XLIV INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

Proceedings of Papers

Volume 1

Sofia, 2009

XLIV INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

ICEST 2009

organized by

Faculty of Telecommunications,
Technical University of Sofia, Bulgaria

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

Faculty of Electronic Engineering,
University of Niš, Serbia

Ss. Cyril and Methodius University of Veliko Tarnovo, Bulgaria

in co-operation with

- IEEE Bulgaria Section
- IEEE Macedonia Section
- IEEE Serbia Section

The organizing committee would like to thank the sponsors of the conference:

- Bulgarian ICT Cluster
- Telelink EAD
- Bulgarian Section of IEEE
- Regional Organization of the Scientific and Technical Union, V. Tarnovo

TECHNICAL PROGRAM COMMITTEE

Chairman:

V. Poulkov	Technical University of Sofia, Bulgaria

Vice Chairmen:

R. Arnaudov	Technical University of Sofia, Bulgaria
B. Milovanović	University of Niš, Serbia
C. Mitrovski	St. Kliment Ohridski University, Bitola, Macedonia

Members:

E. Altimirski	Technical University of Sofia, Bulgaria
D. Antić	University of Niš, Serbia
G. Arsov	University of Skopje, Macedonia
W. Bock	University of Quebec at Hull, Canada
H. Deliç	Boğaziçi University, Istanbul, Turkey
M.Delimar	University of Zagreb, Croatia
V. Dimčev	University of Skopje, Macedonia
B. Dimitrijević	University of Niš, Serbia
D. Dimitrov	Technical University of Sofia, Bulgaria
D. Dobrev	Technical University of Sofia, Bulgaria
N. Dončov	University of Niš, Serbia
A. Doseva	Technical University of Sofia, Bulgaria
S. Efremidis	Athens Information Technology ,Greece
J. Erfanian	Bell Mobility, Canada
R. Goleva	Technical University of Sofia, Bulgaria
L. Gavrilovska	University of Skopje, Macedonia
N. Gospić	University of Belgrade, Serbia
I. Iliev	Technical University of Sofia, Bulgaria
M. Jevtić	University of Niš, Serbia
D. Janković	University of Niš, Serbia
V. Katić	University of Novi Sad, Serbia
R. Kountchev	Technical University of Sofia, Bulgaria
M. Lutovać	University of Novi Pazar, Serbia
J. Makal	Poland
V. Marković	University of Niš, Serbia
V. Mikarovska	St. Kliment Ohridski University, Bitola, Macedonia
M. Milanova	University of Arkansas at Little Rock, USA
S. Mirtchev	Technical University of Sofia, Bulgaria
G. Paunović	University of Belgrade, Serbia
R. Pranchov	Technical University of Sofia, Bulgaria
Z. Perić	University of Niš, Serbia
I. Reljin	University of Belgrade, Serbia
T. Skubis	Silesian University of Technology, Gliwice, Poland
M. Stojčev	University of Niš, Serbia
L. Stojmenov	University of Niš, Serbia
G. Stoyanov	Technical University of Sofia, Bulgaria
I. Stoyanov	Romania
N. Tapus	Politehnica University of Bucharest, Romania
D. Tasić	University of Niš, Serbia
G. Todorov	Technical University of Sofia, Bulgaria
B. Tsankov	Technical University of Sofia, Bulgaria
St. Valtchev	Faculty of Science and Technology, UN Lisboa, Caparica, Portugal
L. Zieleznik	Oxford Brookes University, Great Britain

CONFERENCE ORGANIZING COMMITTEE

Chairman:

V. Poulkov Technical University of Sofia, Bulgaria

International Coordinators:

V. Georgieva Technical University of Sofia, Bulgaria
B. Milovanović University of Niš, Serbia
C. Mitrovski Kliment Ohridski University, Bitola, Macedonia

Local Coordinators:

Georgi Todorov St Cyril and Methodius University, Veliko Tarnovo, Bulgaria
Margarita Todorova St Cyril and Methodius University, Veliko Tarnovo, Bulgaria
Ivo Dochev Technical University of Sofia, Bulgaria

Members:

G. Iliev Technical University of Sofia, Bulgaria
T. Mitsev Technical University of Sofia, Bulgaria
R. Goleva Technical University of Sofia, Bulgaria
L. Lubih Technical University of Sofia, Bulgaria
K. Dimitrov Technical University of Sofia, Bulgaria
V. Georgieva Technical University of Sofia, Bulgaria
R. Mironov Technical University of Sofia, Bulgaria
P. Koleva Technical University of Sofia, Bulgaria
P. Petkov Technical University of Sofia, Bulgaria
S. Kolev Technical University of Sofia, Bulgaria
J. Tsankova Technical University of Sofia, Bulgaria
Z. Stanković University of Niš, Serbia
D. Mitić University of Niš, Serbia
M. Milijić University of Niš, Serbia
T. Dimitrijević University of Niš, Serbia

CONFERENCE SECRETARIAT

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

Address:

