:50 - 1:1000, but they tend to be nonlinear and are difficult to stabilize. One good way of stabilizing them is to use dual-slope conversion, which is unfortunately also slow and limits the measurement rate and increases dead-time, in some applications such as for example those used time-of-flight particle detectors, laser range finders, and logics analyzers [1,4,9]. In our case, bearing in mind that the time propagation of the ultrasonic signal through the pipe is an order of hundred microseconds [10] the dual-slope TDC represents a good candidate solution. Details concerning TDC based on dual-slope conversion will be given in Section III.

B. Digital TDC-Base on Counter

The counter based digital TDC consists of a Gray code counter running at high speed, which value is sampled when a trigger pulse occurs [1,9]. However, time interval digitalization with sub nanosecond resolution using only a simple frequency counter requires impractical high clock frequencies or long averaging times. In spite of its simple architecture this type of TDC, by our opinion is not a good design choice, primarily due to the prerequisites for high-speed counter operation (order of several GHz).

C. TDC-Based on Vernier Delay Technique

The time resolution of the counter based TDC can be improved significantly by using the gate delay as the basic time unit. The fundamental concept of the delay Vernier technique is that the timing resolution is determined by the difference between two propagation delay values. The interpolation methods implemented in this case is used to interpolate time fractions inside clock cycles. A Vernier structure consists of a pair of tapped delay lines with a flip-flop at each corresponding pair of taps and is presented in Fig. 1. A stop signal propagates through one of the delay chain, while the start signal propagates through the other, clocking the flip-flop at each stage. The difference between the start and stop propagation delays determines the timing between adjacent stages. For more details, see [4].

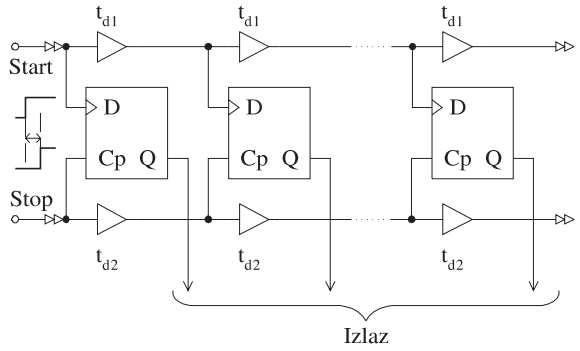


Fig. 1. Typical Vernier delay line

The dynamic range of the TDC based on Vernier delay technique, i.e. the maximum time that can be measured, is limited to $t_{DR} = n \cdot (t_{d1} - t_{d2})$, where n is a number of delay elements of the delay line. However, the range of the TDC can be extended by introduction of a simple counter [3,4,9]. Using Vernier delay technique implemented on commercial available FPGA chips [3] a time resolution of 200 ps can be achieved. However, CMOS process is very temperature sensitive so frequent calibrations with a time reference is required. In spite of the mentioned drawbacks, by our opinion the architecture based on the Vernier principle, represents a very good candidate design solution.

III. Implementation of the TDC Based on Vernier Delay Technique

In order to obtain both high precision and long time measurement we have used the time interpolation technique based on the classic Nutt method [7,11]. It involves splitting the measured interval T_{in} (in our case of order 800 μs), from the rising edge of the start pulse to the rising edge of the stop pulse, into two subintervals. The first subinterval, corresponds to the integer number N_c of the reference clock period T_{CLK} (5 ns), while the second subinterval characterize a duration equal or less than one clock period. The first subinterval, called coarse time interval measurement, is equal to $N_c \cdot T_{CLK}$. It is synchronous to the system clock and is measured with binary counter (*Coarse_Counter*) at 200 MHz clock. More details concerning this problematic can be found in [4,10].

In order to improve a time resolution of the second subinterval, a TDC that use Vernier delay line. In general, the Vernier measurement technique is based on a propagation delay difference between two chains. The chain is realized with delay cells connected in cascade.

In this paper we describe the structure of an analog voltage-controlled delay cell (see Fig. 2). The delay cell represents a basic building block of a delay chain. As can be see from Fig. 2, the delay cell is implemented as a two-stage inverter. Transistors M_1 and M_2 are constituents of the first stage, while transistors M_7 and M_8 form the second stage. Transistors M_3 and M_4 act as current source and current sink, respectively. Voltages $V_{P_{bias}}$ and $V_{N_{bias}}$ regulate currents of M_3 and M_4, respectively. The bias circuit, composed of transistors M_9, M_{10} and M_{11}, provides correct polarization for transistor M_3 and M_4. With order to linearize the transfer characteristic of the delay cell, transistors M_5 and M_6 are used. In Fig. 3, the transfer function, which represents variation of time delay in term of control voltage V_C for the cell pictured in Fig. 2, is presented. As can be see from Fig. 3, for used technology, the variation of a delay is almost linear within the range of 775 ps up to 1025 ps.

The second building block of the Vernier delay line (see Fig. 1) is the storage element. It can be realized as a latch or D flip-flop.

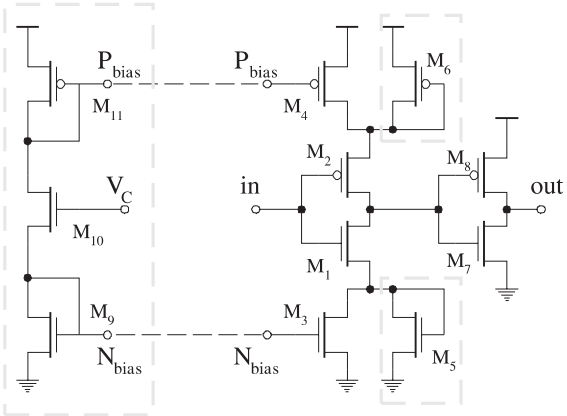


Fig. 2. Analog voltage-controlled delay cell

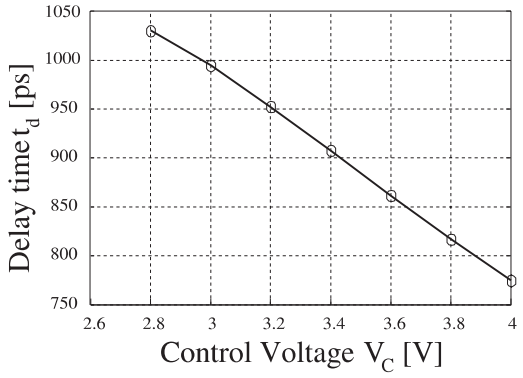


Fig. 3. Propagation delay time in term of control voltage V_C

Two design requirements have to be fulfilled when the Vernier delay line is realized. The first one relate to a measurement resolution (∼10 ps). This resolution can be achieved if the delay difference between the two chains is kept small. The second relate to a wide measurement operating range (up to 5 ns). It can be achieved by installing chain with more then 500 delay cells. Having this in mind, the crucial design challenge now, from aspect of silicon area, can be achieved if we decrease the complexity of both the storage element and the delay cell. Here, we propose one solution where the delay chain of the *start* pulse is modified in respect to [4] such that it includes both, the original delay line (see Fig. 2) and the ladder structure of storage elements, see Fig. 4. The structure of the second chain is composed of delay cells already sketched in Fig. 2.

As can be see from Fig. 4, the delay cell for start pulse consists of three inverters, I_1, I_2 and I_3. Inverters I_1 and I_2 form the chain for the start pulse. At the some time, inverters I_2 and I_3 are connected as latch. In this manner, complexity of the hardware structure given in Fig. 1 is decreased compared to [2-4]. Let note, that I_1 and I_3 are implemented as a three-state drivers, while I_2 as a voltage-controlled delay element.

The structure of the proposed modified Vernier delay line is sketched in Fig. 5. Two control voltages, V_{C1} and V_{C2}, are used. The control voltage V_{C1} is used for delay adjustment of the *start* delay chain, while V_{C2} is intended for delay adjustment of the *stop* delay chain. In our proposal V_{C2} is fixed, while VC1 is determined during calibration phase, see Fig. 6. Namely, during this phase, the inputs *start* and *stop* are driven with pulses of known (in advance defined) phase and frequency. The phase detector generates control signals *up* and *down* by witch it regulates the charging and discharging current of the charge-pump circuit. The voltage at the output of a low-pass filter is then converted by the ADC and then reconverted by the DAC with order to be used as a control voltage V_{C1} during the measurement process.

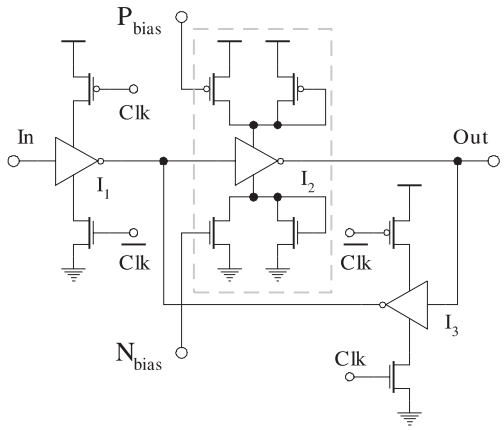


Fig. 4. Hardware structure of a latch with three-state output control and a voltage-controlled delay element

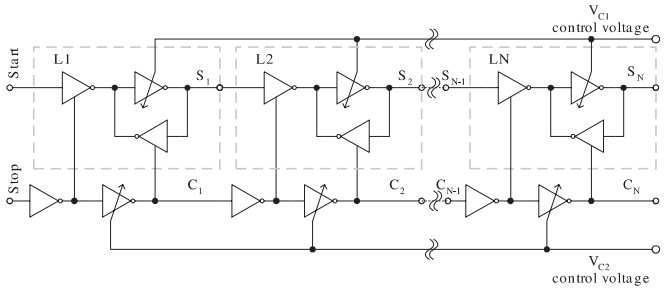


Fig. 5. Modified two chain Vernier delay line

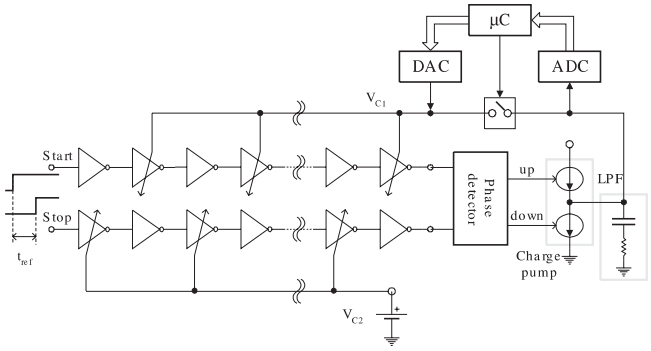


Fig. 6. DLL used in a feedback loop of the Vernier delay line

IV. Simulation Results

Here, we propose a full-custom circuit design approach for implementation of the described TDC architecture based on the Vernier method. Testing and verification of the design was performed using software tool PSpice 9.2 and models of components (transistors, diodes, capacitor, etc.) for 1.2 μm CMOS double-metal double-poly technology. Using simulation method it is possible to evaluate precisely and accurately the propagation time of *start*, S_i, and *stop*, C_i, pulses through both chains (see Fig. 5). For operating system clock frequency of 200 MHz the *coarse_counter* is clocked with pulses of 5 ns duration and its final value corresponds to coarse time interval measurement. From other side, the second subinterval which is less than 5 ns and corresponds to fine resolution interval measurement, have to be estimated with resolution of ∼10 ps. Having this in mind, the hardware structure of the Vernier delay line have to be composed of 500 cells, what corresponds to 5 ns (500·10 ps=5 ns). For better visualization of the simulation results presented in this

paper, a time resolution of 400 ps between two adjustment pulses is adopted for presentation in Fig. 7. Also, the time interval t_x between the *start* and *stop* pulses, in a concrete case, is equal to 4 ns. Under this conditions the waveforms S_i and C_i $(i = 1,,12)$ are generated at the outputs of the corresponding delay cells. As can be seen from Fig. 7, until the delay cell 10 the pulse S_{10} is in advance with respect to the pulse C_{10}. At inputs of cells 11 both pulses arrive at the some moment, while at cell 12 the pulse C_{12} is in advance with respect to S_{12}. This means that the propagation of the *start* pulse is further prohibited. If we analyze the output of the corresponding latch, we see that latches from 1 up to 10 are set to one while latches L_{11} and L_{12} are set to zero.

The principle of operation of both delay chains can be described according to the presentation in Fig. 8. For the some operating condition we see that the delay propagation between the *start* and *stop* signals will be equal after the 10th cell i.e. when the cross-point between two lines appears.

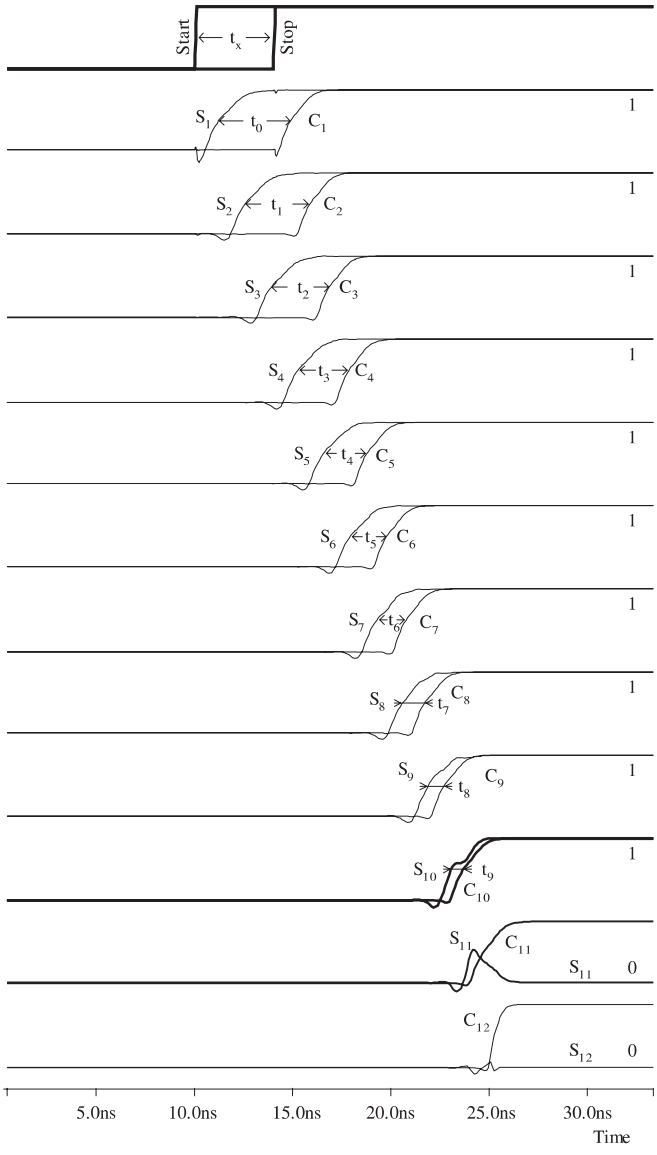


Fig. 7. Propagation of *start* S_i and *stop* C_i pulses

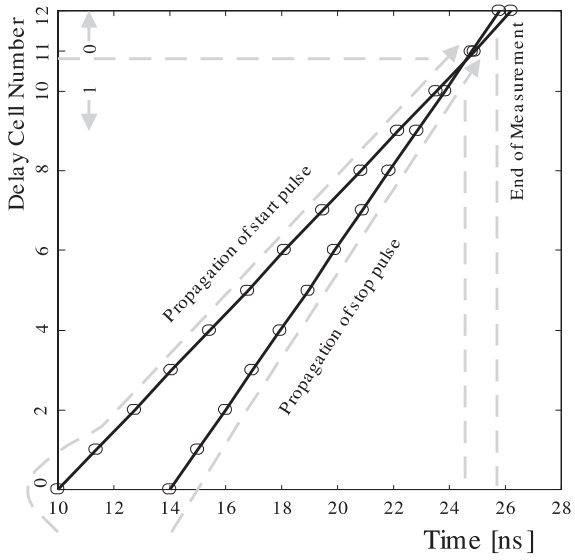


Fig. 8. Number of delay-latch elements set to logic one during fine time resolution measurement

V. Conclusion

An efficient architecture of a time to digital converter, based on the Vernier delay line, intended for high-resolution time interval measurement is presented in this paper. The proposed architecture contains two delay line chains. The first one represents a composition of a delay line and memory cells. The second chain is implemented as a classical chain of a non-inverted stages connected in cascade. The delay of both chains is voltage-controlled. The TDC's operation has been simulated and implemented using models 1.2 μm double-metal double-poly CMOS technology. For operating frequency of 200 MHz and chain composed of 500 delay cell elements a resolution of 10 ps LSB was achieved. Bearing in mind, that the delay of the CMOS cell is sensitive to ambient temperature and supply voltage variations, a delay lock loop is involved in the structure of the TDC in order to compensate these negative effects. For precise and accurate time-interval measurement a calibration procedure is involved. It is primarily intended to achieve a balanced difference in delay between the two voltage-controlled delay chains. To implement this technique a DLL feedback loop is used. The TDC is used as a constituent block intended for precise and accurate time interval measurement in an ultrasonic liquid flowmeter.

References

[1] Porat D.I., "Review of sub-nanosecond time-interval measurement", *IEEE Transaction on Nuclear Sci.*, NS-20, No. 5, pp. 36-51,1973.

[2] Gray C.T., et al., "A sampling technique out its CMOS implementation with 1Gb/s bandwidth and 25ps resolution", *IEEE Journal of SSC*, vol. 29, No. 3, pp. 340-349, March 1994.

[3] Kalisz J., et al., "Single chip interpolating time counter with 200ps resolution and 43-s range", *IEEE Transaction on Instrumentation and Measurements*, vol. 46, No. 4, pp.851-856, 1997.

[4] Dudeck P. et al., "A high-resolution CMOS time-to-digital converter utilizing a vernier delay line", *IEEE Journal of*

Solid-State Circuits, vol. 35, No. 2, pp. 240-246, February 2000.

[5] Sasaki A.E. et al., "1.2GHz GaAs shift register IC for dead-time-less TDC application", *IEEE Transaction on Nuclear Sci.*, vol. 36, pp. 512-516, February 1989.

[6] Stevens A.E., et al., "A time-to voltage converter of analog memory for colliding beam detector", *IEEE Journal of Solid-State Circuits*, vol. 24, No. 6, pp. 1748-1752, December 1989.

[7] Raisanen-Routsalainen E. et al., "An integrated time-to-digital converter with 30-ps single-shot precision", *IEEE Journal of SSC*, vol. 35, No. 10, pp. 1507-1510, October 2000.

[8] Christiansen J., "An integrated high resolution CMOS timing generator based on an array of delay locked loops", *IEEE Journal of SSC*, vol. 31, No. 7, pp. 952-957, July 1996.

[9] Christiansen J., "An integrated CMOS 0.15ns digital timing generator for TDC's and clock distribution systems", *IEEE Trans. on Nuclear Sci.*, vol.42, No.4, pp.753-757, August 1995.

[10] V. Pavlović, et all, "Realization of the Ultrasonic Liquid Flowmeter Based on the Pulse-Phase Method", *Ultrasonics 35* (1997) 87-102.

[11] Nutt R., "Digital time interval meter", *Rev. Sci. Instrum.*, vol. 39, pp. 1342-1345, 1968.

Generation of a Test Strategy for Testing the Analog Part of an Integrated Circuit for Digital Wireless Short-Range Communication

Rumen Iv. Arnaudov[1]and Ivo G. Aldimirov[2]

Abstract – **Subject of this article is the generation and application of a Test strategy for measure and test of Application Specific Standard Product (ASSP) Integrated Circuit (IC) for digital wireless communication. It is given a description of the employed methods for testing the building blocks of the *analog part* of the device. They are based on the way they take place in the industrial test of the mixed-signal IC. Comment and quotation on some existing methods is made. It is emphasized on some potential problems that might influence the test time/cost, the accuracy and the stability of the tests. Suggestions for solving some of those problems are made.**

Keywords – **digital signal processing (DSP), industrial test, mixed-signal Integrated Circuit, test methods, digital tests, device under test (DUT).**

I. Introduction

Fig. 1 shows typical block diagram of a Front-end radio Application Specific Standard Product (ASSP). It is used as a basis for analyses of the methods for test and measurement described hereafter.

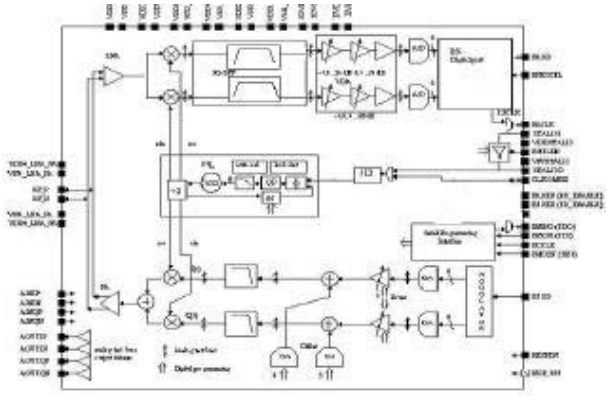


Fig. 1. Block diagram of typical transceiver for digital wireless communication.

The diagram represents typical transceiver with zero intermediate frequency (IF=0). It consists of two parallel I and Q channels in the receive and the transmit part. On the same silicon die is also implemented the digital part, consisting of the corresponding ADCs and DACs, the bank of registers used for initialization and control of the device.

II. Generation of a Test Strategy - Pro and Con.

The test strategy combines in one the test methods employed for measuring the electrical parameters of the functional building blocks of a silicon device. Strictly speaking, generating test strategy is not absolutely necessary. A test engineer can generate a test program entering the test program code based on the device data sheet. There are several problems with this type of undisciplined approach. First, device testability will probably not be identified early enough to allow the addition of test features to the design - design for test (DFT) blocks. The test strategy generation forces the design and test engineers to work through all the details of testing at an early stage in the design cycle. Second, the test engineer may create test-to-test compatibility problems if the details of all tests are not known up front. For example - a clocking scheme that works well for one test may be incompatible with the clocking scheme required of a subsequent test. The first test may then need to be rewritten from scratch so that the clocking schemes mesh properly. Third, test hardware such as device interface board (DIB) and probe interface hardware cannot be properly designed until all test details are known. The test strategy generation helps to identify which hardware resources of the automated test equipment (ATE) will be used and the possible shortfalls in the target testers capabilities.

III. Test Program Structure

Tester languages vary from low-level C routines, to very sophisticated graphical user interface environments. Despite wide differences test programs consist of all or most of the following sections: creation of wave forms and other test initializations, calibrations of the tester hardware, continuity tests, DC parametric tests, AC parametric tests, digital patterns (also known as functional tests), digital timing tests, test sequence control, test limits and binning control.

IV. DSP Based Test Methods [1,2]

Basic principles for forming DSP test strategy.

AC measurements such as gain and frequency response can be measured with relatively simple analog instrumenta-

[1]Rumen Iv. Arnaudov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: rarnaud@vmei.acad.bg

[2]Ivo G. Aldimirov is with the Air Traffic Services Authority (ATSA), Communications department at Plovdiv International Airport, e-mail:aldimirov@yahoo.com

tion. To measure gain, an AC continuous sine wave generator can be programmed to source a single tone at a desired voltage level and frequency. A true RMS voltmeter can then measure the output response from the DUT, and the gain can be easily calculated. The pure analog approach to AC testing suffers from few problems, though. First, it is relatively slow when AC parameters must be tested at multiple frequencies. Second, traditional analog instrumentation is unable to measure distortion in a presence of the fundamental tone. Thus the fundamental tone must be removed with a notch filter, adding to test hardware complexity. Third, analog testing measures RMS noise along with RMS signal, making results unrepeatable unless we apply averaging band-pass filtering.

DSP is a powerful methodology that allows faster, more accurate, more repeatable measurements. DSP based testing is based on the sampling theory. The application of DSP in the industrial test can be briefly described as: time domain captures data (signal sampling), fast Fourier transformations and frequency domain output data (as it is shown on Fig. 2).

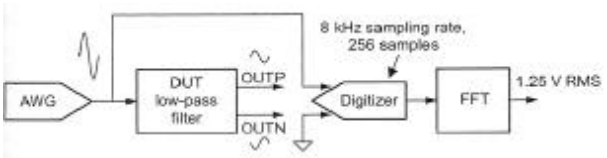


Fig. 2. Configuration of a sampling measurement system.

When testing functional blocks such as LPF, BPF and VGA it has to measure the gain, the THD and the SNR for a number of frequencies in the working range. For that purpose waveforms are software generated and loaded as segments in the memory of an AWG - arbitrary waveform generator, and are applied as a multitone at the input of the analog channel under test. At the output the captured signal is digitized and DSP analyzed using fast Fourier transformation (FFT). The amplitudes and the phases of the spectral bins - elements of the multitone are compared with the ones applied at the input. The spectral bins corresponding to the harmonics of the fundamentals in the multitone are measured too.

The creation of a waveform is based on the sampling theory. Periodical sine signals are described as following:

$$v(t) = A_1 \sin(2\pi\frac{f_1}{f_s} + \varphi_1) + A_2 \sin(2\pi\frac{f_2}{f_s} + \varphi_2) + ... + A_n \sin(2\pi\frac{f_n}{f_s} + \varphi_n)\ , \quad (1)$$

where f_1,...,f_n are the frequencies of the multitone, φ_1,...,φ_n are their corresponding phases, f_s is the used sampling frequency.

When waveform is generated or when capture and reconstruction of a waveform occurs, f_s – the sampling frequency is the clock frequency of the hardware instrumentation. In this case:

$$\Phi = f_{\text{bin}}\frac{f_s}{n_{\text{samples}}}, \quad (2)$$

where Φ is the frequency of interest, n_{samples} is the number of captured samples, f_s is the sampling frequency, and f_{bin} is the corresponding to Φ spectral bin.

The approach to develop DSP based tests should be as following:

1. All the frequencies of interest (Φ) that have to be measured must be known and whenever need have to be corrected within reasonable tolerances for getting correct results out of the DSP computations. The choice of f_s and $n_{samples}$ has to comply with the following: f_{bin} has to be mutually prime with the number of samples. Otherwise there is a risk in the measurement results of increased presence of harmonic distortions components, intermodulation distortions components and quantization noise. Whenever FFT is in use n_{samples} must be power of 2. The n_{samples} has to be chosen according to the required frequency resolution (F_{res}) of the measured spectrum:

$$F_{\text{res}} = \frac{f_s}{n_{\text{samples}}} \quad (3)$$

It has also to be considered how close the measured signal level is to the noise floor of the capture instrument. Measurement with insufficient F_{res} integrates the energy of the neighbouring spectral bins and as a result the noise level is higher. Choosing higher F_{res} solves this problem but increases the test time. In this case the test time is related to n_{samples} and the period of sampling or $1/f_s$ according to:

$$\text{Test time} = T_{\text{hw}} + T_{\text{capture}} + T_{\text{mv}} + T_{\text{calc}}\ , \quad (4)$$

where T_{hw} – time for hardware setup, T_{capture} – time for signal capture, T_{mv} – time for moving captured data, T_{calc} – time for calculations. T_{capture} is described as:

$$T_{\text{capture}} = \frac{n_{\text{samples}}}{f_s} = \frac{1}{F_{\text{res}}} \quad (5)$$

According to (4) and (5) the more the number of samples the longer the test time for the corresponding measurement which could be unacceptable in a number of cases.

2. The choice of n_{samples} must comply with the period of the measured signal according to:

$$\frac{n_{\text{samples}}}{f_s} = \text{int}\ T_i\ , \quad (6)$$

where $t_i = 1/f_i$. Equation (6) expresses the requirement for f_s and the period of the measured signal mutual relation i.e. an integer number of period has to be captured. Otherwise the Fourier analyses of a non-periodical signal will produce smearing in the output spectrum as illustrated on Fig. 3.

3. It is important to choose the value of f_s so that the hardware clock (f_s itself) to comply with (6) for most of the AC measurements - signal frequencies. This way reprogramming during the program run will not be needed and test time and test cost will be saved as a result.

V. DSP Based Tests

They are used for testing analog functional block such as RF receive/transmit, BPF, LPF, and VGA.

For testing BPF, LPF and similar blocks requiring characterization for a numb er of in-band and out-of-band frequencies a multitone is applied [1], [2]. The multitone waveform comprises the specific frequencies in the transfer characteristic of the analog block. The spectral bins of the predefined

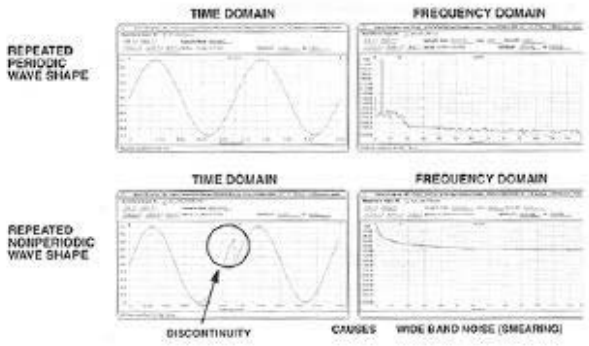


Fig. 3.

frequencies of interest as they appear in the output are subject of analysis. The measurement analysis came across the following problems:

1. The out-of-band frequencies of BPF and LPF are normally highly attenuated and the corresponding amplitudes could be masked under the noise floor of the capture instrument. Increasing the frequency resolution might be unacceptable because of test time requirements (big number of samples). The performed measurements using reciprocal characteristics showed up good results and more in -depth investigation is worthwhile.

2. The base of the DSP based method is the measure of the amplitude of known spectral components (frequency bins). The presence of frequency instability or jitter could have very serious impact on the accuracy and the repeatability of test results. The distribution of the test results in this case might have higher value for σ. In some cases it might become a double distribution. Lowering F_{res} by the purpose of integration the neighboring spectral bins within the effective range of the jitter will lift up the noise level of the measurement. The other way round - seeking for higher accuracy of measuring amplitudes close to the noise levels (poorer F_{res}) will lead to higher inaccuracy and worsen the repeatability. In that kind of cases a compromise should be looked for. One efficient solution would be the implementation of a test technique looking not for the calculated spectral bin of the fundamental but for the maximum amplitude within a certain range in frequency domain. If this frequency range is carefully defined representing with some guard banding the peak values of the jitter, than the higher amplitude within it might be considered as the one of interest. In other words - the so proposed peak-search method could effectively take place for accurate and stable measurements whenever jitter is presented. A major constraint for applying such kind of method is that the jitter maximum value must be less than the spacing between the neighboring elements of interest in the measured spectrum.

When VGA is under test monotone signal is applied at the input. The input amplitude for the different gain levels is calculated by the test program so that at the output to have constant amplitude within the input dynamic range of the capture instrument. This guaranties minim level of introduced noise and harmonic distortions in the measured signal as well highly repeatable measurement results. Not least, avoiding switching the capture instrument input ranges for different amplitudes saves test time - cost.

DSP based methods has taken place by realization of the following tests:

- RF Rx: Rx gain, Rx IIP3, Rx in band power contributed by off band interferers, Rx noise figure;
- RF TX: TX spectrum, TX carrier rejection tuning, TX image rejection tuning;

The choice of the test methods to be employed is not only based on the advantages they give but on the study of a real device measurement, with the purpose to investigate the repeatability of the tests. A criterion for quotation of a particular measurement is the value of C_p - process capability. It is defined according to:

$$C_p = \frac{\text{(USL-LSL)}}{6\sigma} , \tag{7}$$

where USL, LSL are the engineering limits for the corresponding measurement. Criteria for stability of a test result is its distribution to show up $C_p > 20$ after repetition of the test over one and the same device 100 times. This way the widening of the distribution of the test result because of its instability is eliminated and in cases of marginal to the test limits distribution the yield is improved.

VI. Conclusions

This article analysis the employed methods for testing of the main building blocks of the analog part of an ASSP. It is pointed on the advantages of the proposed methods and on the potential problems as well. Approaches for resolving the problems are suggested:

- Approach for development of DSP based tests.
- Considerations for choosing F_{res}.
- Considerations for choosing the test methods and the criteria for their quotation in a particular engineering project.
- Description of some typical engineering solutions for test of specific analog blocks.
- Test method proposed for solving jitter caused test problems.

References

[1] An introduction to Mixed-Signal IC Test and Measurement, Mark Burns and Gordon W. Roberts, *Oxford university press*, 2001.

[2] Catalyst Mixed-Signal Programming, *Teradyne Inc*, March 2000.

[3] "Specification of the Bluetooth System" version 1.1, edited by the "Bluetooth Special Interest Group", Part A, *Bluetooth radio specification*.

[4] *Agilent web site reference materials*.

[5] ASSP product specification C150 version 0.6, *Alcatel Microelectronics*, 2001.

[6] Block review for test [2/12/2001], *Alcatel Microelectronics*.

Generation of a Test Strategy for Testing the Digital Part of an Integrated Circuit for Digital Wireless Short-Range Communication

Rumen Iv. Arnaudov[1]and Ivo G. Aldimirov[2]

Abstract – **Subject of this article is the generation and application of a Test strategy for measure and test of Application Specific Standard Product (ASSP) Integrated Circuit (IC) for digital wireless communication. It is given a description of the employed methods for testing the building blocks of the digital part of the device. They are based on the way they take place in the industrial test of the mixed-signal IC. Comment and quotation on some existing methods is made. It is emphasized on some potential problems that might influence the test time/cost, the accuracy and the stability of the tests. Suggestions for solving some of those problems are made.**

Keywords – **industrial test, mixed-signal Integrated Circuit, test methods, digital tests.**

I. Continuity Test [2]

Continuity test is included in the group of the digital tests because for its execution the digital hardware resources of the mixed-signal tester are used. The continuity tests are executed as a rule in the very beginning of the test program. The purpose of this test is to check the presence of the ESD protection structure of every functional pin of the device. Thus the reliable contact between the device under test (DUT) and the test hardware load board (LB) during the test program run is verified. Although simple this is very important test as it eliminates possible issues caused by misalignment of the docking system - handler to test head. It also prevents activation of potentially device and/or test hardware damaging functions when lack of continuity had occurred.

Major problem with this test is the test coverage versus the test time. Testing devices with big number of pins means multiplied test time hence cost increase.

All the supply pins are connected to GND and current is sourced to the functional pins. Then the voltage drop over the ESD structures towards VDD is measured. Afterwards the current direction is changed and the voltage drop over the ESD structures towards VSS is measured.

In general there are two approaches for performing the continuity test - parallel and serial. Serial test of the pins is a method with 100% coverage but is too slow and thus costly. The parallel method is fast because it uses the so-called per-pin - measurement units of the tester digital hardware.

Unfortunately there is one potential problem that must be considered up front. Whenever there is a short between two neighbouring pins it will result in the expected voltage drop over the ESD structure and the pins will pass the test undetected. The serial approach would not allow this. Third more economical approach should be used balancing between the test cost and the test coverage. Combination of the abovementioned methods is widely used and is executed on two passes. On the first pass every other pin is being tested while the rest of the pins are connected to GND. The next half of the pins is tested on the second pass. The result of the combination of the two approaches is maximum test coverage and acceptable cost.

II. Digital Tests - Iddq Measurement Method. [1-4]

The purpose of this method is to find out the presence of structure defects in the digital part of the device. The number of the used CMOS inverter cells or gates is normally used for description of the complexity of the digital part. Whenever the device has relatively high complexity of the digital, the test of every CMOS cell would be highly time-consuming thus unaffordable or in many cases even unfeasible. To test the presence of defected inverter cells the Iddq test method is widely used.

Iddq is the supply current consumed by the digital part when it had been driven to a static state (no switching inside) with all outputs left open. Automatically generated test pattern (ATPG) driving the digital part is stopped at the so called Iddq vectors where the majority of the P or N gates are in static state - only the P or the N transistors are in ***on*** state.

Fig. 1 shows CMOS inverter cell in a static state.

Obviously the current flowing into the inverter is negligible and is multiplied by the number of the gates in the digital structure. This way measuring the static current consumption after the ATPG pattern had been stopped at the predefined Iddq stop vectors, conclusion for possible defects presence can be drawn.

Whenever a digital part with high complexity has to be tested the Iddq measurement method is used as a rule in the silicon testing industry. The coverage of the method is specified by the digital design that generates the Iddq vectors.

[1]Rumen Iv. Arnaudov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: rarnaud@vmei.acad.bg

[2]Ivo G. Aldimirov is with the Air Traffic Services Authority (ATSA), Communications department at Plovdiv International Airport, e-mail:aldimirov@yahoo.com

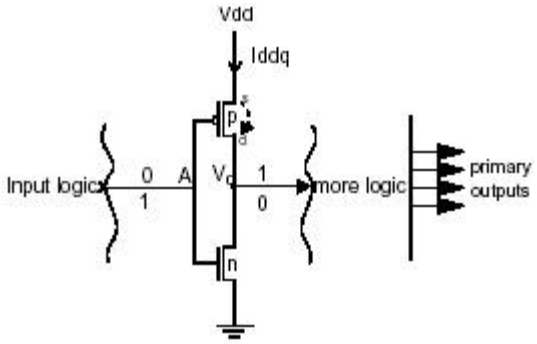


Fig. 1.

Depending of the technology used, the gates number and the temperature, the Iddq value varies and is in the range of few hundreds nano Amperes.

In the modern industrial mixed-signal testers the power supply units are integrated with the measurement facilities for measuring the supplied currents and voltages. In normal working mode the digital part normally consumes several mAmps. Stopping at Iddq vector quickly reduces the consumption down to hundred times.

Applying traditional approach may come across serious problems trying to make quick and accurate measurement of such a quickly switched low value current. Fig. 2 shows in principle the power supply for the digital part of a mixed-signal device and the parasitic capacitance of the supply lines.

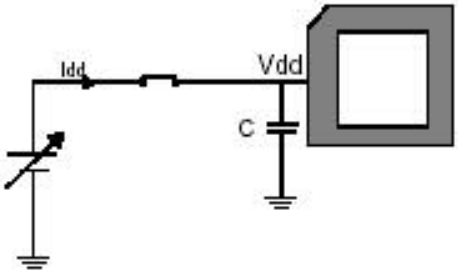


Fig. 2.

For accurate measurement of this many times lower current consumed by the same supply pin, the measurement range of the power supply unit has to be switched over. During this switch over, the inductance of the supply lines and the parasitic capacitance between them cause transients that might last for a number of milliseconds. It has to be added relevant amount of wait time before the real measurement to take place and eventually some filtering and DSP technique have to be used for averaging the measurement result. When digital design had been generated ***n*** numbers of Iddq vectors i.e. ***n*** times run of the test pattern and measure of Iddq, the test time is ***n*** times multiplied including the settle time which in many cases might be unacceptable. Some sources [4] give hardware methods for increasing the low current measuring accuracy, but the major problem - the settling time of the Iddq current that is multiplied by the number of the stop vectors is unresolved. In those cases it is not guaranteed that transients would not occur, more over in some specific cases the Iddq current even shows up oscillations. Possible and with certain efficiency solution would be the implementation of a second current source in parallel to the main one, programmed to provide (and measure) current in the expected range. RC product then could be added on this second supply line to lower the settling time of the switch over process and to produce stable and repeatable test results.

III. ADC/DAC Transfer Curve Tests [2]

There are many similarities testing ADC and DAC and few noticeable differences. The primary difference is the transfer curve - for each input code the DAC generates different output voltages as for a number of input voltages the ADC generates same output codes. Fig. 3 shows the transfer curves of both devices.

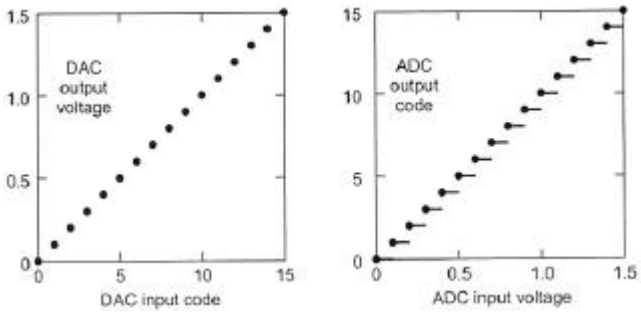


Fig. 3.

IV. DAC test

The transfer curve is being taken during linear (ramp up) increase of the input code. Best definition of the DACs parameters is found by computing the best-fit line. This approach is most preferred, as it is independent of the bit number.

A best-fit line is commonly defined as the line having minimum squared errors between its ideal, evenly spaced samples and the actual DAC output samples. Having this result in hand it is possible to calculate parameters of interest such as:

- **Monotonicity.** Monotonicity testing requires taking the discrete first derivative of the transfer curve, denoted here as $S(i)$, according to

$$S(i) = S(i+1) - S(i)\ ; \qquad (1)$$

  If derivatives are all positive for a rising ramp up input, then the DAC is said to be monotonic.

- **Differential nonlinearity**. The DNL curve represents the error in each step size, expressed in fractions of LSB. DNL is computed by calculating the first derivative of the DACs transfer curve, subtracting one LSB (i.e. V_{lsb}) from the derivative result, then normalising the result to one LSB:

$$\mathrm{DNL}(i) = [S(i+1) - S(i) - V_{lsb}]/V_{lsb}, \mathrm{LSB}\ ; \qquad (2)$$

- Integral nonlinearity. The integral nonlinearity curve is a comparison between the actual DAC transfer curve and the best-fit line.

$$\text{INL}(i) = [S(i) - S_{ref}(i)]/V_{lsb}, \text{LSB} ; \quad (3)$$

V. ADC Test - Linear Ramp Histogram Method

This method implies applying of a rising or falling linear ramp signal to the input of the ADC and collect samples from the ADC at a constant sampling rate. The ramp is set to rise or fall slowly enough that each ADC code is hit several times for example 16 or 32. The number of occurrences of each code is directly proportional to the width of the code. From the so acquired transfer curve the average value of the input voltage for each output code is calculated. The width of each code word - code width (i) in LSB is calculated according:

$$\text{code width}(i) = H(i)/h, \; i = 1, 2, ..., 2\text{pow}N - 2 , \quad (4)$$

where N is the number of the ADC bits, $H(i)$ is the number of the hits for the i-th word and h is the average number of hits for each code word.

For computing the best fit, INL and DNL curves it is used the histogram of the acquired results after their normalisation as it is shown on Fig. 4.

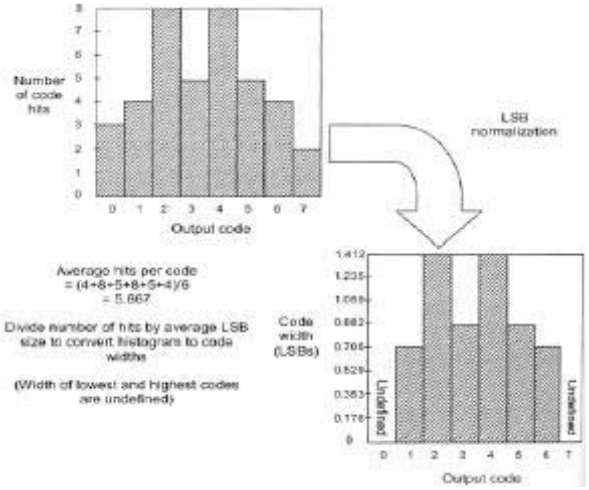


Fig. 4.

Based on the such obtained by the method results INL and DNL are calculated according to:

$$\text{DNL}(I) = \text{LSB}_{\text{code width}}(I) - 1, \quad I = 1, 2, ..., 2\text{pow}N - 2 ; \quad (5)$$

$$\text{INL}(I) = \sum_{k=1}^{I-1} \text{DNL}(k), \quad I = 1, 2, ..., 2\text{pow}N - 2; \quad (6)$$

♢ Local oscilator test - PLL

Although the ”analog” uses of the PLL, from the tester hardware and the employed test method point of view this is purely a ”digital” test. The purpose of the test is to determine the following parameters:

- *lock_time* or the settle time of the VCO for the programmed frequency;
- read out of the DTUNE code which shows up the automated digital tuning;
- CTUNE code representing the automated choice of capacitors bank for compensation of the process spread.

External clock signal is applied and using JTAG protocol it is asserted in the corresponding register the code of the channel (the working frequency). The time till synthesizer_lock goes high is measured - the VCO had settled. Using JTAG protocol the registers containing *DTUNE* and *CTUNE* codes values are read out. The test is executed twice - for the frequencies at the two ends of the working range. The described method practically covers the main PLL parameters and because of its relatively low complexity is cost/test time saving. The use of JTAG as a design for test (DFT) technique for the digital part of the IC provides the opportunity for fast and convenient from test engineering point of view test.

VI. Conclusions

This article analysis the employed methods for testing of the main building blocks of the digital part of an ASSP. It is pointed on the advantages of the proposed methods and on the potential problems as well. Approaches for resolving the problems are suggested:

- It is commented on a balanced approach for execution of continuity tests.
- It is pointed on some possible measurement problems which Iddq current test may come across and on a possible efficient solution for it.
- It is given a description of quick and effective method for testing PLL.
- It is commented on the basic notions of the ADC/DAC test and the way they take place in the industrial tests.

References

[1] *Agilent web site reference materials*.
[2] An introduction to Mixed-Signal IC Test and Measurement, Mark Burns and Gordon W. Roberts, *Oxford university press*, 2001.
[3] Catalyst Mixed-Signal Programming, *Teradyne Inc*, March 2000.
[4] **Keywords** - <*www.ttt.com*> Copyright (c) 1996,1997 T.T.T., Inc.

Lidar Measurements over the City of Sofia

Georgi Kolarov[1], Ivan Grigorov[2], Dimitar Stoyanov[3]

Abstract – **In this paper some results from measurements of aerosol layers in the atmosphere are presented. High repetition CuBr-laser is used in the aerosol lidar, made in the Institute of Electronics at the Bulgarian Academy of Sciences.**

Keywords – **Lidar, Aerosol layers in Atmosphere, Laser Radars.**

I. Introduction

The lidar methods for measurement of the aerosol content in the atmosphere are a powerful up-to-date means of control and analysis of the state of the environment. The aerosol lidars allow real-time tracking and mapping of the movement of aerosol fields over vast territories, without introducing additional perturbation flows in the natural evolution of the transportation processes. For quantitative analysis of the concentrations and chemical composition of aerosols, lidar measurements are being successfully combined with other conventional measurement methods, which allow assessment of the characteristics of the aerosol field at a local scale and subsequent calibration of the data from aerosol lidars. In the meantime, new mobile aerosol lidar systems are being developed (airborne and spaceborne lidars) that significantly increase the maximal spatial dimensions of the analyzed objects. Networks of lidar stations are being set up to research the climatologic characteristics and trans-continental transportation of aerosol layers (EARLINET, Asian lidar network).

In this paper we present some results from measurements of aerosol layers in the atmosphere, made using a high-repetition aerosol lidar. The lidar is constructed and functioning in the Institute of Electronics at the Bulgarian Academy of Sciences. Our previous measurements [1] with a lidar station of this type were aimed at mapping of aerosol fields over a vast region. The measurements were carried out in horizontal directions relative to the ground surface. A standard algorithm for determination of the extinction coefficient of the atmosphere from the lidar data was used [2]. Subsequently the data for the aerosol fields were recalculated for determination of the mass concentrations of dust contaminants after calibration of the lidar with local sampling devices.

[1]Georgi Kolarov is with the Institute of Electronics at the Bulgarian Academy of Sciences, 72, Tzarigradsko chausee, 1784-Sofia, Bulgaria, e-mail: kolarov@ie.bas.bg

[2]Ivan Grigorov is with the Institute of Electronics at the Bulgarian Academy of Sciences, 72, Tzarigradsko chausee, 1784-Sofia, Bulgaria, e-mail: ivangr@ie.bas.bg

[3]Dimitar Stoyanov is with the Institute of Electronics at the Bulgarian Academy of Sciences, 72, Tzarigradsko chausee, 1784-Sofia, Bulgaria, e-mail: dvstoyan@ie.bas.bg

II. The Lidar

Currently the lidar is part of the European scientific research lidar network (EARLINET) for study of climatologic characteristics of the atmospheric aerosols. This network comprises 22 lidar stations situated in a number of European countries: Germany, France, Italy, Great Britain, Spain, Portugal, Greece, Bulgaria, Belarus, Switzerland, Sweden, Poland, Slovenia. EARLINET was a 3-years long project under the 5-th Framework Programme of the Commission for scientific research in the European Union. Our lidar participated in the measurements under two work programmes of the project:

- WP2-Regular measurements – with the goal to establish a comprehensive climatological database of the vertical distribution of aerosol over stations of the network.
- WP7-Observation of special events – observation of specifically high aerosol loads in the lower troposphere, resulting from extreme dust events (transport of Saharan dust, break of forest/industrial fires, intense photoshemical smog episodes, volcano eruption etc.).

For participation in the EARLINET project both the installations and methodology of the lidar were modified. The improvements of the apparatuses allowed for really all day long lidar measurements. The installation of a narrow diaphragm and additional filtering optics contracted the range of vision of the receiving telescope to 0,2 mrad and decreased significantly the registration of the parasitic background daylight in the atmosphere. The usage of a laser with high-frequency of repetition of the impulses, 14 kHz, allows registration of the reflected signals with photomultiplier in photon counting mode for the whole way of sounding. The technical characteristics of the lidar are shown in Table 1.

Table 1. Technical characteristics of the lidar

CuBr-vapor laser	Mean power: 3 W Wavelength: 510 nm, 578 nm Divergence: 3 mrad
Collimator	Output divergence: 0.6 mrad
Telescope	Focus length: 1000 mm Aperture: 200 mm
Photon detector	Photomultiplier EMI9863QB100 in photon counting mode
Acquisition system	Digital correlator "Malvern" K7023 Range resolution: 150 m Multichannel regime: 70 chnls/samples/
Computer	Pentium 4

The measurements under WP2 were carried out regularly twice a week, at noon and sunset on Monday, and at sunset on Thursday. The measurements under WP7 were carried out

upon notification by the programme coordinator for upcoming dust events above the territory of Sofia, based on satellite observations and weather forecasts. The lidar profiles of the laser emission, backscattered in the atmosphere were registered with accumulation time of 1 min. for every sample. In addition, averaging was performed by summation of the data of 30 profiles, thus the effective measurement time for each profile amounting to 30 min. The data processing and calculation of the extinction coefficient of the atmosphere were made by a computer programme in Matlab environment, developed in the Institute of Electronics at BAS. The programme implements an algorithm for lidar data processing by Fernald's method [3].

III. Results

The figures presented below show two types of aerosol layers distribution in the atmosphere. The measurement of 21.10.2002 (Fig. 1) is under WP2 programme, in a day with clear atmosphere, therefore no significant aerosol layers are observed in height, except the Planetary Bounder Layer (PBL) aerosol layer at 1-1,5 km, existing predominantly due to anthropogenic factors. The measurement of 15.07.2002 (Fig. 2) is under WP7 programme, when aerosol layers of Sahara dust have reached Sofia. An increase of the height to 2 km is distinctly observed for the ground aerosol layers. It is due to the sedimentation of aerosol particles from the upper aerosol layer. An additional layer at height 3-3,5 km is also observed, which is most probably in result of the transportation of Sahara dust. The low optical density and relative stability over time are typical for this aerosol layer, which in addition to the prognostic information bear out the conclusion that this is an aerosol layer containing Sahara dust.

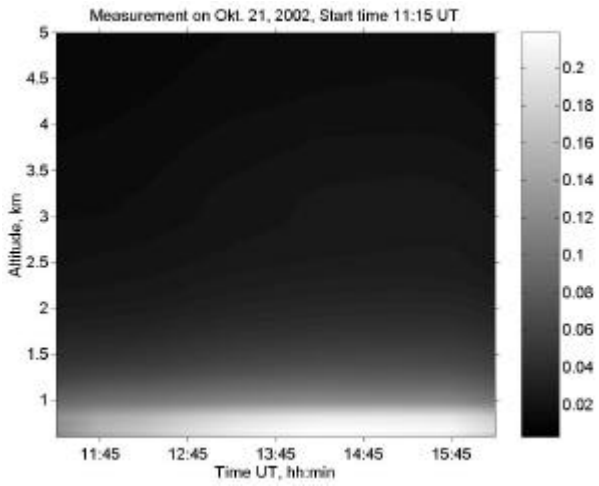


Fig. 1. Lidar measurement in circumstances of clear atmosphere. The scale in the right gives the values of the extinction coefficient in km^{-1}.

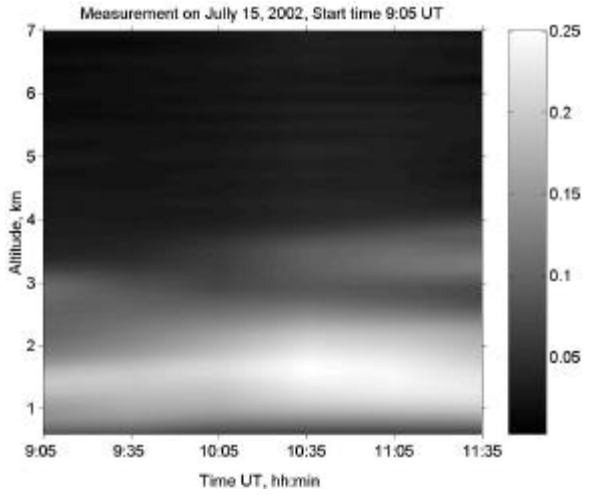


Fig. 2. Lidar measurement in circumstances of multiple aerosol layers in the atmosphere. The scale in the right gives the values of the extinction coefficient in km^{-1}.

IV. Conclusion

The data of presented measurements are stored in the database that has been established by the EARLINET project. This database is now growing continuously and is of very high scientific interest, because it is by far the most comprehensive quantitative data set on the vertical distribution of aerosols on a continental scale.

Acknowledgement

The authors would like to express their gratitude to the Commission for scientific research in the European Union and to principal investigators of the EARLINET project for the financial support of this work.

References

[1] Tz.Mitzev, I.Grigorov, G.Kolarov, D.Lolova, *Investigation of transborder pollution by combining remote lidar sounding and stationary gaz sampling*, SPIE's Int. Conf. EUROPTO'95, SPIE Proc. vol.2506, pp.310-318, Munich, Germany, 1995

[2] J. D. Klett, *Stable analytic inversion solution for processing lidar returns*, Appl.Opt. 20, pp. 211-220, 1981

[3] http://lidarb.dkrz.de/earlinet/scirep1.pdf [*EARLINET: Scientific report for the period Feb. 2000 to Jan. 2001*].

Comparison by Simulation of Torque Control Schemes for Electric Drive Application

Nebojsa Mitrovic[1], Vojkan Kostic[2], Milutin Petronijevic[3]and Borislav Jeftenic[4]

Abstract – **This paper presents results of an investigation into the suitable torque control schemes for high performance induction motor drive application. Three different control schemes for direct torque control (DTC) are considered: classical, modified and twelve sector DTC. A brief overview of the operation of each scheme is presented followed by simulation results**

Keywords – **induction motor drive, direct torque control**

I. Introduction

The method of Direct Torque Control use feedback control of torque and stator flux, which are computed from the measured stator voltages and currents of induction motor [1,2]. As the method does not use position or speed sensor to control the machine and use its output currents and terminal voltages, this is also called as direct vector control scheme. The scheme uses stator flux-linkages control which is directly proportional to the induces emf. The method uses a stator reference model of the induction motor for its implementation, avoiding the trigonometric operations in the coordinate transformations of the synchronous reference frames.

II. Principles of Direct Torque Control

The implementation of the DTC scheme requires flux linkages and torque computations and generation of switching states through a feedback control of the torque and flux directly without inner current loops.

The stator q and d axis flux linkages are

$$\lambda_{qs} = \int (V_{qs} - R_s i_{qs}) dt \tag{1}$$

$$\lambda_{ds} = \int (V_{ds} - R_s i_{ds}) dt \,, \tag{2}$$

where R_s – stator resistance, V_{qs}, V_{ds}, i_{qs}, i_{ds} – qd voltage and current components.

Consider the inverter shown in Fig. 1. The terminal voltage (V_a) with respect to negative of the dc supply is determined by a set of switches, Sa, consisting switchig device T1 and T4 as shown in Table 1. The switching of Sb and Sc sets for line b and c can be similarly derived. The total number of switching states possible with Sa, Sb, and Sc is eight and they are shown in Fig. 2. The stator q and voltages for each state are given by

$$V_{qs} = V_{as} \tag{3}$$

$$V_{ds} = \frac{1}{\sqrt{3}} (V_{cs} - V_{bs}) = \frac{1}{\sqrt{3}} V_{cb} \,. \tag{4}$$

The limited states of the inverter create discrete movement of the stator voltage phasor V_s, consisting of the resultant of V_{qs} and V_{ds}.

For control of voltage phasor both in its magnitude and phase, the requested voltage vector's phase and magnitude are sampled, say once every switching period. The phase of requested voltage vector identifies the neariest two nonzero voltage vectors.

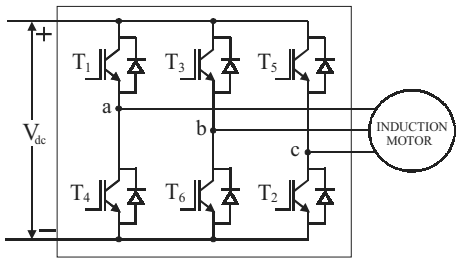


Fig. 1. Power circuit configuration of induction motor drive

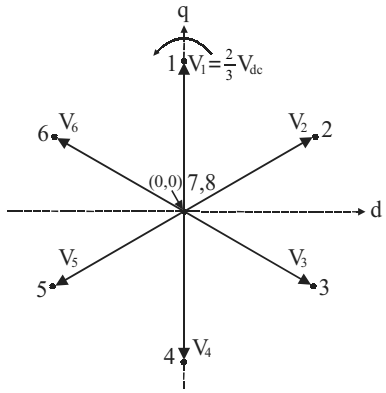


Fig. 2. Inverter output voltages

Table 1. Switching state of inverter phase leg a

T_1	T_4	Sa	V_a
On	Off	1	V_{dc}
Off	On	0	0

[1]Nebojsa Mitrovic is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Nis, Yugoslavia, E-mail: nesa@elfak.ni.ac.yu

[2]Vojkan Kostic is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Nis, Yugoslavia, E-mail: nikola2105@elfak.ni.ac.yu

[3]Milutin Petronijevic is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Nis, Yugoslavia, E-mail: milutin@elfak.ni.ac.yu

[4]Borislav Jeftenic is with the Faculty of Electrical Engineering, Bulevar Revolucije 73, 11000 Belgrade, Yugoslavia, E-mail: jeftenic@etf.bg.ac.yu

III. Classical DTC Model (C-DTC)

A uniform rotating stator flux is desirable, and it occupies one of the sectors at any time, Fig. 3. The stator flux phasor has a magnitude of λ_s with instantaneous position θ_{fs}.