ICEST 2009 Conference
Technical University of Sofia
Faculty of Telecommunications
Kl. Ohridski Blvd. 8,
1000, Sofia, Bulgaria
Phone: (+359 2) 965 3998
Fax: (+359 2) 965 3095
e-mail: icest@tu-sofia.bg

LIST OF ICEST 2009 REVIEWERS

Assoc. Prof. Nikolce Acevski, PhD
St.Kliment Ohridski University, Bitola, Macedonia

Assoc. Prof. Dr. Dimitar Alexiev, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Anna Andonova, PhD
Technical University of Sofia, Bulgaria

Assist. Prof. Aleksandar Atanaskovic
University of Niš, Serbia

Assoc. Prof. Ivaylo Atanasov, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Alexander Bekiarski, PhD
Technical University of Sofia, Bulgaria

Assist. Prof. Boncho Bonev, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Ognian Boumbarov, PhD
Technical University of Sofia, Bulgaria

Prof. Hakan Deliç, PhD
Boğaziçi University, Istanbul, Turkey

Prof. Marko Delimar, PhD
University of Zagreb, Croatia

Assoc. Prof. Vladimir Dimchev, PhD
Ss. Cyril and Methodius University, Skopje, Macedonia

Prof. Dr. Dimiter Tz. Dimitrov, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Dr. Emil Dimitrov, PhD
Technical University of Sofia, Bulgaria

Assist. Prof. Kalin Dimitrov, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Todor Djamiykov, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Nebojsa Doncov, PhD
University of Niš, Serbia

Prof. Sofoklis Efremidis, PhD
Athens Information Technology ,Greece

Assoc. Prof. Elissaveta Gadjeva, PhD
Technical University of Sofia, Bulgaria

Dr. Veska Georgieva, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Georgi Iliev, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Ilia Iliev, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Ivo Iliev, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Michail Iliev, PhD
University of Russe, Bulgaria

Prof. Ratcho Ivanov, PhD
Technical University of Sofia, Bulgaria

Assist. Prof. Jugoslav Jokovic, PhD
University of Niš, Serbia

Prof. Vladimir Katić, PhD
University of Novi Sad, Serbia

Assoc. Prof. Valentin Kolev, PhD
Technical University of Sofia, Bulgaria

Prof. Roumen Kountchev, ScD
Technical University of Sofia, Bulgaria

Assoc. Prof. Andrey Krumov, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Vladimir Lazarov, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Emil Manolov, PhD
Technical University of Sofia, Bulgaria

Assist. Prof. Zlatica Marinković, PhD
University of Niš, Serbia

Assoc. Prof. Marin Marinov, PhD
Technical University of Sofia, Bulgaria

Prof. Georgy Mihov, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Rossen Miletiev, PhD
Technical University of Sofia, Bulgaria

Assoc. Prof. Seferin Mirtchev, PhD
Technical University of Sofia, Bulgaria

TABLE OF CONTENTS

VOLUME 1

ORAL SESSIONS

MICROWAVE TECHNIQUE

ANTENNAS AND PROPAGATION

RADIOCOMMUNICATION SYSTEMS - 1

RADIOCOMMUNICATION SYSTEMS - 2

RADIOCOMMUNICATION SYSTEMS - 3

TELECOMMUNICATION NETWORKS

TELETRAFFIC ENGINEERING

SIGNAL PROCESSING - 1

SIGNAL PROCESSING - 2

SIGNAL PROCESSING - 3

COMPUTER SYSTEMS AND INTERNET TECHNOLIOGIES - 1

COMPUTER SYSTEMS AND INTERNET TECHNOLIOGIES - 2

VOLUME 2

ORAL SESSIONS

ENERGY SYSTEMS AND EFFICIENCY - 1

ENERGY SYSTEMS AND EFFICIENCY - 2

MEASUREMENT SCIENCE AND TECHNOLOGY

MEDICAL INFORMATION SYSTEMS

ENGINEERING EDUCATION

POSTER SESSIONS

POSTER SESSION 1

PO 1.1 Signal Processing

PO 1.2 Medical Information Systems

PO 1.3 Computer Systems

PO 1.4 Engineering Education

POSTER SESSION 2 - ELECTRONICS

POSTER SESSION 3 – MEASUREMENT AND CONTROL SYSTEMS

POSTER SESSION 4 – COMMUNICATION SYSTEMS

Session

MICROWAVE TECHNIQUE

Accurate Algorithm for Calculating Characteristic Impedance of a Square Coaxial Line

Zaklina J. Mancic[1] and Vladimir V. Petrovic[2]

Abstract – **This paper presents an accurate algorithm (of maximal relative error $\delta_{max} = 4 \cdot 10^{-6}$) for calculating the characteristic impedance of a transmission line of a square cross-section. The algorithm is based on known high accuracy approximations of elliptic integrals. The results for capacitance per unit length and characteristic impedance obtained in this way can be used as an accurate benchmark of a priori known accuracy.**

Keywords – ***Square coaxial line, Elliptic integrals.***

I. INTRODUCTION

For testing numerical methods and obtaining reference solutions of electromagnetic (EM) problems it is useful to have simple and accurate solutions of canonical problems (benchmark results). Moreover, it is important to know *a priori* the obtained accuracy [1]–[3]. According to the comparison of the results obtained by numerical methods and benchmark results an estimate can be made of the quality of the applied numerical methods. Such solutions are of the prime importance in all areas of the computational electromagnetics [2-11].

In our previous paper [10] we presented exact benchmark solution (based on exact values of elliptic integrals) and one approximate method for calculating per unit length capacitance (C') and characteristic impedance (Z_c) of the square coaxial line with homogeneous dielectric (Fig.1). This paper is the continuation of the research into more accurate and simple algorithms for this canonical, but also practical, problem.

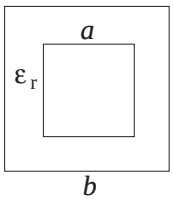


Fig.1. Coaxial line of the square cross section with homogeneous dielectric.

[1]Zaklina J. Mancic is with the School of Electrical Engineering, University of Niš, Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: zaklina.mancic@elfak.ni.ac.rs.

[2]Vladimir V. Petrovic is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia, E-mail: vp@etf.rs.

II. ANALYTICAL SOLUTION VIA ELLIPTIC INTEGRALS

For the coaxial line from Fig.1, method of conformal mapping yields the exact analytical solution for normalized per unit length capacitance, $C'_N = C'/(\varepsilon_0 \varepsilon_r)$, and for characteristic impedance Z_c that can be expressed (for nonmagnetic dielectric) as [4]

$$C'_N = 8 f(k),\ Z_c = \frac{\mu_0 c_0}{\varepsilon_r C'_N} = \frac{149896229}{10000000} \frac{\pi}{\sqrt{\varepsilon_r}\, f(k)} [\Omega], \qquad (1)$$

where

$$k = \left(\frac{\lambda' - \lambda}{\lambda' + \lambda} \right)^2,\ \lambda' = \sqrt{1-\lambda^2},\ \lambda = f^{-1}(s),\ s = \frac{b+a}{b-a}, \qquad (2)$$

$$f(k) = \frac{K(k)}{K'(k)},\ K(k) = \int_0^{\pi/2} \frac{1}{\sqrt{1-k^2 \sin^2 \theta}} d\theta. \qquad (3)$$

Here, $K(k)$ is the complete elliptic integral of the first kind [12], [13], $K'(k) = K(k')$, $k' = \sqrt{1-k^2}$, variable k is called elliptical module, k' complementary module and f^{-1} denotes inverse function. Function $f(k)$, defined by (3), often appears in conformal mapping. It has an useful property,

$$f\left(\sqrt{1-k^2}\right) = f(k') = \frac{1}{f(k)}, \qquad (4)$$

from which follows that if we known the value of f for some argument $k \in (0, 1/\sqrt{2})$, $f(k) = s$, we automatically know the value of f for argument $k' = \sqrt{1-k^2}$, $k' \in (1/\sqrt{2}, 1)$, $f(k') = 1/s$. That means that it is enough to approximate function f only for $k \in (0, 1/\sqrt{2})$. Then, for the rest of the interval, that is $k \in (1/\sqrt{2}, 1)$, formula (4) can be applied. We will denote the range $k \in (0, 1/\sqrt{2})$ the basic range of the function f. The graph of the function $f(k)$ in this range in log-log scale is shown in Fig.2, and in the full range in Fig.3.

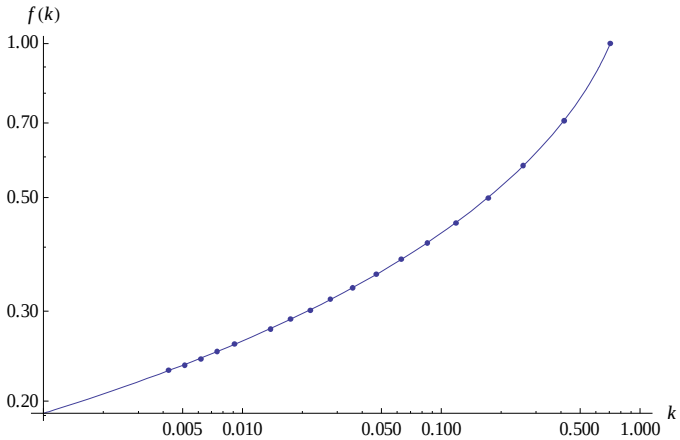


Fig.2. Graph of function $f(k)$ in log-log scale in its basic range, with several exact values marked by dots.