If the stator flux phasor is in sector 2, Fig. 3, the left influencing voltage phasor has to be either V6 or V1. As seen from phasor diagram, in case switching voltage phasor V1, the flux phasor increases in magnitude. In case of of phasor V6, it decrease. This implies that the closer voltage phasor set increase the flux and the farther voltage phasor set decreases the flux and both of them change (rise) the flux phasor in position. Similarly for all other sectors, the switching logic can be developed. A flux error $(\lambda_s^* - \lambda_s)$ thus determines which voltage phasor has to be called, and this flux vector is converted to a digital signal S_λ with hysteresys controller with hysteresis band of $\delta\lambda_s$. The switching logic to realize S_λ is given in Table 2.

Table 2. Switching logic for flux error

State	S_λ
$\lambda_s^* - \lambda_s > \delta\lambda_s / 2$	1
$\lambda_s^* - \lambda_s < -\delta\lambda_s / 2$	0

Torque control is exercised by comparison of the command torque to the torque measured from the stator flux linkages and stator currents as

$$T_e = \frac{3}{2}\frac{P}{2}(i_{qs}\lambda_{ds} - i_{ds}\lambda_{qs}) \,, \tag{5}$$

where P is pole number.

Torque error is processed through hysteresis controller to produce digital outputs, S_T as shown in Table 3. Interpretation of S_T ai as follows: when it is 1 amounts to increasing the voltage phasor, 0 means to keep it at zero, -1 requires retarding the voltage phasor.

Combining the flux error output S_λ, the torque error output S_T, and the sextant of the flux phasor S_θ, a switching

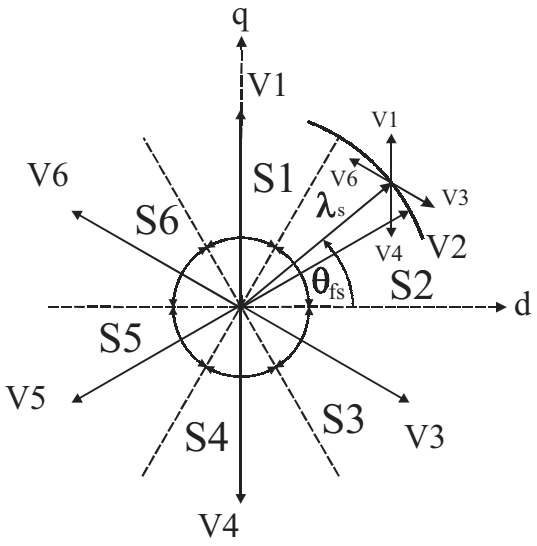


Fig. 3. Division of sectors for stator flux identification (c-DTC)

Table 3. Switching logic for torque error

State	S_T
$T_e^* - T_e > \delta T_e / 2$	1
$-\delta T_e / 2 \leq T_e^* - T_e \leq \delta T_e / 2$	0
$T_e^* - T_e < -\delta T_e / 2$	-1

table can be realized to obtain the switching states of the inverter. The sectors of the stator flux space vector are denoted from S1 to S6. Stator flux modulus error after the hysteresis block can take just two values. Torque error after the hysteresis block can take three different values. The zero voltage vectors V7 and V8 are selected when the torque error is within the given hysteresis limits, and must remain unchanged. Finally, the DTC classical (c-DTC) look up table is shown in Table 4.

In the classical DTC, there are several drawbacks [3,4]. Some of them can be summarized as follows:

- slow response in both start up and changes in either flux or torque,
- large and small errors in flux and torque are not distinguished. In other words, the same vectors are used during start up and step changes and during steady state.

Table 4. Switching states for c-DTC

S_λ	S_T	S1	S2	S3	S4	S5	S6
1	1	V6	V1	V2	V3	V4	V5
1	0	V8	V7	V8	V7	V8	V7
1	-1	V2	V3	V4	V5	V6	V1
0	1	V5	V6	V1	V2	V3	V4
0	0	V7	V8	V7	V8	V7	V8
0	-1	V3	V4	V5	V6	V1	V2

IV. Modified DTC Model (M-DTC)

In order to overcome the mentioned drawbacks, there are different solutions. First idea that comes up, when it is tried to improve the DTC by means of changing the tables, is to use six sectors, as in classical DTC, but changing the zones. Hence, instead of having as a second sector the zone from 0° up to 60°, it will be from 30° up to 90°. It can be observed that in this case, the states not used in the second zone will be V3 and V6 instead of V2 and V5. This novel sector division is shown in Fig. 4.

Control of the flux and torque can be done by the similar procedure as for the classical model.

Table 5 shows the m_DTC look up table for all its six sectors. It can be seen that the states V2 and V5, are not used in the classical DTC (c_DTC) because they can increase or decrease the torque at the same sector depending on if the position is in its first 30 degrees or in its second ones. In the modified DTC (m_DTC), V3 and V6 are the states not used. However, now the reason is the ambiguity in flux instead of

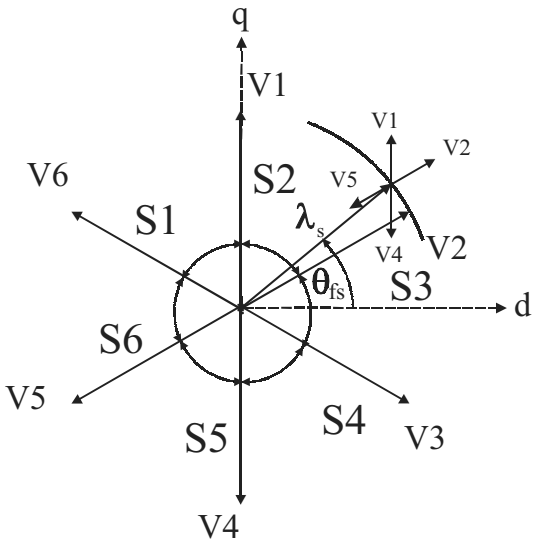


Fig. 4. Modified DTC and its sectors

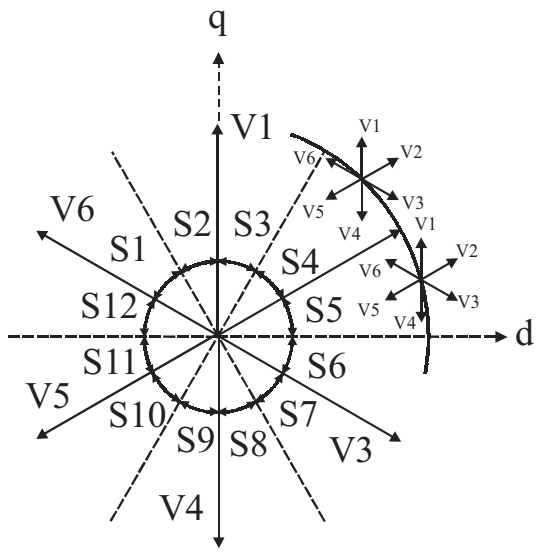


Fig. 5. Twelve sectors DTC

torque, as it was in the c_DTC. This is considered to be an advantage in favour of the m_DTC as long as the main point is to control the torque. Therefore, it is better to loose the usage of two states for flux ambiguity that for torque one [5].

Table 5. Switching states for m-DTC

S_λ	S_T	S1	S2	S3	S4	S5	S6
1	1	V6	V1	V2	V3	V4	V5
1	0	V8	V7	V8	V7	V8	V7
1	-1	V1	V2	V3	V4	V5	V6
0	1	V4	V5	V6	V1	V2	V3
0	0	V7	V8	V7	V8	V7	V8
0	-1	V3	V4	V5	V6	V1	V2

V. Twelve sector DTC model (12_DTC)

In classical DTC there are two states per sector that present a torque ambiguity. Therefore, they are never used. In a similar way, in the modified DTC there are two states per sector that introduce flux ambiguity, so they are never used either. It seems a good idea that if the stator flux locus is divided into twelve sectors instead of just six, all six active states will be used per sector. Consequently, it is arisen the idea of the twelve sector modified DTC (12_DTC). This novel stator flux locus is introduced in Fig. 5. Notice how all six voltage vectors can be used in all twelve sectors. However, it has to be introduced the idea of small torque increase instead of torque increase, mainly due to the fact that the tangential voltage vector component is very small and consequently its torque variation will be small as well.

As it has been mentioned, it is necessary to define small and large variations (S_T=1 - torque increase, S_T=2- torque small increase, S_T=3 - torque small decrease, S_T=4 - torque large increase). It is obvious that V2 will produce a large increase in flux and a small increase in torque in sector $S5$. On the contrary, V1 will increase the torque in large proportion and the flux in a small one.

Therefore, the torque hysteresis block should have four hysteresis levels and eight levels of flux and torque variation. Finally, the look up table is presented in Table 6.

Table 6. Switching states for 12-DTC

		Sector number S_θ											
S_λ	S_T	1	2	3	4	5	6	7	8	9	10	11	12
1	1	5	6	6	1	1	2	2	3	3	4	4	5
1	2	6	6	1	1	2	2	3	3	4	4	5	5
1	0	8	8	7	7	8	8	7	7	8	8	7	V7
1	3	1	1	2	2	3	3	4	4	5	5	6	6
1	4	1	2	2	3	3	4	4	5	5	6	6	1
0	1	4	5	5	6	6	1	1	2	2	3	3	4
0	2	4	4	5	5	6	6	1	1	2	2	3	3
0	0	7	7	8	8	7	7	8	8	7	7	8	8
0	3	3	8	4	7	5	8	6	7	1	8	2	7
0	4	2	3	3	4	4	5	5	6	6	1	1	2

VI. Simulation Results

Simulations have been carried out for the comparison of a described schemes. The simulations were conducted using Matlab/Simulink simulation package. The DTC drive for all simulation were run with speed feedback. The control algorithms are taking into account by the appropriate look-up tables. The system is discredited with sample time $T_s = 2 \cdot 10^{-6}$ s.

Simulation parameters are as follows;
Motor rating: 3 phase, 2 pole, 380V, 37 kW,
Parameters: R_s=0.087 Ω, R_r=0.228 Ω, L_s=35.5 mH, L_m=34.7 mH, L_r=35.5 mH.

Fig. 6 shows actual speed, motor torque and stator flux of c-DTC, m-DTC and 12_DTC schemes, speed reference is set

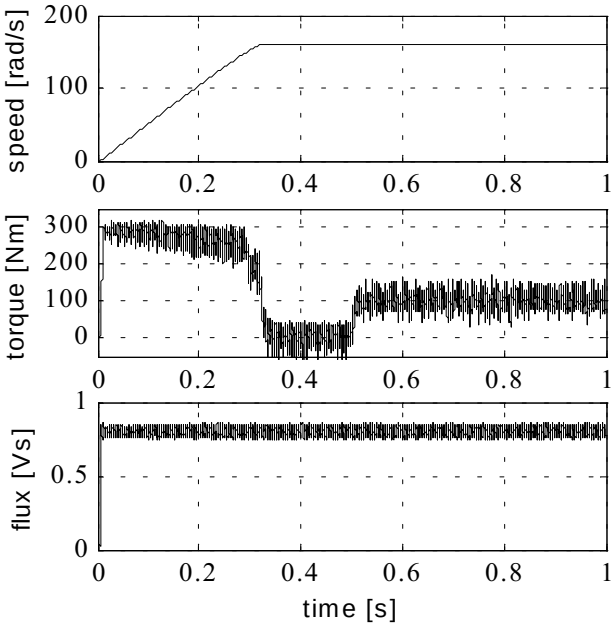


Fig. 6a. Dynamic performance of the drive with c-DTC

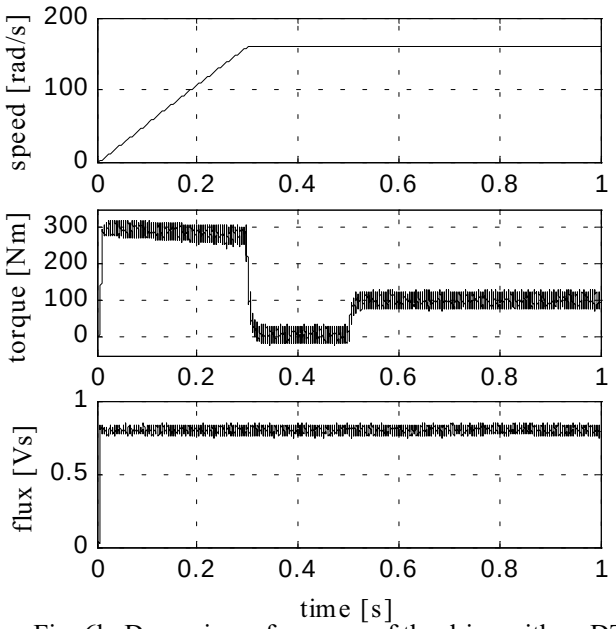


Fig. 6b. Dynamic performance of the drive with m-DTC

to 160 rad/s, torque limit at start-up is $3T_n$. At t=0.5 s load torque is set to 100 Nm.

In all cases it is possible to control directly the stator flux and the torque by selecting the appropriate inverter state. According to the Fig. 6, smaller variation can be seen in the torque and flux with m-DTC and 12_DTC scheme than c-DTC

VII. Conclusion

In this paper, three different schemes for direct torque control for induction motor drives are presented. In all cases it is possible to control directly the stator flux and the torque by

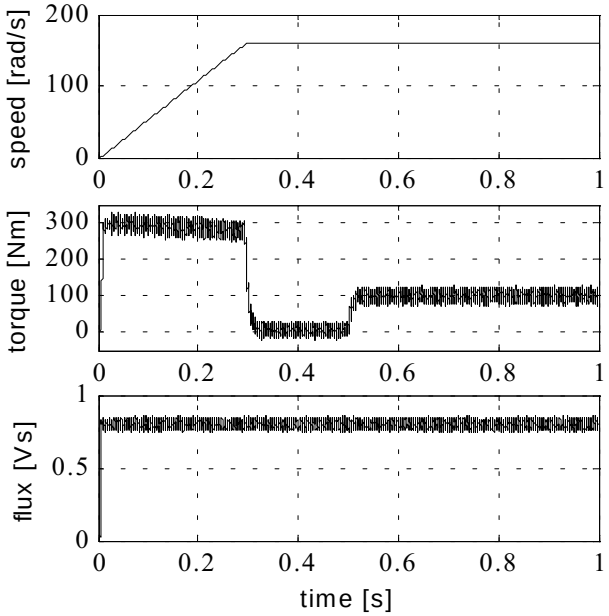


Fig. 6c. Dynamic performance of the drive with 12-DTC

selecting the appropriate inverter state. Its main features are as follows: direct torque control and direct stator flux control, indirect control of stator currents and voltages, approximately sinusoidal stator fluxes and stator currents, high dynamic performance even at locked rotor. Some disadvantages are present: possible problems during starting requirement of torque and flux estimators, implying the consequent parameters identification, inherent torque and flux ripples.

A further publication may show drive behaviour at low speed and dead time influence.

References

[1] I.Takahashi, T.Noguchi,"A new quick-response and high-efficiency control strategy af an induction motor," IEEE Trans. on Ind. Appl., Vol.22, No.5, pp.820-827, 1986.

[2] D. Casadei, G. Grandi, G.Serra, A. Tani," Switching Strategies in direct Torque Control of Induction machine", International Conf. on Electrical Machines, Paris, France, 5-8 Sept 1994.

[3] D. Casadei, G. Grandi, G.Serra, A.Tani," Effect of flux and torque hysteresis band amplitude in direct torque control of induction motor", IECON '94., Bologna, Italy , 5-8 Sept 1994.

[4] T.G. Habatler, F.Profumo, M. Pastorelli, L. Tolbert,"Direct Torque Control of Induction Machines using Space Vector Modulation", IEEE Trans. on Ind. Appl., Vol.28, No.5, pp.1045-1053, Sept/Oct 1992.

[5] N.R.N. Idris and A.H.M. Yatim, "Reduced Torque Ripple And Constant Torque Switching Frequency Strategy For Direct Torque Control Of Induction Machine", *In Conf Rec. IEEE-APE*C, pp. 154-161, vol .1, 2000.

Drivers for High Frequency Power Supply

Iliya Nemigenchev[1], Iliya Nedelchev[2]

Abstract – **The reliable and flawless, as well as steady, work of the high-frequency power devices is provided by the way their power switches have been controlled. The present paper treats a development of MOSFET drivers for controlling switch transistor sets of a high frequency power supply designed for induction heating with working frequency of 1÷1.5 MHz and output power of 750 W, transferred on a resonance load.**

Keywords – **Power Supply, Inverter, Induction Heating.**

I. Introduction

There is a large variety of industrial processes that require reliable, low cost, high frequency power for an induction heating. The technologies for welding and soldering of non – ferrous metals with a thickness of microns are impossible without this induction heating [1-5].

The results from the investigation of frequency power supply for soldering of non-ferrous metals with P_{out} = 750 W; f_{in} = 1 ÷1.5 MHz; U_{ps} = 150 V DC are presented in the report.

The soldering of non – ferrous metals is made with an inductor L_{Load} = 3.9 μH compensated in the resonant parallel circuit from a high frequency capacitor C_{Load} = 5.6 nF. The equivalent resonant resistance is $R_e \approx 572\ \Omega$.

The block circuit of the high frequency power supply is shown in Fig. 1. The impedance matching on the load is made by an output transformer with a transformation ration 5:1.

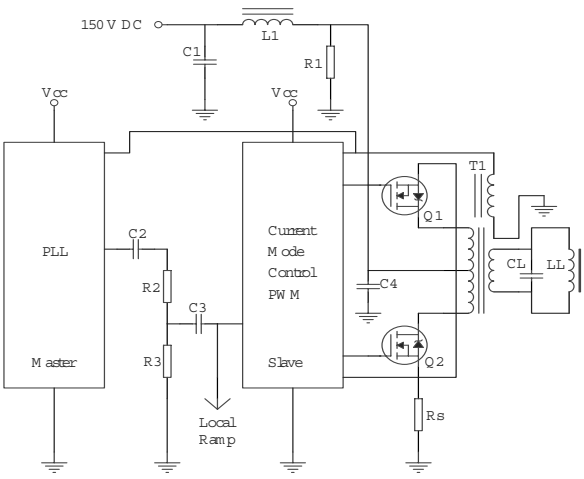


Fig. 1. Block circuit of the high frequency power supply

A push – pull configuration for inverter is selected. In the simultaneously are introduced voltage and current mode controlled to keep up the output voltage and the output current in the MOSFET switches. The output current is stabilized by pulse width modulation (PWM) system.

II. Operating Principle

The requirement in the induction heating for a higher frequency and an improved efficiency has lead to the need for using MOSFETs. The power MOSFETs have several parasitic elements, capacitance and inductances, with inhibit high speed operation and will affect its switching behavior and its power dissipation. Consequently the type of drive circuitry is very important for the switching one power MOSFETs, the reduction of the power dissipation and the operation at high frequency mode.

The electrical circuits with a push/pull drive circuit and a Power MOSFET are given in Fig. 2.

The modelling and the calculation of the gate driver is presented by the simplified MOSFET model and all associated inter connections. The equivalent electrical circuit is shown in Fig. 3 [6,7], when:

- R_D = 0.54 Ω is the resistance of the driver;
- $R_{G,I}$ = 1.8 Ω I - the internal gate mesh resistance;
- R_C = 0.05 Ω - their interconnection;
- C_D = 81 pF - the capacitance of the driver
- C_C = 50 pF- the capacitance of the interconnection,

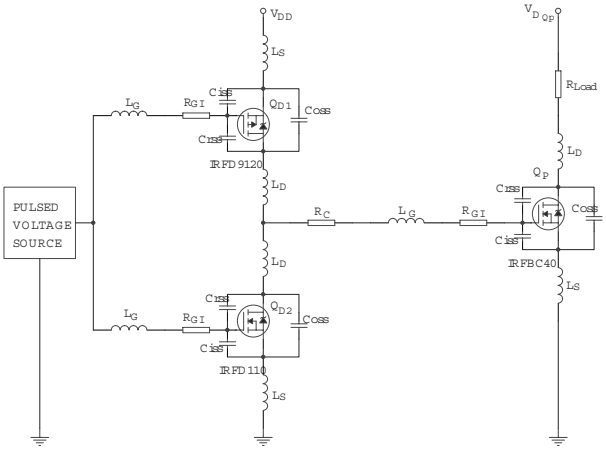


Fig. 2. Electrical circuit – a push/pull drive circuit and a power MOSFET

[1]Prof. Dr. Iliya N. Nemigenchev is with department Communications Technics and Technologies, Technical University, "H. Dimitar" 4, 5300 Gabrovo, Bulgaria, E-mail: nemig@tugab.bg

[2]Iliya V. Nedelchev is with department Communications Technics and Technologies, Technical University, "H. Dimitar" 4, 5300 Gabrovo, Bulgaria, E-mail: ilned@tugab.bg

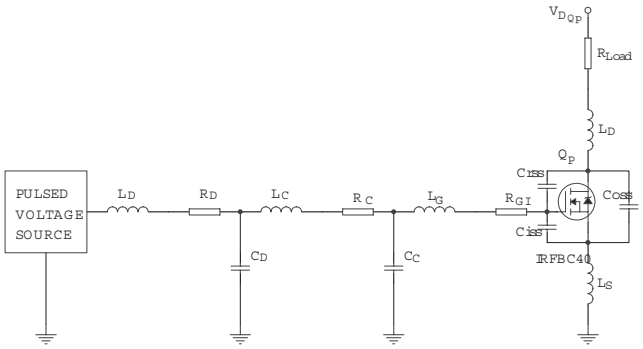


Fig. 3. Equivalent circuit of the driver

- C_{iss} = 1300 pF- the internal capacitance of the Power MOSFET;
- C_{rss} = 30 pF - reverse transfer capacitance;
- L_G = 7.5 nH - the inductance of the gate;
- L_S = 7,5 nH - the inductance of the source;
- L_D = 10 nH - the inductance of the driver;
- L_C - the inductance of the interconnection.

III. Calculation

From the theoretical model and equations, given in [8,9], the next calculation we have:

- resistance of the circuit

$$R_L = R_D + R_C + R_{GI}\ , \quad (1)$$
$$R_L = 0,54 + 0,05 + 1,80 = 2,39\ \Omega\ ;$$

- capacitance of the circuit

$$C_{FET} \cong C_{iss} + C_{rss}\frac{\Delta\left(V_{DS} - V_{GS}\right)}{\Delta V_{GS}}, \quad (2)$$
$$C_{FET} = 1300 + 30\frac{600-8}{8} = 3520\,pF,$$

$$C_L = C_{FET} + C_C + C_D = 3651\,pF\ , \quad (3)$$
$$C_L = 3520 + 50 + 81 = 3651\,pF\ ;$$

- gate switching time

$$T_r = 2R_LC_L, \quad (4)$$
$$T_r = 2 \cdot 2.39 \cdot 3.651 \cdot 10^{-9} = 17.45\ nS\ ;$$

- inductance of circuit

$$L_L \cong \left(\frac{2T_r}{\pi\sqrt{C_L}}\right)^2 = \left(\frac{2 \cdot 17.45 \cdot 10^{-9}}{\pi\sqrt{3.651 \cdot 10^{-9}}}\right)^2 = 33.8\ nH \quad (5)$$

- corrected rise time

$$T_{r(corrected)} \cong \sqrt{\left(\frac{\pi\sqrt{L_LC_L}}{2}\right)^2 + \left(2R_LC_L\right)^2} = \sqrt{\left(\frac{\pi\sqrt{33.8 \cdot 10^{-9} \cdot 3.65 \cdot 10^{-9}}}{2}\right)^2 + \left(2 \cdot 2.39 \cdot 3.85 \cdot 10^{-9}\right)^2} = 24,67\ nS \quad (6)$$

- gate current

$$I_{Gpk} \cong \frac{C_L \cdot V_{DD}}{T_{r(corrected)}} = \frac{3.65 \cdot 10^{-9} \cdot 8}{24.67 \cdot 10^{-9}} = 1.18\ A \quad (7)$$

- gate drive voltage

$$V_{Gpk} \cong \frac{I_{DS}}{g_{fs}} + V_{TH} + R_GI_{Gpk} + L_L\frac{I_{Gpk}}{T_r} = \frac{5}{6.1} + 4.2 + 2.39 \cdot 1.18 + 33.8\frac{1.18}{24.67} = 9.46\ V\ , \quad (8)$$

g_{fs} - forward transconductance;

- gate power dissipation

$$P_{Gd} = C_{(FET)} \times V_{DD}^2 \times f_s = 0.337\ W \quad (9)$$

IV. Experimental Results

The experimental results from the simulation model core shown in Fig. 4. *Protel 99 SE* for the simulation is used. Gate and

Drain voltages are given in Fig. 5.

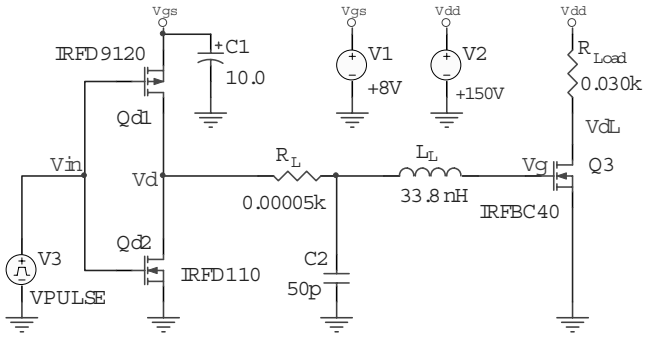


Fig. 4. The simulation model

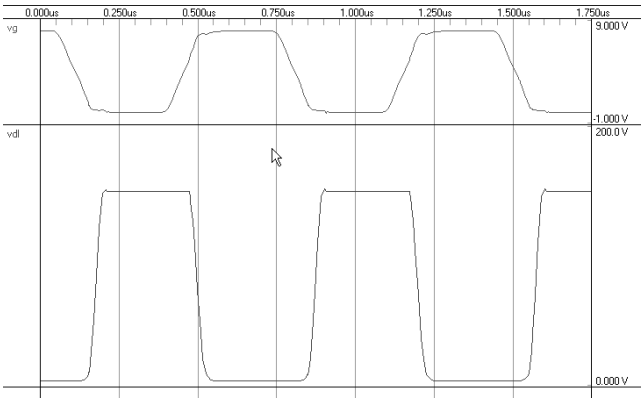


Fig. 5. Waveforms V_{GS} and V_{DS}

The load of drive is R_L = 2.39 Ω, L_L = 33.8 nH and maximum frequency 1,5 MHz. The load of the power MOSFET is R_{Load} = 30 Ω. The driver works steadily as the optimal switch times given by the company producing MOSFET transistors have been reached at.

At inductivity increased to a definite value, a sudden driver hesitation appears follower by unsteady work. The same can be observed when the value of the resistor R_L is increased. This is shown in Fig. 6.

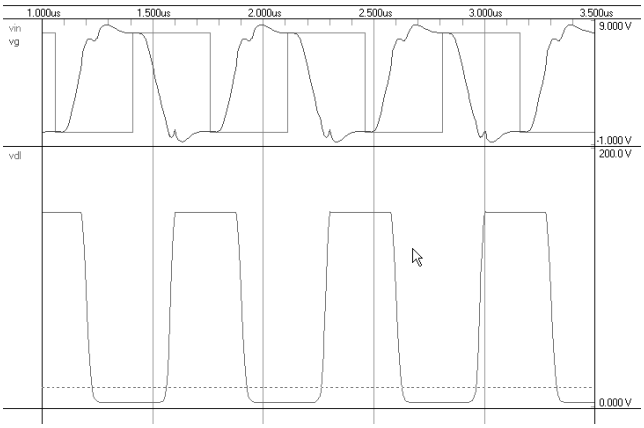


Fig. 6. Unsteady work of the driver

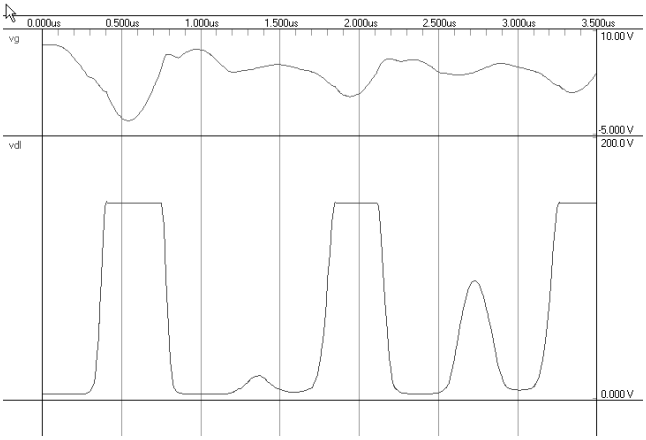


Fig. 7. Breakdown regime of the drive

The impedance increase in the Gate circuit R_L, C_C and L_L, as a result of poor design of a print chip and a poorly executed assembly, may lead to unsteady work of the driver as a whole , an oscillation of harmonious frequency in the driver circuit and destruction of the converter in general. Such a working regime is shown in Fig. 7.

Table 1. Power Consumption of Gate Drive

F [MHZ]	0,8	0,9	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7
P [mW]	180	203	225	248	270	293	315	338	360	383

The research work carried out for the power released in the Gate of the powerful transistor show that the gate depends linearly on the working. The dependence $P_G = f\,(F)$ is given in Fig. 8.

On the basis of the research conducted, a sample model of the High Frequency Power Supply for induction heating was designed, having output power P_{out} = 750 W; f_{out} = 1÷1.5 MHz; U_{PS} = 150 V DC working with resonance load L_{Load} = 3.9 μH; C_{Load} = 5.6 nF, with equivalent resistance $R_e \approx 572\,\Omega$. He work forms demonstrating the driver's work are shown in Fig. 9.

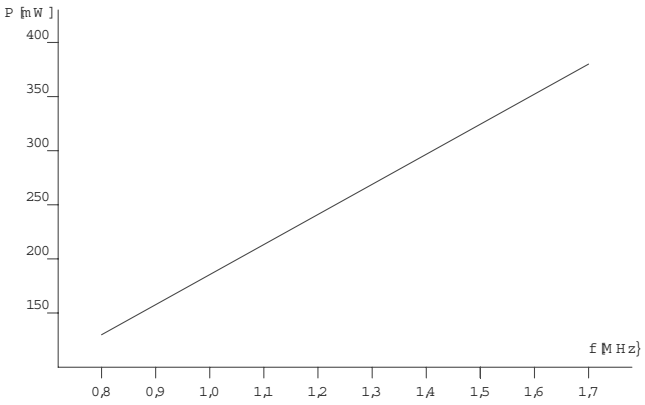


Fig. 8. Dependence $P_G = f\,(F)$

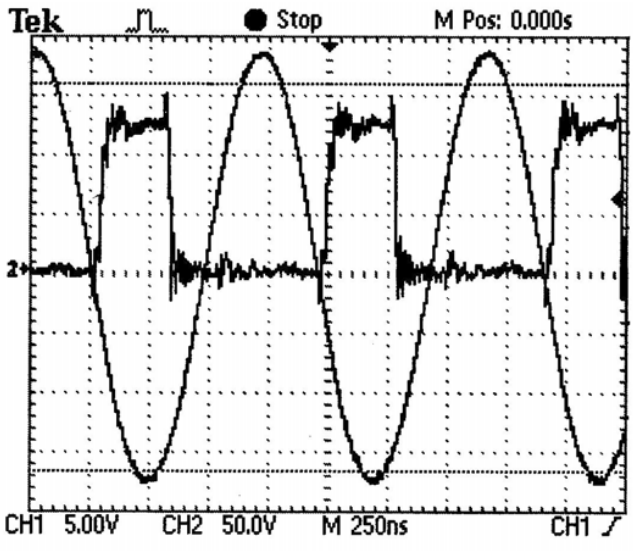


Fig. 9. U_{GS} and U_{Load}

V. Conclusion

The reliable and flawless work of the High Frequency Power Supply depends exceptionally on the precise drivers design. Drivers control the powerful switches, as the reduction of the parasite components appearing as a result of incorrect construction and assembly are of great importance.

References

[1] I. Nemigenchev, I. Nedelchev *High-Frequency Power Supply for Induction Heating* IWKM 2002

[2] L. Woffard, - *New Pulse Width Modulator Chip Controls, 1 MHz Switchers* – U-107; Unitrode Applications Handbook 1987/88.

[3] L. Woffard, - *New Pulse Width Modulator Chip Controls, 1 MHz Switchers* – U-107; Unitrode Applications Handbook 1987/88.

[4] B. Andreycak – *1.5 MHz Current Mode IC Controlled 50 Watt Power Supply*, Proceedings of the High Frequency Power Conversion Conference 1986

[5] L. Dixon, – *Closing the Feedback Loop Section* C1 – Unitrode Power Supply Design Seminar Book, SEM – 500

[6] S.Clemente, B. R. Pelly *Understanding HEXFET Switching Performance* AN 947 International Rectifier

[7] S. Malouyans, *Spice Computer Models for HEXFET Power MOSFETs* AN 975B International Rectifier

[8] K. Dierberger *Gate Drive Design for Large die MOSFETS* ART 9302 PCIM ,93 USA

[9] G. Krause, *Gate Drive Design for Swithch-Mode Application* Application Note IXYS

3D Numerical Analysis of Electromagnetic Systems with Magnetic Circuit Using Impedance Boundary Condition

Marin Dimitrov, Stoimen Balinov and Pavel Mintchev[1]

Abstract – **The rates of convergence of the edge element method (EEM) coupled with impedance boundary condition (IBC) and of the finite element method based on a gauged system of differential equations for the magnetic vector potential and coupled with IBC are compared in this paper in the case of numerical analysis of magnetic systems with concentrating and shielding magnetic cores. It is established that EEM coupled with IBC has a substantial advantage.**

Keywords – **crucible furnaces, edge element method, impedance boundary condition, magnetically shielded systems.**

I. Introduction

The use of impedance boundary condition (IBC) for 3D numerical analysis of the electromagnetic field in systems with eddy currents results in substantial acceleration of the calculations. A numerical model coupling the finite-element method (FEM) with IBC and based on a gauged system of differential equations for the magnetic vector potential (MVP) is proposed in [1]. The gauging of the equations for the MVP improves the convergence but can result in violation of the continuity conditions on the boundaries between regions with different permeabilities [2,3]. A coupling of the IBC and the edge-element method (EEM), for which the continuity conditions can't be violated from a theoretical point of view, is proposed in [4] for the study of an induction heater with a magnetic core for concentrating the magnetic flux. It is established in [4] that the use of EEM coupled with IBC results in a substantial improvement of the convergence rate, providing at the same time good accuracy. The non-linear properties of the ferromagnetic materials are not taken into account in this study.

The purpose of the present paper is a comparison of the rates of convergence of both formulations: EEM coupled with IBC [4] and FEM coupled with IBC [1,5,6] in the case of 3D numerical analysis of systems with concentrating and shielding magnetic cores, taking into account the non-linear properties of the materials.

[1]Marin Dimitrov, Stoimen Balinov and Pavel Mintchev are with the Institute of Metal Science - Bulgarian Academy of Sciences, 67 Shipchensky Prohod St., 1574 Sofia, Bulgaria, E-mail: marin.d@ims.bas.bg

II. Formulation of the Problem for 3-D Numerical Analysis of the Electromagnetic Field by Means of EEM Coupled with ibc

The following equation for the magnetic vector potential (MVP) is used:

$$rot\left(\frac{1}{\mu} rot\, A\right) + j\omega\sigma\, A = J\ , \tag{1}$$

where μ is permeability, $\omega = 2\pi f$, f is frequency, σ is conductivity, J is the source current density and $j = \sqrt{-1}$.

The following relationships can be written in the case of strong skin-effect in the electroconductive details [1]:

$$K = H \times n = \frac{1}{Z_S}(n \times E) \times n \tag{2}$$

$$Z_S = \frac{1+j}{\sigma\delta}\ . \tag{3}$$

Here K is the density of the current on the surface Γ_C of the details, n is the outer normal to Γ_c, Z_S is the surface impedance and δ is the electromagnetic penetration depth. Eq. 2 is IBC.

After applying EEM and taking into account Eq. 2 and the corresponding boundary condition (for the distant points and for the planes of symmetry [1,7]), Eq. 1 is transformed into the following functional:

$$\int_\Omega \left(\mu^{-1} rot A\right)\cdot(rot W)\, d\Omega + \int_\Omega j\omega\sigma A\cdot W\, d\Omega = \\ = \int_{\Gamma_c} W.\ (H\times n) d\Gamma + \int_{\Gamma_S} J.W\, d\Gamma\ , \tag{4}$$

where W is the weight function, Ω is the region of integration and Γ_S is the surface on which is distributed the exciting current.

The region Ω is divided into tetrahedral finite elements while the surface Γ_c is divided into triangular ones.

The weight function for an edge connecting the nodes with numbers i and j is defined as [6]:

$$W = \varsigma_i \nabla\varsigma_j - \varsigma_j\nabla\varsigma_i\ , \tag{5}$$

where ς is the shape function.

The following relations are used to determine A and J within the boundaries of a finite element:

$$A = \sum_{e=1}^{m} A_e W_e \quad J = \sum_{e=1}^{m} J_e W_e\ , \tag{6}$$

where the index e is the local number of the edge, A_e and J_e are circulations of A and J on the edge e, $m = 6$ for a tetrahedral element and $m = 3$ for a triangular element.

The following formulae are used to determine $rotA$ within the boundaries of a tetrahedral element:

$$rot\, A = \sum_{e=1}^{6} A_e rot\, W_e \quad rot\, W_e = 2\nabla\varsigma_i \times \nabla\varsigma_j \,. \qquad (7)$$

Taking into account Eqs. 5, 6 and 7, the functional (Eq. 4) can be transformed into a system of algebraic equations in which the unknowns are the circulations of the MVP A along the finite element edges. The Incomplete Cholesky-Conjugate Gradient (ICCG) Method [10] is used to solve this system. The accuracy of the solution of the system is evaluated by means of the relative Euclidean norm of the residuals ε_i [7].

III. Numerical Algorithm

The numerical algorithm contains an iteration cycle, each step of which includes: calculation of the coefficients of the system of algebraic equations; solution of the system by means of a second iteration cycle which terminates when ε_i reaches a preliminary given value; calculation of the values of the magnetic field strength B for each volume and surface element; correction of the values of the permeability μ for each element using the $B(H)$ curve for the corresponding material.

At the end of the iteration cycle the values of the reactive and active power are calculated.

IV. Numerical Results

Numerical analysis of the field in two electromagnetic systems has been carried out.

The first one is an induction heater for flat surfaces. The heater has a U-shaped magnetic core and two inductors (see Fig. 1). The heated surface is ferrous steel. The working frequency is 4000 *Hz* and the surface density of the inductor current is 2.10^6 A/m. The same system is studied in [4] but the non-linear properties of the ferromagnetic materials are not considered. The studied region, encircled in the figure with a dash line, contains 1/4 of the volume of the whole device.

The second system is an induction crucible furnace-mixer for copper alloys with short inductor, large capacity - above 100 t and working frequency of 50 Hz (see Fig. 2). The crucible of the furnace is filled with melt up to the upper end of the inductor. 30 shielding magnetic cores are installed in the space between the inductor and the furnace mantle, which is made of ferrous steel. The surface density of the inductor current is 2.10^5 A/m. The region of integration, which is dashed on Fig. 2, contains 1/60 of the volume of the furnace.

In the case of FEM, the region of integration is divided into hexahedra, while the surfaces on which IBC is applied are divided into quadrangles. In the case of EEM, each hexahedron is divided into 6 tetrahedrons and each quadrangle – into 2 triangles. On Fig. 3 is shown the division of the area

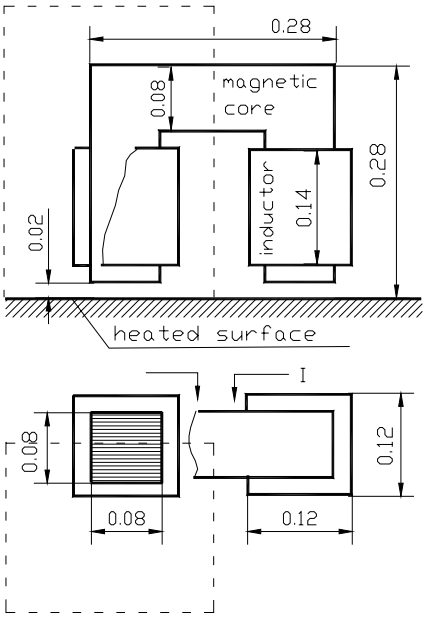


Fig. 1. Diagram of the induction heater for flat surfaces. All dimensions are in m.

of integration into tetrahedrons and triangles for the second studied system – the induction crucible furnace.

Two algorithms are applied for the solution of the problems: EEM coupled with IBC [6] and FEM with gauged MVP coupled with IBC [1] in a non-orthogonal coordinate system [5,6]. The rates of convergence of both algorithms are compared. The EEM coupled with IBC is used to verify the accuracy of the FEM coupled with IBC.

The geometry of the studied devices allows the region of integration to be divided into high-quality hexahedral elements with a shape close to cubical and into almost equilateral tetrahedral elements. That is why high accuracy of the solution of the systems of algebraic equations approximating the differential equations can be expected.

In Table 1 are shown the results from the calculation of the complex power in the region of integration, the necessary number of iterations and the achieved value of the Euclidean norm of the residuals ε_i for the first studied system. As can be seen from the table, the necessary number of iterations for EEM coupled with IBC is nearly 10 times smaller than the corresponding number of iterations for FEM coupled with IBC. Also, the achieved accuracy of the solution of the system of algebraic equations is low: $\varepsilon_i < 0.12$, while for EEM coupled with IBC $\varepsilon_i < 0.02$. The minimum value of ε_i which can be achieved is an important indicator for the rate of convergence of the numerical process.

Despite the relative high value of ε_i in the case of FEM coupled with IBC, the accuracy of the SOLUTION of the problem as a whole is good: the values for the complex power obtained by means of both algorithms are almost identical.

In Table 2 are shown the results from the 3D numerical analysis of the electromagnetic field for the second studied system. The accuracy of the calculation of the value of the complex power by means of FEM coupled with IBC and gauged MVP is high: the results are almost identical with

Table 1. Results of the numerical modelling of the system with concentrating magnetic core

	EEM coupled with IBC [3]	FEM coupled with IBC and gauged MVP [1,4]
Number of iterations	265	2521
Value of ε_i	<0,02	<0,12
Complex power, VA	4139+ j55122	4300+ j56143

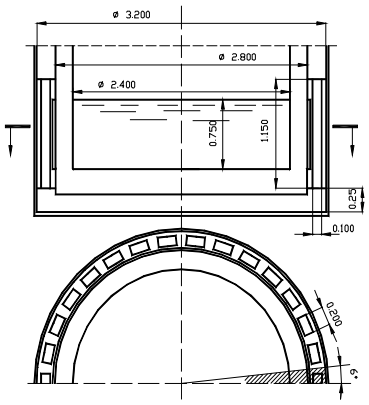


Fig. 2. Diagram of the induction crucible furnace for copper alloys. All dimensions are in m.

Table 2. Results of the numerical modelling of the system with shielding magnetic cores

	EEM coupled with IBC [3]	FEM coupled with IBC and gauged MVP [1,4]
Number of iterations	190	491
Value of ε_i	<0,01	<0,01
Complex power, VA	15568+ j256830	15314+ j241563
Power losses in the mantle's bottom, W	281	300
Power losses in the cylindrical part of the mantle, W	591	620

those obtained by means of the EEM coupled with IBC. The accuracy of the calculation of the power losses in the mantle is satisfactory. Better convergence is observed in the case of EEM coupled with IBC: the number of iterations is 2.5 times smaller than those for FEM coupled with IBC and gauged MVP.

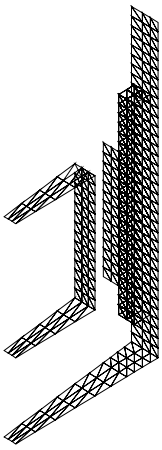


Fig. 3. Division of the area of integration for the induction crucible furnace.

V. Conclusion

It is established that the use of EEM coupled with IBC instead of the gauged formulation of FEM coupled with IBC improves the convergence of the numerical process in the case of 3D numerical analysis of the electromagnetic field, taking into account the non-linear properties of the magnetic materials. In the case of the studied system with a concentrating magnetic circuit, the number of iterations decreases almost tenfold and the accuracy of the solution of the system of algebraic equations approximating the differential equations of the electromagnetic field increases substantially. In the case of the second studied system, which contains shielding magnetic cores, the convergence improves to a lesser degree – the number of iterations decreases 2.5 times while the accuracy of the solution of the system of algebraic equations is the same for both algorithms.

References

[1] Louai, F.-Z., D. Benzerga, M. FelIachi , F. Bouillault, "A 3D finite element analysis coupled to the impedance boundary condition for the magnetodynamic problem in radio frequency plasma devices", *IEEE Trans. on Magn.*, May 1996, pp. 812-815.

[2] Mimoune S.M., J.Fouladgar, G.Develey, "Modeling of 3D Electromagnetic and Heat Transfer Phenomena for Materials with Poor Conductivity", *IEEE Trans. on Magn.*, November 1995, pp. 3578-3580.

[3] Preis, K., I.Bardi, O.Biro, C. Magele, G. Vrisk, K.R. Richter, "Differrent Finite Element Formulations of 3D Magnetostaic Fields", *IEEE Trans. on Mag.*, 28, 1992, No 2, 1056-1059.

[4] Mintchev, P., M. Dimitrov, S. Balinov, "Impedance boundary condition coupled with edge-element method to solve 3-D electro magnetic fields in induction heaters", *XII-th International Symposium on Electrical Apparatus and Technologies*, 31 May – 1 June 2001, Plovdiv. Proceedings Volume I, 237-244.

[5] Димитров М., П. Минчев, "Тримерен числен анализ на електромагнитното поле в индукционни нагреватели с използване на импедансни гранични условия и неортогонална координатна система", *Техническа мисъл*, 1999, є 1-2, 13-20.

[6] Dimitrov M., P. Mintchev, S. Balinov, A. Krusteva, "Choice of Coordinate System in Eddy Current Problems Using 3D Finite Element Numerical Analysis Coupled to the Impedance Boundary Condition", Applied electromagnetics. *Proceedings of the 1st Japanese-Bulgarian-Macedonian Joint Seminar on Applied Electromagnetics*, September 14-15, 1998, Sofia. 2000 Heron Press Ltd., 80-88.

[7] Biro O., K. Preis, "On the Use of the Magnetic Vector Potential in the Finite Element Analysis of the Tree-Dimentional Eddy Currents", *IEEE Trans. on Magn.* Vol. 25, No 4, July 1989, pp. 3145-3159.

[8] Biro O., K. Preis, "On the Use of the Magnetic Vector Potential in the Finite Element Analysis of the Tree-Dimentional Eddy Currents", *IEEE Trans. on Magn.* Vol. 25, No 4, July 1989, pp. 3145-3159.

[9] Golias, N., T.Tsiboukis, "3-D Eddy-Current Computation with a Self-Adaptive Refinement Technique", *IEEE Trans. on Magn.*, 31, 1995, No 3, 2261-2268.

[10] Kershaw, D, "The incomplete Cholesky-conjugate gradient method for the iterative solution of systems of linear equations", *Journal of Computational Physics*, No 26, 1978, pp. 43-65.

Measurements of 50-Hz Electromagnetic Fields in Different Environments

Vladimir Dimcev, Senior Member IEEE, Radmila Sekerinska, Member IEEE[1]

Abstract – **The power companies as well as the public are becoming increasingly interested in the possible link between exposure to power frequency magnetic fields and the human health. Therefore, special attention is given to measurements of the magnetic field level and their comparison to the established or recommended limits according to national standards in different environments**

Keywords – **electromagnetic fields, measurements of the magnetic field, influence to human health**

I. Introduction

The development and the frequent diffusion of the various parts of the electrical energy sector and the consumers' systems are increasing the safety hazards and impose obligations concerning the maximum allowed human exposure to magnetic fields. This has led to standards defining magnetic field reference levels, aimed at providing simpler means of verifying compliance with the basic restrictions and assessing field effects. The CENELEC European Pre-standard ENV 50166-1 defines the magnetic field reference levels for workers and for the general public as follows:

Table 1: Magnetic field reference level - Workers	
Frequency f [Hz]	Magnetic field B
0 - 0,1	2 T
0,1 - 0,23	1,4 T
0,23 - 1	320/ f [mT]
1 - 4	320/ f^2 [mT]
4 - 1500	80/ f [mT] (1,6 mT at 50 Hz)
1500 - 10000	0,053 mT

Table 2: Magnetic field reference level - General public	
Frequency f [Hz]	Magnetic field B
0 - 0,1	0,04 T
0,1 - 1,15	0,028 T
1,15 - 1500	32/ f [mT] (0,64 mT at 50 Hz)
1500 - 10000	0,021 mT

These values are only for an 8-hour time-weighted average and generally allow higher levels for limbs. Nevertheless, even this pre-standard mentions reports stating that electromagnetic fields of lower intensity than the reference levels specified in the standard may have long-term effects on health, but currently available research however has not established adverse effects and does not provide a basis for further restriction of exposure.

The remaining obligation is therefore to design systems or perform measurement campaigns that can analyse the magnetic field levels in both working and residential environments and determine whether they fit in the defined reference levels. Furthermore, it is important to observe the maximum and mean field values, their relationship and their occurrence especially in systems where calculation is not available.

The purpose of this paper is to examine the magnetic fields in residential environments and fields generated by transmission overhead lines. The long time monitoring of the magnetic fields has been achievable with the development of a new generation of microprocessor-based instrumentation. The investigation of the electromagnetic fields (EMF) under power transmission lines are directed towards improving the analytical models for predicting those fields and to find lines with reduce EMF emissions.

The measurements in residential apartments were carried with a field analyser (Wandel & Golterman, Germany) whose three-dimensional isotropic probes are used to record data allowing non-directional measurement. It measures extremely-low frequency (ELF) magnetic fields in the range of 5 Hz to 30 kHz, with the possibility of spectral recording of field components. The main attention during the measurements was given to power frequency (50 Hz) magnetic fields, but rms values of the field in the widest range (5 Hz - 30 kHz) were also recorded for comparison. The EFM measurements under transmission lines were performed with three axis data logging instrument - Emdex C with triggering wheel, made by EFM Company, U.S.A. The Emdex C instrument incorporates technology developed under the sponsorship of EPRI. The magnetic field waveform measurements were conducted with FLUKE scopemeter with EFM 140 magnetic sensor. Those measurements were single axis.

II. Measurements in Residential Environment

The measurements in this environment were carried in 4 apartments in different multi-dwelling buildings, all of them located in the central distribution area of Skopje. Since significantly different measurement results can be obtained using different protocols and instrumentation, the Protocol for spot measurements of residential power frequency magnetic fields, which is a report of the IEEE Magnetic Fields Task Force [1], was followed, meaning measurements were done with appliances left as found; in five rooms frequently used by the occupants (the protocol requires at least three rooms) and near the centre of each room and away from appliances.

The last comment is especially important having in mind the results of the survey, as reported in [2], on the spatial variations of the power frequency magnetic field levels in

[1]Authors are with Electrotechnical Faculty – Skopje, University Ss. Cyril and Methodius, Karpos 2, 1000 Skopje, MACEDONIA e-mail: vladim@etf.ukim.edu.mk

the same room. The correlation between measurements in the centre of the room and at other points varies from 0.642 (in kitchens) to 0.789 (in living rooms), probably due to the dominant presence of appliances in former. Additionally, during the measurement large variations were detected close to appliances that were noted in the mentioned protocol, as well.

The measurement campaign covered a 16-hour period during which the magnetic levels were recorded in every hour. The recordings contain values of the maximum (B_{max}) and root-mean-square value (B_{rms}) of the power frequency magnetic field, as well as the rms value of the magnetic field in the bandwidth of 5 Hz to 30 kHz (B_{wb}).

Table 3 contains the average values for these quantities, while Table 4 contains the average values of the ratio between rms and maximum value as well as the ratio between the rms value for 50 Hz and the wide-band all this in different rooms in apartments #1 to #4. In both tables, overall averages and standard deviations are also presented.

Table 3 – Average values of the measured magnetic field level [μT]

	Apartment #1			Apartment #3		
	B_{max}	B_{rms}	B_{wb}	B_{max}	B_{rms}	B_{wb}
Living	0.080	0.045	0.131	0.149	0.090	0.138
Entry	0.109	0.063	0.142	0.128	0.080	0.131
Kitchen	0.105	0.063	0.138	0.163	0.104	0.150
Bedro.1	0.116	0.069	0.149	0.184	0.122	0.158
Bedro.2	0.091	0.054	0.145	0.198	0.127	0.164
Average	0.100	0.059	0.141	0.164	0.105	0.148
Stan.dev	0.015	0.009	0.007	0.028	0.020	0.014
	Apartment #3			Apartment #4		
Living	0.174	0.111	0.174	0.111	0.174	0.111
Entry	0.223	0.140	0.223	0.140	0.223	0.140
Kitchen	0.205	0.134	0.205	0.134	0.205	0.134
Bedro.1	0.224	0.153	0.224	0.153	0.224	0.153
Bedro.2	0.172	0.109	0.172	0.109	0.172	0.109
Average	0.200	0.129	0.174	0.960	0.675	0.669
Stan.dev	0.025	0.019	0.013	0.373	0.322	0.306

Table 4 – Average ratios [%]

	Apartment #1		Apartment #2	
	Brms/Bmax	Brms/Bwb	Brms/Bmax	Brms/Bwb
Living	55.92	34.47	59.46	62.12
Entry	56.07	42.68	60.61	58.81
Kitchen	59.93	44.87	61.09	64.26
Bedro.1	60.01	47.49	63.72	72.63
Bedro.2	59.03	37.44	63.81	76.31
Average	58.19	41.39	61.74	66.83
Stan.dev	2.04	5.35	1.94	7.36
	Apartment #3		Apartment #4	
Living	64.30	64.79	74.60	100.86
Entry	64.72	80.38	69.25	99.51
Kitchen	65.03	75.67	75.19	99.58
Bedro.1	67.83	84.44	76.38	110.07
Bedro.2	64.68	72.49	56.05	90.87
Average	65.31	75.55	70.29	100.18
Stan.dev	1.43	7.54	8.42	6.82

It is important to note that apartment #4 steps out of the magnetic field range in which the other ones lay, due to closeness of a 110 kV transmission lines.

III. Measurements under Transmission Lines

This part included electric and magnetic fields measurements under transmission and distribution lines. Also, the appropriate computer programs were developed for the calculation of the electric and the magnetic fields using the quasi-static approach.

Field levels were recorded along the way perpendicular to the overhead line and passing through line lowest point, 1m above the ground. The measurements of the electric field must be made with extreme caution, because the electric field is influenced by close objects.

Additional problem could be the exact height of transmission line conductors, because the designed values of the lines heights could be different from real.

The measured and calculated curves of the electric fields under 220 kV and 110 kV line are given on Fig. 1 and 3. The values of the measured curves are smaller than calculated one, as input data for the height of the conductors were used designed values of the transmission lines. The measured magnetic field under 220 kV and 110kV lines are shown on Fig. 2 and 4. For the 220 V line (Fig. 2) the curve 2 is calculated with measured height and curve 3 is calculated with

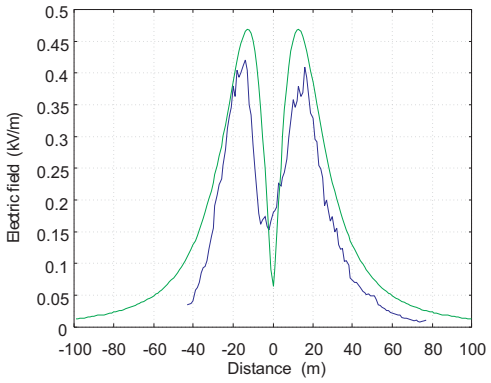


Fig. 1. Measured (1) and calculated (2) curves of electric field under 220 kV line Skopje 1 - Kosovo

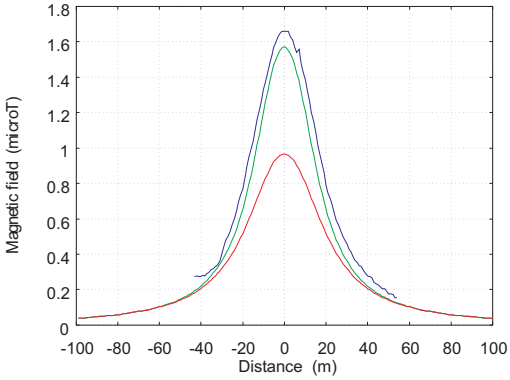


Fig. 2. Measured (1) and calculated (2) and (3) curves of magnetic field under 220 kV line Skopje 1 - Kosovo

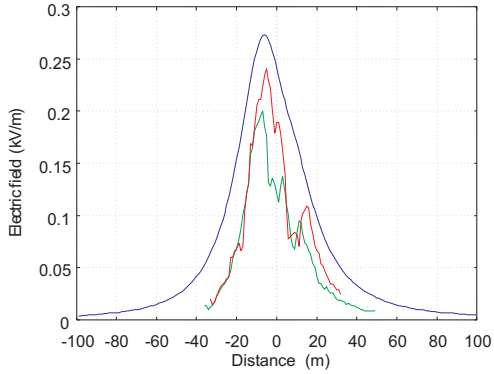


Fig. 3. Measured (1), (2) and calculated (3) curves of electric field under 110kV line Skopje-Tetovo

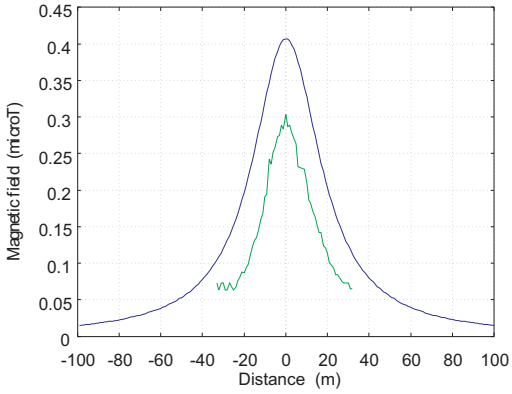


Fig. 4. Measured (1) and calculated (2) curve for magnetic field under 110kV line Skopje - Tetovo

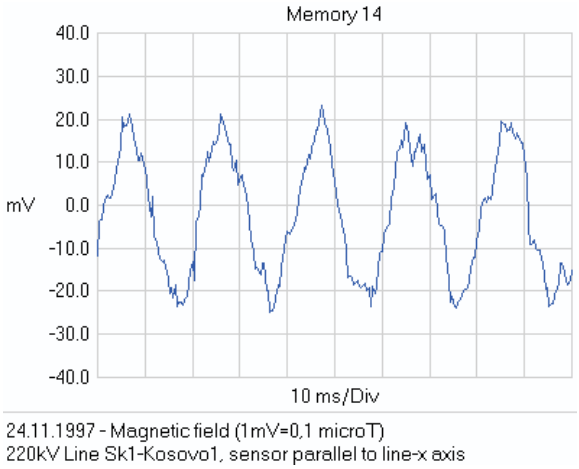


Fig. 5. Magnetic field under 220kV line

designed height of the line conductors.