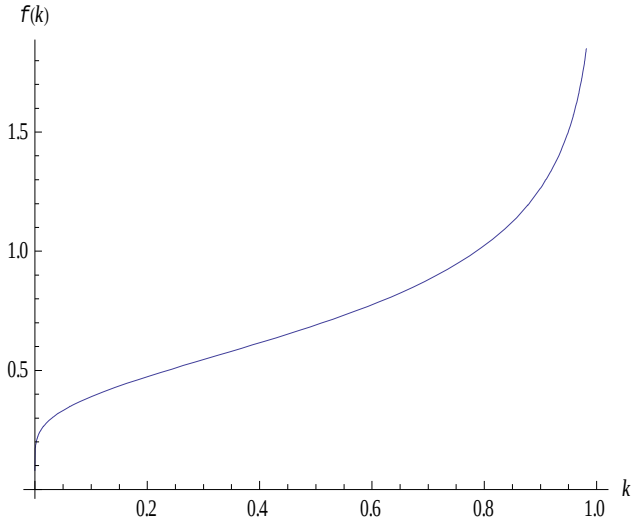


Fig.3. Graph of function $f(k)$ in its full range.

For a number of arguments k, expressed explicitly by the four basic operations, square root and cubic root, function $f(k)$ has explicit (exact) values [12]. Several such values (taken from [14]) are shown in Fig.2.

However, the main practical computational difficulty is calculation of the inverse function, $f^{-1}(s)$ for *arbitrary* argument s, as this function is usually not included in standard function packages (rare exceptions from this rule are [15] and [16]).

III. Accurate and Simple Approximate Solution

To efficiently calculate C' and Z_c of the square coaxial line it is the best to have a simple and accurate analytical formulas for both $f(k)$ and $f^{-1}(s)$ and precise estimation of their accuracy. For this purpose we will apply a simple approximate formula from [1, eq.(17)], whose maximal relative error is $\delta_{max} = 3 \cdot 10^{-6}$,

$$f(k) \approx \begin{cases} \dfrac{\pi}{\ln\left(2\dfrac{1+\sqrt{k'}}{1-\sqrt{k'}}\right)}, & 0 \le k \le \dfrac{1}{\sqrt{2}},\ 0 \le f(k) \le 1 \\ 1/f(k'), & \dfrac{1}{\sqrt{2}} \le k \le 1,\ 1 \le f(k) \le \infty. \end{cases} \tag{5}$$

This formula has a very useful property, that can be explicitly solved for k or for k', thus enabling explicit expression for the inverse function too, as

$$f^{-1}(s) = \begin{cases} \sqrt{1-\left(\dfrac{e^{\pi/s}-2}{e^{\pi/s}+2}\right)^4}, & 0 \le s \le 1 \\ \sqrt{1-\left(f^{-1}(s^{-1})\right)^2}, & s > 1. \end{cases} \tag{6}$$

In this manner, we constructed a complete algorithm for calculating approximate C' and Z_c via elementary functions only, which consists of equations (1)–(3), (5) and (6). Next we proceed to analyze the relative error of this algorithm.

First, the relative error of the formula (5) throughout the full range of k is obtained by comparing the numerical results derived by it with the results of high accuracy calculation done by the Wolfram Mathematica [16]. From these results, relative error of the formula (6) is obtained and, by using the standard calculation of error propagation, the relative error of C' and Z_c for arbitrary ratio a/b is obtained. Results are shown in Fig.4. It can be seen that the maximal relative error of the algorithm is $\delta_{B,max} = 4 \cdot 10^{-6}$ and that the relative error away from this maximum quickly drops to very low values, thus showing the overall high accuracy of the algorithm.

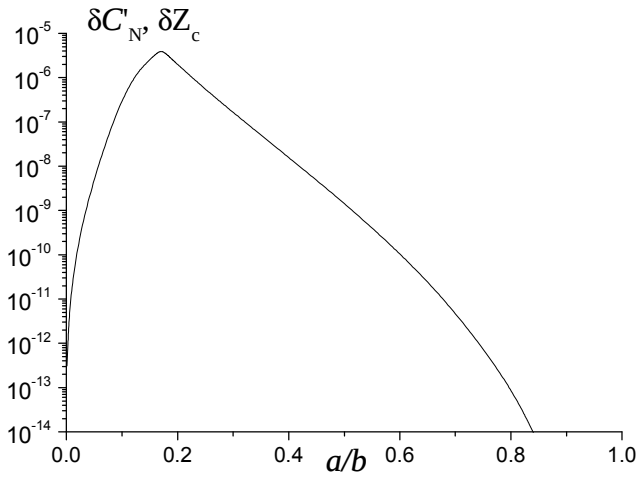


Fig.4. Relative error for normalized per unit length capacitance and for characteristic impedance.

As a further demonstration of the algorithm, values of C'_N and Z_{c0} (for the vacuum line, $\varepsilon_r = 1$) obtain by it are shown in Figs.5 and 6.

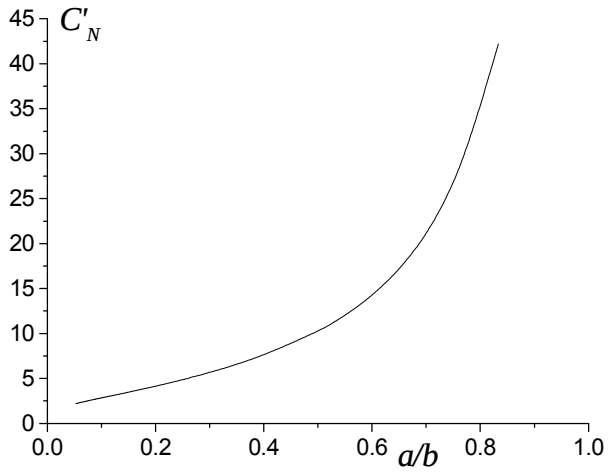


Fig.5. Dependence of C'_N on the ratio a/b.

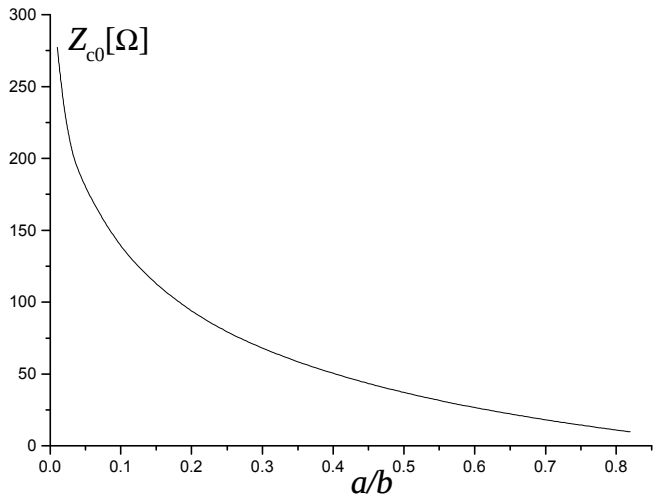


Fig.6. Dependence of Z_{c0} on the ratio a/b.

IV. Numerical results

In order to demonstrate the practical operation of the presented algorithm, Table I shows the intervals of absolute certainty of C'_N for a/b ranging from 0.01 to 0.80. In this table $C'_{N,estim}$ denotes the value calculated by the presented algorithm. In addition, $C'_{N,min}$ and $C'_{N,max}$ denote the interval in which (with 100% certainty) the exact value of C'_N lies. It can be seen that the majority of the presented values for $C'_{N,estim}$ have 7 to 8 accurate decimal digits, which is in agreement with the behavior of the relative error shown in Fig.4.

Table I. Intervals of absolute certainty for C'_N.

a/b	$C'_{N,min}$	$C'_{N,estim}$	$C'_{N,max}$
0.01	1.391559	1.391559	1.391559
0.02	1.643946	1.643947	1.643946
0.03	1.839053	1.839053	1.839053
0.04	2.008149	2.008149	2.008149
0.05	2.162367	2.162368	2.162368
0.06	2.307132	2.307132	2.307132
0.07	2.445556	2.445556	2.445557
0.08	2.579625	2.579626	2.579626
0.09	2.710702	2.710703	2.710704
0.10	2.839775	2.839777	2.839778
0.15	3.476790	3.476799	3.476808
0.20	4.134479	4.134487	4.134495
0.25	4.844417	4.844420	4.844422
0.30	5.632827	5.632828	5.632829
0.35	6.527455	6.527457	6.527458
0.40	7.561524	7.561524	7.561524
0.45	8.777792	8.777792	8.777792
0.50	10.234093	10.234093	10.234093
0.55	12.012490	12.012490	12.012490
0.60	14.234880	14.234880	14.234880
0.65	17.092054	17.092054	17.092054
0.70	20.901582	20.901582	20.901582
0.75	26.234915	26.234915	26.234915
0.80	34.234907	34.234907	34.234907

V. Conclusion

The paper presents an accurate and simple algorithm for calculating per unit length capacitance and characteristic impedance of a square coaxial line. This problem can be regarded as one of the important canonical problems of the 2D electromagnetics. It can be applied for validating numerical methods and results of high and *a priori* known accuracy can serve as an excellent benchmark. Our algorithm is based on the known approximation of elliptic integrals that is computationally very simple and also enables derivation of both elliptic functions required for the computation. Maximal relative error of the results obtained by this algorithm is found to be $\delta_{max} = 4 \cdot 10^{-6}$.

References

[1] Wolfgang Hilberg, "From Approximations of Exact Relations for Characteristic Impedances", IEEE Transactions Microwave Theory Tech., Vol.17, No.5, May 1969.

[2] Branko Kolundzija, "Accurate Solution of Square Scatterer as Benchmark for Validation of Electromagnetic Modeling of Plate Structures", IEEE Transactions Antennas and Propagation, Vol.46, No.7, July 1988.

[3] Branko Kolundzija, Andrej M. Hofman, Vladimir V. Petrovic, and Antonije R. Djordjevic, "Evaluation of Capacitance with Prescribed Accuracy Adaptive Method Based on Exact Error Estimation", IEEE Antennas and Propagation Magazine, Vol. 40, No.6, December 1998.

[4] S.W. Conning, "The Characteristic Impedance of Square Coaxial Line (correspondence)", IEEE Transactions Microwave Theory and Tech., Vol.12, No.4, July 1964.