The measured values of magnetic fields are well below some recommended limits [6], this is due low line load in the moment of measurements. The measured data were used for calculating the normalized values that correspond to the maximum load. The average normalized density peaks mounted to the 16,96 μT for 110 kV line, 18,35 μT (220kV line) and 34,18 μT (400 kV line).

The influence of the soil resistivity on the magnetic

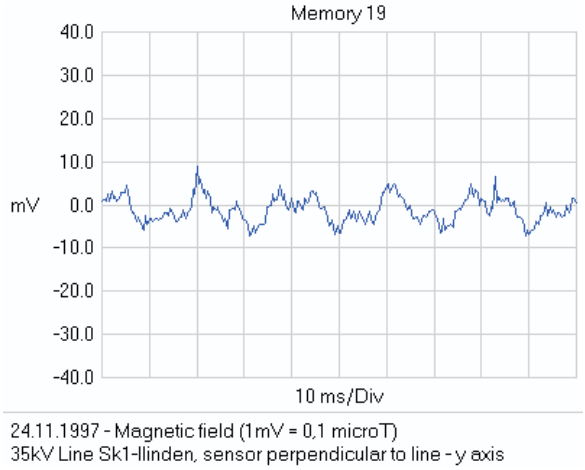


Fig. 6. Magnetic field under 35 kV line

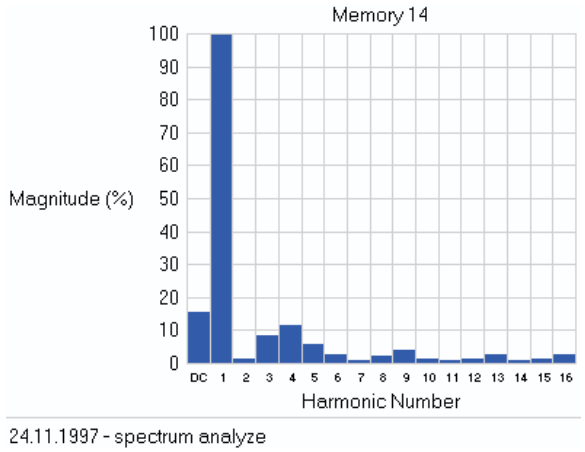


Fig. 7. Spectrum content of magnetic field under 220kV line

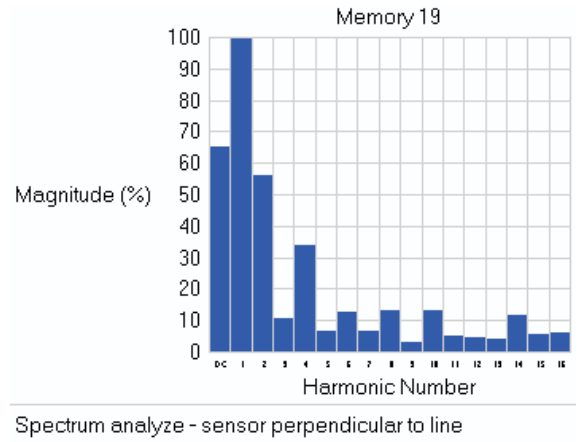


Fig. 8. Spectrum content of magnetic field under 35kV line

fields was also considered and it had no importance for the distances near power lines. Single-axis waveform measurements of the magnetic field were made under 220kV transmission line (Fig. 5) and under 35kV distribution line (Fig. 6). Spectrum analyses were performed for measured waveforms (Fig. 7 and 8). The harmonic content is richer

for 35kV line, it seems that phase currents in this line is less balanced.

Two cases of private houses in the vicinity of transmission lines were closely observed. In both cases, the limits defined in standards were not exceeded, the normalized peak values reached 30,7 μT (on the estate very close to 400kV line) and 5 μT (under 110kV line), which is hardly recommendable for permanent environment and it is difficult to estimate the long-term effect of the EMF on humans health.

IV. Conclusions

This part of the measurement campaign covered residential environments and transmission lines. Generating stations, distribution substations, large consumers and offices will be included in the second phase.

The residential environments are characterized by a low magnetic field level, way below the standard limits. Even the apartment with the highest level has an average almost 1000 times smaller then the limit prescribed by CENELEC standard (0,675 μT as opposed to 0,64 mT).

The highest values of B_{rms} peaks in all apartments were recorded in kitchens. However, this is not the case with their average values, due to significant sags characteristic of kitchen.

Also, the three axis and the single axis measurements of the EMF under the power transmission and distribution lines have been done. The measurements are compared against computer calculations, a close correspondence between them has been observed. Observation of the spectrum content of the magnetic field under transmission and distribution lines was made.

References

[1] A report of the IEEE Magnetic Fields Task Force, *"Magnetic Fields From Electric Power Lines"*, IEEE Trans. on Power Delivery, Vol. 3, No. 4, October 1988.

[2] A report of the IEEE Magnetic Fields Task Force, *"A Protocol for Spot Measurements of Residential power Frequency Magnetic Fields"*, IEEE Trans. on PWRD, Vol. 8, No. 3, July 1993.

[3] J. Swanson, *"Magnetic fields from transmission lines: comparison of calculations and measurements"*, IEE Proc.-Gener. Transm. Disturb., Vol.142, No. 5, Sept. 1995.

[4] R.G. Olsen, D.C. James, *"The Performance of Reduced Magnetic Field Power Lines Theory and Measurements on an Operating Line"*, IEEE Trans. on PWRD, Vol. 8, No.3, July 1993.

[5] W.L. Cotten, K. Ramsing, C.Cai, *"Design Guidelines for Reducing Electromagnetic Field Effects from 60-Hz Electrical Power Systems"*, IEEE Trans. on Industry Applications, Vol. 30, No. 6, Nov/Dec 1994.

[6] IEEE Std 1460-1996, *IEEE Guide for the Measurement of Quasi-Static Magnetic and Electric Fields*, March 1997.

[7] CENELEC ENV 50166-1, *"Human Exposure to Electromagnetic fields, Low frequency (0 Hz to 10 kHz)"*, January 1995.

Calculation of leakage Fluxes in Solid Salient Poles Synchronous Motor by Finite Element Method

Mirka Popnikolova Radevska[1]

Abstract – **Solid Salient Poles Synchronous Motor (SSPSM) is well known for its simple construction as well as for its operational reliability due to absence of short circuit cage. Object of investigation in this paper is SSPSM, product of MAWDSLEY with rated data: P_1=3.520 kW, U_n=240 V, I_f=5.5 A, cosφ=0.97 and 2p=4. Software package FEM for 3D calculation in magnetostatic case is applied in order leakage fluxes of stator and rotor windings as well as leakage reactances to be calculated.**

Keywords – **SSPSM, FEM 3D, leakage fluxes, leakage reactances**

I. Introduction

Numerous classical methods exist for analytic calculation of motor parameters through empirical equations by simplifying electromagnetic processes inside the machine. In this paper is used completely different approach for motor parameters calculation using software package FEM 3D which calculates magnetic field in 3D motor domain, as well as leakage fluxes and consequently enables calculation of leakage reactances in stator and excitation winding of SSPSM.

II. Calculation of Leakage Fluxes in Stator Winding

Calculation of leakage fluxes of stator winding is made in active parts of stator winding, part of the winding placed in stator channel or part of the winding placed in first axial layer of motor mathematical model Fig. 1 and winding overhangs – second, third and fourth axial layer of motor mathematical model. Mathematical model is generated by dividing motor domain per z axis in five layers.

In first layer are placed active parts of stator and excitation winding while winding overhangs are placed in second third and fourth axial layer. Each layer is divided into sub domains with local coordinate system. FEM 3D automatically generates mesh of finite elements and calculates the values A in each node of finite element mesh.

Local coordinate system is placed regarding winding sub domains as it is shown on Figs 2 and 3.

- Leakage flux in first axial layer

Leakage flux in active part of motor windings is determined from the value of magnetic vector potential **A** which includes active parts of stator and rotor windings (first axial layer per z coordinate of motor model).

[1]Mira Popnikolova Radevska is with Technical Faculty, Ivo Lola Ribar b.b., 7000 Bitola, Macedonia, E-mail: mirkara@mt.net.mk

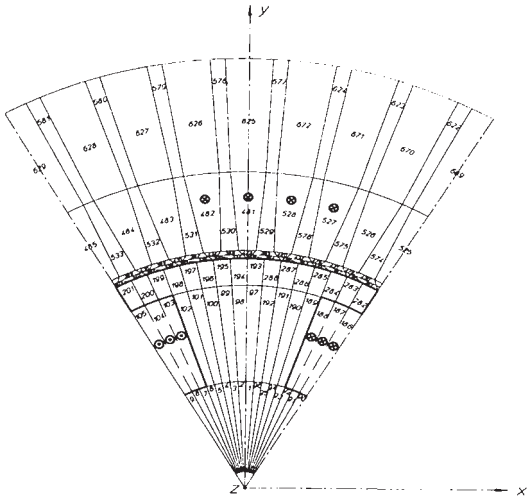


Fig. 1. Cross section of first axial layer with current densities

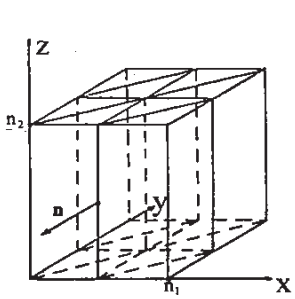


Fig. 2. Local coordinate system per z coordinate

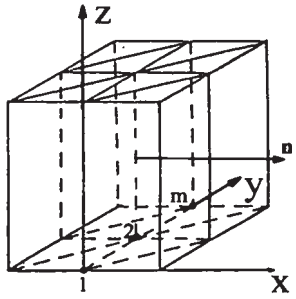


Fig. 3. Local coordinate system per y coordinate

Leakage flux is calculated when current is in direction of z-axis, thus normal vector of magnetic induction which creates leakage flux is in the direction of x-axes.

Calculation of leakage flux in active part of motor winding is done according to Eq. (1).

$$\Delta\Psi_n = 2 \cdot \left(\Delta\Psi_{1n} + \sum_{i=1}^{m} \Delta\Psi_{in} \right) \tag{1}$$

Calculation of leakage flux of stator winding in first axial layer is done for excitation current $I_f = 0$ and rated current I_{an} in phase A of stator winding. Leakage flux per pole for one channel is calculated according to:

$$\Delta\Psi_{Iki} = l_s \frac{N_k}{4} \frac{\Delta A_i}{\Delta y} \tag{2}$$

I - denotes first axial layer, k - channel leakage, i - channel number, Δy - radial channel length, N_k - number of conduc-

tors.

Total leakage flux in first axial layer or in active part of stator winding per pole is read out from the output file POT3D.DAT of FEM 3D and its value is:

$$\sum \Delta\Psi_{Iki} = \sum_{i=481}^{482} \Delta\Psi_{Iki} + \sum_{i=527}^{528} \Delta\Psi_{Iki} \qquad (3)$$

$\sum \Psi_{Iki} = 0.02248$ Vs

- Leakage flux in second and third axial layer

Calculation of leakage flux in second and third layer of stator winding represents the leakage flux in winding overhangs. Method of calculation is identical as in the first layer. Now local coordinate system is placed in sub domains which approximately represent winding overhangs. In equation 1 only lengths of stator winding overhangs in second and third layer are replaced and following results are gained respectively:

$$\sum \Delta\Psi_{IIki} = \Delta\Psi_{IIk1153} + \Delta\Psi_{IIk1154} + \Delta\Psi_{IIk1199} + \Delta\Psi_{IIk1200} \qquad (4)$$

$\sum \Psi_{IIki} = 7.676 \cdot 10^{-5}$ Vs

$$\sum \Delta\Psi_{IIIki} = \Delta\Psi_{IIIk1825} + \Delta\Psi_{IIIk1826} + \Delta\Psi_{IIIk1871} + \Delta\Psi_{IIIk1872} \qquad (5)$$

$\sum \Psi_{IIIki} = 3.449 \cdot 10^{-5}$ Vs

- Leakage flux in fourth axial layer

Now, local coordinate system is placed as it is shown on Fig. 4.

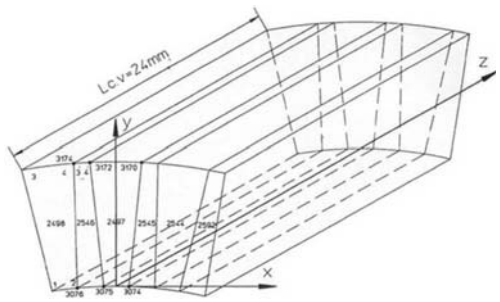


Fig. 4. Local coordinate system in fourth axial layer. l_{SIVcv}=24 mm – axial length of fourth layer; b_k=55.9 mm – channel width; t_1=99.7 mm pitch channel-tooth; b_z=43.8 mm – width of stator tooth; N_k – number of conductors per channel

Leakage flux is determined from equation **B**=rot **A**.

a) Leakage flux per **k** ort

1. For y=const, $l_{SIVc.v}$=24 mm

$$\sum \Delta\Psi_{IVay} = l_{sIVc.v}\frac{N_k}{4}(\Delta A_{2497} + 3\Delta A_{2498} + + 3\Delta A_{2543} + + \Delta A_{2544} + 2\Delta A_{2546} + \Delta A_{2592}) \qquad (6)$$

$\sum \Psi_{IVay} = 0.236 \cdot 10^{-3}$ Vs

2. For x=const

Following the same logic calculated value of leakage flux is:

$\sum \Psi_{IVax} = 0.222 \cdot 10^{-3}$ Vs

Total leakage flux in stator winding per **k** ort is calculated froom root square of $\sum \Psi_{IVay}$ and $\sum \Psi_{IVax}$ and its value is:

$\sum \Psi_{IVa} = 0.324 \cdot 10^{-3}$ Vs.

b) Leakage flux per **i** ort

1. For y=const Δx=b$_k$

$$\sum \Delta\Psi_{IVbky} = b_k \cdot \frac{N_k}{4}(\Delta A_{2497} + 3\Delta A_{2498} + 3\Delta A_{2543} + \Delta A_{2544}) \qquad (7)$$

$\sum \Delta\Psi_{IVbky} = 0.00596$ Vs

2. For y=const Δx=b$_z$

$$\sum \Delta\Psi_{IVbzy} = b_z \cdot \frac{N_k}{4}(\Delta A_{2546} + \Delta A_{2544}) \qquad (8)$$

$\sum \Delta\Psi_{IVbzy} = 0.00399$ Vs

Total leakage flux in winding overhangs for y=const per **i** ort is calculated as:

$$\Sigma\Delta\Psi_{IVby} = \Sigma\Delta\Psi_{IVbky} + \Sigma\Delta\Psi_{IVbzy} = 0.0096 \text{ Vs}. \qquad (9)$$

Total leakage flux in winding overhangs for z =const per **i** ort is:

$\Sigma\Delta\Psi_{IVb} = 0.0102$ Vs.

Total leakage flux in stator winding overhangs per pair of poles is found from:

$$\Sigma\Delta\Psi_{IVc.v} = 2\,(\Sigma\Delta\Psi_{IVa} + \Sigma\Delta\Psi_{IVb}) = 0.021046 \text{ Vs}. \qquad (10)$$

III. Calculation of Leakage Fluxes in Excitation Winding

Similar to the calculation of leakage fluxes in stator winding, calculation of leakage fluxes in excitation winding is made in winding active part-first axial layer of mathematical model and winding overhangs- second, third and fourth axial layer of mathematical model.

- Leakage flux in first axial layer

It is assumed that that rated current flows through excitation winding $I_f = I_{fn} = 5.5\,A$ and no current flows through stator winding $I_{an} = 0$. Local coordinate system is placed as it is shown on Figs. 2 and 3. Again using software FEM 3D calculation of magnetic vector potential in complete motor domain is made and data are read out from output file.

Normal vector of **B** is in direction of x-axis. Leakage flux is also in the direction of x-axis. Calculation is made according to Eq. 1. Value of leakage flux per pair of poles is:

$$\sum \Delta\Psi_{If} = \sum_{i=103}^{105} \Delta\Psi_{Ifi} + \sum_{i=186}^{188} \Delta\Psi_{Ifi} \qquad (11)$$

$2\sum \Delta\Psi_{If} = 0.0143$ Vs

- Leakage flux in second axial layer

Leakage flux is calculated regarding sub domains which define winding overhangs of excitation winding in second axial layer Fig. 5. Again local coordinate system is placed regarding this sub domain. Values of **A** in each node of this sub domain are calculated.

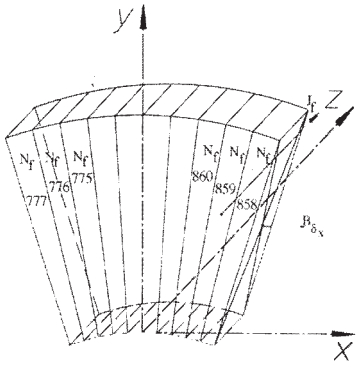


Fig. 5. Local coordinate system in second axial layer of excitation winding overhangs

Sum of all leakage fluxes in second layer of excitation winding is:

$$\sum \Delta\Psi_{IIf} = \sum_{i=775}^{777} \Delta\Psi_{IIfi} + \sum_{i=858}^{860} \Delta\Psi_{IIfi} = 0.01456$$

$$\sum \Delta\Psi_{IIf} = 0.01456 \text{ Vs} \tag{12}$$

- Leakage flux in third axial layer

Mathematical model of excitation winding overhangs in third axial layer is presented on Fig. 6

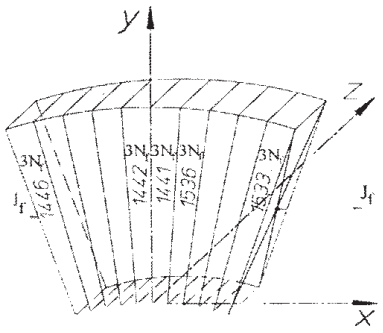


Fig. 6. Local coordinate system in third axial layer of excitation winding overhangs

Leakage flux is calculated from:

$$\sum \Delta\Psi_{IIIc.v.} = \frac{N_f}{4} l_{III} \Big(\sum_{i=1441}^{i=1446} \Delta\Psi_{IIIi} + \sum_{i=1533}^{i=1536} \Delta\Psi_{IIIi} \Big) \tag{13}$$

and its value per pair of poles is:

$$2 \cdot \sum \Delta\Psi_{IIIc.v.} = 0.02442 \text{ Vs}$$

IV. Calculation of Leakage Reactances in Stator Winding

In above sections is explained method of calculation of leakage fluxes in active parts as well as in winding overhangs.

Leakage fluxes close around the winding itself. These fluxes determine leakage reactances of SSPSM.

- Leakage reactance in first axial layer

In order leakage reactance to be calculated value of leakage flux which closes around winding active part must be known. Using result from Eq. 3 inductance is calculated as:

$$L_{Ik} = \frac{\sum \Delta\Psi_{Iki}}{I_{an}} = 0.00446 \text{ H}, \; I_{fn} = 0 \text{ A} \tag{14}$$

Leakage reactance in first axial layer of stator winding is calculated from:

$$X_{Ik} = \omega L_{Ik} = 1.4005 \, \Omega \tag{15}$$

- Leakage reactance in second and third axial layer

Considering that leakage fluxes in second and third axial layer of stator winding from Eqs. (4) and (5), inductances and reactances are calculated as:

$$L_{IIk} = \frac{\sum \Delta\Psi_{IIki}}{I_{an}} \tag{16}$$

$$X_{IIk} = \omega L_{IIk} = 0.0241 \, \Omega \tag{17}$$

$$L_{IIIk} = \frac{\sum \Delta\Psi_{IIIki}}{I_{an}} \tag{18}$$

$$X_{IIIk} = \omega L_{IIIk} = 0.0108 \, \Omega \tag{19}$$

- Leakage reactance in fourth axial layer

Considering the value of leakage flux from Eq. (10) inductance and reactance in fourth axial layer of stator winding are calculated as:

$$2L_{IVc.v.} = \frac{\sum \Delta\Psi_{IVc.v.}}{I_{an}} = 4.176 \cdot 10^{-3} \text{ H} \tag{20}$$

$$X_{IVc.v.} = 0.656 \, \Omega$$

Total leakage reactance of stator winding is calculated as sum of leakage reactances in all four layers :

$$X_{\sigma a} = X_{Ik} + X_{IIk} + X_{IIIk} + X_{IVc.v.} \tag{21}$$

or in per units:

$$X^*_{\sigma a} = 0.0431 \text{ p.u.}$$

V. Calculation of Leakage Reactances in Excitation Winding

- Leakage reactance in first axial layer

Considering the value of leakage flux in active part of excitation winding from Eq. (11), inductance and reactance in this part of winding are calculated as:

$$L_{If} = \frac{\sum \Delta\Psi_{If}}{I_{fn}} = 0.0026 \text{ H} \tag{22}$$

$$X_{If} = \omega L_{If} = 0.816 \, \Omega \tag{23}$$

- Leakage reactance in second axial layer

Leakage flux in second axial layer of excitation winding is calculated according to Eq. (12). Consequently inductance and reactance in this part are calculated:

$$L_{IIf} = \frac{\sum \Delta\Psi_{IIf}}{I_{fn}} = 2.647 \cdot 10^{-3}\ \text{H} \tag{24}$$

$$X_{IIf} = \omega L_{IIf} = 0.831\ \Omega \tag{25}$$

- Leakage reactance in third axial layer

Leakage flux in third axial layer of excitation winding is calculated according to Eq. (13). Inductance and reactance in third axial layer of excitation winding are calculated:

$$L_{IIIc.v} = \frac{\sum \Delta\Psi_{IIIc.v}}{I_{fn}} = 2.22 \cdot 10^{-3}\ \text{H} \tag{26}$$

$$X_{IIIc.v} = \omega L_{IIIc.v} = 0.697\ \Omega \tag{27}$$

Total leakage reactance in excitation winding is:

$$X_{\sigma f} = X_{If} + X_{IIf} + X_{IIIc.v} \tag{28}$$

or in per units:

$$X_{\sigma f} = 0.0491\ \text{p.u.} \tag{29}$$

VI. Conclusion

Using novel Finite Element Method calculation of motor leakage fluxes can be done due to what motor reactances can be calculated more accurately than by analytic methods since there is no simplification of electromagnetic processes inside the machine. Value of calculated leakage reactance of stator winding $X^*_{\sigma a} = 0.0431$ p.u. is compared with value gained from analytic calculation $X^*_{\sigma a} = 0.0694$ p.u. Compared result show reasonable agreement. This proves methodology as adequate one also for calculation of excitation winding leakage reactances as well as self reactances or machine reactances per d and q axises.

References

[1] M.Popnikolova Radevska, M.Cundev, L.Petkovska, " From Macroelements to Finite Machine Field Analyses", ISEF International Symposium on electromagnetical Fields in electrical engineering p.p. 346-349, Thessaloniki, Greece, 1995

[2] M.Popnikolova Radevska, M.Cundev, L.Petkovska, " Modelling of Three-dimensional Magnetic Fiels in Solid Salient Poles Synchronous motor" ISTET international Symposium on Electromagnetic fields in p.p.70-73 Thessaloniki, Greece, 1995

[3] M.Cundev, L.Petkovska, M.Popnikolova Radevska " An Analyses of Electrical Machines Synchronous Type Based on 3D-FEM", ICEMA International Conference on Electrical Machines and Applications", p.p. 29-32, Harbin, China, 1996

[4] M.Popnikolova Radevska, M.Cundev, L.Petkovska " Electromagnetic Field Analyses and Computation of Electromagnetic Characteristics of Solid Salient Poles Synchronous Motor" ICEM'98, p.p. 707-709, Turkey, 1 998.

[5] M.Poponikolova Radevska, M.Cundev, L.Petkovska, "Modelling the Configuration of Solid Salient Poles Synchronous Motor for 3D-FEM", Scientific Bulletin of Lodz Technical University Nr. 788, p.p. 178-186 Elektrika Lodz, Poland, 1998

[6] M.Popnikolova Radevska, V.Sarac, M.Cundev, L.Petkovska, "Computation of Solid Salient Poles Sunchronous Motor Parameters by 3D Finite Element Method", EPNC Symposium on Electromagnetic Phenomena in Nonlinear Circuits, p.p. 111-114, Belgum, 2002

Assessment of the Performances of the Grounding System of the Transmission Lines

Nikolce Acevski[1] and Risto Ackovski[2]

Abstract – **Step and touch potentials at transmission line (TL) structures are sometimes important parameters in the design of the grounding system (GS). To have confidence in conventional (deterministic) methods of potential calculation, worst-case values of the design parameters must be assumed. This often leads to overly pessimistic results. A probabilistic approach enables more realistic values to be obtained with the same degree of confidence. A study has recently been performed to calculate the probability distributions of step and touch potentials in 400 kV transmission line Shtip - Dubrovo.**

Keywords – **transmission line, grounding system, probability distribution, and touch (step) voltage**

I. Introduction

The recent ten years have been characterized by a worldwide trend in using stochastic approaches for calculation of potential differences at touch or step in the surrounding of EE objects, instead of using the custom deterministic approach. Applying the deterministic approach, the maximum values of the potential differences at step or touch to which a human being could be exposed in case of an error in the grounding system of (in this case) a transmission line were calculated. This means that the worst case was presented. The results of investigations of numerous cases have shown that these maximal values occur very rarely. Namely, the probabilities for occurrence of such potential differences are very low. On the other hand, using stochastic approach, the probabilities for occurrence of certain potential difference - respectively the frequencies of occurrence - can be calculated. Using this approach, so called "frequency histograms" (FH) of the potential differences at touch or step for each tower grounding can be drawn, as well as, in general, the cumulative probabilities for the whole grounding system of the considered transmission line. This approach can be applied not only to choose a solution for the GS of the TL, but also for assessment of the risk when operating with running transmission lines. This means that the safety conditions from too high potential differences at touch or step – as defined with the regulations – can be tested in a more realistic way by using a stochastic approach.

[1]Nikolce Acevski is with the Faculty of Technical Sciences, Bitola, Macedonia, -mail: acevski@hotmail.com

[2]Risto Ackovski is with the Faculty of Electrotechnical Sciences, Skopje, Macedonia, E-mail: acko@ieee.org

II. Risks from Dangerous Electrical Strokes Close to Transmission Lines

Coincidence of numerous stochastic events, each one with different probability for occurrence, is necessary for occurrence of an accident caused by an electrical stroke close to an energetic object (in our case the towers of a transmission line):

1. A fault to ground has occurred on the transmission line.
2. In the moment of the fault to ground, a person bridging certain potential difference at touch or step is located close to one of the towers.
3. The electrical currents through the body of the person are strong enough and are of sufficient duration to cause ventricular fibrillation.

Let P_{KV} be the probability that at the moment of observation, a fault to ground occurs on the observed TL. Then we have [3]:

$$P_{KV} = \frac{N_{KV} \cdot r_{KV}}{T_{god}} , \qquad (1)$$

where N_{KV} is the average number of faults to ground of the line per year, r_{KV} is the average duration of a fault to ground (h), and T_{god} is a one year period (T_{god} =8760 h). The probability P_d that the observed person touches a TL tower, and, in that way, bridges a potential difference at touch will be:

$$P_d = f_d \cdot t_d . \qquad (2)$$

Applying a similar reasoning, the probability P_c that an observed person is close to a certain tower in a certain moment, and, in that way, is bridging a potential difference at step will be:

$$P_c = f_c \cdot t_c . \qquad (3)$$

In the relations (2) and (3), the frequencies at touch and at step are marked with f_d and f_c, and the average duration of a touch or a step is marked with t_d and t_c respectively. These amounts can be obtained as a result of long lasting observations of the stochastic approaches of persons to a TL, as it was done, for ex., in [4]. Based on these observations, the average yearly figures N_d and N_c of approaching towers and exposure to touching a TL by casual passers-by and the respective frequencies are f_d and f_c.

$$f_d = N_d/T_{god}; \qquad f_c = N_c/T_{god} \qquad (4)$$

Taking into consideration the information given above, according to the probability distributions of compound events,

the risk R can be calculated, i.e. the probability a person to be a victim of an electric stroke caused by a fault to ground of a TL:

$$R = P_{KV} \cdot \left(\sum_d P_d \cdot R_d + \sum_c P_c \cdot R_c \right) \quad (5)$$

In (5), R_d and R_c are the death probabilities (ventricular fibrillation) caused by an exposure to a dangerous voltage touch, step respectively. The summing in (5) is made for all possible touches (d) and all possible steps (c) close to an observed TL.

III. Calculation of Risks for Death Caused by a Dangerous Touch (Step) Voltage Due to a Short Circuit on Any Part of the Line

At first, the distribution of the phase currents of a one-phase short circuit for an error on any location along the line will be calculated, using, for example, the procedure elaborated in [1]. After the calculation of the values of the currents J_k, J_{p-k}, J_{q-k} $(k = 1, n)$, these values will be used for calculation of the potentials along the longitude of the transmission line φ_{ki} $(i = 1, n)$, in case of arising a one-phase short circuit on any part k (tower no. k) of the TL.

For calculation of the distribution of the phase currents J_k, J_{p-k}, J_{q-k} during a fault to ground on any part k along the line (see Fig. 1), it is sufficient to know the standard data which are output of custom calculations of short circuit currents in the EE systems. Those are the phase current in a three-phase and the phase current in a one-phase short circuit for the three characteristic periods: sub transient, transient and permanent. This procedure is repeated n times, for each tower place k of the TL $(k = 1, n)$. The results of the calculations are put in a square matrix $[\varphi]$ with dimensions *nxn*. In that case, the elements of row k of this matrix, will consist of the potentials φ_{ki} $(i = 1, n)$ that will occur at all n towers in case of a one-phase short circuit on the tower k. Similarly, the elements of the column i of this matrix will contain the potentials φ_{ki} that will occur on different locations of the error k on the tower i.

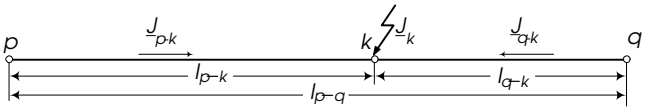


Fig. 1. Distribution of phase currents during a fault to ground

Let us observe the case of a one-phase short circuit on one of the towers of a TL. Let $P_{KV}(i)$, $(i = 1, n)$ be the probability for occurrence of a short circuit on the i-th tower of the TL. We assume that this probability is equal for all n towers of the observed TL (this assumption is approximately true on flat ground where the towers have uniform resistance to ground and uniform heights), meaning that they are equally exposed to discharges from the atmosphere and to recurrent skips of the isolation. Then:

$$P_{KV}(i) = \frac{P_{KV}}{n} \quad (6)$$

Let us assume that a one-phase short circuit has occurred on a certain part of the observed TL and let t be the time of elimination of the short circuit, i.e. the duration of the error. If φ_{ki} is the potential that will occur on the tower i, under condition that the error has occurred on the tower k, then the risk for occurrence of a dangerous touch voltage that would cause death of the person who in that moment has been touching the tower i, will be:

$$R_d(i) = \sum_{k=1}^{n} P_{KV}(k) \cdot PA_{Ed}(\varphi_{ki}, t_{isk}) = \\ = \frac{1}{n} \cdot \sum_{k=1}^{n} PA_{Ed}(\varphi_{ki}, t_{isk}) \quad (7)$$

when the tower i have a GS of type A, and respectively when the towers i have a GS of type B:

$$R_d(i) = \sum_{k=1}^{n} P_{KV}(k) \cdot PB_{Ed}(\varphi_{ki}, t_{isk}) = \\ = \frac{1}{n} \cdot \sum_{k=1}^{n} PB_{Ed}(\varphi_{ki}, t_{isk}) \quad (8)$$

In the relations given above, the following symbols were used: N – total number of towers; φ_{ki} – potential (V) that occurs on the tower number i during a fault to ground that has occurred on the tower number k; $R_d(i)$, $R_c(i)$ – death occurrence risk caused by a too high touch and step voltage, respectively, on the tower i, during the occurrence of a short circuit on any place k of the observed tower; $PA_{Ed}(\varphi_{ki}, t_{isk})$, $PB_{Ed}(\varphi_{ki}, t_{isk})$ – risk (probability) for death occurrence caused by a dangerous touch voltage at a GS of type A, and at a GS of type B respectively, when the duration of the error is t_{isk} seconds, and the voltage of the tower equals to φ_{ki}; $PA_{Ec}(\varphi_{ki}, t_{isk})$, $PB_{Ec}(\varphi_{ki}, t_{isk})$ – risk (probability) for death occurrence caused by a dangerous step voltage at a GS of type A, and at a GS of type B respectively, when the duration of the error is t_{isk} seconds, and the voltage of the tower equals to φ_{ki}.

A. *Types of Grounding Systems Encircling High Voltage Towers*

The contour GS are the most frequently used ones for grounding of high voltage towers ($U_n \geqslant 10$ kV). The foundation of these towers is composed of 4 symmetrically lined

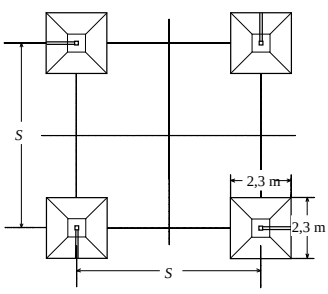


Fig. 2.a Appearance of the GS of type A (view from above)

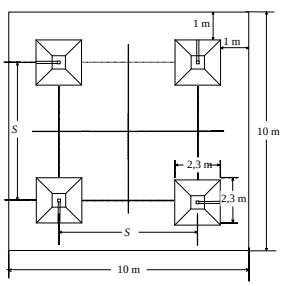


Fig. 2.b Appearance of the GS of type B(view from above)

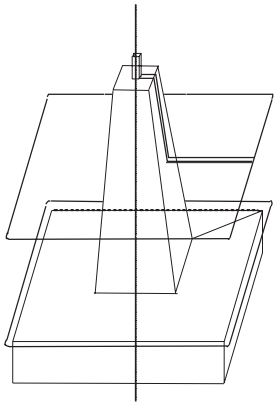


Fig. 3. Appearance of the GS encircling each foot of the tower

feet (Fig. 2.a, 2.b). The distance among the foundation feet S and their dimensions, depend on the height of the towers and on the capacity of the foundation. A common characteristic of all contours GS is that two square rings (contours) are being placed around each of the four feet. The first contour is placed at depth of 0.5 m, and the depth at which the other contour will be placed, depends on the depth of the foot. It depends on the dimensions of the foundation (Fig. 3). Besides the contours around each foot, for some types of GS, placement of an additional contour, encircling the feet of the foundation at depth of 1 m, is foreseen. The shape of this contour is a square with variable longitude of the side. The dimension of the side depends on the distance among the feet, i.e. - once again - it depends on the height of the tower and the capacity of the ground. The contour is positioned in a way that the horizontal distance from the sides of the contours that encircle the foundation is 1 m (Fig. 2.b).

According to the adopted indication, the marks for the GS without a common square contour encircling the feet of the tower start with A. The GS with a common square contour encircling the feet of the tower start with B. They are used for tower places that can be easily reached by people (or animals) (e.g. close to a road, settlement, etc.). The insertion of a joint contour encircling the four feet brings to reduction of the resistance along the GS and to shaping the potential in the surrounding of the tower, aiming at reduction of the risk for occurrence of too high touch voltages, or step voltages respectively.

When dealing with 400 kV towers, two families of standard GS– type A and type B - are used in our country. All contours are made of steel wires coated with zinc, 10 mm in diameter. Most frequently, the distance among the feet S varies from 2.5 m for the shortest, to 6 m for the highest towers. The average longitude of the side of the two square contours encircling the foundation foot is 2.35 m. The average depth of burying the lower contour is 2 m. The average longitude of the side of the square which is a joint grounding system encircling all feet is 10 m.

B. Probability for Death Occurrence in Case of a Short Circle on a Transmission Line

The death caused by a too high touch voltage caused by a person touching the tower $i$in the moment when a short circle arises, is a compound event, for which coincidence of two events is necessary: 1. There is a person close to the tower i, who decides to touch it in a certain moment, and 2. The touch voltage to which the person is exposed is too high (dangerous). The probability of the first event is marked as $P_d(i)$, and the probability of the second event is $R_d(i)$. Accordingly, we will get the probability $P(i)$ for a fatal result of the observed event that has occurred on the tower i, by multiplying the probability $P_d(i)$ in the critical moment when the person touches the tower i, with the probability $R_d(i)$ that the voltage to which the person will be exposed will be too high (dangerous), namely:

$$P(i) = P_d(i) \cdot R_d(i);\ \ i = 1, n\,. \tag{9}$$

Having calculated the risks $P(i)$ for each tower place i of the long-distance power line, the total probability for death occurrence caused by a dangerous touch voltage on the power line can be calculated. Introducing indications: $p_i = P(i)$ and q_i = 1-p_i, i = 1, n. Then, according to the probability laws, the probability Q that no death will occur at any tower place of the TL will be:

$$Q' = \prod_{i=1}^{n} q_i = \prod_{i=1}^{n} (1 - p_i) \tag{10}$$

The complement of the probability Q, $P' = 1 - Q'$, gives the probability for occurrence of at least one death at one of the towers, due to a short circle on the TL. If we know the average yearly number of one-phase short circles that occur on a power TL N_{KV}, then the total probability Q that no death case will occur during the year caused by a dangerous touch voltage will be:

$$Q = (Q')^{N_{KV}} \tag{11}$$

This means that the total risk R_d for occurrence of at least one death case during the year due to a dangerous touch voltage will be (12). The same principle is applied to calculate the risk for an accident for one year, caused by dangerous step voltage R_c.

$$R_d = 1 - Q = 1 - (Q')^{N_{KV}} \tag{12}$$

IV. Example

The procedure for calculation of the risk for death occurrence caused by a dangerous touch voltage will be illustrated in the case of the long-distance power TL Dubrovo – Shtip, which currently works at a 110 kV voltage. The data used for calculation are presented in [9]. Due to the short period of functioning of the TL, the average yearly number of one-phase short circles N_{KV} that occur on this TL is not known sufficiently precise. Also, there is no information on the probabilities $P_d(i)$ and $P_c(i)$ of exposure of persons to touch and step voltage in the surrounding of the towers of this long-distance power TL. For these reasons, the calculations were made with assumed values, close to the values on these parameters found in the literature. Parametric analysis of the total risks R_d and $R_c$for different values of the parameters N_{KV}, $P_d(i)$ and $P_c(i)$ was performed.

The 400 kV long-distance powers TL use two types of GS: A and B. The basic characteristics are presented in the table below.

Table 1. Characteristics of the Grounding type A and B

	Resist. R_{100} , Ω	Poten. diff. of touch E_d (%)			Poten. diff. of step E_c(%)		
type		max	med	min	max	med	min
A	4,58	24,2	15,8	12,1	17	2,84	0
B	3,91	13,5	9,4	7,1	13	2,62	0

The distribution of the potentials in different directions in the surrounding of the GS is different. It is more favorable in case of GS of type B. This is illustrated in Fig. 4. Conclusion can be drawn that, from the point of view of security, use of type B would be recommended for towers whose surrounding is charac-terized with higher frequency of movement of people or animals.

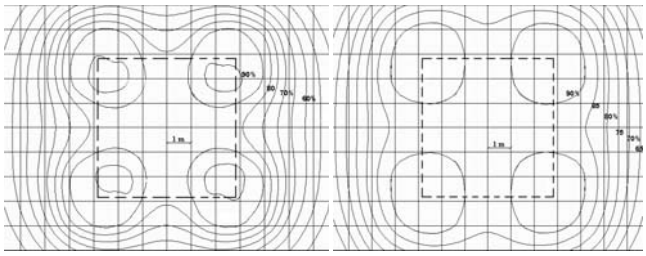


Fig. 4. Equipotent lines on the surface of the ground in the surrounding of the GS: type A to the left, type B to the right

The probability curves for death occurrence due to dangerous touch voltage (R_d), in relation to the potential on the tower and the duration of the short circulation for the GS of type A and of type B are presented in Fig. 5 and 6. According to the professional departments of the ESM for relay protection, the cutting duration due to short circuits on this long-distance power TL is:

a) 0.5 s for short circuits that have occurred on the first 20% and on the last 20% of the longitude of the TL;

b) 0.1 s for the rest (in between) of the transmission line.

This means that along approximately 40% of the TL, there is an increased risk for death occurrence. Using of B-type GS can lower the risk.

In Fig. 7 the numeration of the towers of the TL goes from the TS Dubrovo 400/110 kV to the TS Shtip 110/35 kV. The data on the geometry of the head of 400 kV towers used for calculation of the single electric parameters of the line (z and y) are given in [1] and [9], as and currents of short circuits in TSs.

V. Conclusion

A stochastic procedure for assessment of the risk (probability) for death occurrence caused by high touch and step potentials during a short circuit on a TL has been presented in this paper. In difference from the classical methods with application of conventional methods for assessment of these

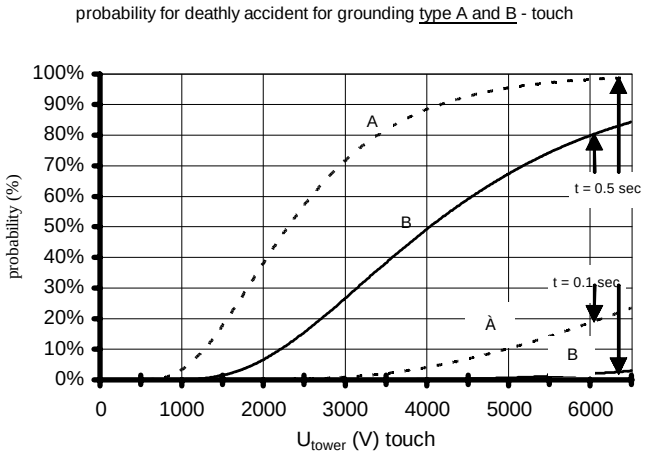


Fig. 5. Probability for death occurrence caused by dangerous touch voltage, depending of the potential of the tower

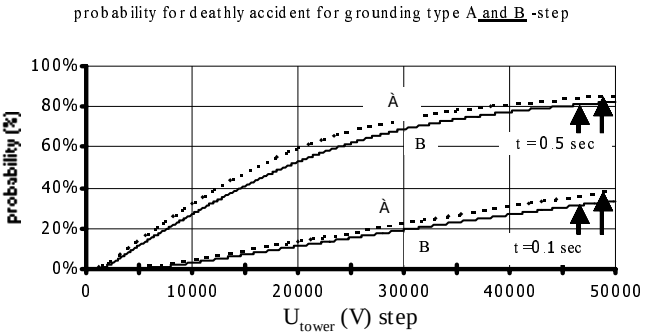


Fig. 6. Probability for death occurrence caused by dangerous step voltage, depending on the potential of the tower

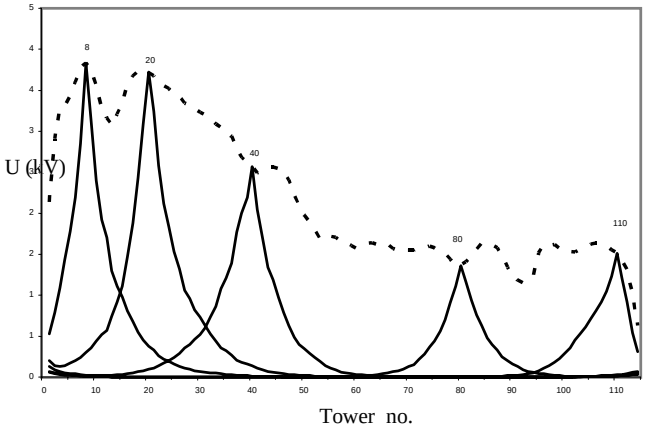


Fig. 7. Potential U on tower k, where a one-phase short current has occurred, k = 8, 20, 40, 80, 110, and distribution of U on the remaining towers i

potentials, where the worst scenario, taking into consideration the maximum values of these parameters is applied, the stochastic approach gives more realistic results. The more realistic estimation of the risks contributes to improvement of the implementation of GS offering upgraded tech-economic implementation of GS on the TL, as well as gives a clear picture on the level of security at work for running TL. The stochastic procedure was applied for risk assessment of the

400 kV long-distance TL Dubrovo – Shtip. The results of the analysis showed that at the starting and at the ending parts of the line, there is an increased risk for high touch potentials. On these parts of the line, the security system cuts the high touch potentials with delay of 0.5 s. The highest risk has been assessed in the vicinity of the TS Dubrovo, where, due to the high value, the short circuits on these towers cause high potentials on the GS. The probability for death occurrence decreases if a B-type GS is used for these towers. In any case, for towers on locations with increased frequency of moving and residence of people and animals, B-type GS is recommended. Certainly, for assessment of the total public risk, the calculations have to be based on precise information on the probabilities for occurrence of short circuits along those parts of the TL, as well as on information about the probability for presence of people.

References

[1] R. Ackovski, N. Acevski, K. Naumoski. “Distribution of Currents at a Ground in the Grounding System of the Transmission Lines”. Fourth Conference MAKO-SIGRE. 2003.

[2] J. Nahman, V. Mijailovic. Selected Chapters in the Field of High Voltage Installations. ETF-Belgrade. Belgrade, 2002 (book).

[3] J. Nahman, V. Mijailovic. High Voltage Installations. BEOPRES. Belgrade, 2000 (book)

[4] M. Zlatanoski. “Accidents Risk in Different High Voltage Installations”, Doctoral Thesis. ETF-Skopje, 1991.

[5] Wang, W., Y.Gervais, D.Mukhedar. ”Probabilistic Evaluation of Human Safety near HVDC Ground Electrode” (85 SM 318-1); T-PWRD Jan 86, pp. 105-110.

[6] M. A. El-Kady, P. W. Hotte, M. Y. Vainberg, ”Probabilistic Assessment of Step and Touch Potentials near Transmission Line Structures”, IEEE, Transactions on Power Apparatus and Systems, Vol. PAS. 102, No. 3. March 1983, pp. 640-645.

[7] M. Zlatanoski, “Assessment of Probability of Exposure to Danger in the Surrounding of High Voltage Installations”, JUKO CIGRE, XX Conference, Neum, 22-26 April 1991.

[8] N. Acevski, R. Ackovski. “Analysis of the Characteristics of Grounding Elements and Safety Criteria’s of Towers of Overhead Lines Using Monte Carlo Simulation”. XXXVII International Conf., ICEST 2002. Nis, 2-4 October, 2002.

[9] EMO – Ohrid, Institute of Energetic HEP – Skopje. Main Project for a 400 kV Long-Distance Transmission Line Dubrovo – Shtip, 2001.

Electric Characteristics Of Barrier Electric Discharge

Peter Dineff[1], Diliana Gospodinova[2]

Abstract – **The external or static volt-ampere characteristic describes the behavior of the barrier discharge at the various stages of development and regimes of application. The technological regime of plasma-chemically active oxygen-containing plasma is examined. The effects of the non-uniformity of the electrical field of discharge, of the barrier gauge and of air gap size on the no-load regime of the discharge are shown.**

Keywords – **plasma treatment, cold plasma, oxygen-containing plasma, barrier discharge, external characteristic.**

I. Introduction

The barrier discharge has many technological advantages that impose its application to the technology of textile and textile fibers, the electronics and microelectronics, the printing industry [1].

The barrier technological discharge burns in air or in various gases and vapors at atmospheric pressure. The absence of a vacuum technological system is one of the great advantages of the barrier discharge in comparison to vacuum discharges used as sources of cold technological plasma, namely the *RF*- and glow discharges.

The great number of ionization and chemical processes going on simultaneously during burning of barrier discharge create certain difficulties not only for the description of this discharge, but also with respect to its control.

The experimental investigations [2] conducted for a continuous time period allow searching for a new integral description and control of the barrier discharge through its external characteristic expressing the relationship between the average value of the electrical current passing through the discharge and the effective value of applied voltage, Fig. 1.

Moreover, it has turned out that this characteristic can be presented by a broken polygon of three linear sectors corresponding to:

- ♢ the stage before discharge ignition or the so-called non-operating regime;
- ♢ the stage of existence of a cold ozone- and oxygen-containing plasma;
- ♢ the stage of existence of a cold plasma that contains nitrogen oxides, Fig. 1.

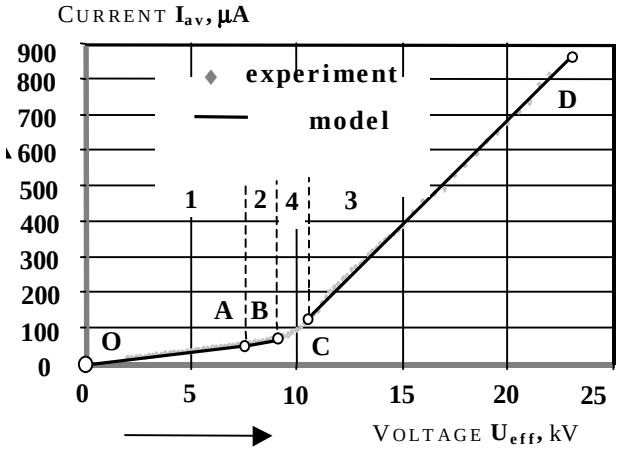


Fig. 1. Operating sectors on the external characteristic of electrical barrier discharge, namely the relationship between the average value of current I_{av} and the effective value of applied voltage U_{eff}: OA – non-operating sector; AB – first operating sector – a cold technological plasma containing ozone and products of its decomposition; CD – second operating sector – a cold technological plasma containing nitrogen oxides; BC – transient area.

THE TASK of the present work consists in examining the behavior at no load, i. e. without any material to be treated in the air gap, of low-frequency (50 Hz) barrier discharge burning in air at atmospheric pressure. The investigations are focused on the first operating part of the external characteristic of the discharge responsible for the generation of oxygen-containing technological plasma.

It is necessary to find new possibilities for increasing the effectiveness of the technological discharge, which is expressed by the high steepness of the working sector and the large intercept on the ordinate axis, Fig. 1.

The experimental investigation is carried out by varying: the non-uniformity of the electric field, the gauge of the glass barrier, and the size of the working air gap of plasma generator.

II. Experimental Investigations

The barrier is a plate with various gauge values that is made of alkaline silicate glass with dielectric permittivity ε=10, electrical volume resistivity $\rho = 10^9$ Ω.m, and $tg\delta = 25$ (at 20°C).

The change in thickness δ of the glass barrier determines the actual change in capacitance C_δ that is introduced by the barrier in the electric discharge circuit. The electric current is of capacitive character.

The values of capacitance C_δ of used glass barriers with various values of thickness δ, which have been calculated

[1]Peter Dineff is with the Faculty of Electrical Engineering, Technical University of Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: dineff_pd@abv.bg

[2]Diliana Gospodinova is with the Faculty of Electrical Engineering, Technical University of Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: dilianang@abv.bg

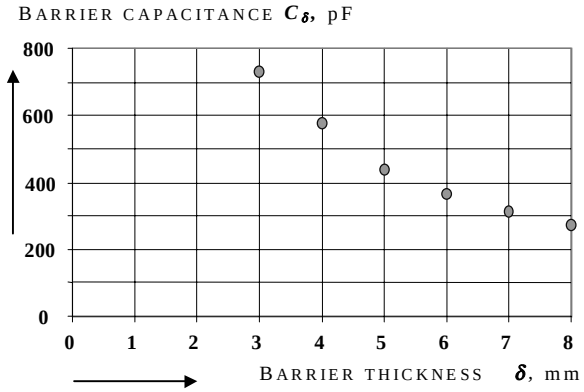


Fig. 2. Change in the electric capacitance C_δ of dielectric barrier - alkaline silicate glass, as a function of the change in its thickness δ.

through the experimentally plotted external characteristics for virtually uniform electrical field of discharge, are shown in Fig. 2.

The virtually uniform electrical field of discharge is realized by using two flat rectangular electrodes placed in parallel to each other at a distance b, forming the working gap in between. The ends of the two electrodes are rounded with a radius of 5 mm in order to diminish the non-uniformity of the field caused by the so-called edge effect.

The non-uniform electrical field is realized by using one of the two large electrodes with active area S_{E1}=651cm^2 and eight flat round electrodes with ∅50 mm and total area $S_{E2} = 8\times$ 19.6=156.7 cm^2. The coefficient of non-uniformity is $\beta = S_{E1}/S_{E2}$=4.15.

These eight electrodes may be situated in a large number of ways with respect to the large rectangular electrode. Two such approaches have been selected and they differ essentially from each other. According to the first one the eight electrodes are placed at distances not permitting their electrical interaction; the second approach requiring that the electrodes are placed in a group forming a maximally dense (hexagonal) package in the plane.

These configurations correspond to two different degrees of non-uniformity of the electrical field despite the same coefficient of non-uniformity β=4.15.

III. Results and Discussion

Based on the experimentally plotted external characteristic, the intensive characteristic of barrier discharge, namely the average value of current density J_{av}, is determined numerically as a function of the effective value of applied voltage U_{eff},Figs. 3÷ 5. It expresses the specific quantity of electricity transferred through the discharge gap in unit time.

The investigation performed shows that the introduction of an increasing degree of non-uniformity of the electrical field leads to a decrease in the intensity of ionization and chemical interactions going on in all sectors of the external characteristic: the inclination of straight sectors diminishes, and the value of intercept or free term goes up.

The cases investigated experimentally may be classified in the order of decreasing intensity of the barrier electrical discharge as follows:

- virtually uniform electrical field;
- eight electrodes placed at considerable distances from one another;
- eight electrodes placed as a group of maximum density.

The technological characteristic of the barrier discharge, however, demonstrates something else: the surface density of the active power of discharge p_a increases with the growing

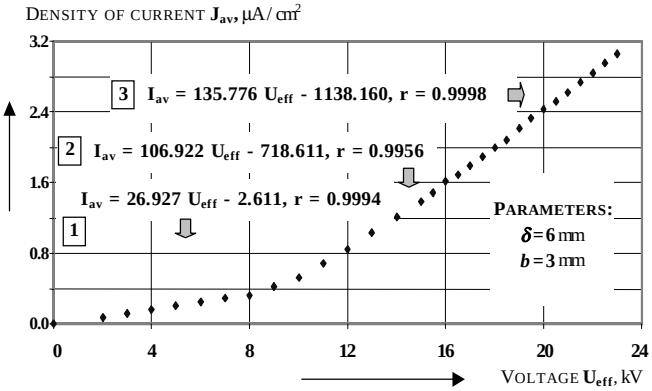


Fig. 3. Characteristic of the intensity of barrier discharge for a homogeneous electrical field. The regression equations of currents describing characteristic parts of the characteristic are presented.

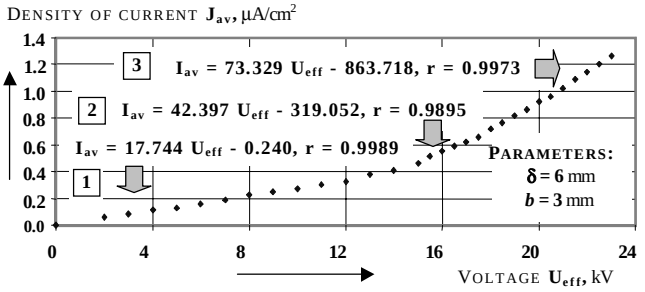


Fig. 4. Characteristic of the intensity of barrier discharge for a nonhomogeneous electrical field with eight autonomous electrodes of ∅50 mm. The regression equations of currents describing characteristic parts of the characteristic are presented.

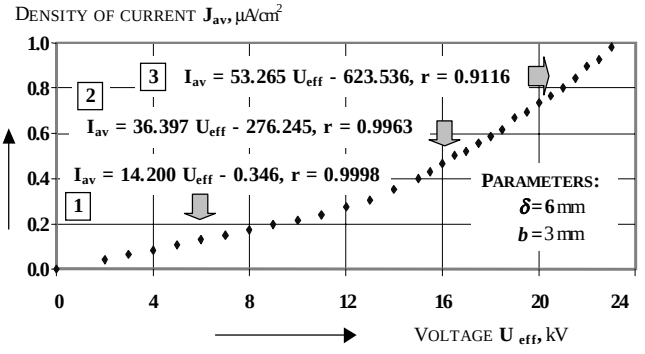


Fig. 5. Characteristic of the intensity of barrier discharge for a nonhomogeneous electrical field with eight grouped electrodes of ∅50 mm. The regression equations of currents describing characteristic parts of the characteristic are presented. These parts correspond to the graphically presented characteristics of the barrier discharge intensity.

non-uniformity of electrical field in the two operating sectors of the characteristic, i.e. in *sector* 1 characterized by the obtaining of oxygen-containing cold plasma, and in *sector* 2 characterized by the obtaining of cold plasma that contains nitrogen oxides, Figs. 6, 7, and 8.

The active power p_a is perceived as a measure for the ionization and plasma-chemical processes going on in the discharge.

The difference observed may be explained by the fact that the surface density of active power p_a accounts not only for the influence of density J_{av} but also for the influence of the crucial parameter of the regime of burning, namely the critical density of current J_{cr}, as well as of the value of the voltage applied to discharge U_S:

$$p_a = (J_{av} - J_{cr})\, U_S = J_p U_S \ .$$

That is why only the corrected current J_p can be a measure for the intensity of the threshold process of burning of barrier electrical discharge.

Rising the non-uniformity of electrical field decreases the critical density of current J_{cr}, and this effect is considerably larger than the decrease in the current density J_{av}, so that the difference $(J_{av} - J_{cr})$ becomes larger. The earlier ignition of the discharge and the displacement of operating sectors to the left determine the change in power density p_a.

The influence of the thickness δ of glass barrier and of the size b of discharge gap on the external characteristic of discharge or on the relationship between the average value of current I_{av} and the effective value of voltageU_{eff} applied to discharge gap is investigated experimentally for the first operating sector, i.e. for the area where the oxygen-containing cold plasma exists. Regression equations modeling the burning process of barrier discharge are obtained in accordance with a well-known methodology.