[5] J.C. Rautio, "An ultra-high precision benchmark for validation of planar electromagnetic analyses", IEEE Transactions Microwave Theory Tech., Vol.42, No.11, November 1994.

[6] L.B. Felsen, "Benchmarks: an option for quality assessment", Antennas and Propagation Society International Symposium, AP-S 1988., Digest, Vol.3, pp. 1344–1346, June 1988.

[7] John L. Volakis, "Benchmarks of Simple, Generic, Shaped Plates for Validation of Low-Frequency Electromagnetic Computational Codes", IEEE Antennas and Propagation Magazine, Vol.35, No.4, August 1993.

[8] P. Alotto, A.V. Kuntsevitch, C. Magele, G. Molinari, C. Paul, K. Preis, M. Repetto, K.R. Richter, "Multiobjective optimization in magnetostatics: a proposal for benchmark problems", IEEE Transaction on Magnetics, Vol.32, No.3, May 1996.

[9] E. Cardelli , A. Faba, R. Specogna, A. Tamburrino, F. Trevisan, S. Ventre, "Analysis Methodologies and Experimental Benchmarks for Eddy Current Testing", IEEE Transaction on Magnetics, Vol.41, No.5, May 2005.

[10] Vladimir V. Petrovic, Zaklina J. Mancic, "Jedno referentno (benchmark) rešenje za dvodimenzionalne elektrostatičke probleme", (in serbian) YUINFO Conference, Kopaonik, 2009.

[11] Zaklina J. Mancic, Vladimir V. Petrovic, "Analiza konvergencije i tačnosti metoda konačnih elemenata za proračun vodova sa TEM talasom", (in serbian) YUINFO Conference , Kopaonik, 2009.

[12] E.T. Whittaker, G.N. Watson, *A Course of Modern Analysis*, Cambridge University Press, Cambridge, England, ch.22, p.525, 1927.

[13] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series 55, Issued June 1964, Tenth Printing with corrections, December 1972.

[14] http://functions.wolfram.com/EllipticIntegrals/EllipticK/03/03/

[15] *Maple 12*, MAPLEsoft (www.maplesoft.com).

[16] *Wolfram Mathematica 7*, Wolfram Research (www.wolfram.com).

Training Data Pre-Processing for Bias-Dependent Neural Models of Microwave Transistor Scattering Parameters

Zlatica Marinković[1] and Vera Marković[2]

Abstract –**Frequency and bias dependence of scattering parameters of microwave transistors can be successfully modelled by artificial neural networks. But, sharp changes in the frequency dependence of angle of a scattering parameter may result in inappropriate modelling accuracy in the vicinity of the frequency when sharp change occurs. In this paper we are discussing pre-processing of the training data in order to make modelling accuracy better. The proposed approach is illustrated by a suitable example.**

Keywords – **Microwave transistors, scattering parameters, artificial neural networks.**

I. INTRODUCTION

Development of accurate and reliable models of microwave transistors is a very important issue in design of modern communication systems. Scattering (S-) parameters of a microwave transistor are dependent on frequency and on operating conditions. Therefore, it is important to include these dependences into a transistor model. In this paper dependence on frequency and bias condition are considered.

Artificial neural networks (ANNs), widely known as structures with huge modelling and generalization capabilities, [1], have been successfully used for modelling of S-parameters of microwave transistors, [2], [3]. One of the models, a basic one, consists of an ANN trained to predict S-parameters for the given frequency and bias conditions. Quality of the data used for the ANN training directly affects the accuracy of the model. Sharp changes in the frequency dependence of angle of a scattering parameter may result in inappropriate modelling accuracy for frequencies near the frequency when sharp change occurs. Therefore, special attention has to be paid to the training data pre-processing in order to increase modelling accuracy. In [3] a possible solution is given, training data have to be pre-processed, and the angle of each S-parameter has to be expressed in the angular range where sharp change does not exist. This kind of pre-processing requires detailed analyses of the training data, which could make modelling process more complex. Here, we are investigating a possible approach for accuracy improving based on training data pre-processing.

The paper is organized as follows: after the introductory section, a brief theoretical background on microwave transistor S-parameters and on artificial neural networks is given in Sections II and III, respectively. Section IV starts with the description of the basic ANN microwave transistor S-parameters' model, then earlier proposed improvement is discussed and the proposed way of data pre-processing aimed for accuracy improving was presented. The proposed model and modelling procedure are illustrated on a suitable modelling example. Modelling results are given in Section V and main conclusions in Section VI.

[1]Zlatica Marinković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: zlatica.marinkovic@elfak.ni.ac.rs

[2] Vera Marković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: vera.markovic@elfak.ni.ac.rs.

II. MICROWAVE TRANSISTOR S-PARAMETERS

Microwave transistors operating under small-signal conditions can be characterized by the scattering parameters (S-parameters) which relate the voltage waves incident on the ports to those reflected from the ports (Fig.1).

The scattering matrix, or [S] matrix, is defined in relation to these incident and reflected voltage waves as, [4]:

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \quad (1)$$

or in matrix notation

$$[V^-] = [S][V^+]. \quad (2)$$

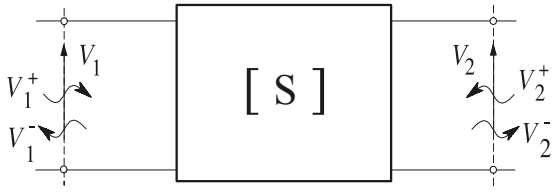


Fig.1. Incident and reflected waves in a two-port network

An element of the [S] matrix can be determined as

$$S_{ij} = \frac{V_i^-}{V_j^+}\bigg|_{V_k^+=0, \text{za } k \neq j} \quad (3)$$

S_{ii} is the reflection coefficient seen looking into the port i when all other ports are terminated in load matches. S_{ij} is the transmission coefficient from port j to port i when all ports are terminated in matched loads.

The S-parameters of microwave transistors are frequency-, temperature- and bias-dependent.

III. ARTIFICIAL NEURAL NETWORKS

A standard multilayer perceptron (MLP) artificial neural network (ANN) consists of neurons grouped into layers: an input layer (layer 0), an output layer (layer N_L) as well as several hidden layers.

Input vectors are presented to the input layer and fed through the network that then yields the output vector. The *l*-th layer output is:

$$\mathbf{Y}_l = F(\mathbf{W}_l \mathbf{Y}_{l-1} + \mathbf{B}_l) \quad (4)$$

where $\mathbf{Y}_l$ and $\mathbf{Y}_{l-1}$ are outputs of *l*-th and (*l*-1)-th layer, respectively, $\mathbf{W}_l$ is a weight matrix between (*l-1)*-th and *l*-th layer and $\mathbf{B}_l$ is a bias matrix between (*l-1)*-th and *l*-th layer. Function *F* is an activation function of each neuron and, in our case, is linear for input and output layer and sigmoid for hidden layers:

$$F(u) = 1/(1+e^{-u}) \quad (5)$$

The neural network learns relationship among sets of input-output data (training sets) that are characteristics of the component under consideration. First, input vectors are presented to the input neurons and output vectors are computed. These output vectors are then compared with desired values and errors are computed. Error derivatives are then calculated and summed up for each weight and bias until whole training set has been presented to the network. These error derivatives are then used to update the weights and biases for neurons in the model. The training process proceeds until errors are lower than the prescribed values or until the maximum number of epochs (epoch - the whole training set processing) is reached. Once trained, the network provides fast response for different input vectors, even for those not included in the training set, without additional optimizations.

IV. BIAS-DEPENDENT ANN MODELS

ANNs have been successfully applied for modeling bias- and frequency dependence of MESFET/HEMT microwave transistor S-parameters, [2], [3]. The basic ANN model of S-parameters consists of an MLP ANN trained to predict transistor S-parameters for bias voltage, bias current and frequency presented at its inputs, Fig. 2. Therefore, the network has three input neurons corresponding to

- bias voltage (dc drain-to-source voltage, V_{ds})
- bias current and (dc drain-to-source current, I_{ds}) and
- frequency f.

The output layer consists of eight neurons corresponding to magnitudes and angles of the scattering parameters.

Number of the hidden layers can be one or two. The network is trained using S- parameters' data referring to certain number of bias points in the operating frequency range. Neural networks with different number of hidden neurons are trained, tested and after their comparison, the network giving the best testing results is chosen to be the bias-dependent model of S-parameters of the modeled device.

Once the model is developed the S-parameters are obtained by calculating the ANN response to the given bias conditions and frequency. Measured values of the device S-parameters are required only for the model development.

As it is said in the introductory section and shown in details in [3], a sharp change in the frequency dependency of the angle of some S-parameter can make modeling of that parameter, and generally ANN accuracy, worse. This can be illustrated using the results given in the Section V in Figs. 6 and 8. Suppose that the dependence of the of S_{22} parameter on frequency is being modeled and that the frequency dependence of its angle has a sharp change from -180° to 180°, as it is in the case shown in Fig. 6. Furthermore, suppose that magnitude of S_{22} parameter is modeled correctly and consider only modeling of the S_{22} angle. The available measured values, represented by full dots, are used for the training process of an ANN, which will be used as a model of this frequency dependence. Since, there are not enough training data in the vicinity of the sharp change, which is often the case, the trained ANN is most likely not to achieve accurate modeling in this region. The ANN continuous response is represented with dotted line. The ANN obtained values for the frequencies corresponding to the reference values are denoted by triangles. It is obvious that satisfying modeling has not been achieved only in the region of the sharp change, moreover, it is very poor. Due to the poor modeling in this region, there is a "circle" in the modeled S_{22} plot in polar diagram, which does not exist in the actual S_{22} plot, Fig. 8.