The *inclination* of the straight line or the current increase rate B, the intercept or the free term of the straight line A, and the correlation coefficient r, taking into account the degree of linear correlation between the current and applied voltage are shown in Figs. 9, 10, and 11.

The respective characteristics for a non-uniform electrical field are not presented because of the fact that the corresponding sector of their external characteristics is characterized by lower parameter values.

The maximal rate of current increase, namely within 200 to 240 μA/kV, is observed for thickness δ=4 mm of the glass barrier (the barrier capacitance being 580 pF) and size of discharge gap within 3 to 12 mm. Such rates of current increase are also observed for glass barrier thickness of 3 mm, but for a discharge gap of 1.5 mm, which is not, however, of great practical importance, Fig. 9.

The intercept B of the straight line modeling the external characteristic in the first operating sector reflects the line

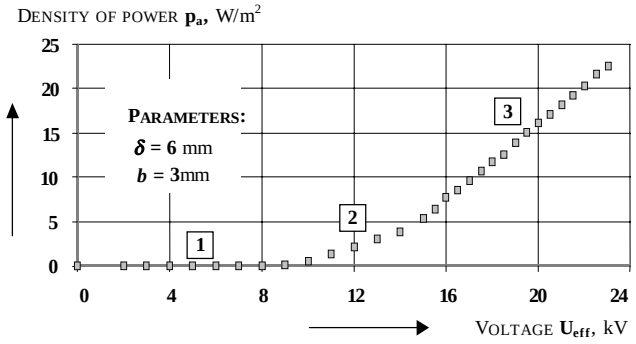


Fig. 6. Technological characteristic of the barrier discharge for a homogeneous electrical field.

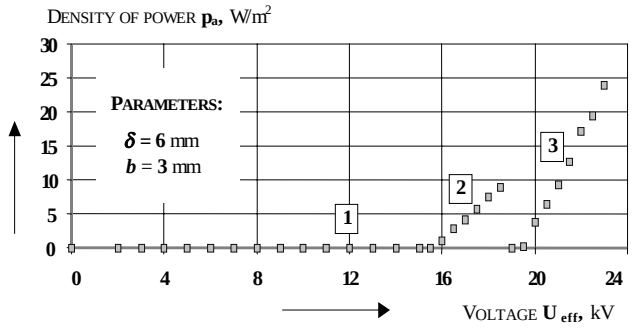


Fig. 7. Technological characteristic of the barrier discharge for a non-homogeneous electrical field with eight electrodes placed at considerable distances from one another.

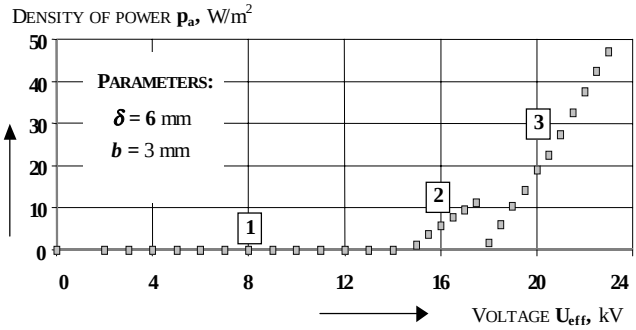


Fig. 8. Technological characteristic of the barrier discharge for a non-homogeneous electrical field with eight electrodes placed in a group.

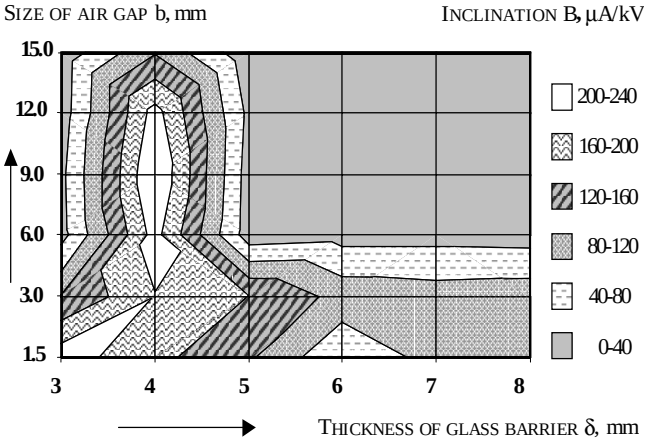


Fig. 9. Effect of thickness δ of the glass barrier and of size b of the discharge gap upon the increase rate B of the average value of current I_{av} with the increase of voltage U_{eff} applied to discharge gap for the first operating sector of the external characteristic.

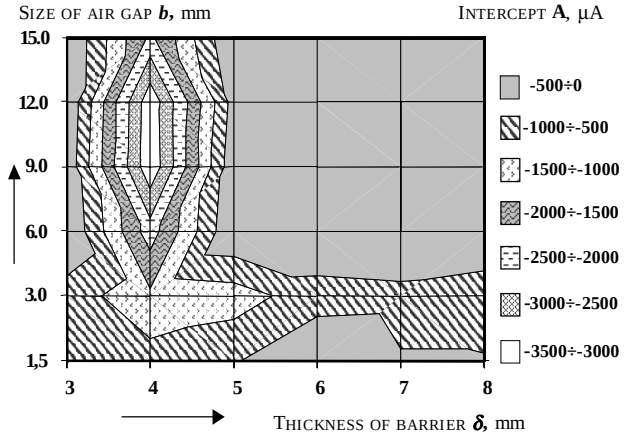


Fig. 10. Effect of thickness δ of the glass barrier and of discharge gap size b on intercept B of the straight line modeling the external characteristic in the first operating sector.

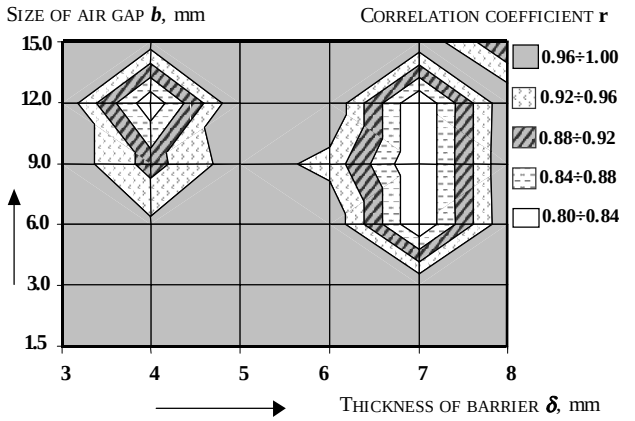


Fig. 11. Effect of the thickness of glass barrier δ and of the size of discharge gap b on the correlation coefficient r of the straight line modeling the external characteristic in the first operating sector.

location with respect to the voltage scale; a larger intercept means displacement of the line to the high values of voltage applied to the discharge gap, and vice versa. From the viewpoint of energy effectiveness it is better to operate at low voltages of discharge burning.

Unfortunately, the large values of intercept A are related to gauge δ=4 mm of the glass barrier or are found where the highest inclination is, Fig. 10.

The inclination is maximally high, but the curve is displaced strongly to the right on the voltage scale, which diminishes the advantage of the high rate of current increase.

It is natural to seek an improvement of the characteristic by augmenting the non-uniformity of the electrical field of discharge for this gauge of the glass barrier.

The value of the coefficient of linear correlation r remains relatively high in the whole region of investigation, except in two small areas, Fig. 11.

For barrier thickness δ=4 mm this relates to large distances between electrodes, namely b=7÷15 mm. For barrier gauge δ=7 mm this concerns nearly the whole range of investigation; it includes values of the discharge gap size b=3÷15 mm.

The observed relatively low values of correlation coefficient r are associated with the characteristic discharge instability in these regions.

IV. Conclusion

The electrical characteristics of the barrier electric discharge with industrial frequency (50 Hz) are obtained on the basis of the external discharge characteristic plotted experimentally.

The average value of the electric current density can perform the role of an intensive parameter of the process of discharge burning, because it does not reflect the threshold character of the process of discharge ignition and the transition to each of the two working parts of the external characteristic.

The growing extent of non-uniformity of the electrical field of discharge increases the surface density of the active power and its rate of increasing with the augmentation of voltage applied to the discharge gap. In such a way the non-uniformity of the electrical field influences positively the electrical and technological characteristics of barrier electrical discharge.

The thickness of the glass barrier exerts an influence on discharge burning through the capacitance it introduces in the electrical circuit of discharge.

References

[1] P. Dineff, D. Gospodinova. Energy-effective plasma-chemical surface modification of polymers and polymeric materials at atmospheric pressure. International conference MED-POWER'02, November 04-06, 2002, Athena, Greece, Proceedings, pp. 1124-1128.

[2] P. Dineff, D. Gospodinova. The surface density of the power as a basic parameter of the plasma-chemical modification of the materials in a barrier electrical discharge, International conference ELMA' 02, September 13-14, 2002, Sofia, Bulgaria, Proceedings, pp. 304-310.

Reduced Admittance Matrix Method for Asymmetrical Load-Flow in Sequence Domain

Ljupco D. Trpezanovski[1] and Vladimir C. Strezoski[2]

Abstract – **In this paper a new linear method for asymmetrical load-flow solution is presented. With 6×6 matrix model of the power system elements, the node-admittance matrix is formed by "overlapping" procedure. Applying the new scaling concept and enhanced bus classification, also synthesizing the low voltage nodes of step-up transformers in the high voltage nodes, the node-admittance matrix dimensions are reduced.**

Keywords – **Node-admittance matrix, Sequence domain, Linear method, Node reduction, Asymmetrical load-flow.**

I. Introduction

Always, the three-phase electrical power systems states are asymmetrical, which deviate more or less from the symmetrical states. The reasons for asymmetrical states are presence of long unbalanced (untransposed) lines and asymmetrical or single-phase loads (as induction furnaces and traction motors etc.). These states cause: negative-sequence currents at generator terminals rise heating in their rotors; malfunctions of protective relays; zero-sequence currents increase greatly the effect of inductive coupling between parallel transmission lines; higher power system loss etc. Therefore, for more precise analysis of three-phase power system states, the asymmetrical load-flow (ALF) analysis are required. Also, ALF calculations are required to study the effects of various phase arrangements of transmission lines, single pole switching, etc.

Because of mutual inductive and capacitive couplings between phases, 6x6 node-admittance matrices in phase domain, which describe the generators, transformers and all lines, are not sparse. But, 6×6 node-admittance matrices which describe the balanced power system elements (practically all generators, transformers and transposed lines) in sequence domain are sparse [1] (all mutual couplings between phases are eliminated). In this domain, only 6×6 node-admittance matrices, which describe untransposed lines, are not sparse. Usually, the solution of ALF problem is performed using methods in phase domain (Newton-Raphson and fast decoupled procedures) [2]. Taking into account that 6×6 node-admittance matrices which represent power system elements are sparse, it is obvious that memory for problem storage and CPU time for problem solution in the phase domain will be much greater against solutions in the sequence domain.

First solutions of the ALF problem in sequence domain are proposed in [3]. In these methods the transformer model in sequence domain is obtained by transformation of it's model in the phase domain. Because of this procedure, the advancements of direct modeling in sequence domain and application of 6×6 sparse matrices are lost. Applying the generator, transformer and transposed line models presented in [1], a new efficient linear method for ALF analysis in sequence domain is established. This method is based on the power system nodal voltage equations.

II. Power System Buses Reduction

Let us consider a three-phase (unbalanced) power system in (asymmetrical) steady state. The system consists of n three-phase buses, i.e. $3n$ phase nodes and the zero potential node R. These buses consist of: N_G generator internal buses (fictitious buses behind the generator synchronous impedances); N_G generator external buses (are buses connecting generators and their step-up transformers); N_G buses of the high voltage sides of step-up transformers; N_L load buses; N_{QV} buses in which synchronous and static compensators, capacitor and reactor units are connected; N_O transfer buses (all buses that do not belong to any of previous five groups).

The most widely used linear power system model is that of the nodal voltage equation. In the phase domain (*abc*) and sequence domain (*dio*), this model says:

$$\underline{\mathbf{Y}}^{abc}_{3n\times 3n}\underline{\mathbf{U}}^{abc}_{3n\times 1} = \underline{\mathbf{I}}^{abc}_{3n\times 1}\,, \tag{1}$$

$$\underline{\mathbf{Y}}^{dio}_{3n\times 3n}\underline{\mathbf{U}}^{dio}_{3n\times 1} = \underline{\mathbf{I}}^{dio}_{3n\times 1}\,. \tag{2}$$

The node-admittance matrix in phase domain $\mathbf{Y}^{abc}_{3n\times 3n}$ has more nonzero elements then the matrix $\mathbf{Y}^{dio}_{3n\times 3n}$ in the sequence domain which is sparse matrix.

The node-admittance matrix would be symmetrical if ideal transformers with complex turns ratios did not appear in the power systems equivalent circuits. It is always the case when the power system is treated in the phase domain, or when it is treated in the sequence domain, but transformed by the New Scaling Concept [4].

Applying this concept for normalization, the phase shifts introduced by the ideal transformers with complex turns ratios in the sequence circuits, are eliminated. Now, each generator and it's step-up transformer can be presented in the sequence domain, separate from the transmission network, as it is shown on the Fig. 1a).

[1]Ljupco D. Trpezanovski is with the Faculty of Technical Sciences, University St. Kliment Ohridski, I. L. Ribar bb, 7000 Bitola, Macedonia, E-mail: ljupco.trpezanovski@uklo.edu.mk

[2]Vladimir C. Strezoski is with the Faculty of Engineering, University of Novi Sad, Fruskogorska 11, 21000 Novi Sad, Yugoslavia, E-mail: streza@eunet.yu

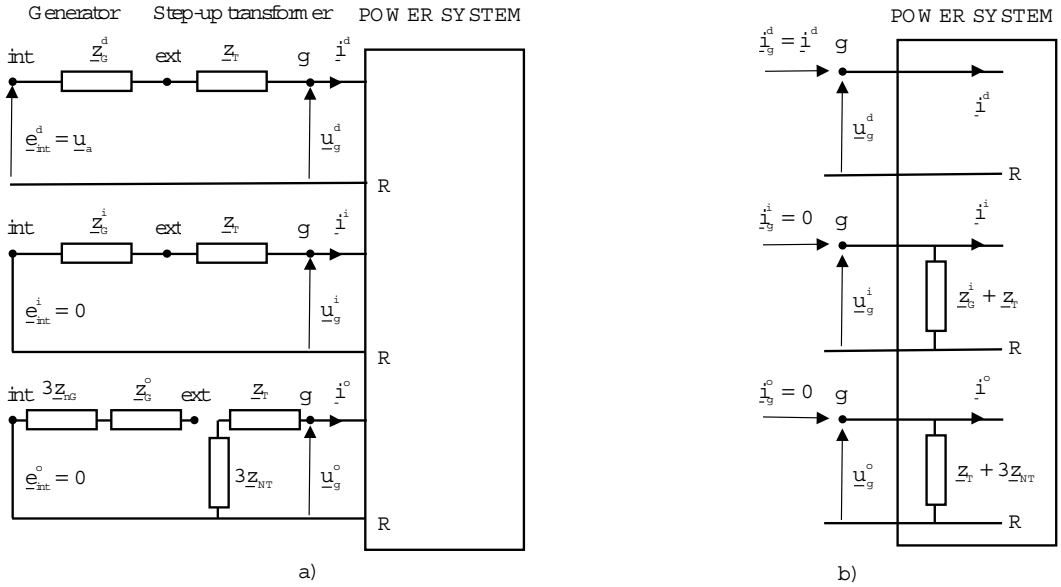


Fig. 1. Scaled sequence circuits of synchronous generator and step-up transformer a) and b) synthesized in the high voltage bus g of the step-up transformer.

The superscript of positive-sequence parameters is d, for negative-sequence is i and for zero-sequence is o; the internal and external generator buses are signed by *int* and *ext*, respectively.

The scaled sequence impedances of the synchronous generator are denoted by $\underline{z}_G^d$, $\underline{z}_G^i$ and $\underline{z}_G^o$; $\underline{z}_{nG}$ represents the generator grounding impedance; the phase a generator open-circuit voltage is denoted by $\underline{u}_a$ and it is equal to the positive-sequence internal bus voltage $\underline{e}_{int}^d$. The transformer positive-sequence and grounding impedances are denoted as $\underline{z}_T$ and $\underline{z}_{NT}$, respectively.

Because the generator internal bus voltage, as well as the voltage drops on the generator and transformer impedances are not of interest simultaneously with values of other power system quantities, the voltage control and the active power control are associated with the high voltage transformer bus g (that is the usual practice). Thus, the circuits presented in Fig. 1a are simplified as those presented in Fig. 1b, where the internal and external generator buses are synthesized in the high voltage bus of the step-up transformer. In this case, parameters of the positive-sequence are omitted in the Fig. 1b, but parameters of the negative and zero-sequence circuits are suppressed in the transmission network. The high voltage bus denoted by g may be of $P_\Sigma V$, θV or $P_\Sigma Q_\Sigma$ type [5].

It is obvious that synthesizing procedure enables power system buses reduction for $2N_G$ buses. Now, the power system can be treated as a system with $r = n - 2N_G$ buses or $3r$ nodes.

III. Reduced Node-Admittance Matrix

The node-admittance matrix is formed for a sequence circuits without ideal transformers with complex turn ratios and for a power system model with reduced number of nodes $3r$. Thus, it will be symmetrical and can be derived very simply by inspection of the power system structure. Applying the power system elements models given by 6x6 matrices [1] and "overlapping" procedure [6], the reduced node-admittance matrix is forming step by step.

Let us consider "overlapping" of two three-phase power system elements. The element E is connected between buses p and j and it's 6×6 node-admittance matrix is signed as $\mathbf{Y}_E$. The element F is connected between buses j and k and it's 6×6 node-admittance matrix is signed as $\mathbf{Y}_F$. The corresponding matrices

$$\mathbf{Y}_E = \begin{bmatrix} \mathbf{Y}_E^{(pp)} & \mathbf{Y}_E^{(pj)} \\ \mathbf{Y}_E^{(jp)} & \mathbf{Y}_E^{(jj)} \end{bmatrix} \text{ and } \mathbf{Y}_F = \begin{bmatrix} \mathbf{Y}_F^{(jj)} & \mathbf{Y}_F^{(jk)} \\ \mathbf{Y}_F^{(kj)} & \mathbf{Y}_F^{(kk)} \end{bmatrix}$$

are consist of four 3×3 sub-matrices. The sub-matrices $\mathbf{Y}_E^{(pp)}$, $\mathbf{Y}_E^{(pj)}$, $\mathbf{Y}_E^{(jp)}$, $\mathbf{Y}_E^{(jj)}$ and $\mathbf{Y}_F^{(jj)}$, $\mathbf{Y}_F^{(jk)}$, $\mathbf{Y}_F^{(kj)}$, $\mathbf{Y}_F^{(kk)}$ are diagonal if the elements E and F are balanced. But if these elements are untransposed lines, these matrices are with non-zero elements.

The connection way of these two elements in the power system is shown on the Fig. 2.

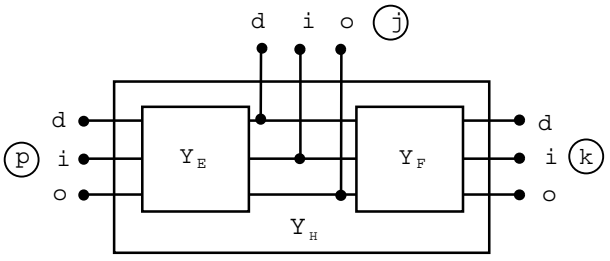


Fig. 2. Elements E and F connection in the power system.

The resultant node-admittance matrix for this part of the power system, consist of these two elements, is calculated as:

$$\mathbf{Y}_H = \begin{bmatrix} \mathbf{Y}_E^{(pp)} & \mathbf{Y}_E^{(pj)} & \mathbf{O} \\ \mathbf{Y}_E^{(jp)} & \mathbf{Y}_E^{(jj)} + \mathbf{Y}_F^{(jj)} & \mathbf{Y}_F^{(jk)} \\ \mathbf{O} & \mathbf{Y}_F^{(kj)} & \mathbf{Y}_F^{(kk)} \end{bmatrix}. \quad (3)$$

Thus, continuing by this procedure, adding element by element, the reduced node-admittance matrix $\mathbf{Y}_{3r\times 3r}^{dio}$ for the power system of $3r$ nodes, can be formed very easy.

IV. Linear Model in Sequence Domain

The nodal voltage equation for the power system represented with reduced number of buses shown on Fig. 1b, is:

$$\underline{\mathbf{Y}}_{3r\times 3r}^{dio}\underline{\mathbf{U}}_{3r\times 1}^{dio}=\underline{\mathbf{I}}_{3r\times 1}^{dio}\,. \tag{4}$$

The dimensions of the linear power system model given by Eq. (1) are $3n\times 3n$. In contrast to this model, the dimensions of the proposed model (Eq. (4)) are reduced to $3r\times 3r$. The proposed model is advanced with respect to the model given by Eq. (2), [3], as follows:

1. The node-admittance matrix is performed for a circuit without ideal transformers with complex turns ratios (thus, it is symmetrical and it can be derived very simply by inspection);
2. The node-admittance matrix dimensions are reduced for all generator internal and external buses;
3. There is no need to introduce small fictitious zero-sequence admittances at Δ-sides of step-up transformers to avoid the singularity of corresponding node-admittance matrices.

If the three-phase bus, type θV is a last numbered bus r, then Eq. (4) represented with sub-matrices, gets the form:

$$\begin{bmatrix} \mathbf{Y}_{11}^{dio}\mathbf{Y}_{12}^{dio}\cdots\mathbf{Y}_{1s}^{dio}\cdots\mathbf{Y}_{1t}^{dio}\cdots\mathbf{Y}_{1r}^{dio}\\ \mathbf{Y}_{21}^{dio}\mathbf{Y}_{22}^{dio}\cdots\mathbf{Y}_{2s}^{dio}\cdots\mathbf{Y}_{2t}^{dio}\cdots\mathbf{Y}_{2r}^{dio}\\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\\ \mathbf{Y}_{s1}^{dio}\mathbf{Y}_{s2}^{dio}\cdots\mathbf{Y}_{ss}^{dio}\cdots\mathbf{Y}_{st}^{dio}\cdots\mathbf{Y}_{sr}^{dio}\\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\\ \mathbf{Y}_{t1}^{dio}\mathbf{Y}_{t2}^{dio}\ \ \mathbf{Y}_{ts}^{dio}\ \ \mathbf{Y}_{tt}^{dio}\ \ \mathbf{Y}_{tr}^{dio}\\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\\ \mathbf{Y}_{r1}^{dio}\mathbf{Y}_{r2}^{dio}\ \ \mathbf{Y}_{rs}^{dio}\ \ \mathbf{Y}_{rt}^{dio}\ \ \mathbf{Y}_{rr}^{dio} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^{dio}\\ \mathbf{U}_2^{dio}\\ \cdots\\ \mathbf{U}_s^{dio}\\ \cdots\\ \mathbf{U}_t^{dio}\\ \cdots\\ \mathbf{U}_r^{dio} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1^{dio}\\ \mathbf{I}_2^{dio}\\ \cdots\\ \mathbf{I}_s^{dio}\\ \cdots\\ \mathbf{I}_t^{dio}\\ \cdots\\ \mathbf{I}_r^{dio} \end{bmatrix} \tag{5}$$

In common terms, the sub-matrix with dimensions $3\times 3$is:

$$\mathbf{Y}_{st}^{dio}=\begin{bmatrix} \underline{Y}_{st}^{dd} & \underline{Y}_{st}^{di} & \underline{Y}_{st}^{do}\\ \underline{Y}_{st}^{id} & \underline{Y}_{st}^{ii} & \underline{Y}_{st}^{io}\\ \underline{Y}_{st}^{od} & \underline{Y}_{st}^{oi} & \underline{Y}_{st}^{oo} \end{bmatrix},\ s,t=1,2,..,r$$

.

If the three-phase element, connected in the power system between buses s and t is balanced, this sub-matrix is diagonal. But, if the element is unbalanced this sub-matrix is with all non-zero elements. The same conclusion can be established for $s=t$. Namely, if any three-phase element connected into bus s (or t) is unbalanced, the sub-matrix $\mathbf{Y}_{ss}^{dio}$ (or $\mathbf{Y}_{tt}^{dio}$) is with all non-zero elements. The sequence circuits voltages of bus s and injected currents in bus s (s=1,2,...,t,...,r) are given in matrix form, respectively $\mathbf{U}_s^{dio}$ and $\mathbf{I}_s^{dio}$.

Depending of the power system buses type, the following values are specified:

a) for the slack bus, type θV, the angle and magnitude of a voltage in positive-sequence circuit:

$$\underline{U}_r^d=\underline{U}_{r,sp}^d=U_{r,sp}^d\angle\theta_{r,sp}^d; \tag{6}$$

b) for the buses type $P_\Sigma V$, values of three-phase injected active powers (P_g^Σ) and magnitude of the positive-sequence voltage:

$$P_g^\Sigma=P_g^a+P_g^b+P_g^c=P_{g,sp}^\Sigma \text{ and } U_g^d=U_{g,sp}^d,\ g\in\{P_\Sigma V\};$$

c) for the bus type PQ, values of three pairs of injected active and reactive powers:

$$P_p^a=P_{p,sp}^a;\ P_p^b=P_{p,sp}^b;\ P_p^c=P_{p,sp}^c,$$
$$Q_p^a=Q_{p,sp}^a;\ Q_p^b=Q_{p,sp}^b;\ Q_p^c=Q_{p,sp}^c,\ p\in\{PQ\}.$$

Because for the slack bus, the complex voltage of the positive-sequence circuit is known (specified), the corresponding equation should be excluded from the linear equations system given by matrix Eq. (5). This exclusion is performed by specific transformation of Eq. (5) presented in [7], without losing the dimensions of the equations system. After the transformation procedure, a new corrected system of linear equations (Eq. (7), with the same unknown values as a system given by Eqs. (4) or (5), is established:

$$\mathbf{Y}_{3r\times 3r}^{(dio)k}\mathbf{U}_{3r\times 1}^{dio}=\mathbf{I}_{3r\times 1}^{(dio)k}. \tag{7}$$

Although, the modules of the complex voltages in the nodes of the positive-sequence circuit for the $P_\Sigma V$ type buses are specified, in this method they are treated as unknown values.

V. Linear Equations System Solution

The main purpose of the asymmetrical load-flow solution is to obtain the complex voltages in all power system buses. With known complex voltages, the injected active and reactive powers into all buses, as well as the powers in all system elements, can be easy calculated.

The solution of the linear equations system, given in matrix form by Eq. (7) is iterative. For the initial iteration (h=0) the initial values are needed. The symmetrical phase voltage "flat profile" is preferred for the initial values: 1 p.u. for the PQ type and specified magnitude values for the $P_\Sigma V$ type buses. Assuming $\cos\varphi_g$ factor (usually nominal) for the block (generator and its step-up transformer), the initial values of the injected reactive powers into $P_\Sigma V$ type buses are:

$$Q_g^\Sigma=P_{g,sp}^\Sigma tg\varphi_g,\ \ g\in\{P_\Sigma V\}\,. \tag{8}$$

The positive-sequence injected currents initial values, in this type of buses, can be calculated from the equation, which expresses the total injected power in the same bus:

$$\underline{U}_g^d\underline{I}_g^{d*}+\underline{U}_g^i\underline{U}_g^{i*}\underline{Y}_g^{i*}+\underline{U}_g^o\underline{U}_g^{o*}\underline{Y}_g^{o*}=(P_{g,sp}^\Sigma+jQ_g^\Sigma/3,\ \in\{P_\Sigma V\}.$$

The negative- and zero-sequence injected currents into $P_\Sigma V$ and θV type of buses, in every iteration are:

$$\underline{I}_g^i=0;\ \underline{I}_g^o=0,\quad g\in\{P_\Sigma V\}\cup\{\theta V\}.$$

Zero valued currents are obtained by suppression of the negative- and zero-sequence admittances (equivalent of generator and it's step-up transformer) into reduced power system node-admittance matrix $\mathbf{Y}^{dio}_{3r\times 3r}$ (section II).

Taking into account all mentioned above, the complex voltages in the sequence circuits nodes, can be calculated in a few steps, for each iteration h.

Step 1. The positive-sequence injected current in buses type $P_\Sigma V$ (with subscript g) are calculated by following procedure:

a) Complex voltages in iteration h, are corrected with specified magnitude and calculated argument, as:

$$\underline{U}_g^{(d)k}(h) = \underline{U}_g^d(h)\, U^d_{g,sp} \Big/ U_g^d(h),$$

b) With the values from iteration h, the injected reactive powers are calculated as:

$$Q_g^\Sigma(h) \cong 3I_m \left[\underline{U}_g^{(d)k}(h)\, I_g^{d*}(h) - \underline{U}_g^i(h)\, \underline{U}_g^{i*}(h)\, Y_g^{i*} - \underline{U}_g^o(h)\, \underline{U}_g^{o*}(h)\, Y_g^{o*}\right].$$

c) The generator internal bus complex voltage $E^d_{int}(h)$ is calculated from the equation for total injected power in the bus g:

$$P^\Sigma_{g,sp} + jQ^\Sigma_g(h) \cong 3\left[\underline{E}^d_{int}(h) - \underline{U}_g^{(d)k}(h)\right]^* \cdot \underline{Y}_g^{d*}\underline{U}_g^{(d)k}(h) - 3\underline{U}_g^i(h)\underline{U}_g^{i*}(h)\underline{Y}_g^{i*} - 3\underline{U}_g^o(h)\underline{U}_g^{o*}(h)\underline{Y}_g^{o*}.$$

d) The new "more correct" value of the positive-sequence injected current is obtained as:

$$\underline{I}_g^d(h+1) \cong \left[\underline{E}^d_{int}(h) - \underline{U}_g^d(h)\right]\ \underline{Y}_g^d.$$

Step 2. In this step the positive-, negative- and zero-sequence injected currents in the buses type PQ are calculated.

a) By the complex sequence voltages $\underline{U}_p^{\,d}(h)$, $\underline{U}_p^i(h)$, $\underline{U}_p^o(h)$, the phase complex voltages $\underline{U}_p^a(h)$, $\underline{U}_p^b(h)$ and $\underline{U}_p^c(h)$ are calculated.

b) The phase injected currents are calculated by the specified powers, with equations:

$$\begin{aligned}
\underline{I}_p^a(h+1) &\cong -\left(P_p^a + jQ_p^a\right)^* \Big/ U_p^{a*}(h),\\
\underline{I}_p^b(h+1) &\cong -\left(P_p^b + jQ_p^b\right)^* \Big/ U_p^{b*}(h),\\
\underline{I}_p^c(h+1) &\cong -\left(P_p^c + jQ_p^c\right)^* \Big/ U_p^{c*}(h).
\end{aligned} \qquad (9)$$

c) If the phase to sequence domain transformation is used, from the phase injected currents (Eqs. (9)), the sequence injected currents $\underline{I}_p^d(h+1)$, $\underline{I}_p^i(h+1)$ and $\underline{I}_p^o(h+1)$ are obtained very easy.

Step 3. The sequence injected currents calculated in previous steps are applied in the right side in Eq. (5). After appropriate transformation [7], the new modified system (Eq. (7)) with same sequence unknown complex voltages is obtained. A very suitable method for this linear equations system solution is Gauss's method of coefficient elimination.

Step 4. If the convergence criteria is achieved, the iteration procedure stops. But, if it is not achieved, a new iteration going on from the step 1.

VI. Method Verification

The ALF solution for several power systems (including the entire power system of the Republic of Macedonia) in several asymmetrical states are performed with above presented method. The results validity are compared with rezults obtained by the well known Newton-Raphson and fast decoupled procedures in phase domain. The results are identical among all methods. Althought the proposed method is linear, the eficiency in occupied CP memory and CP calculation time is on it's side, against the methods in phase domain [2].

VII. Conclusion

In this paper a new linear method for asymmetrical load-flow solution in sequence domain is presented. Several recently published procedures as: power system elements modeling in sequence domain and the negative- and zero-sequence equivalent admittances of the generator and it's step-up transformer suppression into the power system node-admittance matrix, are applied. A new reduced form of power system node-admittance matrix is obtained by "overlapping" procedure. Nodal voltage equations for the power system with reduced number of buses represents a linear model. A procedure for the node injected sequence circuits currents calculation, needed for equations system solution is explained.

References

[1] Lj.Trpezanovski, V.Strezoski, "Power System Elements Modeling in Sequence Domain", XXXVII International Scientific Conference on Information, Communication and Energy Systems and Technologies, Proceedings, Vol. 2, pp. 459-462, Nis, Yugoslavia, 2002.

[2] J.Arrillaga, C.P.Arnold, B.J.Harker: *Computer Modelling of Electrical Power Systems*; John Wiley & Sons Ltd, 1983.

[3] X.-P.Zhang, H.Chen, "Asymmetrical three-phase load-flow study based on symmetrical component theory", *IEE Proc.-Gener. Transm. Distrib.*, Vol. 141, No. 3, pp. 248-252, May 1994.

[4] V.C.Strezoski, "New scaling concept in power system analysis", *IEE Proc.-Gener. Transm. Distrib.*, Vol. 143, No. 5, pp. 399-406, 1996.

[5] V.Strezoski, Lj.Trpezanovski, "Three–phase asymmetrical load–flow", *International Journal of Electrical Power and Energy Systems*, Vol. 22, No. 7, pp. 511-520, October 2000.

[6] F.L.Alvarado: "Formation of Y-Node Using the Primitive Y-Node Concept", *IEEE Trans. on PAS*, Vol. PAS-101, No. 12, pp. 4563-4571, December 1982.

[7] Lj.Trpezanovski, Three-Phase *Asymmetrical Load Flow in Power Systems*, PhD thesis (in Serbian), University of Novi Sad, Faculty of Technical sciences, Novi Sad, July 2000.

Multi-objective Power System Planning

Blagoja Stevanoski[1] and Arsen Arsenov[2]

Abstract – **This paper discusses a multi-objective optimization approach to generation expansion planning. Power system planning, nowadays, must deal with a wide range of options, with a large degree of uncertainty and with conflicting objectives due to the liberalization of the electricity market and increasing concern for the environmental impact. Multicriteria decision-making method is combines with conventional dynamic programming to compare different alternatives. Practical application of proposed approach concern the Macedonian electric system.**

Keywords – **Multiple criteria, uncertainty, risk, decision analysis, integrated resource planning**

I. Introduction

One of the basic objectives of power system planning is to determine the best possible investment options for the location, technology and timing of installing generation facilities, financing such investment and meeting satisfactory operation requirements in order to meet future demand for electricity over a planning horizon. The criteria, usually, are to minimize the total cost and maximize the reliability with different type of constraints. The total cost has two basic components: the investment cost given by construction cost of generating units and interconnection links and the operating cost associated to the fuel cost of the thermal system units.

The major complicating factors in such analysis include simultaneous consideration of demand side resources, uncertainties and risk management. Many planning factors are uncertain during planning process such as [7]:

- demand growth;
- fuel prices;
- interest and inflation rates;
- economic growth;
- environmental constraints;
- financial constraints;
- public opinion;

Etc.

The standard solution approach of the generation expansion problem in some planning models are deterministic by minimizing the total present worth of investment and operation costs subject to various types of constraints. The expansion plan is based on the best available forecasts and takes the optimal investment decision associated to the first stage of this plan (for example, the current year). This approach does not necessarily lead to the most adequate expansion strategy because an investment decision for the current stage is optimal under the assumption that the future conditions will occur as predicted. Used values for the model parameters, usually, are determined by more or less complex estimations. The problem with them is that these estimations have proved to be erroneous most of the times.

The other usual way of introducing uncertainty has been by probabilistic analysis. But, the degree of uncertainty may vary, ranging from items showing stochastic behavior within a known probability distribution to that exhibiting apparently chaotic behavior. Magnitudes may be known, but not frequency or timing and it is really difficult to assign probabilities to any of the different considered events. Although probabilistic analysis may be considered suitable for short-term uncertainties, it is certainly not the case for most uncertainties implied in a long-term planning process. For long-term uncertainties, we cannot assign probabilities, but rather possibilities.

Therefore, it becomes necessary to introduce in the decision making process a systematic and consistent treatment of these sources of uncertainty [7]. This task is very complex in metho-dological and computational terms. In contrast with “natural” uncertainty such as hydrological variation or equipment outages, many uncertainties mentioned above are dependent on economics and politics organization. Inclusion of environmental groups, industrial firms and consumer groups into the decision-making process related to environmental quality, reliability and cost of electricity, play a role in choosing a strategy to meet future electric demand. The fact that the decision process includes groups with such different viewpoints makes the choice of a single plan more difficult. This not only requires a wider scope in the methodological tools, but changes in the way results are presented. The concept of a “plan” as an expansion schedule is inadequate. It is necessary to have expansion strategies which take into account the “tree” of possible future scenarios and the dynamics of the decision making process.

The traditional objective function must be reformulate because the use of only one objective (usually cost) not adequate represent conflicting objectives such as, for example, economic costs and environmental impacts and etc.

Therefore, power system planning is decision process , which attempts to resolve multiple-conflicting objectives [2]. It is often not possible to identify a single plan, which simul-

[1]Blagoja Stevanoski is with the Faculty of Engineering, I. L. Ribar bb, 7000 Bitola, Macedonia, E-mail:blagoj.stevanovski@uklo.edu.mk

[2]Arsen Arsenov is with the Faculty of Electrical Engineering, Karpos II bb, 1000 Skopje, Macedonia, E-mail: aarsenov@etf.ukim.edu.mk

taneously optimizes all objectives. The solution is the selection of a "robust" plan, which may not be the best plan in some future, but is a good plan in most futures.

Uncertainty imposes risk and each type of uncertainty has different implication for decision-makers and analysts.

Risk management and evaluating of risk management strategies is now an important part of the integrated resource planning process [3]. Competitive forces are adding new risks that make responsible decision-making even more difficult. Planners are shifting from simply optimizing resource investments assuming a certain future to a different mode planning, assuming uncertainty.

Each type of uncertainty has different implication for decision-makers and analysts. At the broadest level, three classes of prescriptions exist. For uncertainties in the operating environment of the decision makers (uncontrollable exogenous variable) technical investigation is in order. For uncertainties about guiding values of the decision-makers (weighting factors on the objective function), consultation about policy priorities is needed. Finally, for uncertainties about future decisions on related agendas, coordination among actors is crucial.

II. Methodology

The framework and methodology presented in this paper accepts the reality that there is no optimal solution, in that the future is essentially unknowable. For these reason the framework is based on the comparative analysis of multiple scenarios concerning alternative futures. The framework allows the resource planner to utilize existing accepted planning and financial tools to develop the information upon which the trade-off analysis is based. High speed and inexpensive computational capabilities make the generation and evaluation of multiple scenarios possible

The methodology tries to integration the following characteristics [2]:

- considers multiple criteria;
- based on optimization techniques;
- takes into account the preferences of the different social interest group;

The core of the methodology is WASP III Plus production-cost simulation model developed by the International Atomic Energy Agency and the Argonne National Laboratory. Used in conjunction with other analytic methods, a wide range of options and uncertainties can be evaluated.

By evaluating how different supply strategies perform under a variety of possible futures, robust strategies can be identified. Capacity expansion strategies are evaluated against a range of possible changes in electric demand, fuel prices, and fuel availability. The comparative performance of various strategies over the range of possible future events identifies the most robust or least vulnerable strategies with respect to price, reliability, environmental emissions and other important measures.

The first phase of the proposed method [2] is the selection and characterization of technologies and fuels, both in the generation and demand sides, which may be available for the electricity system for the planning horizon.

The second step requires the generation of scenarios that incorporate all the uncertainties to which the planning process is subject (technical parameters, macroeconomic data, regulatory measures and etc.). Given the interrelation among many of these parameters, it should be possible to generate a small number of scenarios, which cover the whole range of uncertainties. These scenarios should be generated by means of interaction between analysts and decision-makers.

Than, the preferences of the decision-makers regarding the criteria considered have to be estimated using the analytic hierarchy process. It is based on a pairwise comparison of the criteria considered, and assignment of values for this comparison from a lexicographic scale. In addition, this method may be extended for the estimation of group preferences, for example by the weighted arithmetic mean method.

The preferences held by decision-makers may vary depending on the range of attribute values and therefore this information has to be presented to them before they elicit their preferences. The usual approach is to present them with payoff matrices. Payoff matrices are matrices where the values of the attributes of the problem are shown for the optimal solutions obtained for every one of the criteria considered. These matrices help understand the trade-off among conflicting criteria, and show the ideal and anti-ideal values for each of the attributes or criteria. Payoff matrices are built by running traditional single criteria optimization models for each of the criteria considered.

Once the criteria have been weighted, the generation of the efficient strategies for each scenario is undertaken by means of compromise programming theory. Compromise programming is based on the assumption that the preferred solution will be the one whose distance to the ideal point (the one in which all criteria considered reach their optimal level) is minimal.

The efficient strategies generated up may only be considered in an economic sense. When other uncertainties are introduced, it is necessary to incorporate risk analysis into the decision-making model.

III. Case Study

The multi-objective criteria were used to evaluate the relative impacts of some capacity expansion scenarios affecting to the development of Macedonian power system over a period of twenty year. The three criteria considered were:

- economic cost;
- fuel import vulnerability;
- risk of plant disaster;

At present, the major characteristic of Macedonian electric power system is domination of thermal power plants, which produced about 85% of total electricity demand. The

whole installed capacity is 1440 MW distributed as fellows: (1) Steam power pants: 795 MW; (2) Fuel oil power plants 210 MW; (3) Hydroelectric power plants 435 MW [5].

As were described before, the first step of the proposed method consisted in selection and characterization of technologies and fuels expected to be available during planning period. The remaining reserves of coal and lignite fuel in Macedonia are quite limited. Additional coal-fired generating capacities are based on imported coal. There is a significant natural gas supply available through a pipeline from Ukraine and Russia. Several types of gas-fired thermal power plants are considered as candidates. Nuclear power is also considered as one of the potential long-term option for electricity generation. The new capacity thermal units are described in Table 1.

Table 1. Thermal Candidate Units

Name	Net Capacity (MW)	Fuel Cost ($/GJ)	Overnight Cost ($/kW)	Fuel Type
Bitola Rehabilitation	207	1,42	810	Lignite
Imported Coal	207	1,78	1450	Lignite
Gas Turbine (GT)	122	2,86	280	Natural Gas
Combined Gas Turbine	220	2,86	620	Natural Gas
Cogeneration (COGN)	175	2,86	670	Natural Gas
Nuclear	323	0,43	1860	Uranium

The second step consisted on the generation of small, but consistent number of scenarios, which might account for the uncertainty related to socio-economic aspects. The generation of scenarios was based, mostly on rehabilitation of thermal power plants “Bitola”, imported coal and involved a natural gas-fired plants. For this planning exercise three presumed scenarios were considered: Power engineers scenario (Case I), Energy economist scenario (Case II) and Public scenario (Case III).

As was mention above, three criteria was considered: (1) Costs as a total present worth cost including capital, fuel and O&M cost; (2) Fuel import vulnerability as a cumulative power production (MWh) for each fuel type and (3) Risk of plant disaster as a cumulative installed capacity (MW) [1]. In order to obtain the preference weights from each social interest group the trade-off was done between attributes for the different scenarios considered. This was done by using of the payoff matrices, which were build with a single criteria classical generation expansion model, by which the optimal solution for each of the three criteria considered was determined. Each group compared different criteria and thus the individual preferences were obtained. This individual preferences were aggregated using the weighted arithmetic mean method and are presented in Table 2.

Table 2. Relative weights among preferences

	Cost	Plant disastrous risk	Fuel import vulnerability
Power engineers	0,66	0,21	0,13
Nuclear engineers	0,68	0,15	0,17
Energy economists	0,75	0,11	0,16
Public	0,35	0,52	0,13

When the preferences given before were introduced into the multiple-criteria optimization model, the different efficient solutions under every scenario were obtained. The multiple-criteria optimization model was basically a classical generation expansion model in which the objective function was formed by adding all the objectives considered, previously normalized and weighted, according to the compromise programming theory.

As should be expected, the introduction of additional criteria (besides from economic costs) generates a more expensive solution depending on the preferences of the decision maker groups toward the balance between these conflicting objectives.

It is important to note that this solution will be modified when other uncertainties are introduced into the analysis, in such a way that the efficient strategies will be different under different scenarios. The determination of which of these efficient strategies is the best requires evaluating their behavior under every scenario considered and using a decision rule, which incorporates the attitude of the decision-maker toward risk. The optimal strategies for each scenario, first, should be obtained, for each set of decision-maker preferences. As an example, the values of attri-butes of the optimal planning strategies for power engineers set of preferences, under each of scenarios are presented in Table 3.

Table 3. Attributes for the optimal strategies under each

	Cost 10^6 ($)	Plant disastrous risk (MW)	Fuel import vulnerability 10^3 (MWh)
Case I	1719	2681	156
Case II	1772	2590	142
Case III	1854	2825	135

The values of the different attributes are different across the considered scenarios. This reflects that the optimal expansion under one scenario may be bad one under another. To select robust strategies was used decision theory (the Savage criterion) which minimizes regret across all scenarios. Regret was calculates as the Manhattan distance between the studied solution and the ideal solution for each scenario. The

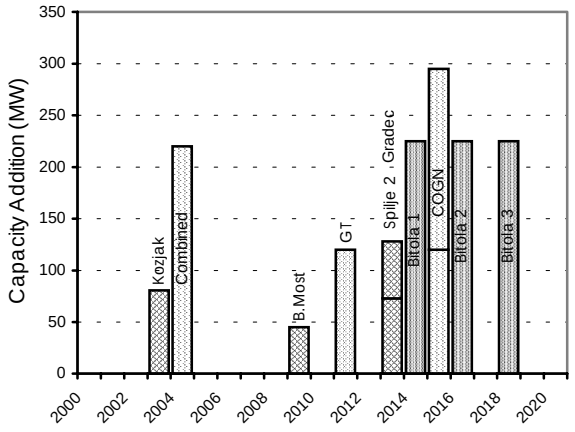


Fig. 1. Optimal expansion plan for the Case I

most robust strategy for power engineers weights is shown in Fig. 1.

IV. Conclusion

The use of a multiple criteria in power system planning process is useful in assessing how different options are suited to preparing for an uncertain future. Different perceptions of relative value of competing attributes allows decision-makers to weigh and constructively discuss trade-off associated with any one decision.

This model incorporates the preferences of different groups of decision-makers, so that the results of the model may be interpreted in terms of the preferences of society toward these conflicting objectives. The results obtained from the multiple-criteria model are assessed under scenarios, which cover a full range of uncertainties. By the application of classical decision rules, such as the Wald or Savage criteria, the most flexible and robust strategy can be obtained.

Results shown that the application of this methodology achieves large reduction in risk with small increments in cost, while allowing the society to express their preferences toward any of the risk considered.

In order to ensured the flexibility and robustness of the solution, a detailed study should be performed and a large number of scenarios should be generating by expanding the set of option and uncertainties, particularly a more explicit treatment of demand-side management alternatives.

References

[1] H.T. Yang, S.L. Chen, “Incorporating a multi-criteria decision procedure into the combined dynamic programming/production simulation algorithm for generation expansion planning”, IEEE Trans. on Power systems, Vol.4, No.1, pp. 165-174, 1989.

[2] P. Linares, “Multiple criteria decision making and risk management tools for power system planning”, IEEE Trans. on Power systems, Vol.17, No.3, pp. 895-900, 2002.

[3] C.J.Andrews, “Evaluating Risk Management Strategies in Resource Planning”, IEEE Transaction on Power systems, Vol.10, No.1, pp. 420-426, 1995.

[4] S.Majumdar, D.Chattopadhyay, “A Model for Integrated Analysis of Generation Capacity Expansion and Financial Planning”, IEEE Transaction on Power systems, Vol.14, No.14, pp. 466-471, 1999.

[5] ***, “Eksperten elaborat za razvoj na energetikata i energetskata infrastruktura vo Republika Makedonija do 2020 god”, AD EMO-Ohrid, Institut za energetika, Skopje, 1999.

[6] W.J.Burke, F.C.Schweppe, B.E.Lowell, M.F.McCoy, S.A.Monohon, “Trade off methods in system planning”, IEEE Trans. on Power systems, Vol.3, No.3, pp. 1284-1290, 1988.

[7] B.G.Gorenstin, N.M.Campodonico, J.P.Costa, M.V.Pereira, “Power system expansion planning under uncertainty”, IEEE Trans. on Power systems, Vol.13, No.1, pp. 129-136, 1993.

[8] R.D.Tabors, S.R.Connors, C.G.Bespolka, “A Framework for Integrated Resource Planning: The Role of Natural Gas Fired Generation in New England”, IEEE Transaction on Power systems, Vol.4, No.3, pp. 1010-1016, 1989.

[9] M.Yehia, R.Chedid, M.Ilic, A.Zobian, R.Tabors, J.Lacalle-Melero, “A Global Planning Methodology for Uncertain Environments: Application to the Lebanese Power System, IEEE Transaction on Power systems, Vol.10, No.1, pp. 332-338, 1995.

[10] E.O.Crousillat, P.Dorfner, P.Alvarado, H.M.Merrill, “Conflicting Objectives and Risk in Power System Planning”, IEEE Trans. on Power systems, Vol.8, No.3, pp. 887-893, 1993.

[11] B.F.Hobbs, P.M.Meier, “Multicriteria Methods for Resource Planning: An Experimental Comparation”, IEEE Transaction on Power systems, Vol.9, No.4, pp. 1811-1817, 1994.

[12] P.S.Neelakanta, M.H.Arsali, “Integrated Resource Planning Using Segmentation Method Based Dynamic Programming”, IEEE Transaction on Power systems, Vol.14, No.1, pp. 375-385, 1999.

A Fuzzy Method of Distribution Energy Losses Calculation

Miodrag S. Stojanović[1] and Dragan S. Tasić[2]

Abstract – **There are two main difficulties in calculating the distribution energy losses if, beside the total energy losses, knowledge of losses structure is required. First one is lack of input data for network node loads, and second one is large number of necessary load flow calculations. A method, based on the deterministic load estimation, which requires only one fuzzy load flow, is presented in this paper. The method is applied to calculate annual energy losses in the real distribution network. Accuracy of the presented method is tested comparing to a simulation results.**

Keywords – **Distribution losses, Fuzzy approach, Estimation, Consumer category.**

I. Introduction

Power and energy losses are inevitable consequence of energy transmission from generation points to consumers. The losses sometimes make ten or more percents. Therefore, it is important right assess losses, as well as find methods for their reduction. Basic items of mentioned problems are: dispense technical and non-technical losses, determine structure of losses (distribution of losses throughout the network elements), locate critical elements from aspect of losses, and select optimal methods for losses reduction.

Knowledge of load curves for each particular element is needed for exact calculation of energy losses. That is not possible, because of measurements are made only at some locations in network, so many different approaches to distribution losses assessment are developed. All these approaches can be classified in two basic groups deterministic and probabilistic.

As a result of the fact that loads are not exact known, it is developed fuzzy load flow methods [1,2], as well as the fuzzy method for estimation of node loads [3]. Fuzzy estimation method considers daily load profiles of different consumer categories. Based on that estimation method, distribution losses can be calculated. However, this approach requires large number of the fuzzy load flow calculations. That is why a new method for distribution losses calculation, which requires only one fuzzy load flow, is presented in this paper.

At the first, in this paper is described a method for distribution network simulation. After that, it is presented a method of forming fuzzy numbers that represent loads. The presented method is based on the deterministic load estimation. Annual energy losses of the test system are then obtained using the formed fuzzy loads.

[1]Miodrag S. Stojanović is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Yugoslavia, E mail: miodrag@elfak.ni.ac.yu

[2]Dragan S. Tasić is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Yugoslavia, E mail: dtasic@elfak.ni.ac.yu

II. Simulation

Simulation of distribution network have been made under following assumptions:

- there are three consumer categories with specified daily load profiles,
- variation of maximal daily load during the year is different for different consumer categories,
- participations of any consumer category for each load node are known with accuracy within ±15%. Deviation is random and with normal probability distribution.
- annual peak demand of any load node is known with accuracy within ±15%. Deviation is random and with normal probability distribution.
- variation of root node voltage is known.

In Fig. 1 the possible forms of daily load profiles for three consumer categories are shown.

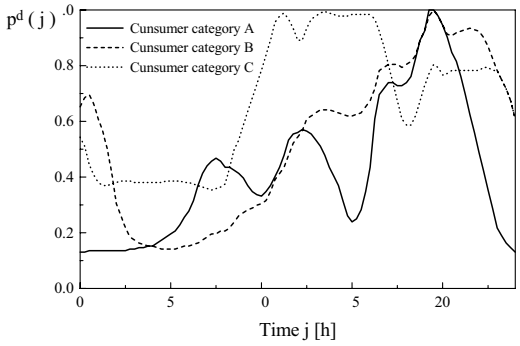


Fig. 1. Hourly load patterns for different consumer categories

Variation of relative daily peak load during the year for three consumer categories are shown in Figs. 2,3 and 4. Based on these curves it is possible to determine load of each node, for any hour during the year, as well as calculate the power and energy losses. Energy losses calculated in this way can be considered as the accurate ones, and they will be used for comparing the results obtained by estimation methods.

Since, in practice, the current of the first feeder section is usually known, the annual current curve of this branch is kept as result of the simulation.

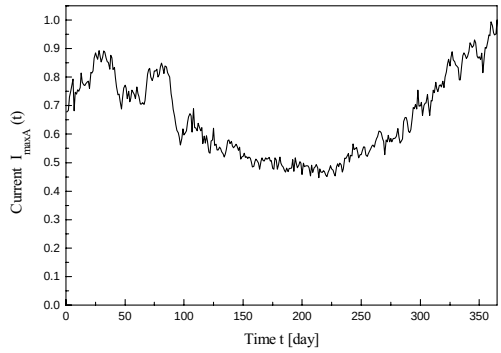


Fig. 2. Variation of daily peak load for A consumer category during the year

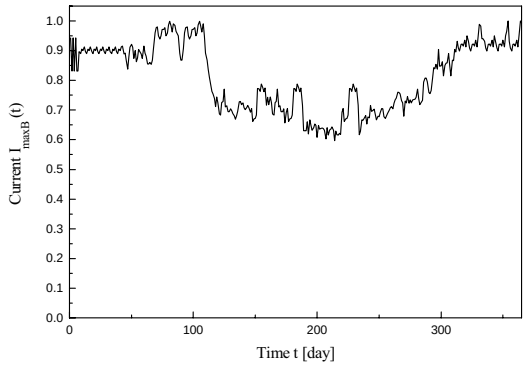


Fig. 3. Variation of daily peak load for B consumer category during the year

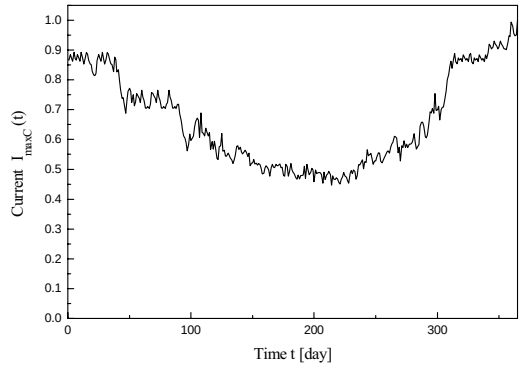


Fig. 4. Variation of daily peak load for C consumer category during the year

III. Fuzzy Approach

Fuzzy approach starts from the real fact that loads (powers) for many load nodes are not complete known, but they are assessed in some way. This is the consequence of the fact that in distribution networks measurements are made only for industry consumers, and on the HV/MV substations. Since the loads (powers) are not strictly known, it is appropriate to consider their as fuzzy numbers.

In this paper the fuzzy numbers that represent the load are formed on the following way.

Firstly the values of relative powers $p_A^d(j)$, $p_B^d(j)$ and $p_C^d(j)$ are taken from Fig. 1, where j is hourofday. The current of i-th distribution transformer, for t-th hour-of-year (corresponds to j-th hour-of-day) can be estimated as:

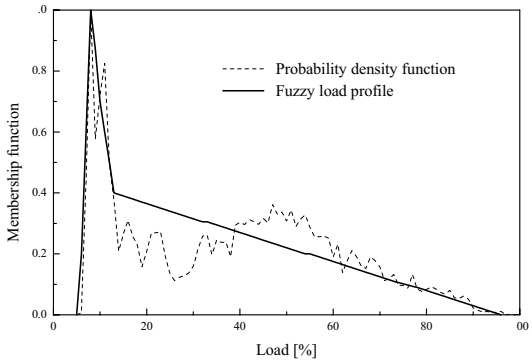


Fig. 5. Fuzzy load profile of A consumer category

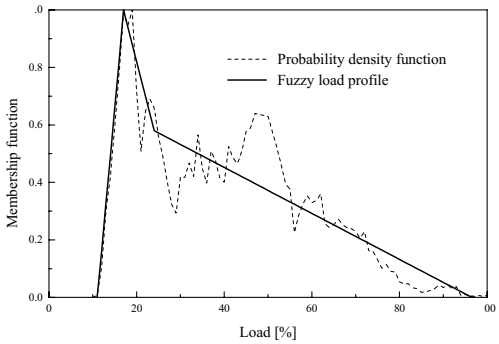


Fig. 6. Fuzzy load profile of B consumer category

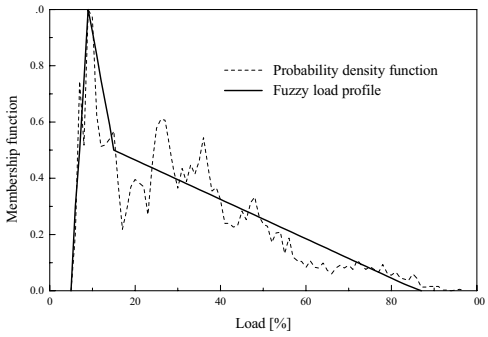


Fig. 7. Fuzzy load profile of C consumer category

$$I_i(t) = k(j) I_{fs}(t) k_i I_{ni} [k_{A_i} p_A^d(j) + k_{B_i} p_B^d(j) + k_{C_i} p_C^d(j)] , \quad (1)$$

where coefficient $k(j)$ is:

$$k(j) = \frac{1}{\sum\limits_{i \in \alpha_L} k_i I_{ni} [k_{A_i} p_A^d(j) + k_{B_i} p_B^d(j) + k_{C_i} p_C^d(j)]} , \quad (2)$$

and: α_L – set of load nodes, k_{A_i} – participation of A consumer categories in i-th node loads, k_{B_i} – participation of B consumer categories in ith node loads, k_{C_i} – participation of C consumer categories in ith node loads, $I_{fs}(t)$ – current of first feeder section at tth hour, $k_i I_{n_i}$ – maximal current of distribution transformer installed in i-th node.

For this reason, current of all consumers that belong to x

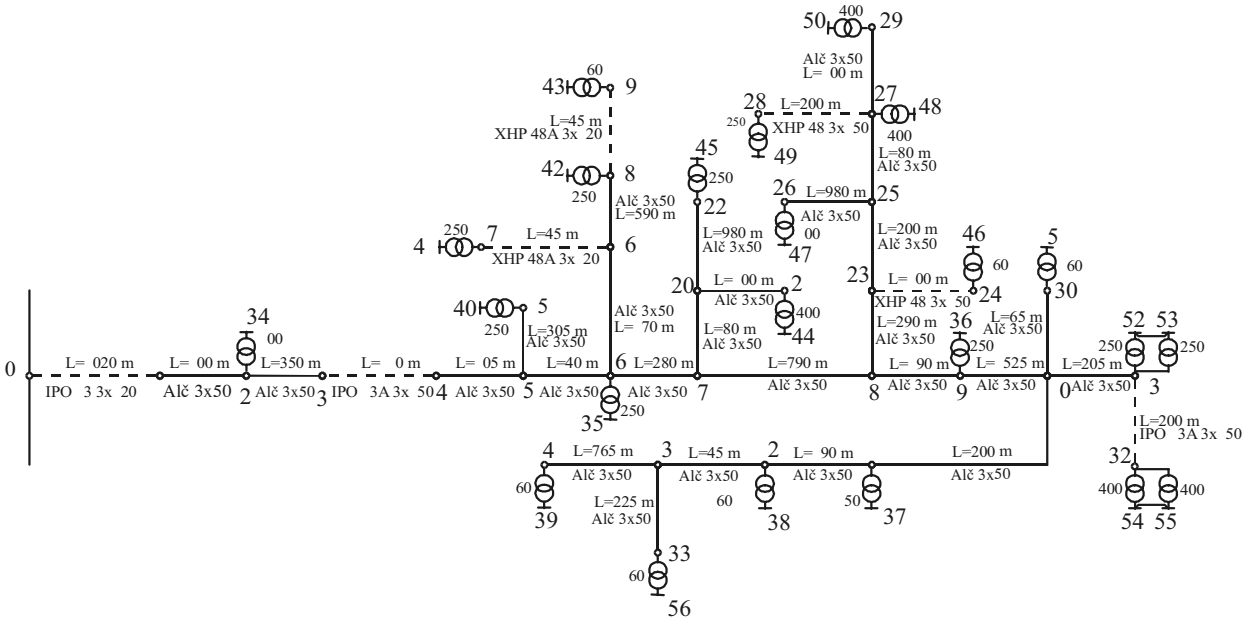


Fig. 8. Test network

consumer category, for t-th hour-of-year is:

$$I_x(t) = k(j)\, I_{fs}(t)\, p_x^d(j) \sum_{i \in \alpha_L} k_{x_i} k_i I_{ni},\ \ x \in \{A, B, C\}. \quad (3)$$

Obtained current curves are then normalised using equation:

$$I'_x(t) = \frac{I_x(t)}{\max\limits_t I_x(t)}, \qquad x \in \{A, B, C\}. \quad (4)$$

Based on normalised curves, probability distribution of consumer category current is obtained and converted to fuzzy membership function according to a possibility-probability consistency principle. Fuzzy numbers that represent load for each consumer categories with three lines ($\tilde{I}_A(t)$, $\tilde{I}_B(t)$ and $\tilde{I}_C(t)$) are shown on Figs. 5, 6 and 7.

This is not the only way for fuzzy numbers determination but that has been the way chosen by authors.

Current of each node, as fuzzy number, can be determined using relation:

$$\tilde{I}_i = [\tilde{I}_A\, k_{A_i} + \tilde{I}_B\, k_{B_i} + \tilde{I}_C\, k_{C_i}]\, k_i\, I_{ni}\ . \quad (5)$$

Root node voltage is considered as triangular fuzzy number. Power flow calculation is then made on the way presented in [2], and according to fuzzy arithmetic lows [4].

Results of calculation are node voltages and power/current flows as fuzzy numbers. Using the calculated current values, the power losses as fuzzy numbers can be obtained. Defuzzification gives the deterministic value of power losses that multiplied with number of hours for observed period gives energy losses. There are several defuzzification strategies, but authors suggest the bisector method.

IV. Test Example

The presented procedure is used for determining annual energy losses of the test network shown in Fig. 8. In this figure are also shown data needed for calculation, while data about maximum powers of load nodes and participation of any customer categories in each load node are shown on Table 1.

Table 1. Maximum powers and participation of any customer categories

Node number	A (%)	B (%)	C (%)	S_{max} [MVAr]	cos φ
34	70	30	0	0.090	0.96
35	60	20	20	0.220	0.96
36	00	0	0	0.240	0.95
37	20	20	60	0.020	0.97
38	20	0	80	0. 70	0.98
39	0	20	80	0. 70	0.96
40	20	50	30	0. 80	0.94
4	30	70	0	0. 80	0.93
42	80	20	0	0.200	0.95
43	0	80	20	0.200	0.98
44	60	40	0	0.300	0.96
45	90	0	0	0. 80	0.95
46	0	90	0	0. 70	0.99
47	0	20	70	0.05	0.99
48	0	0	80	0.350	0.96
49	0	20	80	0.230	0.97
50	20	30	50	0.380	0.95
5	20	20	60	0. 60	0.96
52	20	0	80	0.270	0.94
53	20	0	80	0.270	0.94
54	20	20	60	0.380	0.97
55	20	20	60	0.380	0.97
56	0	50	40	0. 60	0.96

In Table 2 the results have following meaning:

- Simulation – Results of simulation described on Section II.
- Estimation 1 – Estimation based on relations (1) and (2).
- Estimation 2 – Estimation based on relations presented in [5]. This estimation assumes that consumer of any load node have been put into consumer category with

Table 2. Results obtained by simulation, estimations and fuzzy

	Total losses		Lines losses		Transformers losses		Line 4-5 losses		Line 9-10 losses		Transformer 30-51 losses	
	MWh	MVArh	MWh	MVArh	MWh	MVArh	MWh	MVArh	MWh	MVArh	MWh	MVArh
Simulation	795.1	716.1	546.5	330.2	248.7	385.9	30.87	18.43	81.75	48.8	9.916	15.61
Estimation 1	780.2	709.2	530.9	321	249.3	388.2	30.98	18.5	75.4	45.02	9.795	15.3
Estimation 2	823.5	749.9	569.2	343.9	254.4	406	31.12	18.58	84.51	50.45	9.637	14.93
Estimation 3	752.5	673.4	511	309.1	241.4	364.3	31.09	18.56	65.87	39.33	8.036	10.82
Fuzzy approach	794.1	735.3	560.7	339.1	255	396.2	33.2	19.8	70.1	41.9	7.8	10.3

the highest percent participation. The consumers in nodes 34, 35, 36, 42, 44 and 45 are A category, in nodes 40, 41, 43, 46 and 56 B category, while other consumers are C category.

- Estimation 3 – Estimation that supposes distribution of first feeder section current proportionally to rated power of distribution transformers.

Comparing the results from Table 2, the following statement can be established. Results obtained using the estimation method presented in this paper are better then ones given by other two estimation methods. However, the usage of this method requires knowledge the participation of different consumer categories in the total load of each node. This requirement is of course weakness of the presented estimation method.

The accuracy of results obtained by the fuzzy approach is satisfying for engineering applications. The accuracy should be obtained comparing with results of estimation presented in this paper. In our test example the error made by the estimation is nullified by the error of the fuzzy approach. For some cases, these errors can be of course summed.

It should be emphasized on the end that required time for fuzzy approach calculation is very short, because of only one load flow calculation is made.

V. Conclusion

The method for distribution losses calculation, based on the fuzzy approach, is presented in this paper. Presented method enables the total energy losses determination as well as determination energy losses of each network element (structural losses analysis). Simulations that have been made by the authors show that satisfying results for energy losses (total and by network elements) are obtained using fuzzy approach. Computation time for this approach is significant smaller comparing to other methods. This statement, as well as not well knowledge of the input data are reasons for using the presented fuzzy approach.

References

[1] V. Miranda, M. A. C. C. Matos, J. T. Saraiva, “Fuzzy Load Flow - New Algorithm Incorporating Uncertain Generation and Load Representation”, PSCC, Graz, 1990., pp. 621-627.

[2] D. Tasić, M. Stojanović, “Proračun gubitaka snage u distributivnoj mreži pri nepotpunom poznavanju snaga potrošnje”, Elektrodistribucija, Br. 1, str. 16-25, 2002.

[3] Hang-Ching Kuo, Yuan-Yih Hsu, “Distribution System Load Estimation and Service Restoration Using a Fuzzy Set Approach”, IEEE Trans., Power Delivery, vol. 8, No. 4, pp. 1950-1957, October 1993.

[4] G. J. Klir, B. Yuan, *Fuzzy sets and fuzzy logic: Theory and Application*, New Jersey, Prentice Hall, 1995.

[5] N. Rajaković, D. Tasić, M. Stojanović, “A Clustering Technique for Distribution Losses Calculation in Deregulated Environment”, $2^{n\,d}$ Balkan Power Conference; Power Industry Restructuring, Belgrade, pp. 30-36, 19-21. June 2002.

Influence of the Soil Thermal Non-Homogeneity on the Cable Current Ampacity

Dragan S. Tasić[1] and Miodrag S. Stojanović[2]

Abstract – **Determination of the thermal rated current of cable is usually made on the assumption that soil is homogeneous in thermal sense. However, rout of long cables sometimes passes through soils with different thermal characteristics, or one part of the cable passes through a duct. All these practical cases, as a rule, can be reduced to a few characteristic ones. The relations for temperature distribution along the cable, for these characteristic cases, are developed in this paper. The influence of the soil thermal non-homogeneity on the cable current ampacity is then determined using developed relations. The analysis is made for the cable XHE48-A 1x1000/95 mm2 64/100 kV.**

Keywords – **cable, current ampacity, soil, drying-out.**

I. Introduction

In order to determine the thermal current ampacity of the cables laid in the ground, general assumption is that the soil is homogeneous in thermal sense. However, the route of long cables may sometimes pass through the soil of different structure and thus different thermal characteristics. There are also examples of the cables placed in pipes and ducts.

In such cases even when the soil is thermally homogeneous, the conditions of cable heat dissipation are changing. The question here is how big is the real influence of soil thermal non-homogeneity on the cable ampacity; for example, short part of the cable passing below a concrete or asphalt surface.

That is why this paper is presenting two general cases, which resemble problems found in practice. If we analyze elementary part of cable conductor and its heat dissipation, following equations will define temperature rise along the conductor.

II. Distribution of temperarature along route of the cable

The analysis of thermal non-homogeneity along the rout of cables on thermal current-carrying capacity is based on the temperature rise along the conductor. Fig. 1 is showing the elementary part of conductor marked as *dx*, with transmitted powers P_1 and P_2, heat conductivity of conductor λ, temperature θ and cross-sectional area S.

[1]Dragan S. Tasić is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Yugoslavia, E mail: dtasic@elfak.ni.ac.yu

[2]Miodrag S. Stojanović is with the Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Yugoslavia, E-mail: miodrag@elfak.ni.ac.yu

$$P_1 = -\lambda S \frac{\partial \theta}{\partial x} , \quad (1)$$

$$P_2 = -\lambda S \frac{\partial}{\partial x}(\theta + \frac{\partial \theta}{\partial x} dx) . \quad (2)$$

P_3 is dissipated heat of elementary part of the conductor, caused by current under normal conditions of exploitation. Beside dissipated heat of the conductor, important parts are also the losses generated in metal sheath. Therefore dissipated heat is calculated as:

$$P_3 = R'_{20}(1 + \alpha(\theta - 20))I^2 dx , \quad (3)$$

where R'_{20} is effective resistance of cable at 20°C, and α is temperature coefficient for the electrical resistance.

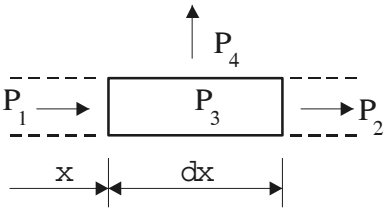


Fig. 1. Elementary part of the conductor

Part of dissipated heat, conducted from the surface of elementary part of the conductor through the cable insulation and surrounding soil, is marked as P_4 in Fig. 1. If we ignore drying-out effect of the soil, conducted power P_4 will be:

$$P_4 = \frac{\theta - (\theta_E + P'_d(T'_{Kd} + T'_4))}{T'_{Ki} + T'_E} dx = \frac{\theta - (\theta_E + \Delta\theta_d)}{T'_{Ki} + T'_E} dx , \quad (4)$$

where: θ_E is temperature of soil with cables not loaded, P'_d is dielectric loss of a cable, T'_{Kd} is fictitious thermal resistance of a cable when considering the dielectric losses, T'_{Ki} is fictitious thermal resistance of a cable when considering the ohmic losses, T'_E is thermal resistance of the ground, $\Delta\theta_d$ is temperature rise of a conductor above ambient due to dielectric losses.

All thermal resistances and powers in Eq. (4) are per unit length. Estimation of thermal resistances is explained in [3,4]. Load has no influence on dielectric losses (eventual variation of voltage is ignored), so temperature rise due to dielectric losses is constant. Power of dielectric losses can be ignored for low-voltage cables.

$$P_4 = \frac{\theta - (\theta_E + \Delta\theta_d) + \frac{\rho_x - \rho_E}{\rho_E} \cdot \Delta\theta_x}{T'_{Ki} + T'_x} dx , \quad (5)$$

where: ρ_E is thermal resistivity of the moist soil, i.e. the moist area, ρ_x is thermal resistivity of the dried-out soil, i.e. the dry area, T'_x is thermal resistance of the dried-out soil, and $\Delta\theta_x$ is limiting temperature rise of the boundary isotherm above ground temperature.

The power balance in operating conditions for the elementary part of the cable is:

$$P_1 + P_3 = P_2 + P_4 \,. \tag{6}$$

If we replace equations for the powers P_1, P_2, P_3 and P_4 into Eq. (6), when the drying-out effect is ignored, we will come up with the following partial differential equation:

$$\frac{\partial^2\theta}{\partial x^2} - \frac{1-(T'_{Ki}+T'_E)\alpha R'_{20}I^2}{\lambda\cdot S(T'_{Ki}+T'_E)}\theta = \\ = \frac{(T'_{Ki}+T'_E)\cdot R'_{20}\cdot(1-20\alpha)I^2 + (\theta_E+\Delta\theta_d)}{\lambda\cdot S(T'_{Ki}+T'_E)} \,. \tag{7}$$

The solution of Eq. (7) has following form:

$$\theta = \theta_{st} + A\cdot e^{a\,x} + B\cdot e^{-a\,x} \,, \tag{8}$$

$$\theta_{st} = \frac{(T'_{Ki}+T'_E)\cdot R'_{20}\cdot(1-20\alpha)I^2 + (\theta_E+\Delta\theta_d)}{1-(T'_{Ki}+T'_E)\alpha R'_{20}I^2} \,, \tag{9}$$

$$a = \sqrt{\frac{1-(T'_{Ki}+T'_E)\alpha R'_{20}I^2}{\lambda\cdot S(T'_{Ki}+T'_E)}} \,, \tag{10}$$

where A and B are appropriate constants.

If drying-out effect of the soil is significant, the steady state temperature is:

$$\theta_{st} = \frac{1}{1-(T'_{Ki}+T'_E)\alpha R'_{20}I^2}\cdot\left[(\theta_E+\Delta\theta_d)\right. \\ \left.+(T'_{Ki}+T'_E)\cdot R'_{20}\cdot(1-20\alpha)I^2 - \frac{\rho_x-\rho_E}{\rho_E}\Delta\theta_x\right] . \tag{11}$$

Factors A and B will be calculated from boundary conditions, for every practical example. We will analyze two characteristic cases shown on Figs. 2 and 3.

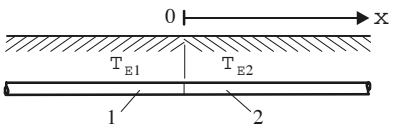


Fig. 2. Example with two different types of soil

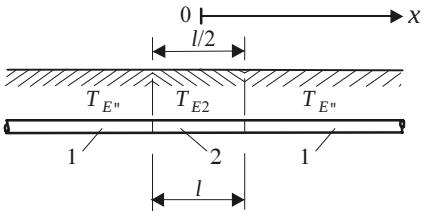


Fig. 3. Thermal non-homogeneity of soil at one part of the cable

According to Eq. (8) for the example shown on Fig. 2 we can calculate temperature distribution along parts 1 and 2:

$$\theta_1 = \theta_{st1} + A_1 e^{a_1 x} + B_1 e^{-a_1 x}, \quad x \leqslant 0 \,, \tag{12}$$

$$\theta_2 = \theta_{st2} + A_2 e^{a_2 x} + B_2 e^{-a_2 x}, \quad x \geqslant 0 \,. \tag{13}$$

If we analyze the two previous relations, temperature will increase indefinitely when $x \to \infty$. Since that is impossible, it is obvious that the factors A_2 and B_1 must equal zero. Factors A_1 and B_2 are determined for x=0:

$$\theta_1(x=0) = \theta_2(x=0) \,, \quad \left.\frac{\partial\theta_1}{\partial x}\right|_{x=0} = \left.\frac{\partial\theta_2}{\partial x}\right|_{x=0} \,,$$

so the factors A_1 and B_2 are:

$$A_1 = \frac{a_2}{a_1+a_2}\,(\theta_{st2}-\theta_{st1}) \,, \tag{14}$$

$$B_2 = -\frac{a_1}{a_1+a_2}\,(\theta_{st2}-\theta_{st1}) \,. \tag{15}$$

Therefore, the temperature distribution along the cable for example shown on Fig. 2 will be:

$$\theta_1 = \theta_{st1} + \frac{a_2}{a_1+a_2}\,(\theta_{st2}-\theta_{st1})\,e^{a_1 x} \,, \quad x \leqslant 0 \,, \tag{16}$$

$$\theta_2 = \theta_{st2} - \frac{a_1}{a_1+a_2}\,(\theta_{st2}-\theta_{st1})\,e^{-a_2 x} \,, \quad x \geqslant 0. \tag{17}$$

For the example on Fig. 3 equation for the temperature distribution is:

$$\theta_1 = \theta_{st1} + \\ + \frac{a_2}{a_1}\frac{(\theta_{st2}-\theta_{st1})\,sh\left(a_2\frac{\ell}{2}\right)}{ch\left(a_2\frac{\ell}{2}\right)+\frac{a_2}{a_1}sh\left(a_2\frac{\ell}{2}\right)}e^{\left(-a_1\left(x-\frac{\ell}{2}\right)\right)} \,, \tag{18}$$

when $x \geqslant l/2$, and

$$\theta_2 = \theta_{st2} - \frac{\theta_{st2}-\theta_{st1}}{ch\left(a_2\frac{\ell}{2}\right)+\frac{a_2}{a_1}sh\left(a_2\frac{\ell}{2}\right)}ch\,(a_2 x) \tag{19}$$

when $0 \leqslant x \leqslant l/2$.

III. Current Ampacity Factor

Magnitude of the thermal current is determined by permissible operating temperature of conductor θ_c. Considering the relation for steady state temperature Eq. (9), for $\theta_{st} = \theta_c$, current ampacity on the parts 1 and 2 as shown on Fig. 1 will be:

$$I_{1\,(2)} = \sqrt{\frac{\theta_c - \theta_{E1\,(2)} - \Delta\theta_{d1\,(2)}}{(T'_{Ki}+T'_{E1\,(2)})R'_{20}\,(1+\alpha(\theta_c-20))}} \,. \tag{20}$$

It is obvious that for the load capacity we will chose the smaller value. Ratio of these two currents gives us current ampacity factor for the thermal non-homogeneity along route of the cable, and practically points out the current efficiency on the part of the cable under the good thermal conditions:

$$f_i = \frac{I_2}{I_1} = \sqrt{\frac{\theta_c-\theta_{E2}-\Delta\theta_{d2}}{\theta_c-\theta_{E1}-\Delta\theta_{d1}}\frac{T'_{Ki}+T'_{E1}}{T'_{Ki}+T'_{E2}}} \,. \tag{21}$$

Regarding Eqs. (18) and (19) for current ampacity factor of the example shown on Fig. 3 we will have:

$$f_i = \sqrt{\frac{b - \sqrt{b^2 - ac}}{\alpha(\theta_c - \theta_{E1} - \Delta\theta_{d1})(T'_{Ki} + T'_{E2})}}\,, \qquad (22)$$

where:

$$\begin{aligned} a =& \alpha\ (1 + \alpha\ (\theta_c - 20))\ (T'_{Ki} + T'_{E1})\ (T'_{Ki} + T'_{E2})\ , \\ b =& \frac{1}{2}\,[\alpha\,(\theta_c - \theta_{E2} - \Delta\theta_{d2})\,(T'_{Ki} + T'_{E1}) \\ &+ (1 + \alpha\,(\theta_c - 20))\,(T'_{Ki} + T'_{E2}) \\ &- p\,(1 + \alpha\,(\theta_{E1} + \Delta\theta_{d1} - 20))\,(T'_{Ki} + T'_{E2}) \\ &- p\,(1 + \alpha\,(\theta_{E2} + \Delta\theta_{d2} - 20))\,(T'_{Ki} + T'_{E1})]\,, \\ c =& \theta_c - \theta_{E2} - \Delta\theta_{d2} + p\ (\theta_{E2} + \Delta\theta_{d2} - \theta_{E1} - \Delta\theta_{d1})\,, \end{aligned}$$

$$p = \frac{1}{ch\left(a_2\frac{\ell}{2}\right) + \frac{a_2}{a_1}sh\left(a_2\frac{\ell}{2}\right)}.$$

IV. Test Example

The influence of thermal non-homogeneity is based on analysis of current load of three single-core cables XHE 48-A 11000/95 mm ^{2} 110 kV in 3-phased system bunched. Thermally permissible current for these cables, when thermal resistivity is $\rho_{E1} = 1.2$ Km/W and temperature of referent ground $\theta_E = 20$°C is $I = 785$ A [6]. Supposing the cable is long enough (according to [6], the length of such cable in Belgrade is approximately 9 km), and that one part of the cable with the length of $l = 10$ m, placed around the central point of rout, is laid in the ground of thermal resistivity $\rho_{E2} = 2$ Km/W. This example is presented on Fig. 3.

According to available information of the cable presented in [6], we can easily define data needed for this analysis: R'_{20}=39.2 $\mu\Omega$/m, $T'_{Ki} = 0.344$ Km/W, T'_{Kd}=0.25 Km/W, P'_d = 0.23 W/m, T'_{E1}=1.9 Km/W, T'_{E2}=2 Km/W. Temperature coefficient for electrical resistance of aluminum conductor at 20°C is $\alpha = 0.00403\ \mathrm{K}^{-1}$, thermal conductivity factor is $\lambda = 230$ W/Km and temperature of referent ground in both sections of cable is $\theta_{E1} = \theta_{E2} = 20$°C.

If load capacity is I=785 A for such defined conditions, we will have steady state temperatures $\theta_1 = 90$°C and θ_2 = 149.6°C, while coefficient a will be a_1=1.23 m^{-1} and a_2=0.903 m^{-1}. If coordinate system is as shown on Fig. 3, for the temperature distribution along the cable, we will have:

$$\theta_1 = 90 + 25.23\,e^{-1.23(x-5)}\ ,\ x \geqslant 5\,\mathrm{m},$$
$$\theta_2 = 149.6 - 0.752\,ch(0.903\cdot x)\ ,\ 0 \leqslant x \leqslant 5\,\mathrm{m},$$

where x is expressed in meters (m), and temperature in °C.

Fig. 4 illustrates temperature distribution on the part of conductor defined by previous equations. We can notice that axial heat conduction practically does not influence temperature rise of thermally critical point. The temperatures θ_{st1} and θ_{st2}, reach their values yet at distance 2 to 3 m from the place of discontinuity. These distances are even lesser for medium-voltage cables. This brings us to conclusion that the limit for current capacity of the cable, regardless of its length,

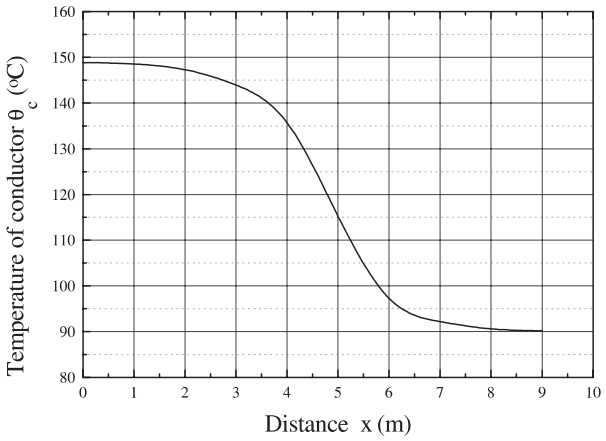


Fig. 4. Temperature distribution along the cable according to the Eqs. (18) and (19)

is determined at the part of the route where thermal conditions are not so good as at the rest of the cable.

The fact that discontinuity effect has influence only on a short part of the cable, we can use Eqs. (16) and (17) instead of Eqs. (18) and (19), which are easier to analyze. Considering these equations, we can place origin of the axis system at the point of discontinuity (which is displaced for $l/2 = 5$ m in direction of x-axis), and temperature rise along the parts 1 and 2 (considering that positive part of x-axis is marked as 1) we will have:

$$\theta_1 = 90 + 25.23\,e^{-1.23\,x},\ x \geqslant 0\,,$$
$$\theta_1 = 149.6 - 34.37\,e^{0.903\,x},\ x \leqslant 0\,.$$

Temperature distribution for this example is shown on Fig. 5. If we compare graphs on Figs. 4 and 5 we can see that basically it is the same curve with different origins of axes system. This points out that even when there are several points of discontinuity, temperature rise can be expressed by Eqs. (16) and (17). The only thing that must be arranged is to place origin of axis system in the point of discontinuity.

From the previous statement for current ampacity factor, we can use Eq. (21), which is simpler than Eq. (22). Using

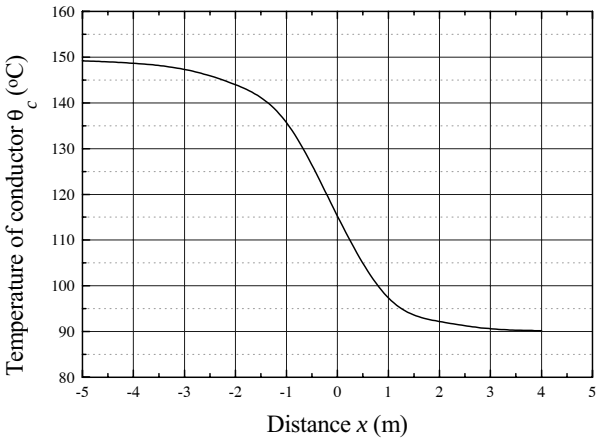


Fig. 5. Temperature distribution along the cable according to the Eqs. (16) and (17)

Eq. (21) for this example, current ampacity factor will be $f_i = 0.797$, and if we use Eq. (22) we will have $f_i = 0.796$. It is obvious that the difference is insignificant, and the usage of Eqs. (16), (17) and (21) is justified.

Existance of the one short part of rout where thermal resistivity of the soil increases from 1.2 Km/W to 2 Km/W leads to decrement of current ampacity for more than 20%. Therefore, instead of 785 A for permissible current we will have 625 A, and most of the cable will be inefficiently exploited. According to these facts we can realize the significance of the special bedding mixtures used in similar situations.

V. Conclusion

This paper presents mathematical model for analysis of soil non-homogeneity on current ampacity of the cables. The analysis of the two general cases, which resemble problems found in practice, clearly points out that the easier method of calculation is efficient enough. Mathematical model based on the less complicated example is appropriate for calculations even when cable has several points of thermal discontinuity.

This analysis proves that thermal non-homogeneity of the soil has significant influence on a cable load capacity. It is obvious that the increase of soil thermal resistivity has essential influence on permissible current. That is why we should know thermal characteristics of the ground before laying a cable. For better current exploitation, in places that are thermally critical, we should use special cable bedding mixtures.

References

[1] H. Brakelmann, G. Anders, "Ampacity Reduction Factors for Cables Crossing Thermally Unfavorable Regions", *IEEE Trans*., *Power Delivery*, vol. 16, no. 4, pp. 444-448, 2001.

[2] D. Tasić, *Termiki aspekti strujne opteretljivosti provodnika nadzemnih elektroenergetskih vodova*, Niš, Elektronski fakultet, 2002.

[3] D. Tasić, *Osnovi elektroenergetske kablovske tehnike*, Niš, Elektronski fakultet, 2001.

[4] ***, *Calculation of thermal resistances*, IEC Standard 287, Part 2-1, 1994.

[5] L. Heinhold, *Power Cables and Their Application*, Berlin, Siemens Aktiengesellschaft, 1990.

[6] B. Lalević, "Struja kroz elektrinu zaštitu 110 kV kabla", Elektrodistribucija, br.2, str. 122-130, 2000.

Application Analysis of GPS Used for Synchronization in Energy Information Networks

Ilia G. Iliev[1], Angel B. Colov[2], Rumen I. Arnaudov[3]

Abstract – **The work analyses the possibility of common GPS receivers' application for one point synchronization of the energy information measurement networks. The sources and types of errors derived in the time defining through GPS are investigated. The speed of variation of the informative quantities in the systems for telemeasurement, control and management analysis is made. An algorithm for synchronization with a common GPS receiver and time for seeing may be proposed on this base.**

Keywords – **GPS time synchronization, energy information networks.**

I. Introduction

There are many publications about GPS application in synchronization of data network, energy information networks, cellular communication networks (GSM, CDMA) etc. There are on the market many devices, for this purpose, working with determined time accuracy. There are also GPS receivers specially designed for uniform time synchronization with high accuracy for computer works, information-measurement system etc. Special algorithms for increasing the accuracy of the defined time are built-in their software (RIAM algorithms). The error is the order of dozens nanoseconds. This advantage is paid by the price. For synchronization purpose may be used receiving GPS signals and synchronization from one point either, or distributed, differential synchronization [1,3].

The work analyses the possibility of common GPS receivers application for one point synchronization of the energy information measurement networks. The sources and types of errors derived in the time defining through GPS, are investigated. The speed of variation of the informative quantities in the systems for telemeasurement, control and management analysis is made. The admissible error of quantization time is determined, from where the admissible time error for registration is computed. An algorithm for synchronization with a common GPS receiver and time for – seeing may be proposed on this base.

[1]Ilia Georgiev Iliev, Dept. of Radiotechnic in Faculty of Communications and Communication Technologies in TU – Sofia, E-mail: igiliev@vmei.acad.bg

[2]Angel Belchev Colov, Dept. of Radiotechnic in Faculty of Communications and Communication Technologies in TU – Sofia, E-mail: abc@vmei.acad.bg

[3]Rumen Ivanov Arnaudov, Dept. of Radiotechnic in Faculty of Communications and Communication Technologies in TU – Sofia, E-mail: ra@vmei.acad.bg

II. Common GPS Reciever Synchronization Error Analysis

All GPS satellites have internal atom clocks (cesium or rubidium frequency standard) on the board.

They are synchronized with the system time of GPS by the system control segment. The clock deviation is transmitted as a navigation message.

Transmitted data for the time of GPS is computed in relating to the zero time of the system (05.01.1980 00:00:00.0000h). It is intersecting that the system time of GPS is Universal Time Coordinated and the difference from synchronization is taken into account as a correction in the navigation message with accuracy of 90 ns.

In the common GPS receiver is generated the same pseudo-random sequence of C/A code (Gold code), transmitted by a satellite in the line of sight. This sequence is synchronized with the output clock signal, obtained every one second – 1PPS. The cross correlation function (CCF) between it and the received code (C/A) is computed in the receiver. The CCF has a maximum in a moment corresponding to the time difference Δt between the receiver and satellite clock signals. The spectrum of the received signal is unspread, by the recovered pseudo random sequence in the receiver. As a result is derived the navigation message. The receiver adjusts its own clock generator and synchronizes it with the given moment from the navigation message. Thus the signal 1PPS is synchronized with the data frame beginning of the navigation message. Because of many factors, the receiver clock generator has a time delay of reader of ms in relevance with the satellite generator. The receiver computes the difference between both clock signals in the following way:

$$dt' = \Delta t - (L/c + dDp + dI + dT) - dD + dC \ , \quad (1)$$

where Δt – time deviation between the receiver and satellite clock signals, L – geometric distance between satellite and GPS receiver antenna, c – light velocity, dDp – Doppler effect deviation, dI – time delay from the ionospheric propagation, dD – time delay in the receiver, dC deviation of the clock signals of the receiver in relevance of GPS time.

With a single GPS receiver, the time is obtained after accomplishing the necessary computations in formula (1). The time for accomplishing the complete cycle of computation in the initial starting of the receiver is of order of 10 minutes and depends on the type of the receiver and the manufacturer. The time synchronization is separated in the following way:

1. The receiver synchronizes its own clock signal 1PPS

with the first bit of the navigation message;

2. Multiple (from 6 to 15 times) times per a second is computed the cross correlation between the received and self generated pseudo random code and average the result;
3. The mean quantity is corrected with all the rest time delay corrections (formula (1));
4. Finding the mean quantity is doing again, but within the frame work of 13 minutes of the obtained time deviation between the receiver and the satellite cock generators;
5. It is corrected the clock signal 1PSS with the obtained deviation.

This method for synchronization is simple and does not require additional data, besides the data obtained by the GPS receiver.

The basic error sources in the synchronization, according to formula (1), and divided by the feature of its source are given in Table 1.

Table 1. Error per second

No	Error source (1s)	C/A code	P code
	Satellite Errors		
1.	Satellite atom clock	10 ns	10 ns
2.	Error in the satellite coordinates	15 ns	15 ns
	Propagation medium Errors		
3.	Ionospheric propagation Effects	15 ns	7.5 ns
4.	Tropospheric propagation Effects	3.0 ns	3.0 ns
5.	Multipath propagation Effects	10 ns	10 ns
	Receiver Errors		
6.	Error in the receiver antenna coordinates	33 ns	33 ns
7.	Time delay in the receiver hardware	5.0 ns	5.0 ns
8.	Time delay of the beginning of the software processing in the receiver	5.0 ns	5.0 ns
9.	Receiver noise	50.0 ns	5.0 ns
	Total error	65 ns	40 ns

The errors in setting the receiver and satellite coordinates and Doppler effect errors influence on the distance L computation. Besides, the time synchronization error depends on the inaccuracy of the receiver antenna coordinates and quantity of received and averaged data. For example, if the inaccuracy of the receiver position is of order 10 m, the mean error is about 40-65 ns for a period of 13 s. If a measurement is carried out for some days, the error decreases to some tenths of the nanosecond.

The synchronization error with P code receivers (military and other with special application) is less than using receiver that work with the C/A code. The reason for this, is that receivers with P code measure the signal time delay in two frequency band width and after a special processing decrease the ionospheric and tropospheric propagation error and the noise level in them is about 10 time less.

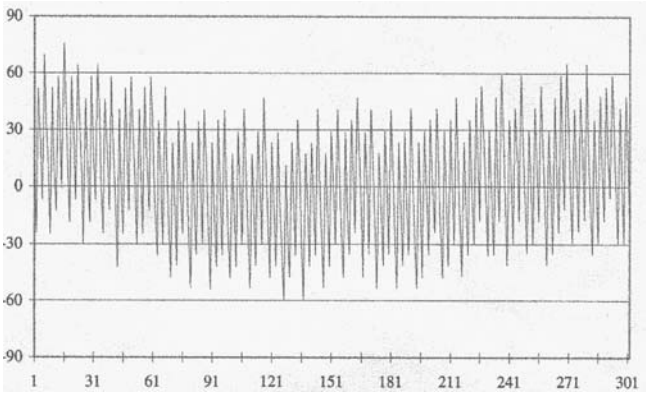


Fig. 1.

Fig. 1 shows the error obtained in the comparison between common GPS receiver and an atom clock.

It is obvious, that the time error obtained by the system is of order of 60-500 ns and depends on the work regime of the receiver. Easy built-in and implementation and easy support of the GPS receiver make it an unique device for providing exact uniform time, synchronized with the World Universe Time. In addition GPS requires less investment by the consumer for achieving the purpose in comparison with other synchronization methods.

III. Applicability of GPS in Telecontrol Power Electrical System Synchronization

A. Processes and Necessity

Fig. 2 shows some of the general system automatics in power electrical system (PES), in which the synchronization necessity (work in an uniform time) of the automatic control system (ACS) is of primary importance.

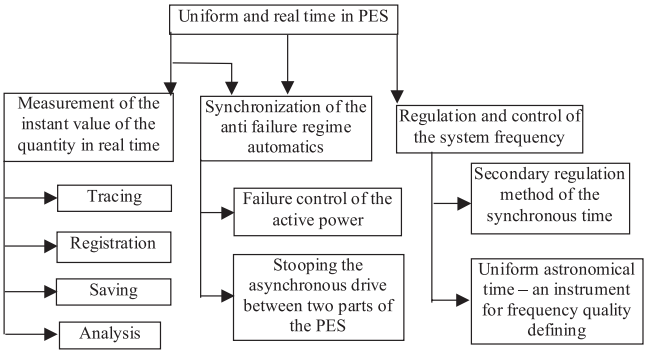


Fig. 2.

B. Increasing the Stability of the Electrical Transmission Lines

The accurate defining of the phase difference δ between the voltages of two PES parts (Fig. 3), connected by a distribution line W, may be done only in the uniform time of measurement in both parts.

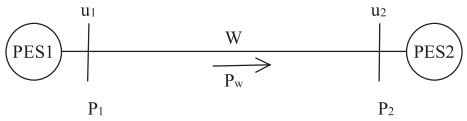


Fig. 3.

The value of the angle δ or its derivate may be used as direct criterion, describing the electrotransmission line stability. The active power in distribution line is:

$$P_w = (U_1 U_2 / X_s) \sin \delta \,, \tag{2}$$

where $X_\Sigma = X_1 + X_W + X_2$ is the sum reactance. At the phase $\delta = \pi/2$ [rad], maximum power P_W is carried and the electrotransmission line is on the border of the stability work. For ensuring the required static stability reserve, that must be not less than 20% in normal work, it is not allowed to carry power higher than the allowed Pa:

$$Pa = P \max / (1 + k) \,. \tag{3}$$

k is the given reserve coefficient of static stability and it may be decreased in some cases. It is in limits of 0.05 – 0.2. The stable work is saved when the condition is true:

$$P_W < Pa \,. \tag{4}$$

The anti failure automatic for control of the active power, whose functions may be done by ACS, uses condition (4) on the base of the measured voltages and the angle δ.

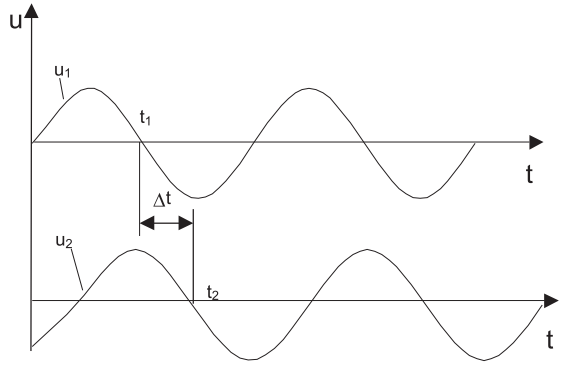


Fig. 4. u_1 and u_2 voltages and the phase difference

Consequently, the ACS may be used as an automation for stopping the asynchronous drive between the two parts. The asynchronous regime between two parallel working parts of PES characterizes with periodic three-phase deviating the phase of the voltage, the current and of the active power. The impedance is changing smoothly. The phase difference δ between the voltages U_1 and U_2 in a normal synchronous regime and constant angle velocity ω_n, is:

$$\delta = \arcsin \left(P_W \frac{X_\Sigma}{U_1 U_2} \right) . \tag{5}$$

In the asynchronous regime, the vectors of U_1 and U_2 are rotating with deferent frequencies $\omega_1 \neq \omega_2$. The angle δ is changing periodically from 0 to 2π [rad] with a slipping frequency $\omega_s = \omega_2 - \omega_1$. The deep degrading of the voltage, the flowing of big currents, commensurable with or bigger than the currents of short-circuit and the variation of the active power in asynchronous regime, are serious disturbances, dangerous for the equipment. The asynchronous regime limited to 2-3 cycles and its duration is not longer than 15 - 20 s. For this time the synchronous work must be recovered or the systems must be disconnected. The main indications of the asynchronous regime are the availability of ω_1 and ω_2 and the periodic variation of the angle δ. Based on this, ACS may be programmed in such way that the defensible object is disconnected and the asynchronous drive is stopped.

Another possible application of the uniform astronomic time in PES is for forming the difference:

$$\Delta t = t - t_a = \int_0^{ta} \frac{\Delta f}{f_n} dt \,, \tag{6}$$

where: t synchronous time – counting from the synchronous electrical clock; t astronomic time; Δf the system frequency deviation f from nominal value f_n.

The admissible difference between the synchronous and the astronomic time is ± 2 min. The electrical clocks, that count the synchronous time, are derived by a synchronous motor and the measured time depends on the frequency of the PES. It is an integral index for frequency deviation from the nominal and through it is made an accurate estimation for the frequency support. The value Δt_c may be used as an input value in the frequency regulators of rotating in the electric power-station.

C. General Criteria

Regardless of its functions, ACS are scanning and processing in real time many quantities with various features. In this case the electrical signals with industrial frequency are critical. Their registration and storing is made synchronous with the internal clock of the ACS.

The uniform time makes possible the comparison between the instant values of the electrical quantities of the remote energy objects. The values, transmitted through the communication channels, are connected with concrete time values and may be compared with the corresponding values, recorded in the same time in the other object recipient.

For fast developing processes (short-circuits) and for the most transient processes, typical time constant of the aperiodic component is 50 – 100 ms, and the order of the informative harmonics, necessary for analysis and control is maximum 15 (750 Hz for the 50 Hz frequency of the basic harmonic).

In the comparative analysis of fast developing failures at remote objects, without synchronization of the registration time of the quantities, the error may approach 100%. By the standard way of analog signal measurement- microcontroller with build-in 10 bit, 8-channel ADC, one cycle of conversion takes 26 clock times. For a typical clock frequency of the microcontroller of 2 MHz (clock generator 16 MHz), the necessary time for a sample is about 18μs. It is included the time for selection the measurement channel and the time for controlling the Sample/Hold, either.

The microcontroller clock generator guarantees the necessary stability and accuracy of the internal base time pulses. A difficulty is raised by the requirement for a synchronous work of set controllers, which are distributed irregularly on the PES area. Besides this, for guaranteeing reliability of the registered events, the error in giving the uniform time at the PES objects, must not be greater than the necessary time for scanning of the quantities instant values.

Usually the clock quartz generators have relative short-time instability of the order of $10^{-4} - 10^{-5}$. Another important parameter – the long-term instability of the frequency depends on a row of parameters, schematics, constructive decision, operational environment, adjustment accuracy etc. The researches chow, that the quartz generators, used in the microcontrollers, have comparatively big relative long-term instability of the frequency. The main reasons are the accuracy of the initial adjustment and the different operation environment. So that it can not rely only on internal clock quartz generators for achieving the necessary accuracy in the measurement system (or control system) working in the uniform time.

IV. Conclusion

The accuracy of the described means for giving the uniform time (synchronization) with GPS is satisfactory for the PES needs. It is necessary to analyze the possibilities for creating an algorithm for processing and simplification of the informative package from GPS, so that it takes minimum time of the microcontoller for synchronization, without any lost of information about the controlled object condition.

The optimal variant is a combination of hardware and software mean for task distribution in ACS.

References

[1] Sangeeta Nagrare, M. R. Sivaraman, Synchronisation for WAAS over Indian Airspace using GPS, The Asian GPS Conference 29-30 October 2001, New Delhi.

[2] Pratap Misra, Brian P. Burke, Michael M. Pratt, GPS Performance in Navigation, Proceedings of the IEEE, Vol. 87, No. 1, January 1999

[3] Enge, P.K. Global positioning systems: signals, measurements, and performance, *International J. Wireless Information Networks* 1(2)

[4] John F. Hauer, Jeff E. Dagle, Pacific Northwest National Laboratory, White Paper on Review of Recent Reliability Issues and System Events, Transmission Reliability Program U.S. Department of Energy

[5] G. Gross (UIUC), A. Bose (WSU), C. DeMarco (UWM), M. Pai (UIUC), J. Thorp (Cornell U) and P. Varaiya (UCB) PSERC, White Paper on Real-Time Security Monitoring and Control of Power Systems, Transmission Reliability Program U.S. Department of Energy

[6] H. Quinot, H. Bourlès, T. Margotin, “Robust Coordinated AVR+PSS for damping large scale power systems ”, article accepted for publication in the IEEE PES Transaction

Control of the Voltage Regime of Electric Power Supply Systems in Industrial Enterprises

A. Pachamanov[1], D. Bibev[2], D. Pachamanova[3]

Abstract – **This report discusses the structure and organizational principles of a data acquisition and control system for maintaining an optimal regime of the voltages in an electric supply power system (ESPS) in an industrial enterprise. The voltages at the terminals of the consumers can be controlled through the regulator in the main adjustable transformer substation (MATS) as well as the compensating devices for reactive electric power in the ESPS. To this end, the assigned electric loads in the ESPS are measured periodically, and an optimization problem is solved in real time.**

Keywords – **Electric Power supply, Voltage Control**

I. Introduction

Maintaining an optimal regime of the voltages in the electric power supply systems of industrial enterprises is important for the improvement of the operating regime of electric devices that are sensitive to deviations of the voltage from its nominal value. Devices containing electromagnetic systems (electric motors, inductive ballasts of discharge lamps, start-stop devices with magnetic core) are sensitive to positive deviations of the voltage – the saturation of their magnetic cores is a reason for loss of active electric power, which leads to overheating and intensive wearing out of the insulation, increased noise, and unutilized additional loss of electricity. Significant deviations of the voltage below its nominal value are also undesirable. When there are lighting loads, such deviations reduce lighting yields (lm/W) because of inefficient operation of the lamps. The start moment of rotation of asynchronous motors decreases when the voltage decreases, which is dangerous for the work of important machines – if the basic machine fails and there is need for automatic start of a backup machine, the latter machine may not be able to start if the voltage is low. This necessitates defining the following allowable deviations of the voltage at consumer terminals: a) for lighting load with incandescent lamps +/-2,5%; b) for lighting loads with discharge lamps +/5%; c) for asynchronous motors +10%/-5%. Regulating devices and appropriate algorithms are used in order to ensure that these constraints are satisfied in an ESPS [1,2].

[1]Angel Pachamanov, Electric Power Supply and Equipment Department, Technical University – Sofia, pach@tu-sofia.bg

[2]Dimitar Bibev, Electric Power Supply and Equipment Department, Technical University – Sofia, dbib@tusofia.bg

[3]Dessislava Pachamanova, Babson College, Babson Park, MA 02457, USA, dpachamanova@babson.edu

II. Essence of the Problem

A. Optimal Regime of the Voltages in an ESPS

The dynamics of the active electric loads requires the voltage regime in ESPS in industrial enterprises to be determined in real time through periodic control of the loading of the transformer substations and of the motors that set in motion machines in power departments – compressor and thermal power departments, pump stations, etc. Since the loads at this level of the ESPS are symmetric, the converters of the electric parameters work in a single-phase mode [2]. Through telecommunication means [3], the information from the converters is sent to the dispatcher station of the industrial enterprise (Fig. 1), and the most appropriate deviations of the voltage of the transformer at the MATS as well as the optimal distribution of the compensating reactive power devices in the ESPS are determined.

If the active electric loads in the industrial enterprise are the approximately equal every hour of the day and night, the optimal voltage regime can be easily achieved. This is due to the fact that in the design of an ESPS, the selected deviations of the voltages in the department transformers correspond to the maximum values of the active electric loads. However, if the active electric loads vary significantly, the addition of deviations of the voltage in the MATS is not always sufficiently effective in managing the voltages at all terminals of the MATS. This is why, for example, for maintaining the optimal voltage of lighting loads, special devices such as transformers and electronic regulators are attached to the low voltage side of the department substation (Fig. 1). For

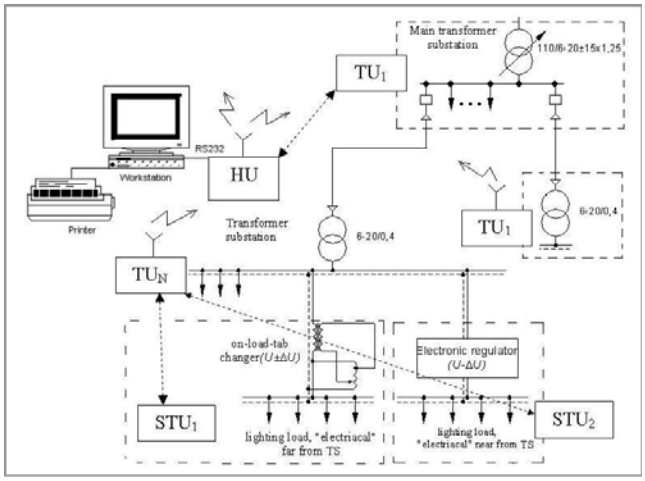


Fig. 1. Voltage control system in industrial enterprises

department substations supplying mainly asynchronous motors, the voltage regime can be influenced through redistribution of the reactive loads. Fig. 3a and Fig. 3b illustrate the dependence of the additional consumed active (respectively, reactive) power on the level of supplied voltage.

It is well known that the stable operating regimes of electric circuits with sinusoidal alternating current can be described by a system of linear algebraic equations with complex values – impedances of lines and devices, voltages at the nodes, assigned electric loads in the ESPS. These algebraic equations are also called state equations, and are constructed according to the substitute circuit of the ESPS. Given the voltage at the entrance of the industrial enterprise and current measurements of the magnitude of the assigned electric loads in the ESPS, the most appropriate method for solving for the stable voltage regime is the matrix method [1,6]. The optimal position of the voltage regulator in the MATS and the optimal distribution of the compensating devices for reactive electric power are determined by solving a concrete optimization problem for the given configuration in the ESPS [4]. The problem is solved in real time through a computing device (a controller or a computer) every one or one half hour. The obtained solution is implemented through control commands to devices – a voltage regulator in the MATS (the so-called Jansen regulator), compensating reactive power batteries in the ESPS, inducing coils of synchronous motors in power departments that produce reactive electric power.

The main constraints in the optimization problem are the allowable deviations of the voltage at ESPS nodes at which there are assigned electric power loads. The objective function is selected according to the concrete requirements of the ESPS – minimize the loss of active electric power while transferring reactive electric power, minimize the total deviation of the voltages at the nodes of the ESPS without considering the loss of active electric power in transferring reactive electric power, and so on. In [4], the optimization problem is solved with objective function "Minimize the consumed total electric power used by the self-service consumer department given hourly constraints on the power factor". This implies the greatest possible overloading capability of the MATS and minimal additional losses of electric energy in it. The decision variables in the optimization problem are the supply voltage at the entrance of the MATS and the magnitude of the assigned electric loads in the ESPS.

In addition to information about the current values of the entrance voltage and the assigned electric loads, the following data are necessary for solving the optimization problem:

- The current position of the Jansen regulator and the available degrees for regulating the voltage in the MATS;
- The maximum compensating reactive power in each department substation and the current state of the compensating devices for reactive power (batteries and synchronous motors);
- The static characteristics of the active and reactive electric power in the lines with assigned electric loads.

B. Static Characteristics of the Consumers

Lighting installations are most sensitive to increases in the supplied voltage, because the life of lighting sources can be greatly reduced, and the loss of active electric power in the ballast increases. In the case of discharge lamps, the reactive electric power also increases because of the increase of electric current through the inductive ballast.

Fig. 2 shows the change in active electric power of lighting installations with different lighting sources for voltage deviations between -10% and +10% [5]. Similar static characteristics of asynchronous motors are shown in Fig. 3 (for given load coefficient $Kn = P/Pnom$). It is recommended that the static characteristics for each consumer (substation) be recorded individually, because the type of the electric loads in each department can be different.

C. Input Information and Use

For effective management of the voltage regimes, it is important to determine the type and the quality of the information reaching the dispatcher station. The construction of a control system with different types of measuring devices (for voltage, electric current, active and reactive electric power), in addition to being expensive, is also unnecessary. The main question is the accuracy of the received information. The measurements from different devices frequently provide contradicting information when verified by theoretical means. Our experience shows that it is more appropriate for all necessary parameters (the effective value of the voltage, the effective value of the electric current, and the active electric power) to be obtained from a single measuring device. By recording the current values of the electric current and the voltage (through scanning using microprocessor devices), it is possible to calculate all other parameters mentioned above [2].

As was already discussed, in order to compute the voltages at the nodes of the ESPS using the matrix method, it is sufficient to have information about the voltage at the entrance of the plant, the parameters of the substitute circuit, and the assigned electric loads in the ESPS. The solution to the system of equations gives the voltage-drops up to the terminals of the consumers: $dU = Z.I$, where dU is a vector containing the total voltage-drops at the nodes of the ESPS relative to the voltage at the entrance of the plant (assumed to be zero); Z – a square matrix containing the coefficients of the con-

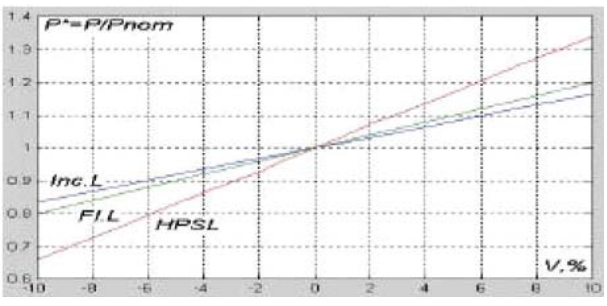


Fig. 2. Static characteristics of the active electric power of different types of lamps, $P* = f(V)$

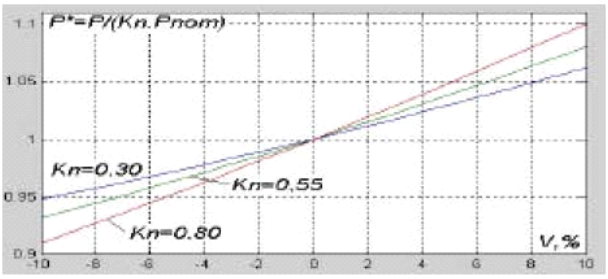


Fig.3 a. Static characteristics of the active electric power P*=f(V,Kn) of asynchronous motors (P=Kn .Pnom.P*)

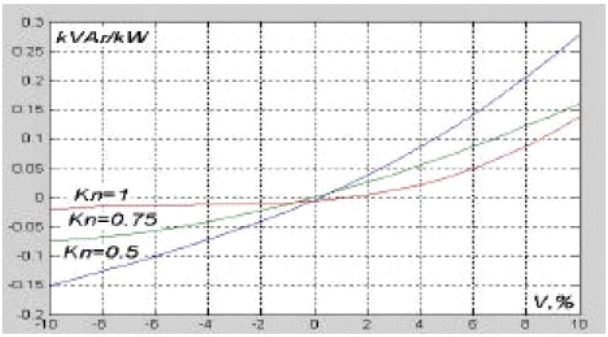


Fig.3b. Additional reactive electric power dQ=f(V,Kn), [kVAr/kW] for asynchronous motors (Q=Qnom+ dQ*Pnom)

tour impedances; I – a vector containing the contour electric currents in the ESPS [6].

For open radial schemes, which is the case with most electric supply power systems in industrial enterprises, the measurement of the effective values of the electric current, the voltage, and the active electric power is usually done at the exits of the MATS radial lines. The recorded active electric power is higher than the actual electric power consumed by the devices because of the losses of active power in the lines. High-voltage lines are designed to be heat resistant and their cross-section is big, so the loss of active electric power in them is small and can be ignored. However, if the deviations of the voltage have to be calculated for the low-voltage side of transformer substations 6-20/0.4 kV, the loss of active power is more significant and should be accounted for. In the latter case, the computation for the voltage at the terminals of the consumers can be done with reasonable accuracy using the algorithm suggested in [4].

D. Description of the Optimization Problem

The input data for the optimization problem are:

- The number of nodes in the ESPS: m;
- The number of lines (arcs) in the ESPS: n (for open graphs $n = m$);
- Deviation of the voltage at the entrance of the MATS (as a percentage of the nominal voltage): $V[0]$, %;
- Current additional deviation of the voltage of the Jansen regulator of the transformer in the MATS: Eqr_c, %;
- Available additional deviations of the voltage in the MATS: $Eqr[25] = -15.0;\ -13.75\ ;..;\ 0\ ;..;\ +13.75\ ;+15.0$;
- The fixed additional deviations of the voltage in the lines containing transformers that cannot be regulated when loaded: $E[j],\ j = 1..n$;
- Maximum values of the assigned active electric loads in the ESPS: $Pmax(j), Qmax(j),\ j = 1..n$;
- Currently measured values of the assigned electric loads in the ESPS: $Pmeas[j],\ Qmeas[j];\ j = 1..n$;
- Load coefficients: $Kn = Pmeas[j]/Pmax[j],\ j = 1..n$;
- Maximum and minimum values of the reactive electric load coming from the compensating devices in the ESPS: $Qkmax(j),\ j = 1..n;\ Qkmin(j),\ j = 1..n$;
- Static characteristics of the active and the additional reactive electric power for the assigned electric loads, described by polynomials for P and dQ as a function of the deviation of the voltage V given load coefficient $Kn[j]$: $P[j] = f(dU, Kn), dQ[j] = f(V, Kn)$.

The objective function and the constraints of the optimization problem are described as follows:

a) Total losses of active electric power in the lines and devices of the ESPS: $dP = Sum\{dP[j],\ j = 1..n\} = \min$;

b) The total power at the entrance of the MATS has to be smaller than the nominal power of the transformer: $S^2 = P^2 + Q^2 \le Snom\_tr$;

c) Active electric power at the entrance of the MATS: $P = Sum\{(Pcur[j] + dP[j]),\ j = 1..n\}$;

d) Reactive electric power at the entrance of the MATS: $Q = Sum\{(Qcur[j] + Qkcur[j] + dQ[j]),\ j = 1..n\}$;

e) Power factor at the entrance of the MATS: $\cos FI = \cos(\text{arctg}(Q/P))$;

f) Lower limits on the power factor for some hours of the day: $\cos FI = \cos FI\min[hour]$;

g) Upper limits on the power factor for the remaining hours of the day: $\cos FI \le \cos FI\max[hour]$;

h) Allowable range for the current compensating reactive electric power in line j: $Qkmin[j] \le Qkcur[j] \le Qkmax[j]$; or $0 \le Qkcur[j] \le 0.62 \cdot Pcur[j],\ j = 1..n$;

i) Current value of the active electric power of line j: $Pcur[j] = Pmax[j] \cdot P(j)$;

j) Change in the active electric power for given load coefficient and deviation of the voltage: $P = P(Kn, V)$ – calculated from polynomials;

k) Current value of the reactive electric power of line j: $Qcur[j] = Qmax[j] + Pmax[j] \cdot dQ$;

l) Additional reactive electric power for given load coefficient and deviation of the voltage: $dQ = dQ(Kn, V)$ – from polynomials;

m) Deviation of the voltage at node i: $4V[i] = 100 * (U[i] - Unom[i])/Unom[i])$, where $U[i]$ is the current value of the voltage at node i, computed using the matrix method for the current value of the voltage $U[0]$ at the entrance

of the MATS, the selected additional deviation of voltage Eqr_c, the current values of the assigned electric loads $(Pcur, Qcur)$, and the selected distribution of compensating reactive powers $Qkcur$ at the low voltage side of the transformers and the 6 kV side of the MATS;

n) Allowable range for the deviation of the voltage at node i: $Vmin[i] \leq V[i] \leq Vmax[i]$;

o) Losses of reactive power in line j: $dQ[j] = X[j] * (Pcur[j]^2 + Qcur[j]^2)/(100 + V[i])/100)^2$;

p) Losses of active electric power in line j: $dP[j] = dPa[j] + dPr[j]$;

q) Losses of active electric power, caused by the active power in line j: $dPa[j] = R[j] * Pcur[j]^2/(100 + V[i])/100)^2$;

r) Losses of active electric power, caused by the reactive power in line j: $dPr[j] = R[j] * Qcur[j]^2/(100 + V[i])/100)^2$.

E. Applications of the Suggested Method

The implementation of the described method has been done in Matlab (with a concrete objective function) for the electric supply power system of the self-service consumer department of a thermo-electric power plant – the completed preparatory work (collecting current information, study and verification of the substitute circuits in the ESPS, representation of the substitute circuits as a graph) is described in [4]. The system is planned to have a role as an advisor to a human operator (as opposed to an independent controller) because of the big responsibility associated with the safety of running the processes the thermo-electric power plant. A printout with the optimal regime will be presented to the dispatcher every hour. The printout will contain: a) recommended position of the Jansen regulator; b) recommended distribution of the compensating reactive electric loads in the MATS; c) sequence of the commands the dispatcher needs to transmit in order to achieve the regime recommended in a) and b). The visualization of the ESPS is done through a PC monitor: the current levels of the voltage at the nodes of the ESPS are continuously updated on the screen.

III. Additional Remarks

Significant preparatory work is needed for the implementation of the suggested algorithm. The first stage consists of the creation of the substitution circuit scheme in the ESPS, its representation as a planar graph, and the development of tools for calculations necessary for determining the optimal voltage regime. A very important part of the research is recording the static characteristics of the assigned electric loads in the ESPS and describing them via polynomials. The next step is designing and developing a data acquisition and control system. The main part of this system is installing measuring devices (for the effective values of the electric current, the voltage and the active electric power in the lines) and execution devices (for control of the compensating reactive electric loads). This part is easy to implement through available and convenient for exploitation microprocessor systems and telecommunication means [2,3].

IV. Conclusion

The most significant part of systems for optimization of the voltage regime, namely obtaining with sufficient accuracy the static characteristics of the assigned electric loads, can be done relatively easily if the implementation of the system begins with its data acquisition part. While installing the execution devices and developing the algorithms for control, the already implemented database is continuously updated with all random changes in the voltage, in the active and the reactive electric power, and in the connected compensating reactive electric devices in the ESPS. After statistical analysis of the recorded data, confidence intervals for a desired level of significance can be created for the static characteristics of change in the active and reactive electric power of the assigned electric loads given a change in the deviation of the voltage. At the final stage of implementation, the system is re-programmed and turned from data acquisition to data acquisition and control system.

Acknowledgement

The authors would like to thank Prof. S. Siderov, chairman of the Electric Power Supply and Equipment department of the Technical University in Sofia, for his responsiveness and his valuable advice in the development of the program for analysis of the voltage regimes that was used in the implementation of the optimization algorithm.

References

[1] Siderov S., A. Pachamanov, D. Bibev. An algorithm and a computer program for analysis of the voltage regime in industrial enterprises. Smolian'2001

[2] Bibev D., A.Pachamanov. Control and monitoring of voltage for lighting systems in industrial enterprises. ELMA02, Sofia, September 13-14, 2002

[3] Matanov N.,R.Pachamanov, D.Bibev, A.Pachamanov. Using GSM-modules for control of street lighting and industrial electric power supply systems. Energetic and information systems and technologies. 2003, October 16-18, TU-Sofia

[4] Bibev D., D.Pachamanova, A.Pachamanov. An optimization-model for determining the optimal voltage regime in industrial electric power supply systems. Energetic and information systems and technologies. 2003, October 16-18, TU-Sofia

[5] Kungs, Ia. A. Automatization of the control of electrical lighting, Energoatomizdat. Moscow, 1989

[6] Melnikov, N. A. Matrix method for analysis of electric circuits, Energia, Moscow, 1972

An Optimization Model for Determining the Optimal Voltage Regime in Industrial Electric Power Supply Systems

D. Bibev[1], D. Pachamanova[2], A. Pachamanov[3]

Abstract – **We solve for the optimal voltage regime in an industrial electric supply power system (ESPS) of the self-service consumer department of a thermo-electric power plant in real time. The objective is to minimize the total electric power of the consumers given hourly constraints on the power factor. The decision variables are the level of the supply voltage at the entrance of the main adjustable transformer substation (MATS) and the magnitude of the assigned electric loads in the ESPS. The constraints are the allowable ranges for the voltages at the terminals of the consumers and the maximum available reactive compensating electric power.**

Keywords – **Electric Power Supply, Voltage Optimization**

I. Introduction

When the terminals of consumers of electric power are fed nominal voltage, they work under their optimal regime. The deviation of the voltages from their nominal values for the nodes with assigned electric loads (consumers of electric energy) is restricted according to the requirements of the electric equipment – lighting installations, electric motion devices, technological processes. Finding an optimal regime of the voltages in an ESPS is a complex task that is solved using state-of-the-art measuring, transfer, and processing devices. After processing the operating information, actions on the ESPS devices are undertaken [1,2]. The dynamics of the active loads in an ESPS requires the voltage regime to be determined in real time through periodic control of the loading of the transformer substation and of the high-voltage motors that set in motion powerful machines (compressors, pumps, fans). The stable operating regimes can be described by a system of linear algebraic equations that are determined from the substitute circuit of the ESPS. The matrix method [3] is the most appropriate for solving for the stable voltageregimes when the supply voltage at the entrance of the MATS is given and the assigned electric loads can be measured in real time. The appropriate position of the voltage regulator of the MATS and the distribution of the compensating devices for reactive electric power are obtained by solving an optimization problem every one (or half) hour.

[1]Dimitar Bibev, Eng., Electric Power Supply and Equipment Department, Technical University – Sofia, dbib@tusofia.bg

[2]Dessislava Pachamanova, Babson College, Babson Park, MA 02457, USA, dpachamanova@babson.edu

[3]Angel Pachamanov, Electric Power Supply and Equipment Department, Technical University – Sofia, pach@tu-sofia.bg

II. Essence of the Problem

A. Characteristics of the ESPS

The main requirements of the voltage regimes in the ESPS of industrial enterprises are commented upon in [2-4]. This

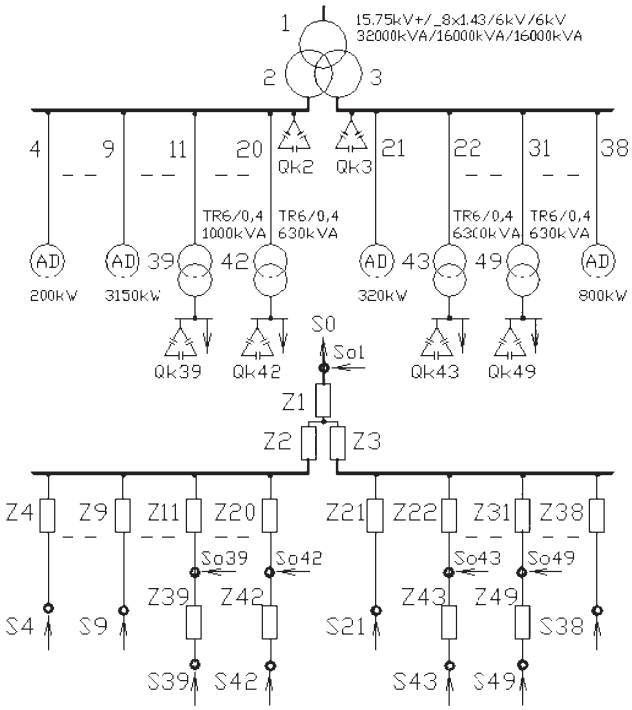


Fig. 1. Single-wire and substitute circuits of the ESPS

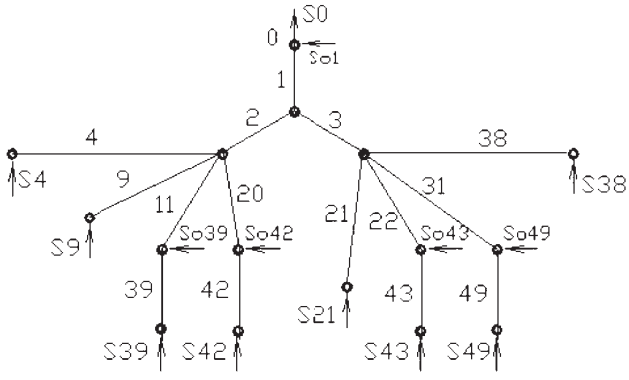


Fig. 2. Representation of the ESPS circuit as a graph diagram

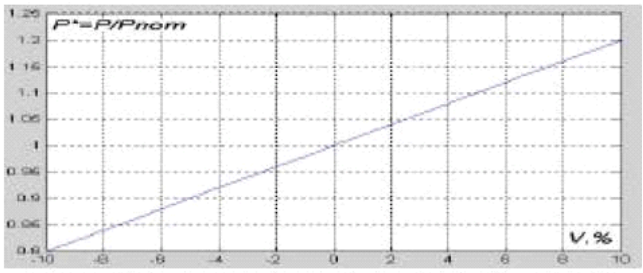


Fig.3a. Static characteristics of the active electric power of nodes 39-49 of the ESPS, P*=f(V)

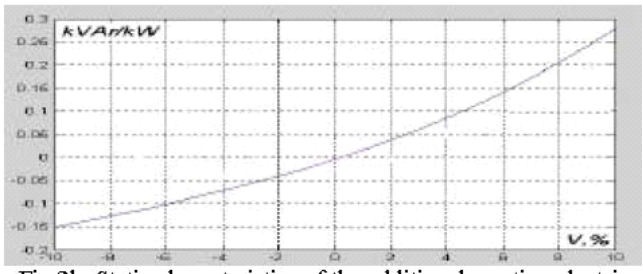


Fig.3b. Static characteristics of the additional reactive electric power of nodes 39-49 of the ESPS (dQ=f(V) [kVAr/kW], where Q=Qnom+dQ*Pnom)

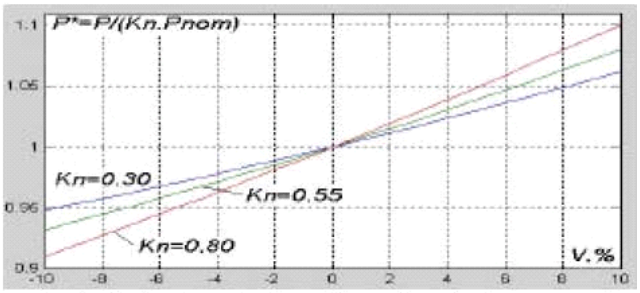


Fig.4a. Static characteristics of the active electric power of nodes 4-38 of the ESPS, P*=f(dU,Kn)

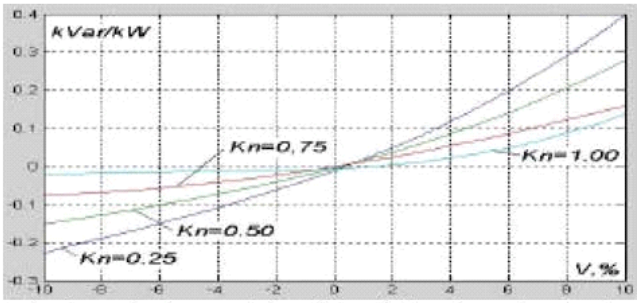


Fig.4b. Static characteristics of the additional reactive electric power of nodes 4-38 of the ESPS, (dQ=f(V,Kn) [kVAr/kW], where Q=Qnom+dQ*Pnom)

report describes the specifics of the algorithms for control of the voltage regimes of a particular ESPS. The following preliminary work is done in preparation for the implementation of the suggested algorithm: a single-wire and substitute circuits of the ESPS (Fig. 1), as well as a graph diagram of the connections (Fig. 2), are developed; the accuracy of the information obtained from the existing information system is verified; and the available information is entered in a database, where it is also sorted in order to be used by the computer serving as an advisor to the human operator.

An important part of the research is obtaining the static characteristics of the assigned electric loads. The latter research can be conducted without interrupting the normal operation of the electric devices – the change in voltage is done through the regulator in the MATS in the allowable ranges for the electricity consumers. This work is not completely finished yet, but examples of the static characteristics are given in Fig. 3-4.

B. Optimization Problem Objective and an Algorithm for Obtaining the Deviations of the Voltage

The main constraints in the optimization problem are the allowable deviations of the voltage at ESPS nodes at which there are assigned electric power loads. The objective function is to minimize the total electric power used by the self-service consumer department given hourly constraints on the power factor. This objective function is similar to the objective function described in [2] (minimize the loss of active electric power while transferring reactive electric power), but now the goal is to force the MATS electric consumers to use minimum total power. This allows a larger part of the energy produced by the thermo-electric power plant generators to be used in the national energy system. In addition, this guarantees minimal loss of power in the MATS.

The decision variables in the optimization problem are the supply voltage at the entrance of the MATS and the magnitude of the assigned electric loads in the ESPS.

In order to compute the voltages in the nodes of the ESPS using the matrix method [3], it is sufficient to have the supply voltage at the entrance of the MATS, the parameters of the substitute circuit, and the assigned electric loads in the ESPS. The solution of the system of algebraic equations is of the form $dU = Z.I$, where dU is a vector containing the total voltage-drops at the nodes of the ESPS, recorded relative to the entrance of the MATS; Z – a square matrix containing the coefficients of the contour impedances; I – a vector containing the contour electric currents in the ESPS [6]. For an open radial scheme, which is the scheme considered in this report, the measurement of the effective values of the electric current, the voltage, and the active electric power, is done at the exits of the MATS radial lines. In this case the recorded active electric power is higher than the actual electric power consumed by the devices because of the losses of active power in the lines. The latter losses, however, are not large because the lines are designed to be resistant to heat (the cross-section of the lines is big). This is particularly true in the case of high-voltage motors (6 kV). However, if a transformer substation 6/0,4 kV does not work at full capacity, the losses of active electric power in the transformer winding should be accounted for. When the loading of a transformer substation is determined mainly by a group of asynchronous motors with approximately equal loads, the deviations of the voltages at the terminals of the motors and the natural reactive electric power can be computed with reasonable accuracy using the following algorithm:

a) At the beginning it is assumed that the active power load $Pmeas$ and the reactive power load $Qmeas$ of the transformer substation are measured at the low voltage side of the transformer;

b) The deviation of the voltage at the low voltage side of the transformer is computed;

c) The load coefficient $Kn[j] = Pmeas[j]/Pmax[j]$ is determined. The additional reactive power dQ for that Kn is found (Fig. 4b). The current reactive power $Qcur = Qmax + dQ * Pmax$ at the low voltage side of the transformer is calculated;

d) The current compensating reactive power at the low voltage side of the transformer, $Qkcur = Qmeas - Qcur - Qo$, is calculated. Here Qo denotes the magnetizing reactive power of the transformer;

e) The losses of active power in the line and the transformer are calculated using the measured values of the active and the reactive loads ($Pmeas$ and $Qmeas$). The current value of the active power at the low voltage side of the transformer, $Pcur' = Pmeas - dP$, is computed;

f) The so-determined values for the loads at the low-voltage side of the transformer, $Pcur$, $Ocur$, and Qk, are then used as input to the optimization problem that solves for the optimal regime of voltages in the ESPS.

C. *Calculation of the Parameters of the Optimization Problem*

The relationship between the decision variables and the parameters in the ESPS can be described by the equations:

a) Total electric power at the entrance of the MATS: $S^2 = P^2 + Q^2$;

b) Active electric power at the entrance of the MATS: $P = Sum\{(Pcur[j] + dP[j]),\ j = 1..n\}$;

c) Reactive electric power at the entrance of the MATS: $Q = Sum\{(Qcur[j] + Qkcur[j] + dQ[j]),\ j = 1..n\}$;

d) Power factor at the entrance of the MATS: $\cos FI = \cos(\text{arctg}(Q/P))$;

e) Deviation of the voltage at node i: $V[i] = 100 * (U[i] - Unom[i])/Unom[i])$, where $U[i]$ is the current value of the voltage at node i, computed using the matrix method for the current value of the voltage $U[0]$ at the entrance of a MATS, the selected additional deviation of voltage Eqr_c, the current values of the assigned electric loads $(Pcur, Qcur)$, and the selected distribution of compensating reactive powers $Qkcur$ at the low voltage side of the transformers and the 6 kV side of the MATS;

f) Current value of the reactive power of line j: $Qcur[j] = Qmax[j] + Pmax[j] \cdot dQ(Kn, dU)$ – using the polynomials for dQ as a function of the deviation of the voltage V given a load coefficient Kn where $Pmax[j]$ is the maximum value of the active power of line j;

g) Allowable range for the deviation of the voltage at node i: $Vmin[i] \le V[i] \le Vmax[i]$;

h) Load coefficient of the assigned electric load of line j: $Kn[j] = Pmeas[j]/Pmax[j]$;

i) Allowable range for the current compensating reactive power of line j: $Qkmin[j] \le Qkcur[j] \le Qkmax[j]$; or $0 \le Qkcur[j] \le 0.62 \cdot Pcur[j]$;

j) Losses of active electric power in line j: $dP[j] = dPa[j] + dPr[j]$;

k) Losses of active electric power, caused by the active power in line j: $dPa[j] = R[j] * Pcur[j]^2/(100 + V[i])/100)^2$;

l) Losses of active electric power, caused by the reactive power in line j: $dPr[j] = R[j] * Qcur[j]^2/(100 + V[i])/100)^2$;

m) Losses of reactive power in line j: $dQ[j] = X[j] * (Pcur[j]2 + Qcur[j]^2)/(100 + V[i])/100)^2$.

The input data for the optimization problem are:

- The number of nodes in the ESPS: m;
- The number of lines (arcs) in the ESPS: n (for open graphs $n = m$); Deviation of the voltage at the entrance of the MATS (as a percentage of the nominal voltage): $V[0]$;
- Current additional deviation of the voltage of the Jansen regulator of the transformer in the MATS: Eqr_c;
- Available additional deviations of the voltage in the MATS: $Eqr[17]$=-11,44; -10,01;...; 0;...; +10,01 ;+11,44;
- The fixed additional deviations of the voltage in the lines, containing transformers that cannot be regulated when loaded: $E[j]$, $j = 1..n$;
- Maximum values of the assigned active and reactive electric loads in the ESPS: $Pmax(j)$, $Qmax(j)$, $j = 1..n$;
- Current values of the assigned electric loads at the nodes of the ESPS: $Pcur'(j)$, $j = 1..n$;
- Maximum and minimum values of the reactive electric load coming from the compensating devices in the ESPS: $Qkmax(j)$, $j = 1..n$; $Qkmin(j)$, $j = 1..n$;
- Values of the power factor for each hour of the day with upper limit $\cos FI\max[hour]$ and lower limit $\cos FI\min[hour]$ for $hour = 1 - 24$;
- Load coefficient: $Kn[j] = Pmeas[j]/Pmax[j]$;
- Static characteristics of the active and the additional reactive electric power for the assigned electric loads, described by polynomials for P and dQ as a function of the deviation of the voltage V given load coefficient $Kn[j]$: $P[j] = f(dU, Kn)$, $dQ[j] = f(V, Kn)$.

III. Optimization Problem Formulation

The optimization problem formulation is as follows:

Minimize	$S^2 = P^2 + Q^2$	
Subject to	$P = Sum\{(Pcur[j] + dP[j]), j = 1..n\};$	(C1)
	$Q = Sum\{(Qcur[j] + Qkcur[j] + dQ[j]), j = 1..n\};$	(C2)
	$Pcur[j] = Pcur'[j] * P[j];$	(C3)
	$P[j] = f(V, Kn, P_{30}, P_{55}, P_{80});$	(C4)
	$cosFI <= cosFImax[hour];$ $cosFI >= cosFImin[hour];$	(C5)
	$V[i] = 100 * (U[i] - Unom[i])/Unom[i]);$	(C6)
	$Vmin[i] <= V[i] <= Vmax[i];$	(C7)
	$Qcur[j] = Qmax[j] + dQ[j]$ $dQ[j] = f(V, Kn, Q_{25}, Q_{50}, Q_{75});$	(C8)
	$P_{[39-49]} = 0,02 * V + 1,0$ $P_{30[4-38]} = 0.00005 * V^2 + 0.0057 * V + 1.0$ $P_{55[4-38]} = 0.00006 * V^2 + 0.0074 * V + 1.0$ $P_{80[4-38]} = 0.00005 * V^2 + 0.0095 * V + 1.0$	(C9)
	$dQ_{[39-49]} = 0.0000353 * V^3 + 0.00096 * V^2 + 0.0277 * V - 0.0102$ $dQ_{25[4-38]} = 0.0000353 * V^3 + 0.00096 * V^2 + 0.0277 * V - 0.0102$ $dQ_{50[4-38]} = 0.00002 * V^3 + 0.000686 * V^2 + 0.0195 * V - 0.0043$ $dQ_{75[4-38]} = -0.0000033 * V^3 + 0.00043 * V^2 + 0.0121 * V - 0.00043$ $dQ_{100[4-38]} = 0.000044 * V^3 + 0.00065 * V^2 + 0.0036 * V - 0.00583$	(C10)
	$Qkmin[j] <= Qkcur[j] <= Qkmax[j];$ $or\ (0 <= Qkcur[j]\ 0 <= ,62.Pcur[j]);$	(C11)
	$dP[j] = dPa[j] + dPr[j];$	(C12)
	$dPa[j] = R[j] * Pcur[j]^2/(100 + V[i])/100)^2$	(C13)
	$dPr[j] = R[j] * Qcur[j]^2/(100 + V[i])/100)^2;$	(C14)
	$dQ[j] = X[j] * (Pcur[j]^2 + Qcur[j]^2)/(100 + V[i])/100)^2$	(C15)

The decision variables in the optimization problem are S, P, Q, $\cos FI$, U, V, $Pcur$, $Qcur$, dQ, $Qkcur$, dP, dPa, dPr. Constraints C1-C9 and C13-15 were discussed in Part C. From equations C9, using interpolation, one can determine the change in active electric power given the change in the deviation of voltage. From equations C10, using interpolation, one can determine the additional reactive electric power dQ for a particular value of the deviation of the voltage. Constraint C12 describes the available compensating reactive power at the nodes with assigned electric loads.

A. *Applications*

The suggested algorithm is implemented in Matlab in a reduced form since the research related to recording the static characteristics of the concrete ESPS is ongoing. At this stage, we do not have complete results about the behavior of the model in the real conditions of the ESPS in consideration.

IV. Additional Remarks

Concrete results about the behavior of the model will be presented after implementing the system. At the moment, we are working on algorithms and programs in C++ (for regime calculations) and in Delphi (for visualization and control). The calculations are designed to be done in real time through an embedded controller of the voltage regulator in the MATS. The regime of the voltages will be determined based on the variation in the assigned electric loads in the ESPS. The controller will transmit the information about the optimal regime to an IBM personal computer for visualization, and will retransmit the commands of the dispatcher for regime implementation.

V. Conclusion

The goal of this report was to describe the specifics of the model. There is a substantial amount of work to be completed for the model's full realization. Its implementation depends to a large extent on the management of the ESPS. If the necessary funding can be found and the system is implemented, a verification of the suggested algorithms (and possible corrections and changes) can be done.

Acknowledgement

The authors are very grateful to the executive manager of thermo-electric power plant "Bobov dol" for his cooperation and for allowing us to record data necessary for the research. We would also like to thank the on-duty and maintenance personnel of the plant, as well as our colleagues from the Electric and Automatic Devices Laboratories for their responsiveness and help in recording the static characteristics of the self-service consumer department of the thermoelectric power plant.

References

[1] Matanov N.,R.Pachamanov, D.Bibev, A.Pachamanov. Using GSM-modules for control of street lighting and industrial electric power supply systems. Energetic and information systems and technologies. 2003, October 16-18, TU-Sofia

[2] Pachamanov A., D.Bibev, D.Pachamanova. Control of the voltage regime of electric power supply systems in industry. Energetic and information systems and technologies. 2003, October 16-18, TU-Sofia

[3] Melnikov, N. A. Matrix method for analysis of electric circuits, Energia, Moscow, 1972

[4] Siderov S., A. Pachamanov, D. Bibev. An algorithm and a computer program for analysis of the voltage regime in industrial enterprises. Smolian'2001

Using GSM-Modules to Control the street lighting AND Industrial Electric Power Supply Systems

N. Matanov[1], R. Pachamanov[2], D. Bibev[3], A.Pachamanov[4]

Abstract – **The possibilities of using the GSM technology for controlling the outdoor lighting systems and electric power supply objects are analyzed in this paper. Opportunities for data exchange using GSM network are discussed – between regional informational systems and power supply objects by a local area network. The costs of using the services of a GSM operator in Bulgaria are evaluated as an alternative of using a single radio channel and conventional systems for data exchange.**

Keywords – **Control, GSM-modules, Electric Power Supply**

I. Introduction

A. Control Systems Using a H.F. Radio Channel

A system for centralized control of the street lighting of Sofia using a high-frequency radio channel was developed during the period 1982-1988 [1,2]. A centralized radio control system removes the need for a large number of photocells and clocks with astronomical devices on site that are normally used for turning on and off the street lighting systems. Selective control based on the natural illuminance (with thresholds 10, 20, 40, 60 and 80 lx for the different categories of streets) improves traffic safety and reduces electricity consumption. Operating under the so-called "midnight regime" (the lamps work at 50% of their nominal flux after 10pm) reduces the amount of electric power consumed by an additional 30% [3].

Since the system is idle for a large part of the day, it is also used for control of central stations for heating installations in Sofia. One main disadvantage of the system is the lack of feedback regarding the execution of the given commands the installation of sending-receiving devices instead of receiving devices on site increases the cost of the system, and the system was not completed in that respect [1]. That is why the protocol used is to send the commands many times. This reduces to a minimum the possibility that a command will not be executed.

The controlling computer in the dispatcher station is connected through RS232 with a central station. The latter transmits information to re-transmitting stations that cover the desired region (Fig. 1). The controlling software carries out

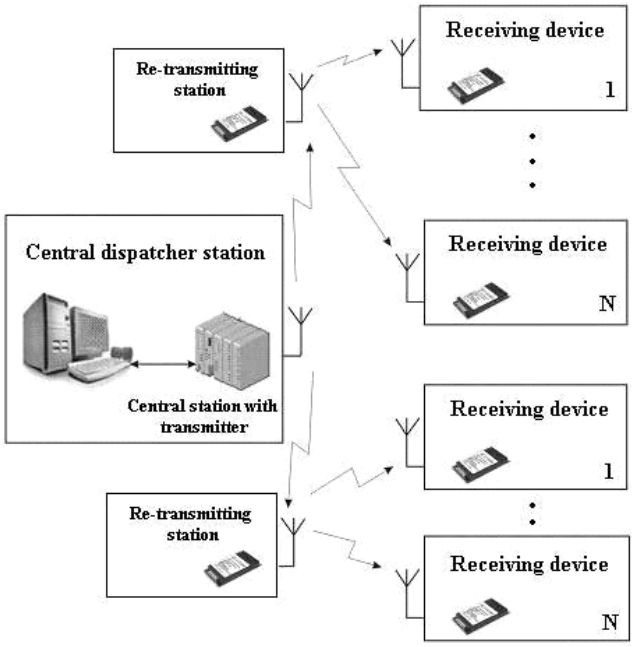


Fig. 1. Block diagram of the centralized H.F. radio channel system for street lighting control in Sofia [1]

three types of control of the street lighting installations [2]: a) simultaneously for all street lighting installations, covered by the system; b) according to specific regions (25 in Sofia); c) according to categories of streets (major avenues, large streets, small local streets); under three operating regimes: a) automatic; b) software-controlled (according to the astronomical calendar and system clock); c) manual.

The automatic regime uses a signal from a central photocell that measures the natural illuminance. The signal turns the street lighting installation on when the natural illuminance moves between 100 and 4 lx, and turns the lighting installations off when the natural illuminance moves between 4 and 100 lx. The threshold level is selected according to the category of street. If a signal is not received for a given time period, the system becomes software-controlled. Both software and automatic control require specifying a threshold of illuminance. The latter threshold determines the astronomic calendar used for control.

B. Using GSM for Controlling Outdoor Lighting Systems and Substations

During the years after developing the system for radio management the information technologies improved a lot. One of them is the GSM which in a case of a little number of re-

[1]Nikolay Matanov, Electric Power Supply and Equipment Department, Technical University – Sofia, nsm@tu-sofia.bg

[2]Radostin Pachamanov, Faculty of Communications and Communication Technologies, Technical University – Sofia, radostin_ap@abv.bg

[3]Dimitar Bibev, Electric Power Supply and Equipment Department, Technical University – Sofia, dbib@tusofia.bg

[4]Angel Pachamanov, Electric Power Supply and Equipment Department, Technical University – Sofia, pach@tu-sofia.bg

ceivers (for example when turning on street lighting not from electric boxes, but from transforming stations) makes unnecessary developing of local area radio networks with preserved radio frequency. Using the GSM-modules for transmitting data is very convenient for long distances – for example constructing regional controller's posts of similar power objects. The main advantage consists in the fact that already built network of an acting operator is used. So it is unnecessary to make investments for creating your own network. Nowadays when the GSM network is well developed it is quite convenient to use it for transmitting information between the observed objects and a controlling post with a random location as there is a good cover almost everywhere in the country.

II. The Essence of the Problem

A. Special Features when Using a GSM-Network for Exchange Information between Controller's Posts and Subordinated Stations

The exchange of information is realized through GSM-modules controlled by microcontrollers or computers. There are two main possibilities to transfer data between GSM networks.

The first one is by means of SMS (SMS - Short Message Service) which are up to 160 symbol messages. Through these 160 symbols it is possible to transfer information between supervising post (with computer or a central station type SCADA – Supervisory Control And Data Acquisition) and controlled posts which are situated in the transformation stations in the built-up areas or in the substations in power objects in the industry. The advantage of this variant is that the GSM-standard is used which provides the correct receiving of the SMS – i.e. the received message is identical with the transmitted. Another advantage is that this type of service is very cheap. Disadvantage of using SMS is the fact that this service doesn't have a priority, i.e. a message in a GSM network is going to be sent only if the network is not overloaded (the transfer of SMS is realized through signalization channels which are usually used for controlling the calls). Because of that it is possible an SMS sent to be delayed or even not received. Therefore using SMS for transferring information is convenient only when most the SMS-s transfer similar information, for instance when presenting information concerning the status of power objects – in that case the possible loss or delay of some messages is not crucial for describing the process (systems for telesignaling). With improvement of the mobile networks of the GSM-operators the quality of this service (SMS) could be increased and therefore the reliability of this type of transferring information in systems for tele-signaling could be increased.

The second way for information transferring is by using the service "data transfer" of GSM-operators. The GSM-phone/module needs a modem in order to use this service. The so-called "data number" is preliminary paid to the operator and "transparent" mode of transferring information in the GSM network is provided (the system doesn't change your information when operating with it). The transferring information sides get access to the network as the modem provides the information to be in conformity with the GSM standard. The network commutes calls between subscribers and the following transmitting of the information differs from the standard conversation by the fact that the analog to digital conversion is missing. For this reason this service is much cheaper than the standard conversation. After realizing the connection the data exchange can be provided by random protocol, because of the "transparent" mode of data transfer. Therefore it is possible to create more complicated protocol for data exchange, different from the previous variant for transfer through SMS with only 160 symbols. The GSM-modules offered now at the market (modem + GSM transceiver) [6,7] permit building the controlling system on subscribers principle (Fig. 2).

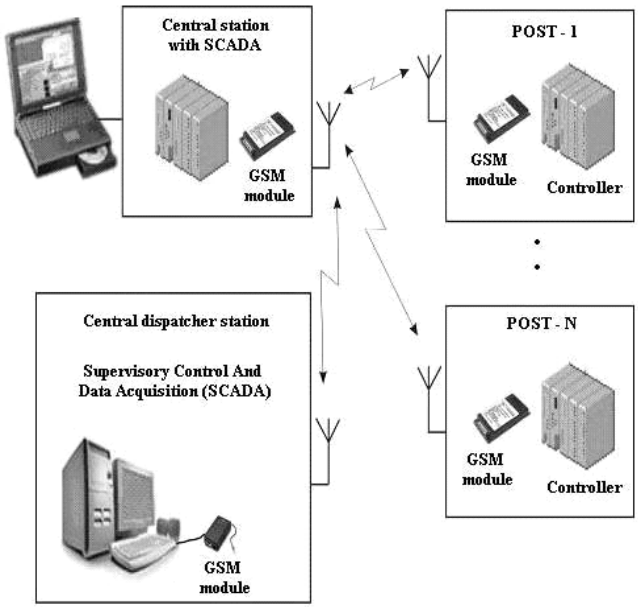


Fig. 2. Diagram for controlling distant power objects (substations) by GSM-modules

B. Data Exchange through GSM-Network between Dispatcher Station and Local Systems for Data Exchange

Local area networks for information exchange between controlled devices and local base station (Fig. 3) are developed for power objects in industry as well as for road devices. These objects are situated in separate substations in industry and in road areas they could be found at comparatively big distances – for example road tunnels with lighting installation, ventilation, fire and other safety systems.

The configuration of the system shown on Fig. 3 is applicable for medium distance between the controlled posts. For example distances between the devices in local control systems in road tunnels are from one to several kilometers. The system, shown on Fig. 4, is applicable for industrial enterprises or centralized management of equipment in trafostations or regional substations where the local systems have dozens or hundreds meters line length.

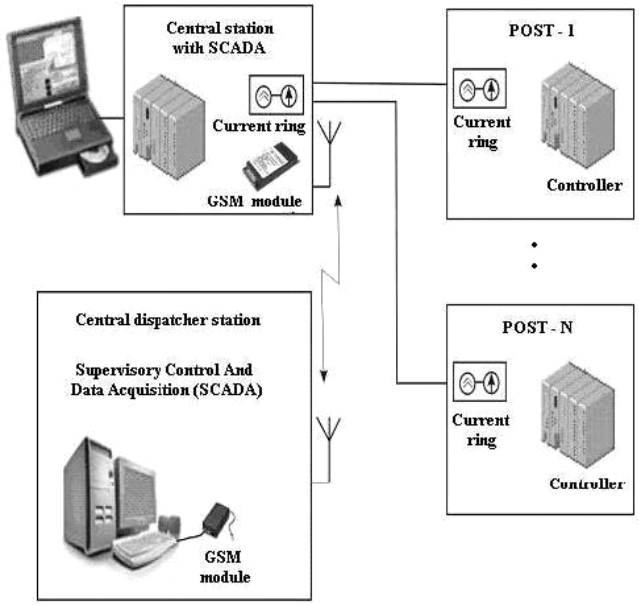


Fig. 3. Variant for using GSM-modules as connection between local base stations and regional dispatcher station – the local base stations contain a central station SCADA type, exchanging information by a current ring with controlled posts situated several kilometers away.

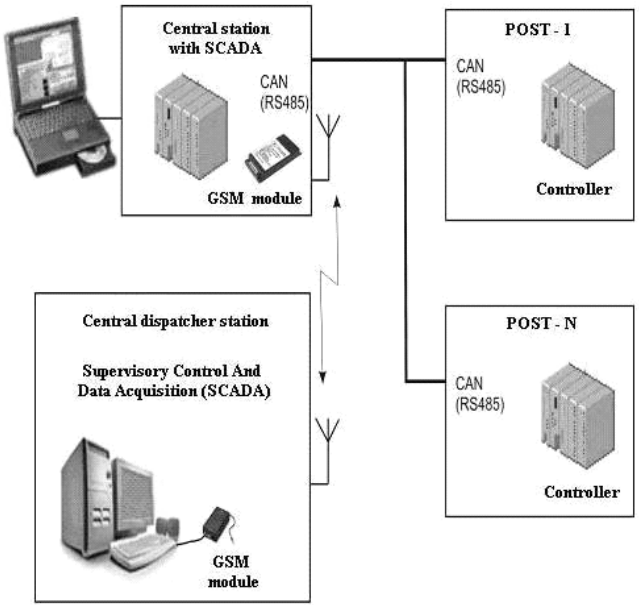


Fig. 4. Variant for using GSM-modules when connecting local base stations with regional dispatcher's station - the local base stations contain central station type SCADA exchanging information with the controlled posts situated on dozens (CAN) or hundreds (RS485) meters.

In both cases the connection between the central stations of local area networks and the dispatcher's point is by using a GSM network and the distance is not significant. A typical feature of the discussed schemes is the fact that the local systems fulfill their functions independently and the staff at the dispatcher's point is involved only when an extraordinary circumstance appears. So the main function of the dispatcher's station is to collect information on purpose of observations of the processes.

C. An Example for Developing a Regional Dispatcher's Point for Road Tunnels

There are four road tunnels each of them with two independent tubes on "Hemus" highway. Two of them ("Vitinia" and "Praveshki hanove") have local information control systems built [8]. During the following years similar systems will be created on the tunnels "Topli dol" and "Echemishka". Constructing of regional dispatcher's station for the four tunnels (Fig. 3) is convenient to be established at one of the command rooms in the tunnel substations which are situated in 30 kilometer section of the "Hemus" highway. The latter will allow decreasing the staff on duty to two persons (for the four tunnels) and the off duty staff can be retrained to maintaining the equipment, working during the day. Thus the preliminary designed schedules for increasing the energy effectiveness of the tunnel lighting on "Hemus" highway can be fulfilled.

D. Centralised control and Voltage Regime Controlling in Industrial Enterprises, Example

In some industrial enterprises some of the transformer substations are not included in dispatcher control system. At the voltage regime control among on-load-tap-changer in the main substation it is often to happen to the local power factor correction capacitors to be switched on or switched off. If the transformer substation has additional voltage regulators (Fig. 5) their managing also can be an object of dispatcher's (controller's) intervention. When telephone line to the object is missing it is appropriate to use GSM-network for realising the necessary switching over and data collection. In this case the data exchange expenses will be minimum, because the dispatcher's intervention is random and the GSM-module from the transformer substation takes initiative only in case the observed parameters exceed the limits.

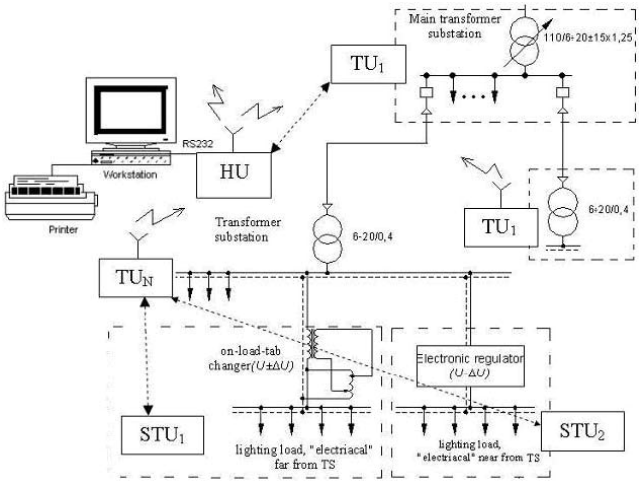


Fig. 5. Control of lighting systems voltage in industrial enterprises

III. Additional Remarks

In cases when controlled post is built as an information system the local system information is sent only if at demand or when changing the controlled parameter.

In the first case receiving tele-signaling information and possibly delivering tele-control commands will be done on behalf of the GSM-device in the controller post. In the second case renovating information to the controller post will be on behalf of the GSM device in the local system. As the demand for tele-signaling information comes from the dispatcher's post payments become minimum for most of the local stations because using prepaid exchange SIM-card (about 20 BGN annually for GSM-module). When the installations in some of the controlled energy object frequently change their status and their GSM-module is programmed to renovate the information, then paying monthly permanent tax is preferable (about 120 BGN per year + 0.05 BGN per minute for data exchange). In all cases that refers to dispatcher's post GSM-module, because the bigger part of its data exchange is on its account (recurrently inquiring information about the status of the objects in the local systems).

IV. Conclusion

The suggested organization of GSM-communications among local area networks of energy objects with random location and central dispatcher's post has significant priorities compared to the H.F.-radio-channel with preserved frequency. First it refers to using an already built network which covers the whole country. This permits connection of many self-working local systems into general regional systems – for road lighting control, traffic light systems, regional road tunnels, etc. In that case the distances between the dispatcher's post and local systems of the energy objects are not significant. Second, the data exchange could be organized so that the expenses to be minimum (one recurrently inquiring GSM-module in the dispatcher's post with monthly permanent tax and many GSM-modules in the local controlled systems with a prepaid annual tax). Last but not least we must take into account the high reliability of that kind of systems.

References

[1] [1] Vassilev N.I. and .... System for feedback of H.F. radio channel system for street lighting in Sofia. Design of devices and software. Report 1234/1989, Research Sector, Technical University-Sofia

[2] [2] Pachamanov A., K. Zaharinov, N.Vassilev. Data acquisition and control H.F. radio channel system for street lighting. Osvetlenie'93, 21-23 September 1993, Varna

[3] [3] Pachamanov A.S. About regimes of work of street lighting. "Energetika", No 3-4/ 1992, Sofia.

[4] [4] Pachamanov A., Hr.Vassilev, N.Matanov, B.Bojchev, I.Angelov. Controller for monitoring of street lighting according to the astronomic calendar. Electric power forum '98, June 10-12 1998, MDU - Varna, book II, p. 324-328

[5] [5] Bibev D., A. Pachamanov. Control and monitoring of voltage for lighting systems in industrial enterprises. ELMA02, Sofia, September 13-14, 2002

[6] [6] http://www.motorola.com

[7] [7] http://www.siemens.com

[8] [8] http://www.tunnellighting.com

Statistical Analysis and Optimization of Voltage Regulator Circuit Using IESD and ORCAD Environment

Galia I. Marinova[1] and Dimitar I. Dimitrov[1]

Abstract – **The paper presents the development of a specialized methodology for statistical analysis and optimization for a voltage regulator circuit. The study is performed in the environment of the simulators ORCAD PSpice 9.2 and IESD. Optimization steps and results from circuit simulation are presented. Conclusions for implementation of the voltage regulator circuit are deduced.**[1]

Keywords – **statistical optimization; voltage regulator circuit.**

I. Introduction

The general methodology for statistical analysis and optimization in electronics is presented in [2]. However different circuits types demand specific implementation of the general methodology. The paper presents a specific implementation of the general methodology for statistical analysis and optimization on a voltage regulator circuit. The tools applied are the statistical simulator IESD, described in details in [2] and ORCAD PSPICE 9.2 [3]. The case study is performed for a linear voltage regulator circuit which is used for power supply in antenna preamplifiers and it is presented on Fig. 1. The circuit is described in details in [1]. The study which is performed in the statistical design environment and the specific methodology developed allows to verify the circuit parameters from [1] , to determine optimal part values and tolerance values with 100% yield over predefined constraints in the Goal function and to enlarge the circuit application area.

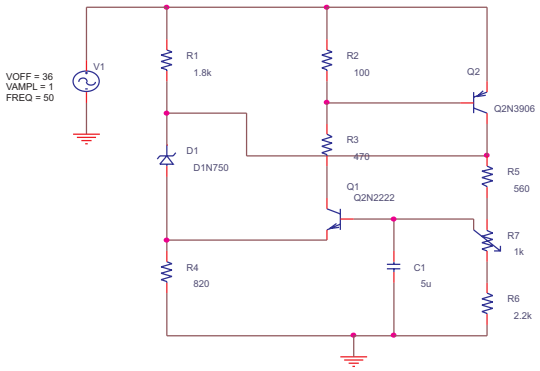


Fig. 1. Voltage regulator circuit

[1]Galia I. Marinova and Dimitar I. Dimitrov are with the Faculty of Communications and Communications Technologies in Technical University of Sofia, 8, Kliment Ohridski street, Sofia - 1000, Bulgaria. E-mail: gim@vmei.acad.bg, ddim@vmei.acad.bg .

II. Steps of the Specific Methodology for Statistical Analysis and Optimization of the Voltage Regulator Circuit

The specific methodology developed for the statistical analysis and optimization of the voltage regulator circuit includes the following steps:

- Specification and constraints definition ;
- Nominal analysis and initial choice for part values;
- Statistical analysis of the voltage regulator circuit in ORCAD PSpice 9.2 and in IESD;
- Definition of the Goal function for statistical optimization;
- Statistical optimization of the voltage regulator circuit in IESD with optimal tolerance determination;
- Nominal and statistical simulation of the voltage regulator circuit in a realistic system.

The specific methodology steps are described in details.

III. Description of the Specific Methodology Steps

A. *Specification and Constrains Definition*

The specification of the voltage regulator circuit is:

- The input voltage applied at node IN can have 3 definitions for DC value and pulsation amplitude: 36 V$\pm$1 V, 29 V$\pm$1.2 V and 39 V$\pm$0.8 V. Two frequency values are possible: 50 Hz and 60 Hz.

The constraints are the following:

- The voltage obtained at the output node STAB should be with DC value:24 V with pulsation amplitude inferior than $\pm$10 mV. So the pulsation amplitude at the output should be at least 100 times inferior than the input pulsation amplitude.

B. *Nominal Analysis and Initial Choice for Part Values*

Nominal time analysis in PSpice is performed in order to define the optimal values for R5 and R6 which replace the resistors R5, R6 and the variable resistor R7 from Fig. 1 insuring the constraints from point 2.1 and corresponding to the transistors and diodes chosen. Fig. 2 presents the voltage regulator circuit with new values of R5 and R6. The results from nominal transient analysis in PSpice are shown on Fig. 3.

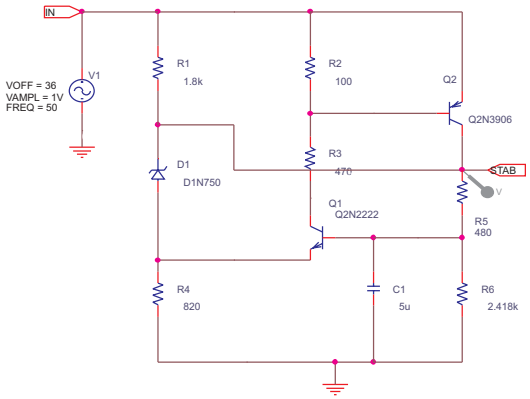


Fig. 2. R5 and R6 values defined from nominal simulation in ORCAD PSpice 9.2

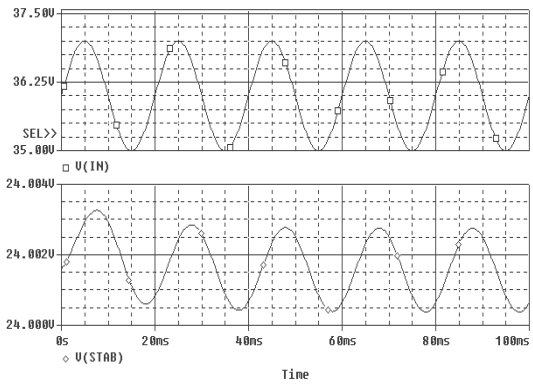


Fig. 3. Applied and stabilized voltages for the circuit from Fig. 2

The nominal analysis is performed for an open circuit. Six cases for the input source V1 definition are estimated: the analysis is performed for two frequencies – 50 Hz and 60 Hz and three groups of DC and amplitude pulsation values : 36 V±1 V, 29 V±1. 2V and 39 V±0.8 V. The results from nominal simulation in these 6 cases are presented in Table 1.

Table 1.

Case	Frequency FREQ	Input voltage V1 source V(IN)		Output stabilized voltage V(STAB)	
		DC value	Pulsation amplitude	DC value	Pulsation amplitude
1	50Hz	36V	1V	24.0010V	± 1.5mV
2	50Hz	29V	1.2V	23.9685V	± 1.5mV
3	50Hz	39V	0.8V	24.0143V	± 1mV
4	60Hz	36V	1V	24.0016V	± 1mV
5	60Hz	29V	1.2V	23.9685V	± 1.5mV
6	60Hz	39V	0.8V	24.0143V	± 0.5mV

The results from Table 2 confirm the good performance of the studied voltage regulator circuit both on 50 Hz and 60 Hz input source. The DC values of the stabilized voltage are similar for both frequencies and the pulsation amplitudes are even lower for the 60 Hz case. The output voltage parameters are good for the 3 cases for DC values and pulsation amplitudes of the input voltage source.

Table 2.

Tolerances for all R and C	DC value variation	Pulsation amplitude variation
1%	-0.4V +4.031V	1.2-1.23mV
5%	-2.747V +4.031V	2-3mV
15%	-4.9V +3.744V	2-3.06mV

C. *Statistical Analysis of the Voltage Regulator Circuit in ORCAD PSpice 9.2 and in IESD*

The circuit from Fig. 2 is analyzed statistically in ORCAD PSpice 9.2. A full Monte Carlo analysis in Time area is performed with 1%, 5% and 15% for all R and C values in the circuit. The variations for the DC value and for the pulse variation for the stabilized voltage V(STAB) is estimated through the option PERFORMANCE ANALYSIS. The DC value variation is calculated with YatX(V(STAB), 30.03 m) and the pulsation amplitude variation is calculated with SWING(V(STAB),20 ms,40 ms).

The results in Table 2 show the very strong stability of amplitude pulsation values . Their value cover the constraints with no dependence from the tolerance values.

A statistical analysis of the voltage regulator circuit in IESD with estimation of the statistical behavior of the transistor Q1 is performed. The transistor 2T6552D is used for Q1. This transistor has a statistical model in IESD and Monte Carlo simulation in performed with the use of this statistical model.

Table 3 present the variations for the stabilized voltage parameters. Fig. 4 presents results from statistical processing in IESD. Fig. 4a presents the histogram for DC value of the stabilized voltage for 1% tolerances on all R and c and statistical model for Q1 and Fig. 4b. presents the linear correlation between the variations of the DC value and the Pulsation amplitude for V(STAB).

Table 3.

Tolerances for all R and C	DC value variation	Pulsation amplitude variation
1%	22.43V 24.8V	0.1 – 1mV
5%	13.367V 25.266V	0.15 – 2mV

The results from the statistical simulation of the voltage regulator circuit in IESD show that the statistical behavior of the circuit is strongly depending from the statistical behavior of the transistor. The transistor 2T6552D is not suitable for the realization of this voltage regulator circuit. Other remark is that there is an inverse linear correlation between the variations of the DC value and the Pulsation amplitude for V(STAB) which corresponds to practical experiments with the circuit and described in [1].

Table 4.

Parts	Tol Step1	Tol Step2	Tol Step3	Tol Step4	Tol Step5	Tol Step6	Tol Step7 Optimal results
R1	**1%**	**2%**	**3%**	**5%**	**10%**	**15%**	**15%**
R2	**1%**	**2%**	**3%**	**5%**	**10%**	**15%**	**15%**
R3	**1%**	**2%**	**3%**	**5%**	**10%**	**15%**	**15%**
R4	**1%**	**2%**	**3%**	**5%**	**10%**	**15%**	**10%**
R5	**1%**	**2%**	**1%**	**1%**	**1%**	**1%**	**1%**
R6	**1%**	**2%**	**1%**	**1%**	**1%**	**1%**	**1%**
C1	**1%**	**2%**	**2%**	**5%**	**10%**	**15%**	**15%**
V(STAB) DC Value variation	**+0.283V -0.169V**	**+0.558V -0.365V**	**+0.280V -0.169V**	**+0.290V -0.186V**	**+0.394V -0.229V**	**+0.511V -0.270V**	**+0.462V -0.300V**

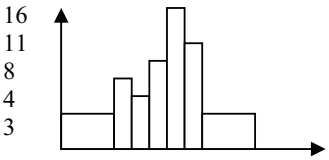
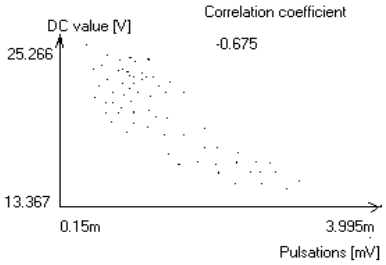


Fig. 4a. Histogram for the DC value of the stabilized Voltage with 1% tolerances for R and C and statistical model for the NPN transistor in IESD, 31% being discarded as too sharply differing values



d

Fig. 4b. Linear correlation between the variations of the DC value and the Pulsation amplitude for V(STAB)

D. Definition of the Goal Function for Statistical Optimization

Taking in consideration the results from the statistical analysis of the voltage regulator circuit described in previous paragraph the statistical optimization task is defined as follows:

- The objective is to perform optimal tolerancing for the circuit from Fig. 2 with 100% yield following the specification constraints defined in point A;
- tolerancing is limited only on R and C elements and statistical models of transistors and diods are not taken in consideration.

The goal function for statistical optimization of the voltage regulator circuit is to find the maximal tolerance values of all R and C in the circuit which guarantee 100% yield on the constraint for DC value variation for the stabilized voltage V(STAB) ± 0.5 V. The goal function is formulated as follows:

– Max Tol (R1, R2, R3, R4, R5, R6, C1),

– V(STAB) [V(1)=DC 36 V ; Pulse amplitude=$\pm$1 V] : DC 24 V$\pm$0.5 V; Pulse Amplitude $\leqslant \pm 0.01$V]

E. Statistical Optimization of the Voltage Regulator Circuit in IESD with Optimal Tolerance Determination

Since the response to constraints for pulsation amplitude is proved to be independent from tolerances only, the constraint for the DC value variations is considered in the optimization process. Table 4 presents the DC value variation at the different optimization steps for the tolerances of R and C in the circuit for the prescribed tolerance values.

At any step, a full Monte Carlo time domain analysis is performed in IESD. 100 transient analysis are executed for the randomly generated R and C values. The constraint for the DC value of the stabilized voltage is verified for each one of these circuits. If there is not a failing circuit all tolerances are increased If there is a randomly generated circuit which fails, the random values for R and C are inspected and those with the biggest deviations from the nominal values are identified. The corresponding tolerances are decreased at their previous values and they are no more increased up to the end of the optimization procedure. The procedure stops when all tolerances are definitely defined.

F. Nominal and Statistical Simulation of the Voltage

Fig. 5 presents the voltage regulator circuit included in realistic system.

The circuit from Fig. 5 is simulated nominally and statistically. Fig. 6 presents the nominal response – Input voltage V(IN), Voltage on C2 (V(C2:2)) and Stabilized voltage V(STAB). The statistical simulation implements the optimization results from point E (Step 7 in Table 4). The capacitor C2 is considered with 15% tolerance value. Fig. 7 presents the results from statistical simulation for V(STAB) and the histogram for the DC value of V(STAB). The variations for the DC value of V(STAB) and for the pulsation amplitude of V(STAB) are estimated from the statistical simulation results on Fig. 7, as follows:

DC value variation calculated with YatX(V(STAB), 30.03 m) : -0,3781 V to +0.2877 V (DC values are: 23.6221 V to 24.2877 V.

Pulsation amplitude variation calculated with SWING(V(STAB),20 ms,40 ms):1.02 mV to 1.45 mV.

These results confirm that the optimization results are good enough for system implementation of the voltage regulator circuit.

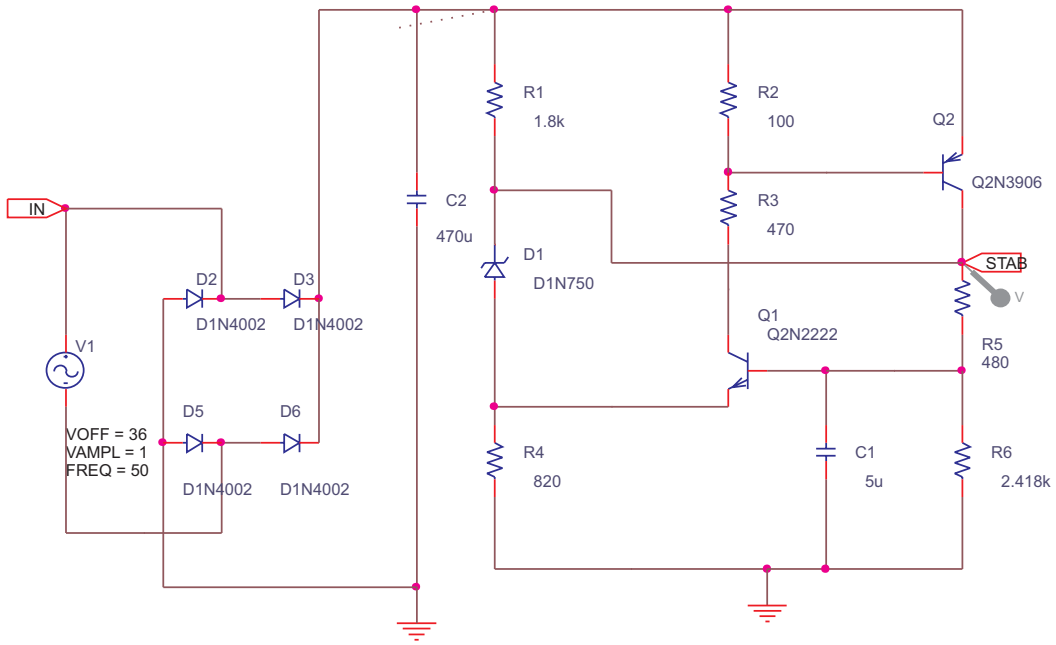


Fig. 5. Voltage regulator circuit in realistic system

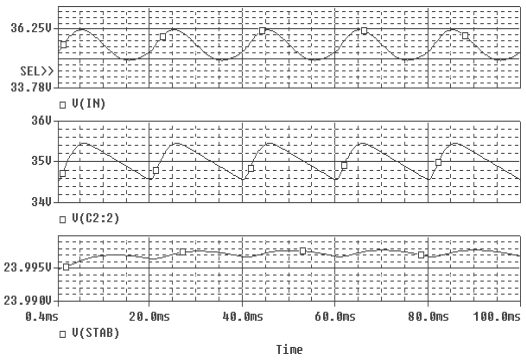


Fig. 6. Nominal response – Input voltage V(IN), Voltage on C2 (V(C2:2)) and stabilized voltage V(STAB)

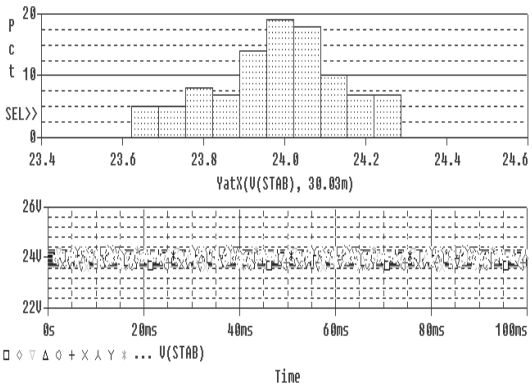


Fig. 7. Statistical simulation for V(STAB). Histogram for the DC value of V(STAB).

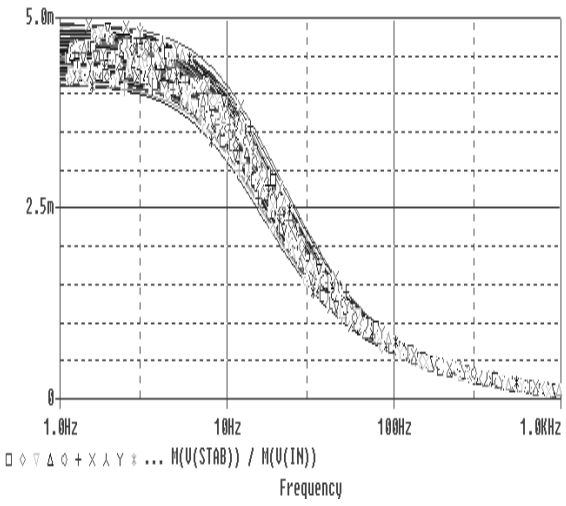


Fig. 8. Statistical simulation of the gain curve

IV. Conclusion

The paper presents a specially developed methodology for statistical analysis and optimization of a voltage regulator circuit.

The voltage regulator circuit studied in the paper is implemented in the TV sets OSOGOVO (SOFIA 11). The application of this methodology permitted to study and to confirm several characteristics of the voltage regulator circuit and to enlarge its application implementation area:

The voltage regulator circuit is suitable for an enlarged voltage supply range both for the European energy net (50 Hz, 220 V±15%) and for the American energy net (60 Hz, 110 V±15%). This conclusion is coming from nominal simulation results in Table 1 and statistical simulation results on Fig. 8 and Table 5.

The statistically optimized voltage regulator circuit with tolerances defined at Step 7 in Table 4 can be realized with hybrid IC technology. If better accuracy is demanded

Table 5.

Tolerances for all R and C	DC value variation	Pulsation amplitude variation
1%	22.43V 24.8V	0.1 – 1mV
5%	13.367V 25.266V	0.15 – 2mV

a monolithic technology could be applied.

The voltage regulator circuit is suitable for implementation in large band amplifiers and antenna amplifiers. With a little modification, decreasing the stabilized voltage at 15 V, it could be used for positive power supply for operational amplifiers.

The methodology developed will be implemented in further research on load curve, enlarged inverse current characteristic and security chain activation in the voltage regulator circuit.

The experiences performed confirm the interest in specific methodologies development for different types of circuits.

Acknowledgement

The authors acknowledge Dr. Vassil Guliashki for his help in the experimental work. Special acknowledgement are expressed to the National Science Fund for its support to the project I-1203/2002.

References

[1] D. Dimitrov, *"Voltage regulator in the TV receiver OSOGOVO"*, "Radio, Televizia, Elektronika", Book N°6, 1974, pp.167-168

[2] G. Marinova, *"Statistical approach in electronics design"*, Ph.D. thesis, Technical University of Sofia, Bulgaria, 1994

[3] *ORCAD 9.2* – Users' guide, Cadence Corp., USA, 2000

Graphics and Media support at Oxford Brookes University

Leslaw W. Zieleznik[1]

Abstract – **The paper will introduce graphics and multi-media teaching at OBU. The university's main computing facilities are presented and its limitations for specific graphics applications discussed. Finally, we take a look at the latest developments in graphics support for teaching and research.**

Keywords – **Teaching of graphics and multimedia, Graphics Workshop.**

I. Graphics Oriented Courses

Graphics and multimedia are taught at Oxford Brookes University in many departments and in a wide range of subjects.

We can generalize that graphics are taught on two types of courses, where graphics are either:

- part of a graphics-oriented course,
- the principal subject of a course.

By graphics-oriented courses we mean courses for which graphics are not the main goal of the course, but are extensively used during that course. For example, on the Geographical Information Systems course, graphics are heavily used for the presentation of analysis results, but are considered purely as a tool. A similar approach to graphics is employed for Mapping and Cartography, Urban Design, and other courses [1].

Graphics as a subject in its own right is taught in a few departments, e.g. in the Departments of Computing, Engineering and Architecture. In the Computing Department, graphics programming and multimedia are taught as the main subject - multimedia is also part of a MSC postgraduate course. The Department of Engineering runs Media Technology and CAD Engineering courses, where there is a more technical and user-centered approach to graphics. In the Department of Architecture, there is extensive use of graphics for design and visualization, but computer graphics are taught and used as a tool for getting the job done.

II. Computing Facilities

The university's computing facilities have evolved greatly in the past ten years. Initially systems were based on supermini and graphics terminals, then a number of CAD/graphics systems were introduced, based on UNIX workstations supplied by Apollo, DEC, HP and Sun Microsystems. But as the power and functionality of PC stations improved, and at the same time more graphics and CAD packages were implemented for the Windows OS, a large PC network was set-up and most of the Unix systems were withdrawn from service.

There are currently 690 PCs on the network, available in 33 open-access rooms, spread across three university campuses. These rooms, called pooled computer rooms, are centrally funded and run by Computer Services [3].

In addition there is a central UNIX system, accessible from PC stations via x-terminal emulation clients, and a specialized room for Geographical Information Systems. A few departments do run their own systems, but on a much smaller scale.

As far as software is concerned, initially many graphics applications were written in-house and were based on graphics libraries like GHOST, NAG and UNIRAS. The current graphics applications are mainly based on commercially available software from all leading suppliers. There are packages for CAD applications, drawing, presentation graphics, multimedia, desktop publishing and also graphics libraries.

III. Specialized Graphics Support

The general computer network, is not however suitable for more sophisticated graphics applications and other requirements. For example:

- packages for solid modeling, rendering and image processing need to be run on specialized graphics stations,
- packages for digital video processing and multimedia applications do require some additional equipment,
- specific graphics packages do not work properly on networked stations and need to be installed locally.

In addition, the complexity of recent packages means that help needs to be provided even for advanced users. In the case of students from departments where there is no formal graphics teaching, some sort of introduction and supervision is necessary.

We therefore we came to the conclusion that a graphics resource, separate from the general network, with suitable equipment and software, and available for the whole university, should be created.

IV. Graphics Workshop

The Graphics Workshop is intended mainly for those cases where graphics projects can not be completed in the university pooled rooms for reasons explained above. It is also aimed at students and staff from departments where there is a

[1]Leslaw W. Zieleznik is with Computer Services, Learning Resources, Oxford Brookes University, Gipsy Lane, Oxford OX3 0BP, Great Britain, Email: lwz@brookes.ac.uk

lack of graphics facilities and specialized support. In particular, students with research projects and final course projects are encouraged to use the facilities.

Currently the Graphics Workshop can offer services and facilities in the following areas:

- digital video editing
- CAD drawing and rendering 3D graphics for modelling and simulation
- creative graphics and desktop publishing
- multimedia and presentation graphics
- image storage and editing
- CD/DVD writing and authoring.

Present hardware is based on SGI workstations, Apple Macintosh G4 and PCs. In addition, there is equipment for video capture, a high quality flat bed A3 scanner and scanners for 35mm film and transparencies, a digital camera and digital camcorder. Access is provided to colour photo-quality and large format printers. More detailed information can be found on the Workshop web-site [2].

Similar facilities are only available at two other UK universities, i.e. at Manchester University and at Edinburgh University.

V. Conclusion

Since opening just over two years ago, the Graphics Workshop has proved to be a very successful and popular resource.

The majority of students using the workshop come from the Schools of Architecture, Arts, Humanities and Technology. There is also a significant number of students from schools that are not normally associated with computer graphics, such as the Schools of Healthcare and Business.

As far as usage is concerned, a short session typically lasts few hours, but longer projects may take several weeks or more to complete.

The most popular types of sessions are: video editing, preparation of presentation graphics and multimedia, desktop publishing and the rendering of architectural projects. As far as equipment is concerned, the scanners and digital cameras are heavily used.

Looking ahead, as well as keeping the Graphics Workshop technologically up-to-date, we plan to extend and develop its pedagogical provision. The most likely initiatives will be:

- the preparation of materials for a university-wide streaming media service,
- the introduction of further training sessions to include rendering, animation and 3D modelling.

References

[1] Oxford Brookes University, Academic Schools www.brookes.ac.uk/schools/schools.html.

[2] Oxford Brookes University, Graphics Workshop www.brookes.ac.uk/graphics.

[3] Oxford Brookes University, Computer Services www.brookes.ac.uk/services/cs.

Automated Term Time-Table for Universities and its Application in the Intranet

Mariya Nikolova[1] and Mariya Eremieva[2]

Abstract – **The automated term time-table is programmed in Borland Delphi v.5. My SQL and PHP are used for the application of the time-table in the Intranet. The students and professors can read the time-table and make some queries about the disciplines studied; professors' work load etc**

Keywords – **Automated Time-Table, Distance Education, Intranet Application**

I. Introduction

The last and final stage for the organization and carrying out of the study process in every school is the creation of the document "Time-table". It consists of the distributed lessons in each one of the disciplines studied according to the training plain. There are two kinds of time-tables according to the specificity of the educational process in every school:

- weekly (fixed) – when the lessons are planned for the same discipline on a definite day of week and at a definite hour for some study unit (group, class);
- term time-tables – when the time-table is daily for the particular groups and courses during the whole training period (term, semester, etc.).

The daily time-tables are unique and they are applied above all in the higher military schools, where the learning process is conformed to the military training.

When the learning process in the higher schools is being planned, it is possible to separate a block of the professors, as well as a block of the groups or classes of study and a block of the halls where the lessons are carried out. What the stage of the next details is depends on the tasks, which must be realized for a definite process and on the nature of the process itself.

II. Description of the Time-Table Programming

A review of publications [2] has been made about the approaches and methods used for solving the task of making the term time-table of the lessons in higher schools. Based on this review a conclusion has been reached that this task is multicriterial and multidimensional. For that reason the using of clear mathematical models is not appropriate, because the receiving of the optimal solution of the task for compilation of the time-table cannot guaranteed

The right method to use for the task solution is the heuristic method. Practically the time-table is developed in two stages.

The first stage includes the solving of the following basic tasks:

- assignment of the necessary intensity of studying of the different disciplines during the term;
- assignment of the necessary order for carrying out the lessons within the frames of the different disciplines;
- making of the necessary links between the disciplines, and, in some cases, between the separate lessons for different disciplines,
- assigning priorities of the carrying out of lessons. The priorities are algorithmically computed weight coefficients, which determine the order of lessons' distribution in the respective disciplines. The priorities are changed dynamically in the process of lessons' distribution according to the temporary computed results.

In the second stage the whole distribution of the lessons is made by days and weeks within the frames of each week, according to the priorities determined. Here a change of the priorities and an estimation of the resulting variants of the time-table is made.

The task for creating of a term (weekly) time-table belongs to the class "Task for ordering". Meanwhile the task of time-table refers to the time-consuming and difficult to structure multicriterial optimization tasks. To solve it a calculation of many logical checks and work with dynamically changing coefficients and priorities is necessary. Therefore the programming is made in DELPHI [5], which is based on Object Pascal with the goal to increase the speed of the multiple calculations. Another positive characteristic of Delphi is the programming and working with databases. For these reasons a program product named "Automated term time-table" is developed in Delphi. In the process of creating of that product, it was made clear that the names and the codes, given to the professors, the kinds of lessons studied, the training halls and units (groups, subgroups, classes, streams of students) must be included in the local database and the access to it must be realized by an application program. Therefore a Dbase for Windows (local database) – included in Delphi - is used for creating and maintaining the file system. The access to the database is realized by means of BDE (Borland Database Engine), which is composed of libraries (*.DLL) and utilities. Program modules have been developed for maintaining

[1]Mariya Nikolova is with the Department of "Mathematics and Informatics", Naval Academy "N. J. Vaptsarov", 73 Vassil Drumev St., 9026 Varna, Bulgaria, E-mail:mpn@ultranet.bg

[2]Mariya Eremiva is with the Department of "Mathematics and Informatics", Naval Academy "N. J. Vaptsarov", 73 Vassil Drumev St., 9026 Varna, Bulgaria, E-mail:memaer@netbg.com

the database in Object Pascal. These program modules convert the current database of DBF format in Pascal files of type record. These files are inputting for the program, realizing the term time-table. Other program modules are started after finishing the final variant for the lessons' distribution, which convert the files of type record into DBF format These programs are developed especially for using the output data from Delphi in the Intranet, based on Apache server in the University. Delphi, the program product "Automated term time-table and database are installed only on the computer in the department of "Planning of the training process".

To use information for the automated time-table in the Intranet in our University, it is necessary to export the output data from Delphi to the Apache server [7] with Linux operating system, MySQL [8] for operations with database and PHP [9] for creating applications for output queries and processing the data. Such software configuration is used not only in our Naval Academy, but in many Bulgarian Universities because of the free software for Linux, MySQL and PHP, which are downloaded from Internet [7-9].

III. Application of the Term-Table in the Intranet

The process of loading the data for the time-table from Delphi in a relational database maintained under MySQL, passes the following stages:

1. Tables (*.dbf) in format Dbase 5.0 are created in Delphi. The tables contain the data of the time-table for the lessons, the names and codes of the disciplines studied, the data for the professors and the lessons, a calendar for the term's weeks. For this purpose program modules in Object Pascal are developed.
2. Data from the tables are converted in Text Files tab delimited, for filling the data directly in MySQL tables. In the text files the text must not be quoted and the data of type Date must be saved in format YYYY-MM-DD, by default in MySQL. The converting can be processed in MS EXCEL, MS ACCESS or by another program.
3. The structures of tables in MySQL are created on Apache server. These tables have a number of fields and type of data corresponding to the same fields and data types in Delphi. For that purpose it is convenient to use MySQL Control Center– a program product with Graphical User Interface (GUI) for working with MySQL. It can also be downloaded free from the site of MySQL [8].
4. Filling the data in the MySQL tables from the text files is made automatically by using the command mysqlimport, as the name of the text file must be identicall with the name of the table in the MySQL database.

The database created in this way can be used for making queries for reading of the time-table in the Intranet, where the professors' computers and computers in the training halls are connected. Each student, online in Intranet can make a query about his work load in the corresponding training period (week, term); about the place, where the lessons will

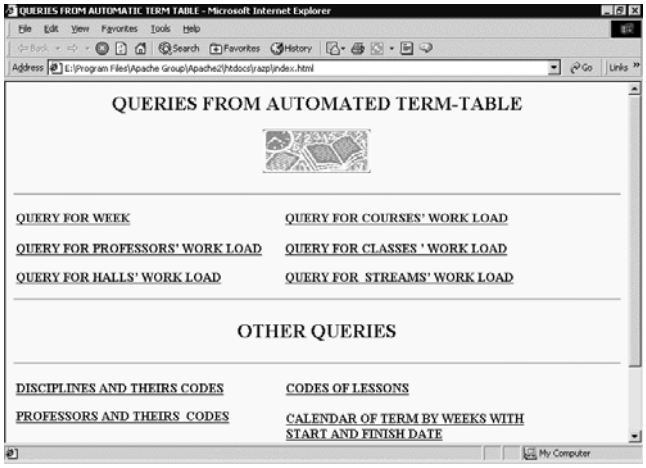
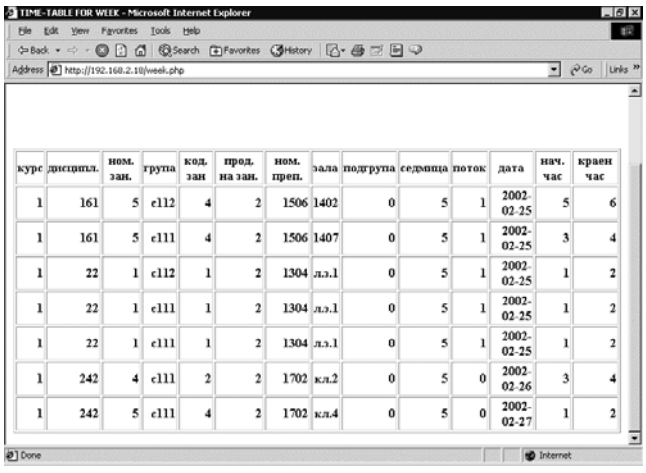


Fig. 1. First Web page of site for the time-table

be carried out (lecture hall, class hall, study room, laboratory), as well as about their form (lecture in a stream, lecture in a class, exercise, laboratory exercise, seminar, exam) and their duration. The professors can also receive information about the students' data (group codes streams, courses, disciplines), form and order of the lessons carried out, their duration, the hall where the lessons are carried out. The queries are summarized in a Web site, whose first page is visualized on Fig. 1.

After pressing on the hyperlink Query for week the Web page-form is opened for choice of week from the term, for which the user will want to read the time table. After choosing the week, a new Web page is started, which visualizes the time-table for the respective week. A time-table for the lessons is shown on Fig. 2, where the 5-th week of the term was chosen. The data are for the summer term of 2002 year.



курс	дисципл.	ном. зан.	група	код. зан	прод. на зан.	ном. преп.	зала	подгрупа	седмица	поток	дата	нач. час	краен час
1	161	5	с112	4	2	1506	1402	0	5	1	2002-02-25	5	6
1	161	5	с111	4	2	1506	1407	0	5	1	2002-02-25	3	4
1	22	1	с112	1	2	1304	л.з.1	0	5	1	2002-02-25	1	2
1	22	1	с111	1	2	1304	л.з.1	0	5	1	2002-02-25	1	2
1	22	1	с111	1	2	1304	л.з.1	0	5	1	2002-02-25	1	2
1	242	4	с111	2	2	1702	кл.2	0	5	0	2002-02-26	3	4
1	242	5	с111	4	2	1702	кл.4	0	5	0	2002-02-27	1	2

Fig. 2. Time-table of lessons for a chosen week of the term

The information of the time-table, printed in the table on Fig. 2 includes:

- number of course – a number from 1 to 5;
- code of the discipline studied – a three digit number, with which each discipline is coded;

- consecutive number of the lesson – a two digit number, with which the consecutive number of the lesson is given according to the training program of the discipline studied in the corresponding specialty;
- code of the group, formed by the trainee students or cadets;
- code of lessons' form – a number, indicating the carrying out of the lecture, exercise, test etc.,
- duration of the lesson– a number, indicating the number of hours for carrying out the lesson;
- code of the professor – a four digit number, which identifies each professor in the University;
- a hall where the lesson will be carried out;
- code of a subgroup of the students – a number, indicating which half of the student's group (visualized in the fourth' column of the table on Fig.!2) will attend the lesson, when it is necessary to divide the group for the exercises;
- consecutive number of the week of the term, for which term-table is output;
- number of stream – a number, indicating the stream of students. The stream is a combination of groups, learning equal training material in separate disciplines;
- data of the lesson;
- starting hour of the lesson – a number from 1 to 10;
- ending hour of the lesson – a number from 1 to 10;

All that data, visualized on Fig. 2 are read on the database of the time-table, placed on the Apache server. The program language PHP [3,4,6] is used for their processing and visualization in the browser. The queries, described below are also output with .PHP files:

Query for professors'work load is started after choosing the respective hyperlink from the initial Web page. A form is opened in which the user has to choose the number of week (optional) and professor, about whose work load the information is demanded. The query includes the same data as shown at Fig. 1, but this time the data refers only to one professor in the chosen week. If no week is chosen, information about the whole term work load of the respective professor is visualized.

Query for halls'work load shows the number of the hall, the date, the starting and ending hour of the lessons in the hall. The user can also enter the number of week in which he chooses to find information about the halls' work load.

Query For Courses' Work Load, Query For Classes' Work Load and Query For Streams' Work Load are constructed in the same way. The first query visualizes the work load of the respective course, the user choosing the number of the course. The second query shows on the display the work load of the group, chosen by the user from the form. The query for the streams workload includes the data from the work load connected only with the respective stream, chosen by the user. The number of week the information is about can be chosen in all three queries.

The rest of the queries summon data from tables, including information about:

- codes and names of disciplines;
- codes of professors and their names;
- codes and kind of lessons;
- calendar of term by weeks with a beginning and ending date – a table including the number of week initial and final date of the week and the number of days, it consists of.

No corrections in the database is allowed in thus created Web site, although this is possible by means of PHP. The reason why this option is not included is that, once created by means of Delphi, the work load can be subject to small changes. They are connected with preplanning of lessons, because of holidays or accidental absence of some professor. Such changes can easily be realized in the MySQL Control Center by an employee in the lesson planning department. A second programming by means of Delphi of the preplanned work load for small changes is not advisable, because this can change the organization of the planning process, which can lead to a non-optimal variant of the time-table.

Presently the Web site is located on Intranet Web server, but its transfer to an Internet Web server is no problem. In this way the time-table information can be received by the home computer of the users. The software product does not demand particularly powerful computers for using. The presence of a browser (Internet Explorer, Netscape etc.) is sufficient.

All forms and the initial Web page have been programmed on HTML v. 4.0 [1]. All queries using information from the database can be programmed on PHP making use the functions it supports to work with database and MySQL [3,4,6].

IV. Conclusion

From the started above the following conclusions can be made:

- using Delphi to program lesson planning is appropriate because the time-consuming computations in the process of designing a term time-table and the presence in Delphi of opportunities to create and maintain databases, the process of their processing being carrying out on a high-level computer language – Object Pascal;
- the Web site created for getting information about the time-table in Intranet is convenient because the time for reading the time-table can be decreased compared to the paper-based information used so far;
- the queries realized assist professors and students in reading their work load;
- students and cadets find it easy to spot the professor by viewing his work load. This is particularly convenient for students, living outside Varna.

The access to the Web site in the Internet together with other sites used for distance learning in which there is a hyperlink for reading the lesson time-table is convenient for extra-mural students, travelling on ships all over the world.

References

[1] Denis Tailor, Microsoft Front Page 2000, vol. 1 and 2, InfoDAR, Bulgaria, 1999.

[2] Eremieva Mariya, The comparative analysis of automatic systems for building of the term time-tables, Marine Scientific Conference, vol. 4. Mechanical engineering and mathematics. Information technology, Naval Academy "N. J. Vaptsarov", pp. 277-280, Varna,. Bulgaria, 2001.

[3] Jay Greenspan, Brad Bulgar, MySQL/PHP database applications, AlexSoft, Bulgaria, 2001.

[4] Jesus Castanetto, Harish Rawat, Sasha Schuman, Chris Scollo, Deepak Veliath, Professional PHP Programming, SoftPress, Bulgaria, 2001.

[5] Kent Risdorph, Borland Delphi 4 in 21 Days, vol. 1 and 2, InfoDAR, Bulgaria, 1999.

[6] PHP 4 Bible, AlexSoft, Bulgaria, 2000.

[7] www.apache.org

[8] www.mysql.com

[9] www.php.net

Information System for Preliminary Testing

Rumen I. Arnaudov[1], Ivo N. Dochev[2]and Milena A. Atanasova[3]

Abstract – **Contemporary education methods are more and more frequently employed in education. One of these methods is distant learning (e-learning). Its high efficiency rate, regardless of the distance between the teacher and the students is one of its main advantages. In the following article the program for preliminary testing and evaluating employed in the "Telecommunications measurements" course in the Technical University of Sofia is described. The program is based on a Linux operating system, Apache web server, MySQL database, HTML, PHP and JavaScript languages.**

Keywords – **Education, Distance learning education, Pretest.**

I. Introduction

Contemporary education methods are more and more frequently employed in education. One of these methods is distant learning (e-learning) [3-11]. A part of that education is the preliminary examination. It aims to check the students' knowledge of theory before having a lab session. The key advantage of that kind of examination is the unbiased evaluation and the speed of data processing.

II. Architecture

The software employed in the system in question is Linux OS [12], Apache web server [16], MySQL database [18], HTML, PHP [17] and JavaScript languages. The architecture of the computer and software configuration is displayed in fig.1. It consists of: users, Internet, local area network (LAN), web server, database (DB), firewall, HTML pages, PHP and JavaScript (JS) modules.

The users connect to the web server through the local network or the Internet. A firewall is used to restrict access to services that are to be used only in some of the laboratories. This aims both to protect information and maintain a more veritable record of the data received from laboratories. In the database are kept the usernames, the respective passwords, the test questions, the correct answers and the scores received from the test. The test questions are displayed in HTML pages, as for every question there are a number of answers with only one of them being correct. The JavaScript module is used to build a real-time countdown clock, used to display the time remaining to finish the test; the usage of JavaScript allows avoidance of network slowdowns at the initial starting of the clock. The PHP module serves as the link between the database and the HTML pages, as well as the test results processing. The use of a PHP module does not allow the users to view the source code of the application.

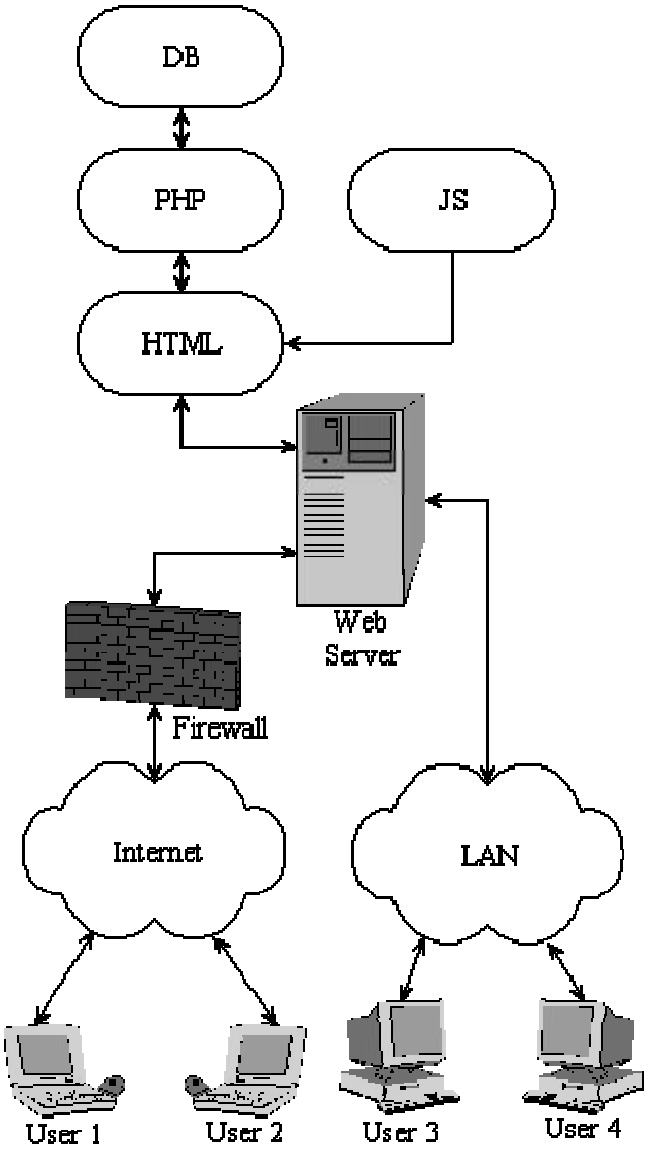


Fig. 1. The architecture of the computer and software configuration.

III. Algorithm

The algorithm of using the application is shown in Fig. 2. The link to the Web page of the "Measurements in Com-

[1]Rumen I. Arnaudov is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ra@vmei.acad.bg

[2]Ivo N. Dochev is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: idochev@vmei.acad.bg

[3]Milena A. Atanasova is with the Faculty of Communicationsand Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: alfirin@mail.bg

munications" course is through http://mc.vmei.acad.bg . The home page of the application is built by two main frames: a navigation frame (1) and information frame (2) (Fig. 3). The navigation frame contains the following links: Home, Introduction, Registration, Labs and Gradebook. The Information frame contains information for the selected link in the Navigation frame.

In order to access the application's resources the students must initially register as users. This can be accomplished through the "Registration" navigation menu (Fig. 4). The following information is required:

- faculty number
- full name
- study group
- password

After clicking on the "Register" button the data filled in is verified. Should there be any incorrect piece of data, an error message is displayed and the registration process must be restarted.

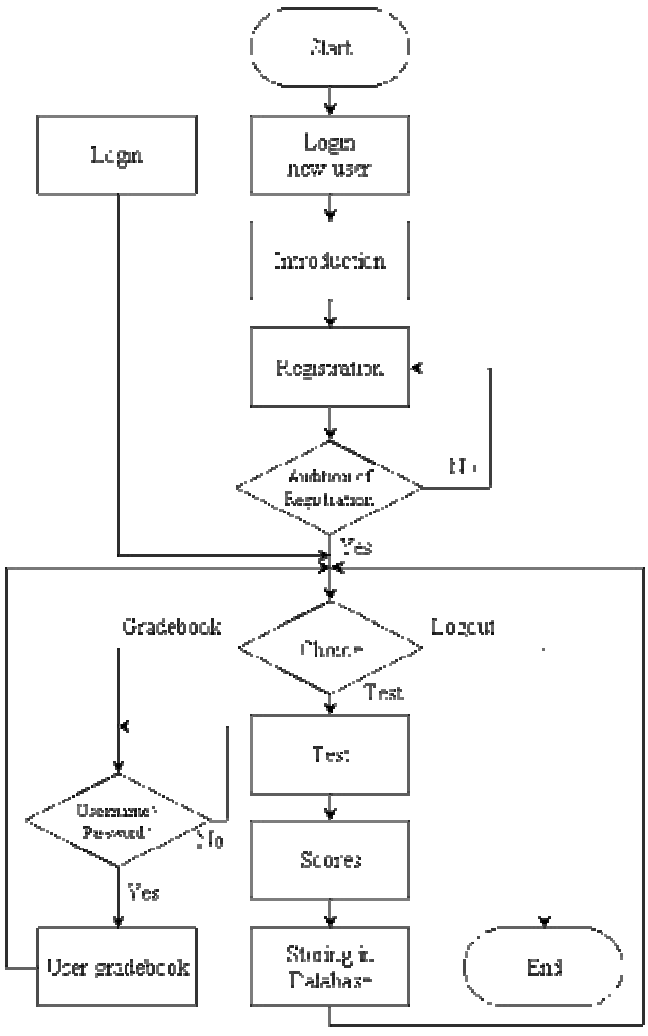


Fig. 2. The algorithm of using the application.

Filling in a preliminary laboratory test is accomplished by selecting the desired lab practice from the navigation menu and by clicking on "Test" on the Information frame(Fig. 5).

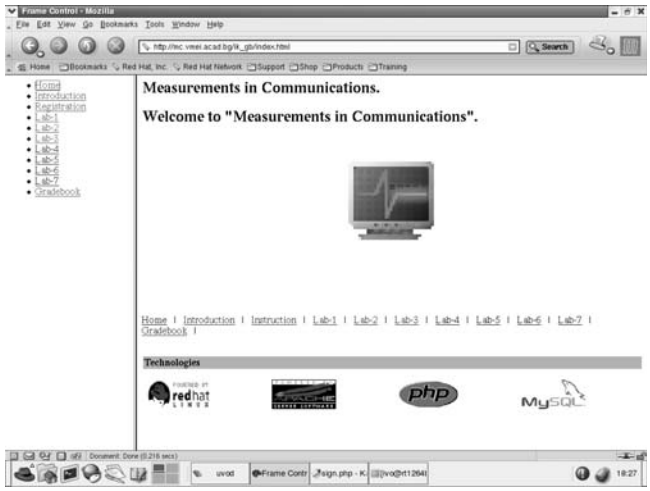


Fig. 3. The home page of the application.

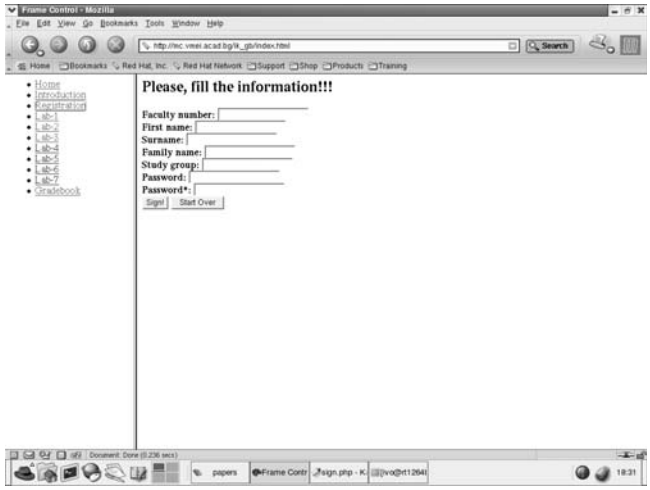


Fig. 4. The registration form.

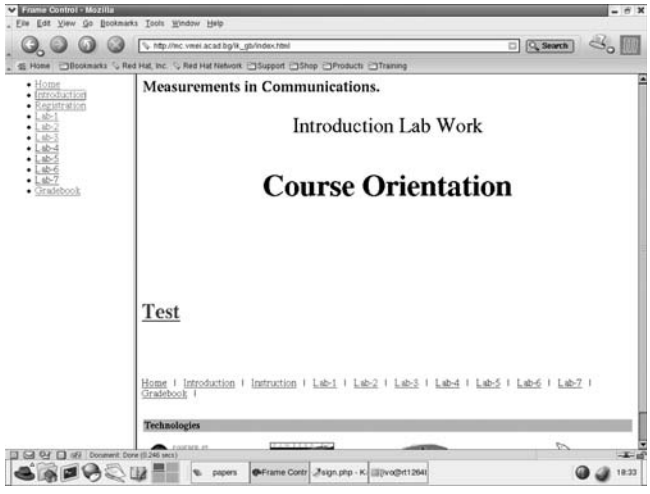


Fig. 5. The laboratory work home page.

After logging in (with faculty number and password Fig. 6) the student is admitted to the questions of the test. The Information frame of the test consists of two parts (Fig. 7):

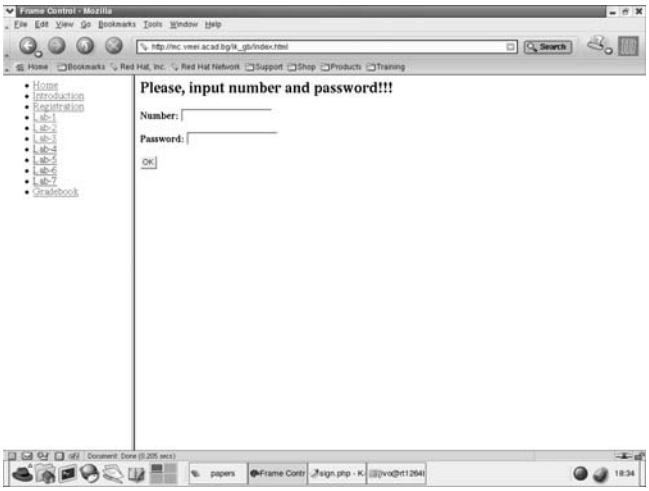


Fig. 6. The test logging registration form.

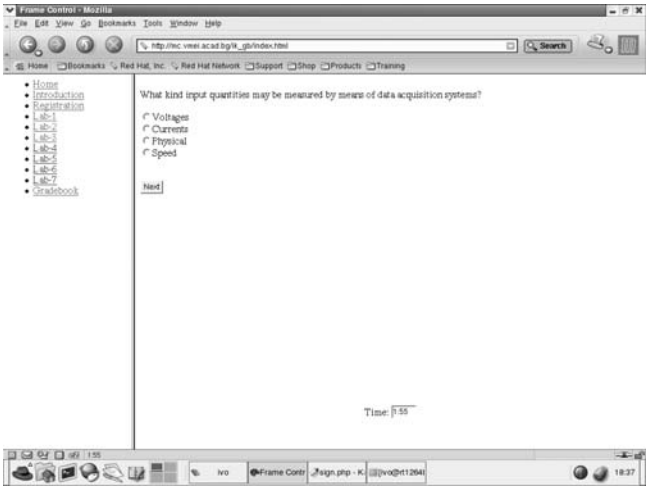


Fig. 7. The information frame of the test consists.

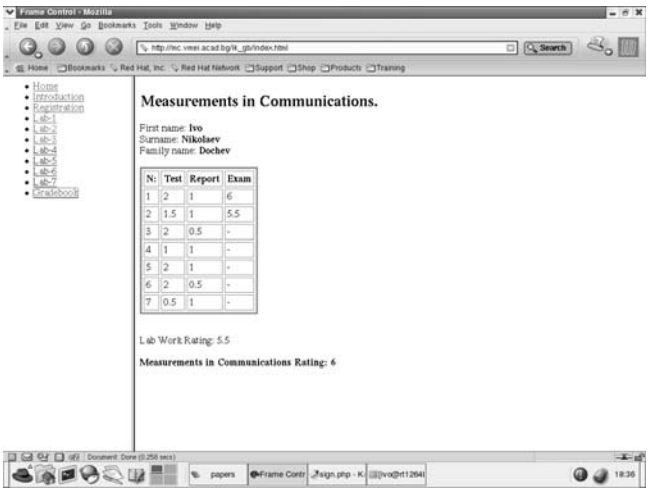


Fig. 8. The Gradebook.

questions to be answered and timer. For every question there are four possible answers, of which only one is correct. The timer shows the time left for answering all questions in the test. If the student does not finish the test in the obligatory time limit, he or she gets no points for the lab practice.

Checks of received scores can be made through the ”Gradebook” link on the Navigation frame (Fig. 8). The Information ”Gradebook” frame contains the following information:

- student’s full name
- grades received on preliminary tests
- grades received from lab practice
- grades received from in-term tests
- Final Lab Practice Grade
- Final ”Communications Measurement” Course Grade

IV. Conclusion

In this article we presented an information system for preliminary testing and examination used in the ”Measurements in Communication” course held in the Technical University of Sofia. The computer architecture and the software were presented in detail – the software based on Linux operating system, Apache web server, MySql database, HTML, PHP and JavaScript languages.

The system benefits by:

- offering a lower price: Linux operation system, Apache web server, MySql database, HTML, PHP and JavaScript languages are all open-source software products and therefore for free distribution
- unbiased evaluation: all students are evaluated on identical criteria
- swift data processing and grading: the received results are automatically processed by the software application
- access through the Internet to the application’s resources: every user can obtain information on the grades on lab practice and displayed skills throughout the course.
- Protection of the source code: by using a PHP module the source code of the application is unable to be viewed by the users.
- Different Web browsers to link the Web page of the course may by used: the software program has been test bay Internet Explorer Browser, Netscape Navigator, Mozilla Web Browser, Konqeror Web Browser, Galeon and Opera.

References

[1] J. Greenspan, B. Bulger, *MySQL/PHP.* IDG Books Worldwide Inc., Foster City, California, USA, 2001.

[2] O. Kirch, T. Dawson, *Linux Network Administrator's Guide*, Second Edition. O'Reilly & Associates, Inc, 2000.

[3] J. Pullen, "Applicability of internet video in distance education for engineering", *Frontiers in Education Conference, 2001* . 31st Annual, Volume: 1 , 2001, pp: T2F -14-19 vol.1

[4] A. Bittencourt,D. Carr, "A method for asynchronous, web-based lecture delivery", *Frontiers in Education Conference, 2001*. 31st Annual, 2001, pp: F2F -12-17 vol.2

[5] M. Castro, A. Hilario, S. Acha, J Perez. A. Colmenar, P. Losada, I. Rivilla, J. Peire, "Multimedia design and development for distance teaching of electronics", *Frontiers in Education Conference, 2001*. 31st Annual, 2001, pp: F3F -13-18 vol.2

[6] W. Braga, "A general methodology for engineering education using the Internet", *Frontiers in Education Conference, 2001*. 31st Annual , 2001, pp: F1F -1-5 vol.2

[7] F. King, H. Mayall, "Asynchronous distributed problem-based learning", *Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference on , 2001*, pp: 157 –159

[8] T. Eriksson, A. Goller, S. Muchin, "A comparison of online communication in distance education and in conventional education", *Frontiers in Education Conference, 2001*. 31st Annual, Volume: 1, 2001, pp: T2F -20-5 vol.1

[9] K. Baas, J. Van den Eijnde, J. Junger," A practical model for the development of Web based interactive courses", *Frontiers in Education Conference, 2001*. 31st Annual, Volume: 1 , 2001 pp: T2F -8-T2F-13 vol.1

[10] A. Striegel, "Distance education and its impact on computer engineering laboratories", *Frontiers in Education Conference, 2001*. 31st Annual, 2001, Page(s): F2D -4-9 vol.2

[11] http://cisco.netacad.net

[12] http://www.linux.org

[13] http://linuxdoc.org

[14] http://www.linux.org/docs/index.html

[15] http://www.redhat.com

[16] http://www.apache.org

[17] http://www.php.net

[18] http://www.mysql.com

Practical Analysis of Signals with Amplitude Modulation

Emil S. Simeonov[1], Veska M. Georgieva[2] and Dimiter C. Dimitrov[3]

Abstract – **A practical exercise for presentation and analysis of signals with different kind of amplitude modulation is described in the paper. A new method for interactive simulation of processes is proposed. The method can be developed and the described exercise can be used for web based distance education.**

Keywords – **signal processing, amplitude modulation, computer simulation, distance education.**

I. Introduction

Modulation is a Physical process, in which the spectrum of a low frequency signal i.e. the carrier of some information, is transferred in the high frequency domain. This is a method of long-distance signal transmitting, whose basics are going to be represented.

One of the parameters $(a_1, a_2, a_3, ..., a_n)$ of a high frequency signal $a(t) = f(a_1, a_2, a_3, ..., a_n)$, called 'Carrier', synchronized with a low frequency signal $s(t)$, that is the carrier of information, is varied in time. As a result a high frequency signal, called 'Modulated Signal' is outputted, which possesses a qualitatively new property. It gets the information that is previously carried by the signal $s(t)$. Usually as a Carrier is used harmonic tremble of this kind:

$$a(t) = A_0 \cos(\omega_0 t + \varphi_0) \,, \tag{1}$$

If just its amplitude varies as time passes, an 'Amplitude-modulated signal' is got, whose mathematical model is as it follows

$$a(t) = A_m \cos(\Omega t + \varphi_\Omega) \,, \tag{2}$$

where the quantity 'm' is called 'Coefficient of modulation'.

In case the low frequency signal (called 'Modulating signal', as well) is a harmonic tremble of this kind:

$$s(t) = A_m \cos(\Omega t + \varphi_\Omega) \,, \tag{3}$$

monotone amplitude-modulated signal is got, described with the following expression:

$$a(t) = A_0[1 + m\cos(\omega_0 t + \varphi_0)]\cos(\omega_0 t + \varphi_0) \,, \tag{4}$$

and the coefficient of amplitude modulation is

$$m = \frac{A_m}{A_0} = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}} \,. \tag{5}$$

Generally this coefficient is taken separately for both positive and negative semi-waves. When the coefficient of amplitude modulation is a small value, the relative alternation of the wrapping curve is small $|ms(t)| << 1$. On condition that $|ms(t)| \approx 1$, the process is called 'Deep amplitude modulation'.

After appropriate mathematical transformations the signal presented with (4) could be represented as a sum of harmonic trembles.

$$a_{AM}(t) = A_0 \cos(\omega_0 t + \varphi_0) + \frac{mA_0}{2} \cos[(\omega_0 - \Omega)t + \varphi_0 - \varphi_\Omega] + \frac{mA_0}{2} \cos[(\omega_0 + \Omega)t + \varphi_0 + \varphi_\Omega] \,. \tag{6}$$

These trembles compose the spectrum of the amplitude-modulated signal, whose amplitude-frequency spectrum density diagram is shown on Fig. 1.

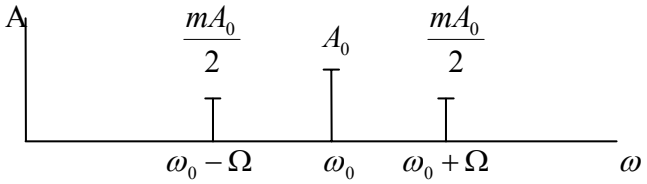


Fig. 1.

The spectrum contains a compound, which frequency equals this of the Carrier signal, and two side compounds with frequencies $(\omega_0 - \Omega)$ and$(\omega_0 + \Omega)$. When the modulating signal is a complex tremble, the spectrum contains as well as the harmonic, that equals the frequency of the Carrier, two side bands of spectrum compounds, as well. The relevant amplitude-frequency spectrum density diagram is shown on Fig. 2.

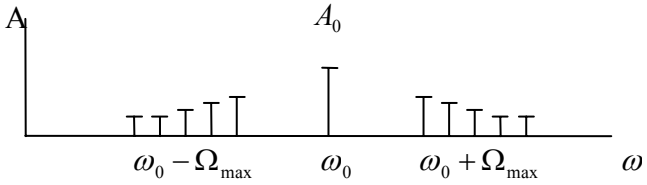


Fig. 2.

The average power of an amplitude-modulated tremble for one period of the Carrier signal is defined by the following expression

$$P_{T_\omega} = \frac{A_0^2}{2}[1 + \cos(\Omega t + \varphi_\Omega)]^2 = P_0[1 + \cos(\Omega t + \varphi_\Omega)]^2 . \tag{7}$$

In case that a modulating signal is absent the power P_0 could be found through (7), when $m = 0$. When $m = 1$ and $\cos(\Omega t + \varphi_\Omega) = 1$ the power is maximal and if $m = 1$ and $\cos(\Omega t + \varphi_\Omega) = -1$, its value is minimal. The average

[1]Emil S. Simeonov is with the Faculty of Computer Systems and Control, TU-Sofia, Kl.Ohridsky str.8, Sofia, Bulgaria, E-mail: zmeibi@vip.bg

[2]Veska M. Georgieva is with the Faculty of Communication, TU-Sofia, Kl.Ohridsky str.8, Sofia, Bulgaria, E-mail: vesg@vmei.acad.bg

[3]Dimiter C. Dimitrov is with the Faculty of Communication, TU-Sofia, Kl.Ohridsky str.8, Sofia, Bulgaria, E-mail: dcd@vmei.acad.bg

power of the radio-signal for one period of the modulating signal is defined by the expression:

$$P_{T_\Omega} = \frac{A_0^2}{2}\left(1+\frac{m^2}{2}\right) = P_0\left(1+\frac{m^2}{2}\right). \quad (8)$$

The method of amplitude modulation has some serious disadvantages in respect of Energy. It is clear that even in a work-mode, when a modulating signal is absent, power is radiated, which is the main cause of the low efficiency of amplitude-modulating devices.

For reducing the energy losses of amplitude modulation, the so-called, 'Balance Amplitude Modulation' (BAM) is utilized. When BAM is used, as it is shown on Fig. 3, the compound, whose frequency equals this of the Carrier, is eliminated. This way 'Null Power' is ensured in a work-

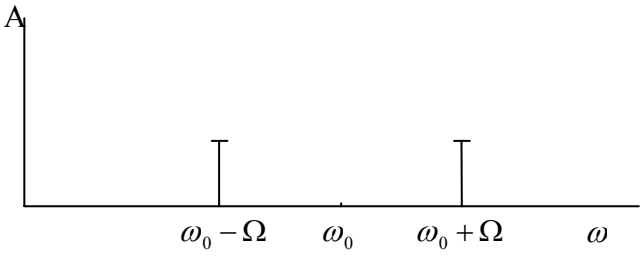


Fig. 3.

mode, when a modulating signal is absent. BAM is rarely used in practice, mainly because of complicated technical problems with receiving devices, as the Carrier has to be restored.

Based on the expression (4), the amplitude-modulated signal could be represented through BAM, where the modulating signal is a harmonic low frequency signal:

$$\begin{aligned} a_{BM}(t) &= mA_0\cos(\Omega t+\varphi_\Omega)\cos(\omega_0 t+\varphi_0) = \\ &= \frac{mA_0}{2}\cos[(\omega_0-\Omega)t+\varphi_0-\varphi_\Omega] \\ &\quad + \frac{mA_0}{2}\cos[(\omega_0+\Omega)t+\varphi_0+\varphi_\Omega]. \end{aligned} \quad (9)$$

Another way of long-distance signal transmitting is the method of 'Amplitude manipulation' (AMn). Manipulated signals distinguish by the fact that their Carriers are continuous ones, but modulating signals are discrete, mainly square pulses. The amplitude of a Carrier varies curtly according to changes in the manipulating signal, so that the manipulated signal is better secured from noise and any kind of interference than amplitude-modulated ones. An amplitude-manipulated signal could be presented with the following expression:

$$a_{AMn}(t) = a_m(t)\cos\omega_0 t,\ a_m(t) = \begin{cases} A_0, & \text{binary } '1' \\ 0, & \text{binary } '0' \end{cases}. \quad (10)$$

II. Practical Analysis of AM, BAM and AMn signal

The results of AM, BAM and AMn could be observed and analyzed in details through the visual tool 'Simulink', which is a part of the programming environment of the 'Matlab' package. 'Simulink' is an outstanding application that implements lots of Matlab's built-in functions, so that their use is easy and powerful and, furthermore, models developed through it resemble real engineering solutions. This will become clear when the structure and the user-interface of the models of AM, BAM and AMn, are discussed. Each of the presented models uses strongly simplified and standard user-interface, which provides full control over the features of real-time simulations. One of the models is presented on Fig. 4.

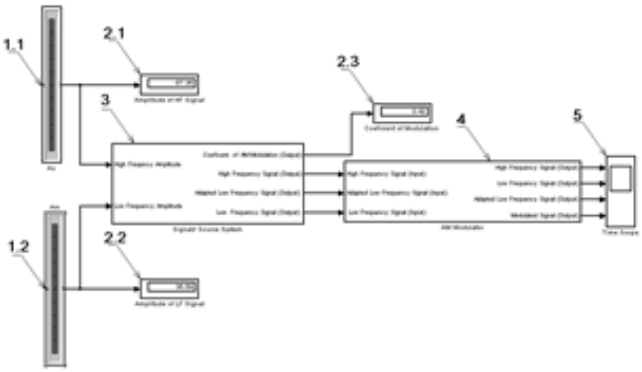


Fig. 4.

One could use such a model to thoroughly analyze Physical processes bonded with these sorts of modulation of continuous signals. An example list of some of the possible tasks of an exercise is presented and detailed explanations are available, as well.

III. Task of the Exercise

1. The process of AM has to be simulated, as the parameters of the Carrier, the Modulating signals and the coefficient of modulation are properly selected. Observe and analyse the graphics of several cases, when $m < 1$.
2. Simulate the processes of Deep modulation and Over modulation and observe the corresponding graphics.
3. Simulate the processes of Balance Amplitude Modulation (BAM) and Amplitude Manipulation (AMn).
4. Observe and analyse in details the Power Spectral Density Diagrams, provided by the PSDAs, for both the modulated/ manipulated and the input signals.

IV. Practical Guide for the Exercize

At first a 'Simulation' has to be run through the 'Play' button, positioned on the main toolbar.

The values of amplitudes of the input signals are set, while a real-time simulation is running, through two ActiveX components (components 1.1 and 1.2 onto Fig. 4), i.e. sliders with discrete range from 0.01 to 100 and step of 0.01, which, practically, means that these amplitudes vary from 0.01 to 100.

Their values could be observed on the digital control displays 2.1 and 2.2 on Fig. 4. Thus both input signals and the coefficient of modulation (estimated in reference to equation (5)), whose value is displayed by the component 2.3 on the same figure, are precisely managed.

The 'Signals' Source System', i.e. component 3 on Fig. 4, generates both the Carrier and the Modulating signal, while doing some minor calculations. For example, it estimates the coefficient of modulation. Inspired by the Object-Oriented Programming (OOP), 'Simulink' makes it possible to create complex structures and hierarchies through its components as it is clear from Fig. 5, where the internal structure of the system 'Signal Source' is presented.

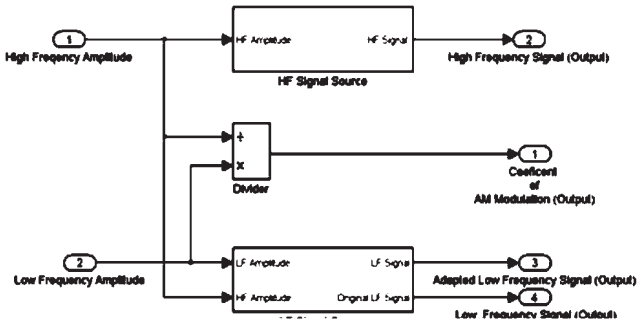


Fig. 5.

An option changing the sort of the continuous input signals (sine, square pulses, saw tooth and random signals are available) and their frequencies is easily found in the subsystems 'Source of the LF signal' and 'Source of the HF signal' (Fig. 6).

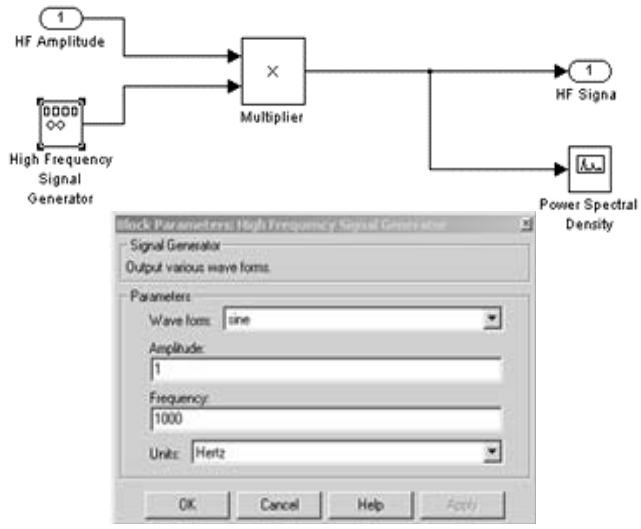
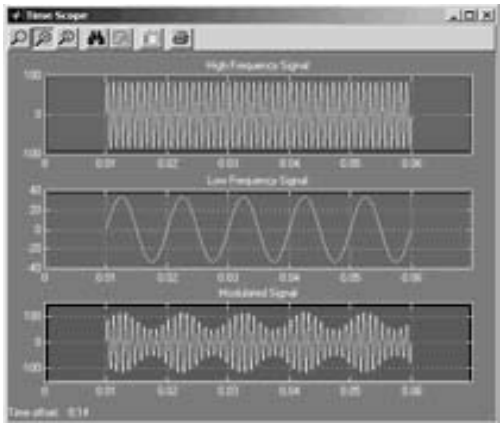


Fig. 6.

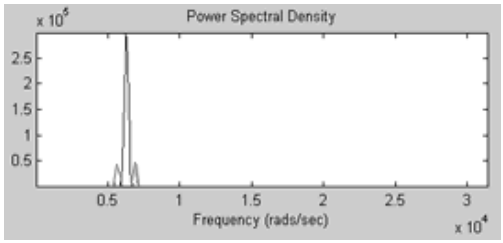
The 'Modulating System', which is component 4 on Fig. 4, modulates/manipulates input signals using the correspon ing expressions for AM, BAM and AMn: (4), (9) and (1(It also passes input and modulated signals to the Time sco (component 5 on Fig. 4), through which they could be observed and analyzed in the time domain. Besides through the three Power Spectral Density Analyzers (PSDAs), positioned within the 'Signal Source' and the 'Modulating System', analyses about the properties of all the signals in the frequency domain could be worked out.

Fig. 7 demonstrates some results of AM, BAM and AMn in the time and frequency domains.

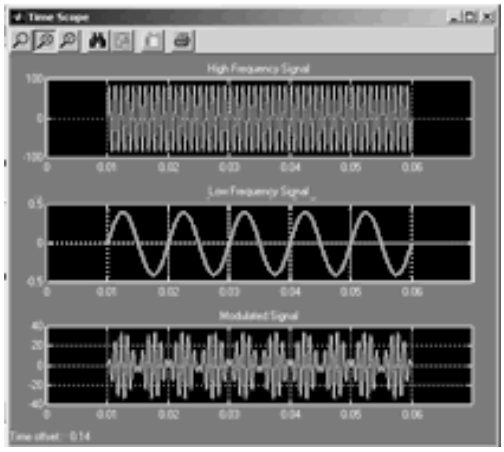
Apparently, having known all about input signals and the modulated signal, i.e. their form, spectrum and the influence of the coefficient of modulation, all the cases from the exam-



AMPLITUDE MODULATION
(TIME DOMAIN, PRESENTED THROUGH A TIME SCOPE)

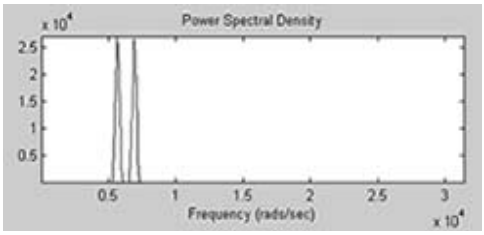


AMPLITUDE MODULATION
(FREQUENCY DOMAIN, PRESENTED THROUGH A PSDA)

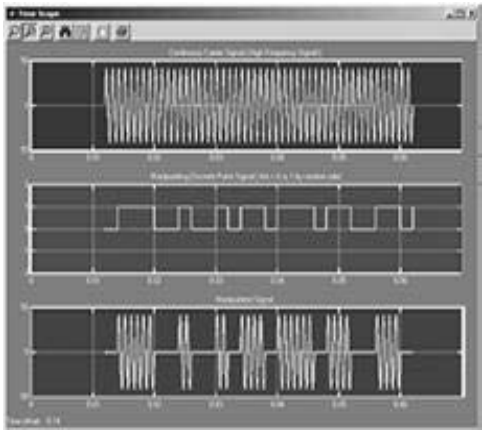


BALANCE AMPLITUDE MODULATION
(TIME DOMAIN, PRESENTED THROUGH A TIME SCOPE)

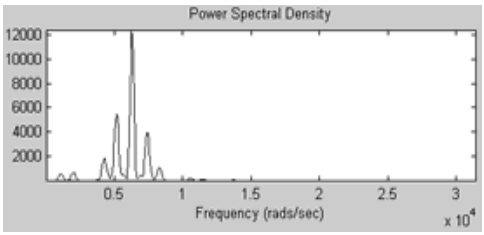
Fig. 7.



BALANCE AMPLITUDE MODULATION
(FREQUENCY DOMAIN, PRESENTED THROUGH A PSDA)



AMPLITUDE MANIPULATION
(TIME DOMAIN PRESENTED, THROUGH A TIME SCOPE)



AMPLITUDE MANIPULATION
(FREQUENCY DOMAIN, PRESENTED THROUGH A PSDA)

ple task list could be easily analyzed. Each of the models, developed through 'Simulink' for this exercise gives the same readiness in use, combined with powerful flexibility in managing with the signals used. To facilitate users all the typical cases, in which the Modulating signal might be sinusoidal, square impulses, saw tooth and even random, have been developed, as well, so that switching to different input signals is as easy as loading an ordinary file in an arbitrary Win32 application.

V. Conclusion

- As the simulations are real-time, when some basic parameters of input signals are altered (the amplitude, the frequency or the form), while a simulation is running, the changes of signals in the time and frequency domain could be observed and analyzed indetails.
- The models are suitable both for the process of searching and developing better and easier to construct real engineering solutions and for education in the field of Signals (Telecommunications, DSP, LAN, WAN etc.).
- Like in the OOP one interface might implement different and complex actions, which means standardization.
- The 'user-friendly' interface, which is one and the same for all the built models, gives users the opportunity to do their job without any preceding training.
- The small size of the model files makes them easily publicized on remote servers and, consequently, they are suitable for distant educational purposes.

References

[1] Dimitrov D., Kamenov C., Georgieva V., Signals and Systems Techniques Laboratory Work, 2001.
[2] Ferdinandov E., Signals and Systems, 2000.
[3] Nenov G., Signals and Systems, 1999.

Methods for 2D Transformations in Modeling of Real Medical Signals

Penka Tz. Isaeva[1]

Abstract – **The types of transformations and their combinations in modeling of real medical signals are given in this article.**

Keywords – **Medical images , 2D medical signals , Education concerning treatment of medical signals , Distance education.**

I. Introduction

The creation of real 2D and 3D medical images is a combination of two fundamental processes – modelling (defining geometry of medical objects) and representing (displaying) these mathematical models. In order to define geometry of medical signals and objects some polygons, curves and surfaces have been used. The process of modelling is connected with thousands of transformations of coordinates of the points, which define relevant images or signals.

II. Types of Transformations

In modelling, a lot of applications have to alter or manipulate medical signals and images by changing their size, position or orientation. Therefore, the basic types transformations are translation, scaling, rotation and shearing.

When it is necessary to translate medical signals, set of points must be shifted in new position. If x and y are coordinates of a point $P(x, y)$, the coordinates of the new point $P'(x', y')$ after the translation are given by:

$$x' = x + t_x$$
$$y' = y + t_y \ , \qquad (1)$$

where t_x and t_y have constant value. This is shown in Fig. 1:

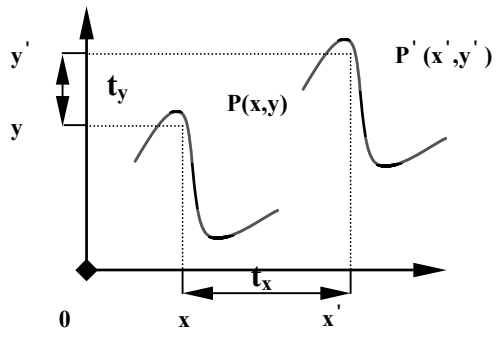


Fig. 1. Translation of signals

The scaling is used to change the size of signals and images. The scaling about origin (0, 0) is multiplication of the coordinates x and y of points respectively with scale factors S_x and S_y.

The coordinates of the new point after that transformation are:

$$x' = x.S_x$$
$$y' = y.S_y \qquad (2)$$

This is shown in Fig. 2:

The possible cases of scaling are two:

1) $S_x = S_y$ – case of symmetric scaling transformation – the size of signals or images changes equally in x and y axes. If $|S_x| > 1$ and $|S_y| > 1$, the size of signals increases. If $|S_x| < 1$ and $|S_y| < 1$, the size reduces.

2) $S_x \neq S_y$ - case of asymmetric scaling transformation. If $S_x < 0$, the medical images reflect in the y axis. If $S_y < 0$ – the signals reflect in the x axis.

The rotation has been used to orientate medical signals and images in 2D plane, as they have been turning on angle α. The coordinates of the point $P(x, y)$, which belongs to the medical signal, are:

$$x = R.\cos\beta$$
$$y = R.\sin\beta \ , \qquad (3)$$

where R is the length of the line, connecting point $P(x, y)$ and point (0, 0). After the rotation the new coordinates of point $P'(x', y')$ are:

$$x' = R.\cos(\alpha + \beta) = x.\cos\alpha - y.\sin\alpha$$
$$y' = R.\sin(\alpha + \beta) = x.\sin\alpha + y.\cos\alpha \ , \qquad (4)$$

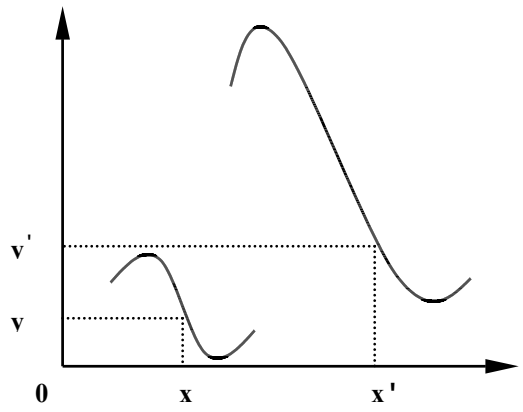


Fig. 2. Scaling

[1]Penka Tz. Isaeva is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia Bulgaria, E-mail: pepi_is@abv.bg

where β is the angle between x axis and the line R.

The rotation is displayed in Fig. 3:

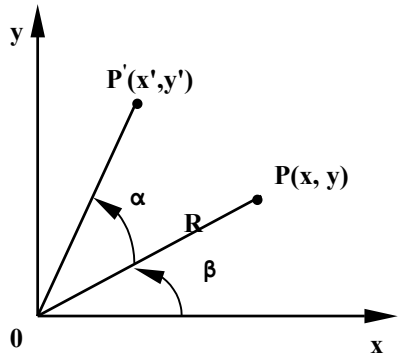


Fig. 3. Rotation

The shear transformation is a type of transformation that has the effect of distorting the signals and images in medicine. The coordinates of the new points, after the shearing, are:

$$x' = x + y.a$$
$$y' = y + x.b\ . \tag{5}$$

If $a \neq 0$ then shear transformation in x axis appears and if $b \neq 0$ then shear of medical signals in y axis appears.

III. Matrices of the 2D Transformations

In the common case, the coordinates of the new point after transformation are:

$$x' = a.x + b.y + c$$
$$y' = d.x + e.y + f\ , \tag{6}$$

a, b, c, d, e, f are constants, x' and y' are linear function of x and y. The matrix forms of the above mentioned expressions are :

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ d & e \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c \\ f \end{bmatrix} \tag{7}$$

or

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}\ . \tag{8}$$

The matrices of the constants a, b, c, d, e, f (for each transformation type) are:

$$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix} \text{ – for translation ,}$$
$$\begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \end{bmatrix} \text{ – for scaling ,}$$
$$\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \end{bmatrix} \text{ – for rotation ,} \tag{9}$$
$$\begin{bmatrix} 1 & a & 0 \\ b & 1 & 0 \end{bmatrix} \text{ – for shearing .}$$

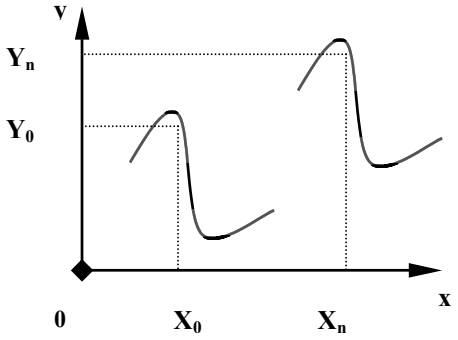


Fig. 4. Object transformation – shift object by (dx, dy)

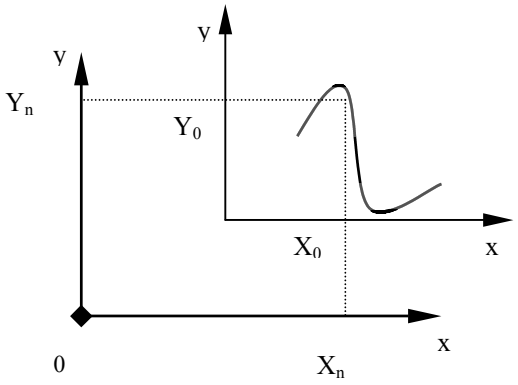


Fig. 5. Axis transformation – shift axes by $(-dx, -dy)$

IV. Object and Axis Transformation of the Medical Signals and Images

The transformations, which have been discussed in the above sections, are object transformations. The object transformation of medical signals is a method, where the signal is transforming, while the axes x and y staying fixed. Another method for transformation of medical signals and images is axis transformations, where the signal has remained fixed, while the axes have been changing. In the first case the set of points, that forms the signal, is shifted by (dx, dy). The transformed points are plotted relative to the same set of axes. The second case shows that the axes are shifted by $(-dx, -dy)$. The points on the medical signal are fixed in the space, but they are shifted by the new axes. These facts can be seen in the next figures – Fig. 4 and Fig. 5. After object transformation coordinates of the new point (x_n, y_n) are:

$$x_n = x_0 + dx$$
$$y_n = y_0 + dy\ . \tag{10}$$

The new coordinates of the same point don't change after axis transformation by $(-dx, -dy)$. Therefore axis transformation is equivalent to an equal and opposite object transformation. This conclusion is valid for all types of transformation. Hence an object scaling transformation with coefficients S_x and S_y is equivalent respectively to an axis scaling transformation with coefficients $(1/S_x)$ and $(1/S_y)$. An object rotation with angle (α) is equivalent to an axis rotation with angle $(-\alpha)$.

The matrices for object and axis transformations of the

medical signals are:

$$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix} \text{ – translate object by } (t_x, t_y) \text{ ;}$$

$$\begin{bmatrix} 1 & 0 & -t_x \\ 0 & 1 & -t_y \end{bmatrix} \text{ – translate } (t_x, t_y) \text{ ;}$$

$$\begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \end{bmatrix} \text{ – scale object by } (S_x, S_y) \text{ ;}$$

$$\begin{bmatrix} 1/S_x & 0 & 0 \\ 0 & 1/S_y & 0 \end{bmatrix} \text{ – scale } (S_x, S_y) \text{ ;} \qquad (11)$$

$$\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \end{bmatrix} \text{ – rotate object by } \alpha \text{ ;}$$

$$\begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \end{bmatrix} \text{ – rotate axes by } -\alpha \text{ .}$$

As a common rule, the inverse of an object transformation is the corresponding axis transformation.

V. Conclusion

The main methods for transformation of points of 2D medical signals and images are translation, scale, rotation and shear. The medical transformations are divided into two groups – object transformations, where the points of signals are transformed, and axes transformations, where the coordinate axes are transformed and the signal points re-expressed relatively to the new axes.

Methods for 2D transformations can be used as in real applications, for observation, research, computation of particular tasks, also for education concerning of treatment medical signals and images.

References

[1] Hearn D. and Baker M. P. , Computer Graphics, Prentice – Hall, 1986.

[2] Blinn J. F. and Newell M. E. , Clipping Using Homogeneous Coordinates, Proceedings of SIGGRAPH , pp. 245 – 251.

[3] Foley J. D. , van Dam A. , Feiner S. K. and Hughes J. F. , Computer Graphics : Principles and Practice (2nd Edition), Addison Wesley, 1990.

Concatenation and Combination of Transformations of 2D Medical Signals

Penka Tz.Isaeva[1]

Abstract – **In this paper the concatenation and the combination between the basic types of transformations of 2D medical signals are given.**

Keywords – **Medical signals, Medical images, 2D signals, Education on treatment medical signals, Distance education.**

I. Introduction

The basic types of transformations of 2D medical signals can be combined at more complex operations. The composite transformations must be expressed by more complex matrices. For the general case a matrix representation of transformations is:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} . \quad (1)$$

In the above matrix equation x and y are the coordinates of a point of the 2D medical signal before the transformation, x' and y' are coordinates of the point after the transformation and a, b, c, d, e, f are the constants for all types of transformations.

It is much easier if the square matrix of the constants is extended at (3x3) matrix and column vectors representing points have an extra entry.

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

If the bottom row of the matrix is [0, 0, 1] w' will be 1 and can be ignored.

II. Concatenation and Combination of 2D Transformations

Consider rotation of an image or a signal about its center (x_c, y_c). This operation includes a number of basic transformations – translation of the signal by $(-x_c, -y_c)$ so that the center point coincides with the origin, rotation about the origin, back translation of the signal by (x_c, y_c) to its first position. Matrix representation of this combination of transformations is:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} x_3 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_c \\ 0 & 1 & y_c \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} \cos\alpha & -\sin\alpha & (x_c - \cos\alpha.x_c + \sin\alpha.y_c) \\ \sin\alpha & \cos\alpha & (y_c - \sin\alpha.x_c - \cos\alpha.y_c) \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (5)$$

Eq. (3) shows translation of the medical signal by $(-x_c, y_c)$, Eq. (4) – rotation by angle α about the origin point $(0, 0)$ and Eq. (5) – translation by (x_c, y_c).

The final effect of the transformations is to arrange point (x, y) onto the point (x_3, y_3). In the particular example, this is represented by matrix multiplication of three basic transformation matrices. Square matrices can multiply together to produce another square matrix on the same dimension. In this way a composite transformations can be represented by a single transformation matrix, which is obtained by multiplying together of the basic matrices. Each point of the medical signals to be transformed is multiplied by this matrix, which performs all the component transformations in one step.

III. Order in Combination of the Transformations

More than one transformation can be combined by multiplication together of the correspond matrices. Matrix multiplication is not commutative operation $\mathbf{M}_1.\mathbf{M}_2 \neq \mathbf{M}_2.\mathbf{M}_1$. Therefore the order of combination of the transformations is important. For example, the possible combinations of the basic transformations translation and rotation are two. In the first combination the medical signal is rotating and then is translating. In the second combination the signal is translating and then is rotating. The effect in the both cases is clearly not the same.

If the transformation matrix $\mathbf{M}_1$ arrange point $\mathbf{p}$ onto point $\mathbf{p}'$:

$$\mathbf{p}' = \mathbf{M}_1.\mathbf{p} \quad (6)$$

A second transformation matrix $\mathbf{M}_2$ can to be combined with $\mathbf{M}_1$, so that $\mathbf{M}_1$ is applied first followed by $\mathbf{M}_2$. In this

[1]Penka Tz. Isaeva is with the Faculty of Communications and Communications Technologies, Technical University, Kliment Ohridski 8, 1000 Sofia Bulgaria, E-mail: pepi_is@abv.bg

case $\mathbf{M}_2$ is postconcatenated with $\mathbf{M}_1$ so that:

$$\mathbf{p}' = \mathbf{M}_2.\mathbf{M}_1.\mathbf{p} \tag{7}$$

In the case when $\mathbf{M}_2$ is applied first then $\mathbf{M}_1$, $\mathbf{M}_2$ is preconcatenated with $\mathbf{M}_1$:

$$\mathbf{p}' = \mathbf{M}_1.\mathbf{M}_2.\mathbf{p} \tag{8}$$

The order in which transformations is applied can be seen by reading outwards from the vector being transformed.

IV. Homogeneous Coordinates

The square (3x3) matrices can be obtained by adding an additional row. In the same time an additional coordinate (w) must be added to the vector of the point. In this way a point from 2D space is presented in 3D homogeneous coordinates. This technique of representing a point in a space whose dimension is one greater than dimension of the point is called homogeneous representation.

The transformed point (x', y', w') on the medical signal is expressed by Eq. (2). On converting a 2D point (x, y) in homogeneous coordinates, the w – coordinate is fixed to 1, giving the corresponding homogeneous coordinate point $(x, y, 1)$. Each of the transformation matrices has bottom row [0, 0, 1]. Therefore w' always is 1 and transformed 2D point is (x', y').

If the matrix's elements of the bottom row (g, h, i) have values resulting in $w' \neq 1$, the effect of this transformation matrix is to transform point $(x, y, 1)$ in the $w = 1$ plane onto point (x', y', w') in the $w' \neq 1$ plane. The plane $w = 1$ is real-word coordinate space and the transformed point must be mapped back onto this plane. Therefore the point (x', y', w') must be projected onto the plane $w = 1$. This operation is known as homogeneous division.

The mathematical effect from the projection is in dividing the x- and y- components of the point by the w- component:

$$\begin{aligned} x' &= x'/w' \\ y' &= y'/w' \,, \end{aligned} \tag{9}$$

Therefore real – world point of the 2D medical signal is (x'', y''), where x'' and y'' are the coordinates of the projected point.

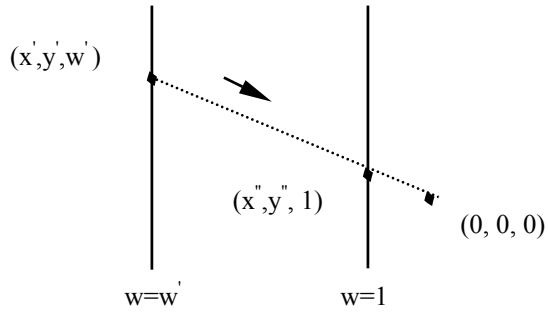


Fig. 1. Homogeneous division

V. Conclusion

The basic types of transformations of the medical signals can be expressed by matrix with dimension (3x3), which is multiplied with the vector of a point on the signal to obtain coordinates of transformed point. Thus different types of transformations of the medical signals can be combined by multiplication together of the corresponding matrices. This mean that the 2D point must be represented as 3D homogeneous point $(x, y, 1)$ to be transformed. After this transformation is obtaining point (x', y', w'). The real 2D coordinates are obtaining by division of the x and y coordinates by the w coordinate.

All these results can be applied as in the education on the medical signals and images so as theoretical basis in developing and researching of the particular problems and in creation of computer programs in this field of study.

References

[1] Foley J. D., van Dam A., Feiner S.K. and Hughes J. F., Computer Graphics: Principles and Practice (2 nd Ed.), Addison Wesley, 1990.

[2] W. Boehm et al., Computer Aided Geometric Design: A Survey of Curve and Surface Methods, 1984.

[3] Gerald Farin, Curves and Surfaces for Computed Aided Design, Academic Press, 2 nd edition, 1990.

High-Performance, Low-Cost EKG Acquisition System for an IBM-PC

Sever Paşca[1], Pablo Sterie Reyes-Turcu[2], Ştefan Radu Andreescu[3]

Abstract – **It is analyzed a low cost - high performance EKG acquisition system thought to be use in private medical cabinets. The EKG amplifier is connected to an IBM-PC. The software implemented in LabVIEW allows an easy to use interface and a good programming environment to perform data acquisition, processing and visualization of biological signal. Two ways of interconnecting the data acquisition system with the PC is possible.**

Keywords – **Data acquisition system, EKG Virtual Instrumentation**

I. Introduction

The EKG is widely spread through hospitals and clinics for diagnosis or/and monitoring of the cardiac activity of the patients. The design of this acquisition system is as cheap as possible without renouncing to its high performances. The most important characteristics are: a high common mode rejection EKG amplifier (> 100); a high-voltage input protection circuit; a bandwidth of 0,05 ... 250 Hz; recording and storage of electrocardiograms in a PC; powered by batteries; small dimension and very light weight; connected by a standard serial interface with a PC; an easy to use software implemented in LabVIEW. All of this makes the board to be cheaper because it does not contain expensive accessories that an EKG usually has as the pagewriter, the waveform display, etc. These are expensive units, some of which have consumable (recording chart paper, ink, etc.). The functions of these units are done by software running on PC.

This acquisition system amplifies the EKG signal and digitalizes it using a microcontroller. The microcontroller also has the job of controlling the system and the communication with PC. The PC acquires the 12 EKG leads, control the acquisition system and display the data acquired.

The software used to make signal acquisition, processing and user interface is developed in LabVIEW programming environment. LabVIEW ave built-in functionality for data acquisition, system control, measurement analysis, and data presentation. With LabVIEW it's easy to create a friendly, intuitive and natural human interface. LabVIEW was selected as the programming languages of this software because it is a graphical programming language based on C witch makes it very easy to debug. In addition, LabVIEW contains libraries for serial port data acquisition [1].

The isolation of the signal is made optically in the digital part of system. All the components of the system are low power devices witch make it possible to be powered by batteries. In this way, it can be assure the conditions of protection imposed by medical instrumentations.

II. Acquisition System Design

The block diagram of the EKG acquisition system is shown in Fig. 1.

Ten **Electrodes** (Ag/AgCl) measure the bio-potentials generated by the heart. **Electrodes Interface** assures the interconnection of the biomedical signal source and the EKG amplifier (Fig. 2). It protect the input of the EKG amplifier from the cardiac defibrillator, the cardiac stimulator or/and the high voltage that can arise from the touching by mistake of the patient of a body or object found under the network voltage. The high voltage that can appear is limited using a simple limitation circuit with BAV45 picoampere diode, needed to preserve the high common mode input impedance realized by the buffers. The Electrodes Interface also protects the patient against electrical shocks. The filters included in the interface attenuate the perturbation signal produce by the RF components (delivered by emitters or/and by electrosurgical equipment).

The **Lead Creation** block using the AD627 instrumentation amplifiers obtains the 12 leads. This is a micro-power, good CMRR amplifier specially designed for low-power medical instrumentation [2]. The EKG leads signals are first amplified with a gain of 10 because it is needed to eliminate the DC component before a higher amplification to assure a linear response. A High-Pass Passive Filter (**HP Filter**) with a cut off frequency of 0.05 Hz eliminates the DC component. A 16:1 CD4067 Multiplexer (**MUX**) select one of the 12 leads witch are amplified by an **Amplifier** (LF353) with an adjustable gain around 100.

A **Low–Pass 4^{th} Order Active Bessel Filter** [3] (Fig. 3) follows to be uses as an anti-aliasing filter with a cut-off frequency of 250 Hz. The Bessel filter is use because in EKG application the phase is very important because a phase error change the shape of waveforms.

The **Microcontroller** acquires the analog signal from the EKG amplifier using only one channel. The analog to digital conversion of the EKG analog waveform is accomplished

[1]Member IEEE, University POLITEHNICA of Bucharest, Department of Applied Electronics and Information Engineering, Splaiul Independentei nr.313, Bucharest, Romania, E mail: Sever.Pasca@elmed.pub.ro

[2]Member IEEE, University POLITEHNICA of Bucharest, Department of Applied Electronics and Information Engineering, Splaiul Independentei nr.313, Bucharest, Romania, E mail: pablo@fx.ro

[3]University POLITEHNICA of Bucharest, Department of Applied Electronics and Information Engineering, Splaiul Independentei nr.313, Bucharest, Romania, E mail: stefan_radu@myx.net

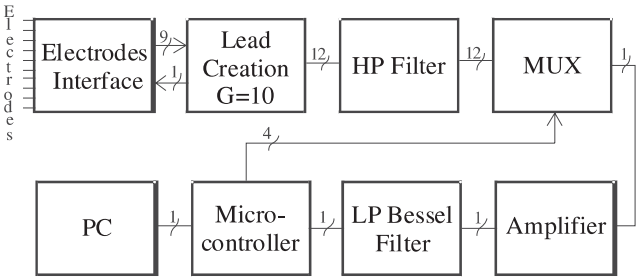


Fig. 1. Block Diagram

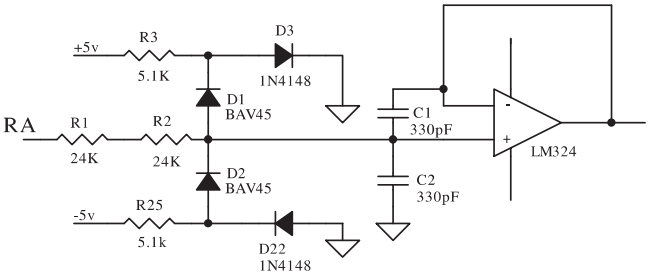


Fig. 2. Electrode Interface

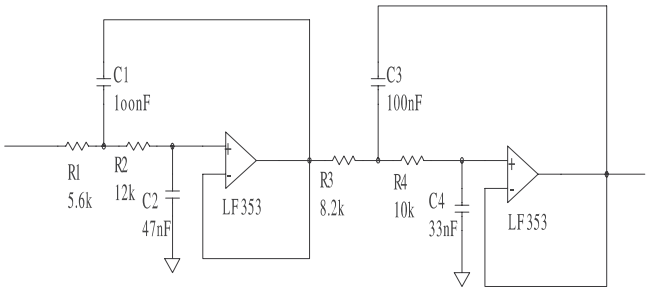


Fig. 3. Low–Pass 4th Order Active Bessel Filter

using a 10-bits ADC found in the **Microcontroller**. The software implemented in assembler assures a sampling rate of 6 KHz. This sampling rate must be achieve because each lead has a minimum sampling rate of 500 Hz, and the number of leads acquire is 12 (500 Hz*12=6 KHz). The **Microcontroller** also control the acquisition time of each of the twelve leads. This is done by establishing the commutation time of each channel of the multiplexer. The **Microcontroller** choused is the Analog Device PIC16F877. The **Microcontroller** transmits the digitalize signal to the **PC** using the USART (The Universal Synchronous Asynchronous Receiver Transmitter) also known as a Serial Communications Interface [4]. The transmission is at a rate of 115.200 BPS.

The communication between the microcontroller and the PC is possible in one of the following ways:

- using the standard serial port and an optocoupler to isolate the digital signal;
- using an infrared emitter in case in witch the PC has an IrDa standard port. In this way, there is no need to use additional ways of isolating the signals, but the acquisition system must be in a short range of the PC (10 to 60 cm from the PC).

The PC communicates the data using its serial port or its infrared port and acquires data raw - the EKG leads. The interface with the user implemented in LabVIEW, accomplish the following demands:

- selection of one of the following data presentation: 3 leads visualization; 6 leads visualization; 12 leads visualization;
- access to the patient file and the possibility to update it;
- storage or/and printing data;
- possibility of visualization the 12 standard lead in two ways: *automatically* (with a fix length of the recording waveform) and *manually* (with different length of recording the leads waveform);
- to monitor the rate of the pulse;
- the possibility of setting the speed of the recording;
- easy to use interface that allows medical personnel without a degree in electrocardiography to work with the system;
- possibility to measure the duration as well as the amplitude and the dimension of the elements of the electrocardiogram (the P and T wave, the PQ and ST segment, the QRS complex, etc);
- the possibility in future research of being able to implement a software witch allows the diagnosis of a number of illnesses of the heart.

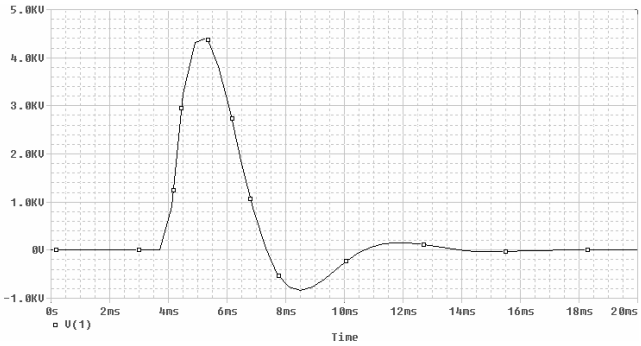


Fig. 4. Defibrillation Impulse.

III. Circuits Simulation

A defibrillation impulse can generate between two electrodes a voltage up to 4.2 KV and duration of 3 ms (Fig. 4). The **Electrodes Interface** gives a protection at the inputs of the EKG amplifier as shown in Fig. 5.

The High-Pass Passive Filter cut the DC component of the leads as shown in the Bode diagram of Fig. 6.

The Bode simulation of the anti-aliasing filter is shown in Fig. 7.

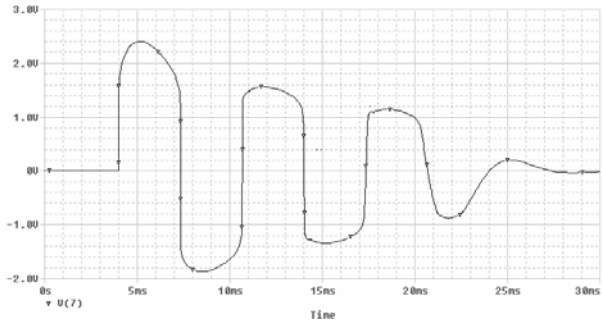


Fig. 5. Electrode Interface result

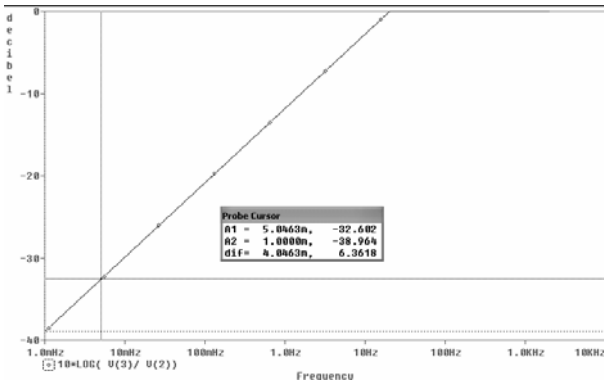


Fig. 6. Bode Plot of High-Pass Filter

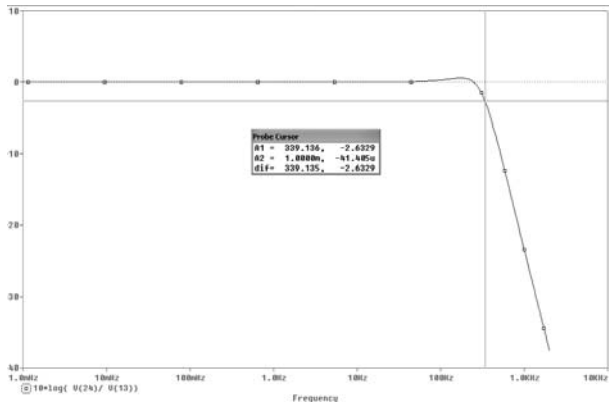


Fig. 7. Bode Simulation of the anti-aliasing filter

IV. Results

The friendly user interface design is shown in Fig. 8:

From the user interface, the doctor can choose one of the three ways of visualization of the leads.

The first visualization mode shows either the bipolar leads (I, II, III), the augmented unipolar leads (aVR, aVL, aVF) or the unipolar precordial leads (V1, V2, V3 or V4, V5, V6). Fig. 9 shows the bipolar leads.

The second visualization mode shows either the bipolar leads and the augmented unipolar leads (I, II, III, aVR, aVL, aVF), or the unipolar precordial leads (V1, V2, V3, V4, V5, V6).

The third visualization mode shows all the 12 standard leads.

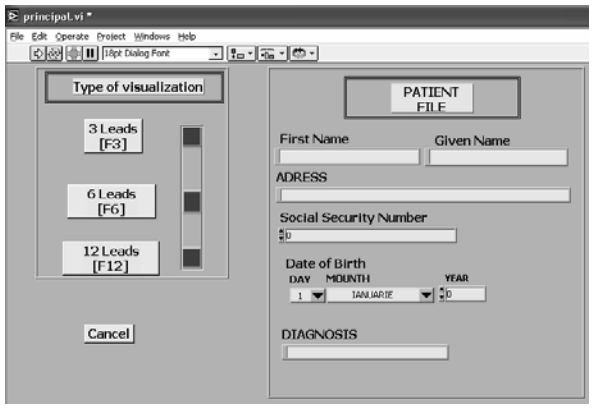


Fig. 8. User Interface

Fig. 9. Bipolar leads (I,II,III)

V. Conclusions

The EKG acquisition system is cheap compare to other similar EKG. This makes it possible for not only hospitals and big clinic, but also to smaller medical facilities to be equipped with an EKG system.

The acquisition system powered with batteries prevents the patient and the device of receiving macro shocks that can put in danger the life of the patient and of the device.

Being such a small and weightless device used with a laptop or home PC, it is easy to make medical checkup at the patient home.

The software implemented in LabVIEW give the possibility of making further software implementation of desire diagnosis, waveform studies, storage of useful information in data bases, etc.

Adding memory capacity to the hardware and changing the firmware software, it is easy to transform this device in a Holter recording system.

References

[1] National Instruments, *LabVIEW User Manual*, National Instruments Corporation, July 2000.

[2] Analog Devices, *AD627 instrumentation amplifier*, Datasheet.

[3] Fedingas Ltd. Lithuania. *Active Filter: Bessel (4th order, 24 dB/octave, Lowpass)*, http://www.fedingas.lt

[4] Microchip, *PIC16F877 microcontroller*, Datasheet.

Parametric Optimization in Noise Reduction of Medical Diagnostic Signals

Veska M. Georgieva[1] and Dimiter C. Dimitrov[2]

Abstract – **The telemedicine systems must deliver a high level of diagnostic quality and that quality must be preserved under viewing conditions that are common in the medical community. Some properties of the Wavelet Transforms are analyzed in the paper as a base for optimal strategy for elimination of noise spectrum components in the case of 2D-diagnostic medical signals. A possibility for parametric optimisation in noise reduction of real medical signals and their restoration has been tested using computer simulation in MATLAB environment. The paper can be used in engineering education in studying this process.**

Keywords – **Telemedicine, medical diagnostic signals, restoration of signals, noise spectrum, wavelet transforms, computer simulation.**

I. Introduction

The medical imaging technologies exploit the interaction between the human anatomy and the output of emissive materials or emissions devices. These emissions are then used to obtain pictures of human anatomy. The most popular technologies are ultrasound, x-rays, x-ray computed tomography and magnetic resonance imaging. These images provide important anatomical information to physicians and specialist upon which can be made diagnoses[5].

The medical signals have different characteristics and are not well matched to the type of natural images that motivate much of the research in telemedicine systems. The Wavelet representation provides a multiresolution – multifrequency expression of signal with localization in both time and frequency.

The Wavelet Transforms present circumstantially the short time differences of the signals, including the noise, too [3,4].

In image processing can be used effective methods in reduction of the noise components.

The program environment of the MATLAB version 6.0 or 6.1 with using the Wavelet Toolbox make possible a practical realization of the difficult Wavelet Transforms [1]. It can be realized computer simulation by investigation of the process to noise reduction of the real medical diagnostic signal. This noise is received by the transmission of 2D-signals during communications systems.

The simulation make possible to trail the process of optimization of some parameters as level of decomposition of the signals, type of the wavelet, parameters of threshold by the noise reduction.

[1]Veska M. Georgieva is with the Faculty of Communication, TU-Sofia, Kl.Ohridsky str.8, Sofia, Bulgaria, E-mail: vesg@vmei.acad.bg

[2]Dimiter C. Dimitrov is with the Faculty of Communication, TU-Sofia, Kl.Ohridsky str.8, Sofia, Bulgaria, E-mail: dcd@vmei.acad.bg

II. Theoretical Aspects of the Problem

One of the basic concept in Wavelet Transforms is to present the signal with two components:

- a broad approximated component
- a precise approximated (circumstantially) component

and graining with the goal to chance the level of decomposition of the signal. This is possible in time and frequency domain, too.

Let the energy of the signal $s(t)$ is $\int\limits_R s^2(t)dt$, limited in the space V, in limited domain R. Continue Wavelet Transform (CWT) of the signal can be assigned in analogy with CFT by calculation of the wavelet coefficients in Eq. 1

$$C(a,b) = \int\limits_{-\infty}^{\infty} s(t)a^{-1/2}\Psi\left(\frac{t-b}{a}\right)dt \qquad (1)$$

Or in limited domain R in Eq. 2

$$C(a,b) = \int\limits_{R} s(t)a^{-1/2}\Psi\left(\frac{t-b}{a}\right)dt\,, \qquad (2)$$

where $\Psi(t)$ is created from basic function $\Psi_0(t)$, and determined the type of the wavelet. It must provide the implementation of following operations:

- a translation in the time axis -$\Psi_0(t-b)$ when $b \in R$
- a scaling packet $a^{-1/2}\Psi_0\left(\frac{t}{b}\right)$ by $a\rangle 0$, where a gives the width of the packet; b gives the place

or in Eq. 3

$$\Psi(t) = a^{-1/2}\Psi_0\left(\frac{t-b}{a}\right) \qquad (3)$$

where b gives the place of the wavelet function and a gives the scaling of the function.

By image processing must be worked with 2D data.They can assigned in the space V, but as function of two variability.

By digital value of a and j, wavelet function can be presented in Eq. 4

$$\Psi_{j,k}(t) = a_0^{-j/2}\Psi(a_0^{-j}t - k) \qquad (4)$$

And DCWT in conclusive description with digital value of a and b is given in Eq. 5

$$C(j,k) = d_{j,k} = \int_{-\infty}^{\infty} a_0^{-j/2} \Psi(a_0^{-j} t - k) s(t) dt \,, \quad (5)$$

where $C(j,k) = d$, are circumstantially coefficients of the wavelet decomposition of the signal on level k. For 2D DCWT the circumstances of discretisation are given in Eq. 6.

$$\begin{array}{l} (j,k) \in \mathrm{Z}^2,\ a = 2^j,\ b = k2^j \\ \Psi_{j,k} = 2^{-j/2} \Psi(2^{-j} V - k)\,, \Phi_{j,k} = (2^{-j} V - k)\,. \end{array} \quad (6)$$

In Wavelet Toolbox the following theoretical methods are realized:

- time domain analysis
- frequency domain analysis
- multiresolution analysis

With help of the corresponding functions, the approximated coefficients A_j of the signal and the circumstantially coefficients D_j on the level j are computed. This representation of the signal is called decomposition of the signal on the level j. The output signal is the signal with the zero mean level of decomposition.

The following algorithm can be used:

1. Computing of circumstantially coefficients in Eq. 7

$$D_j(t) = \sum_{k \in Z} C(j,k) \Psi_{j,k}(t) \quad (7)$$

2. Computing of the signal as sum of his components in Eq. 8

$$s = \sum_{k \in Z} D_j \quad (8)$$

3. Approximation on the level $J - A_j = \sum_{j>J} D_j$
4. String between A_{j-1} and $A_j - A_{j-1} = A_j + D_j$
5. Some decompositions: $s = A_j + \sum_{j \leqslant J} D_j$

So, every level of decomposition of the signal can be represented in Eq. 9

$$\begin{array}{l} S = A_1 + D_1 \\ S = A_2 + D_2 + D_1 \\ S = A_3 + D_3 + D_2 + D_1 \\ S = \ldots\ldots\ldots\ldots\ldots \end{array} \quad (9)$$

If look at Eq. 9 as tree of reconstruction, so it will be exact of the apex. The precision of the reconstruction decreases in downhill. But the spectrum of the signal is narrow band. That means the filtering of the signal and decreasing size of information that is necessary for reproduction of the signal.

In noise reduction with Wavelet Transforms is used another method - the limitation of the level of circumstantially coefficients. The short-time speciality of the signal includes the noise components, which content many fortuitously deviations from the meaning signal. Giving the determinate threshold of their level and breaking on level the circumstantially coefficients, can be reduced the level of the noise, too. But the most interesting aspect of this problem is, that the level of limitation can be determinate separately for every one coefficient. This permits to make adaptive changes of the signal.

III. Experimental Part

Magnetic resonance imaging (MRI) is fast becoming the preferred modality in clinical applications, because it is minimally invasive and can be applied to both hard structures and soft tissues. MRI is based on measuring magnetic properties of hydrogen nuclei [2].

For experiment is using a real MRI of the liver with size 256 by 256 pixels, in JPEG format. Let is presuppose, that the noise is received by transmission during communications systems. As result the signal is received with a Gaussian white noise $e(n)$. The basic model is given in Eq. 10 [1].

$$s(n) = f(n) + e(n) \quad (10)$$

The proceedings to noise reduction consist of 3 stages:

1. Wavelet scanning of the signal on level N. It is made a choice of the type of the wavelet and the level of decomposition, where $N > 0;\ N = 2^n$
2. Circumstantially. For every level, from 1 to N can be choice a definite (hard) threshold and a corresponding (soft) threshold for the circumstantially coefficients
3. Restoration of the signal. Wavelet restoration is based on the output coefficients for approximation on the level $N$and modification of the circumstantially coefficients on level from 1 to N

The simulation makes possible to study the effect on three parameters:

- level of decomposition
- parameter ALPHA, to key the hard threshold for the circumstantially coefficients
- type of the wavelet

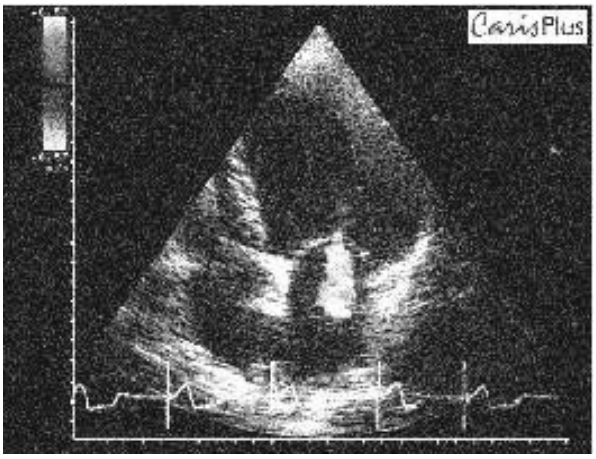
The result of simulation is given in Fig. 1.

On the base of experiments for the parametrical optimization in noise reduction of the 2D signal can make the following concrete conclusions:

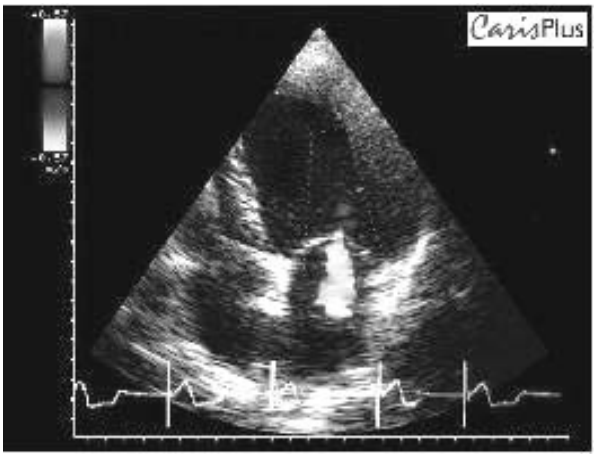
1. The using in Wavelet Toolbox in MATLAB functions give a good results in noise reduction of MRI signal.

2. The parameter ALPHA must be a full date$>$1 and the best result can be obtained when ALPHA$\geq$3

3. By reconstruction of the signal, the basic effect is on the level of decomposition in comparison with type of the wavelet. For the using MRI, $N \leq 8$. Not for all type of wavelet the Converse Wavelet Transforms are possible.



Noised MRI



Denoised MRI

Fig. 1.

4. The threshold of decomposition is realized in three directions: horizontal, vertical and diagonally.

5. The number of level on decomposition is important for the converse transform of the signal. How much it is bigger, the image is unclear. It is received a losses not only in noised as in the original signal. In this case the full restoration of the signal is impossible. For that reason is necessary to search to optimal level of decomposition of the signal.

IV. Conclusions

The parametrical optimization in noise reduction of MRI is a trial to demonstrate the application of Wavelet Transforms by 2D diagnostic medical signals. The MATLAB environment makes possible many experiments with simulation of different type of noise.

They are given as test noise signals in Wavelet Toolbox.

The experiment is realized for MRI, but it can be used successful in case of ultrasound images, x-rays images and xray computed tomography, given a special feature of this signals.

Not of last place can be noticed, that the experiment is expedient for transmission of the signals in Internet. The preliminarily processing of the images with the goal to high level of quality is very useful in succeeding compression by teleconferencing and multimedia systems [6].

The parametrical optimization in noise reduction cannot have only research disposition. It can be used successful in education for studying this problem.

References

[1] Дьяконов, В. Абраменкова, И. MATLAB обработка сигналов и изображений, Специализиранный справочник, Сл Петербург, 2002

[2] G.A. Wright, Magnetic resonance imaging, IEEE Signal Processing magazine, p.56, Jannuary 1997

[3] Teolis Anthony, Compututional signal processing with wavelets, Birkhauser, 1998, www.birkhauser.com

[4] Rao Radhhuveer M, Bopardicar Ajit S. Wavelet Transforms Introduction to Theory and Application, 1998, www.awl.com

[5] M. Smith, A. Docef, Transforms in telemedicine applications, Wavelet, subband and block transforms in communication, Kluwer, 1999

[6] D. Dimitrov, V. Kostadinova, Information system for telemedicine, Conference Proceedings, pp.343-347, EAEEIE, May, 2001, Nancy, France

One Channel ECG System

D.C. Dimitrov[1], Diana J. Vasileva[2]

Abstract – **A simple one channel ECG system is described in the paper. A preamplifier and PC are included in the system. The real analog signals from the human body after amplification can be presented in real time using sound card of PC. It's possible to do different kind of computer treatment of ECG signals, also.The described systems is convinient for application in educational process.**

Keywords – **ECG system, computer treatment, education**

I. Theoretical Introduction

Since 1858 electrical activity has been associated with the contraction of the heart. Precise analysis of the nature of this phenomenon required the ability to measure very rapid changes in extremely small electrical potentials, which is possible with this device. It amplifies the bio-signals and measures them with ADC.

By software processing is possible the "zooming" in time, amplitude or both. It is also possible saving and sending the data through networks (local or world wide).

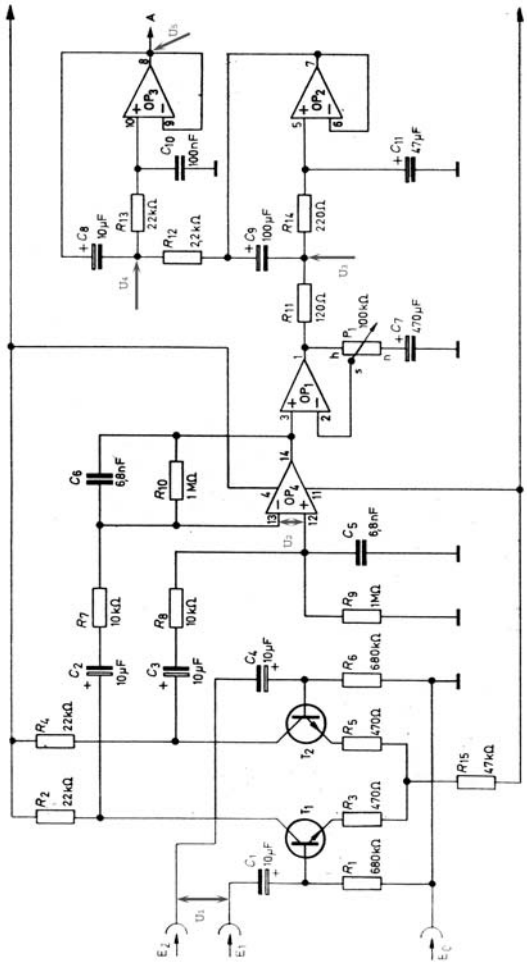
II. Tasks for the students

1. Recognation with the ECG system
2. Analyse of signals in the different points of the scheme /1,2,3,4,5 and 6/.

III. Practical Guide for Exercise

The preamp input block consists of differential amplifier which is built with the transistors T1 and T2 and the resistors R1 through R6 and R16 and capacitor C1 and C4. For better understanding we substitute that in the schematic the T1's base is positive input or E1 and the base of T2 – negative or E2. Through the rest of this text we will mark the ground electrode with E0. Capacitors C1 and C4 are used to ignore DC influence in the input stage. The input resistance is defined by the two resistors R1 and R6, which have high resistance.

The signal at the input of the transistor differential amplifier has level of 70-100 nV and it's form is shown at the first diagram from appendix A. The output signal has the same form as the input signal, but it's level is approx 5 mV (Fig. 2). The differential amplifier built with OP4 is used like selective amplifier suppressing high frequency signals (it's base



Sch. 1

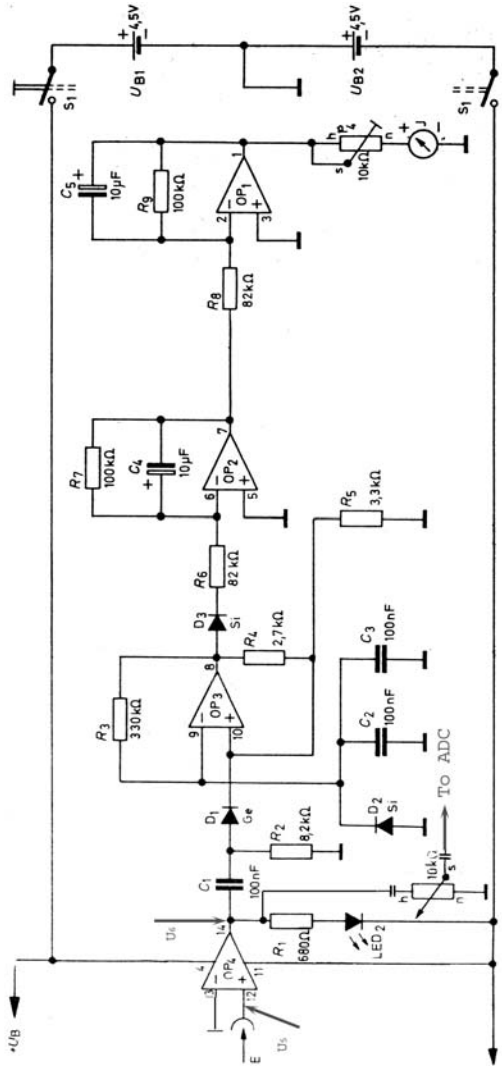
amplification is 100 times), because of the connected it the connected at the loopback circuit capacitor which makes it deep for high frequency signals.

The next amplifier consists of OP1 allows the amplification of AC signals with amplification factor for 1 to 100. With the C7 mounting the amplification of DC generated by the zero drive of each operational amplifier is being suppressed for amplification by the next stages. This stage output's signal is shown at Fig. 3 at the appendix.

OP2 and OP3 are connected by Chebishev filterschematic with cut frequency approx 25 Hz. This stage rejects all signals with frequency over 25 Hz, including the power source embarrass (fig. 4 and 5 shows the decreasing (rejecting) of embarrass signals). The total amplification of the preamp is being regulated by potentiometer P1, and it can be regulated from 500 up to 40000 times.

[1]D.C. Dimitrov is with the Technical University of Sofia dcd@vmei.acad.bg

[2]Diana J. Vasileva is with the Technical University of Sofia, didivassileva@yahoo.com



Sch. 2

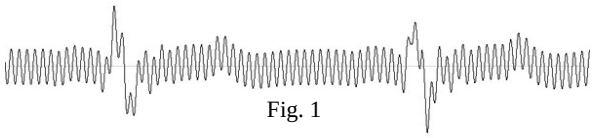


Fig. 1

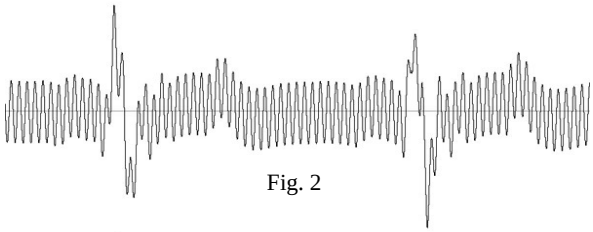


Fig. 2

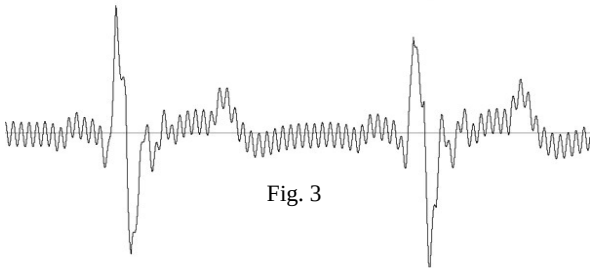


Fig. 3

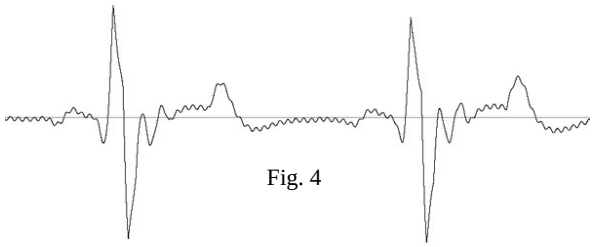


Fig. 4

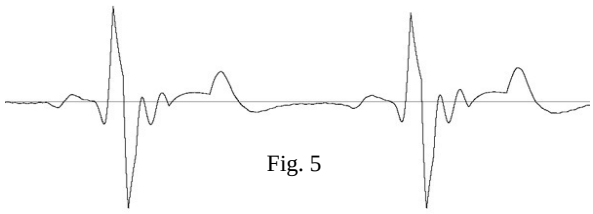


Fig. 5

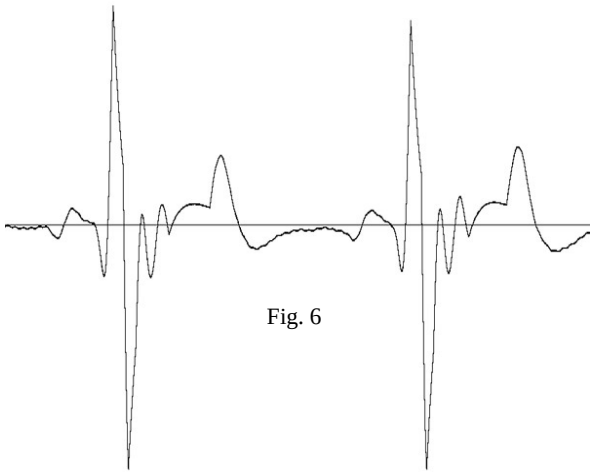
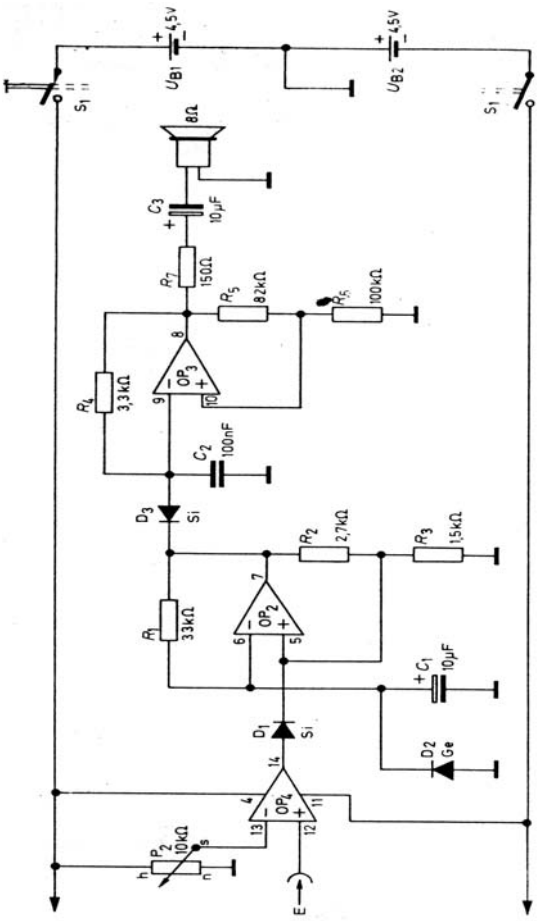


Fig. 6

The operational amplifier at sound signalization unit input is used like a comparator. At the one of it's inputs is connected P2, which provides support voltage, and the other input is being connected to preamp's output. The diode D1 passes only the positive part of comparator's output signal to the trigger connected after it. It is realized with OP2 and drives OP3 which is used like square pulse generator and it's output signal hears through speaker. For the driving is used mono vibrator built with OP2 and D3. If the signal at the comparator input is lower that the trigging threshold at the out put has negative voltage, D3 is opened and the generator cannot operate. If there is positive pulse at the input then at monovibrator's output has positive voltage, D3 is closed and the generator works. After the end of the pulse D3 again is closed and the generation stops.

At this scheme is made and visual pulse indicator and mechanical drawing device driving block. The input part of this stage consists of linear amplifier OP4 with amplification factor approx 100. The signal at it's output is the cardiogram (Fig. 6). This signal level can be regulated by potentiometer and can be digitized by ADC. The connected at it's output is connected LED, which is used to control triggers threshold level. The next two op amps are used to build driver stage for the mechanical drawer.



Sch. 3

IV. Conclusion

1. A simple ECG system, including PC was designed and described in the paper.
2. The ECG system can be used not only in medical practice but in engineering education also as practical exercise.

References

[1] Armin Holz, Bio Elektronik, Verlag Stuttgart, 1993

[2] Laura D. Jantos, Comprehensive Guide to Electronic Health Records, 2000 Edition

[3] James H. O'Keefe, The Complete Guide to ECGs, Physicians Press, 2002

[4] I.Stamboliev, Electro medical devices, Technics, Sofia 1970