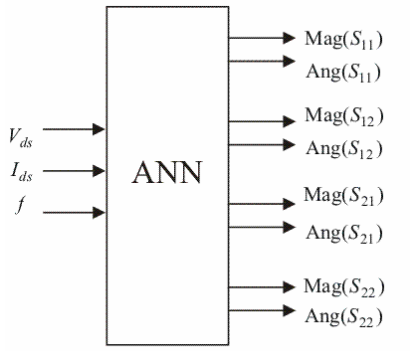


Fig.3. Bias-dependent ANN model

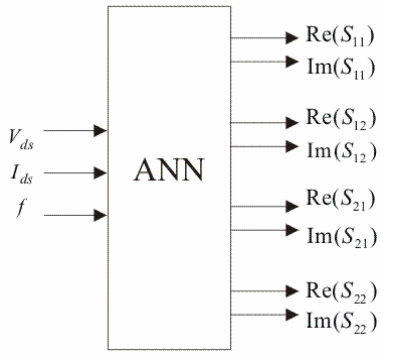


Fig.4. Proposed ANN model

In [3] a possible solution of this problem is proposed: the angle of a S-parameter that has sharp change in the frequency dependence parameter have to be expressed in the range (0°÷360°), and new ANN has to be trained by using the new data. Since the frequency dependence is smooth now, the trained ANN is able to model it with a very good accuracy. Therefore, for each S-parameter, it is recommended to express training and test values of its angle in one of (0°÷360°) or

(-180°÷180°) ranges where there is no sharp change in the frequency depmedence. In that way, modeling accuracy can be significantly increased.

The improvement of modeling proposed in [3] requires at the first a check if the frequency dependence of the angle of each S-parameter is smooth or is with a sharp change and then training data preprocessing if necessary. Here we propose an efficient approach to training data preprocessing.

The basic idea of the proposed approach is to model real and imaginary parts of S-parameters rather than then magnitudes and angles. In that way sharp changes in the frequency dependence can be avoided. As in the case of the basic model, the ANN used in the proposed model has three input and four output neurons, as well as several hidden neurons grouped in one or in two layers. The inputs of the model are the same as the inputs of basic model, while the outputs are real and imaginary parts of S-parameters.

For the model development it is necessary to represent S-parameters in the complex form as Real+j*Imag. The rest of the modeling procedure remains the same.

V. MODELING EXAMPLE

The proposed approach was applied to Hewlett Packard pHEMT device ATF35143. The data used for the model development was taken from the device manufacturer web-site, [5]. They refer to 9 bias points in (0.5-11) GHz frequency range (23 frequency points per one bias point). The data referring to 8 bias points were used as the ANN training data and the data for the remaining bias points were used as the test data.

Firstly, a basic model trained with the original magnitude-angle represented data was developed. It is a model containing ten neurons in each of two hidden layers. Effects of sharp change of angle frequency dependence occur in the case of S_{11} and S_{22} parameter and are shown in Figs 6-9. In Figs. 6 and 7 the angles of the S-parameters were given and polar plots are given in Figs. 8 and 9. Circles represent reference data, triangles neural model prediction and dotted line continuous response of the ANN. As it is discussed in the previous section, the modeling in the vicinity of sharp change is not satisfactory due to sharp changes of the training data.

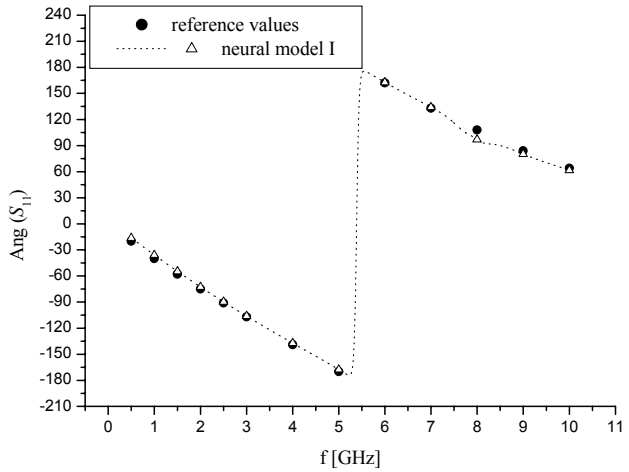


Fig. 5. S_{11} parameter angle

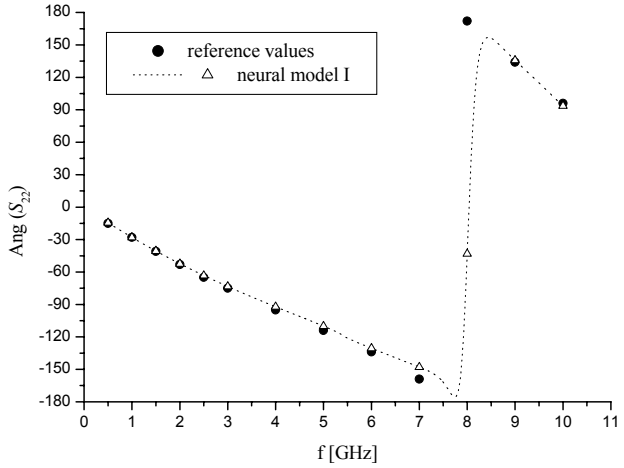


Fig. 6. S_{22} parameter angle

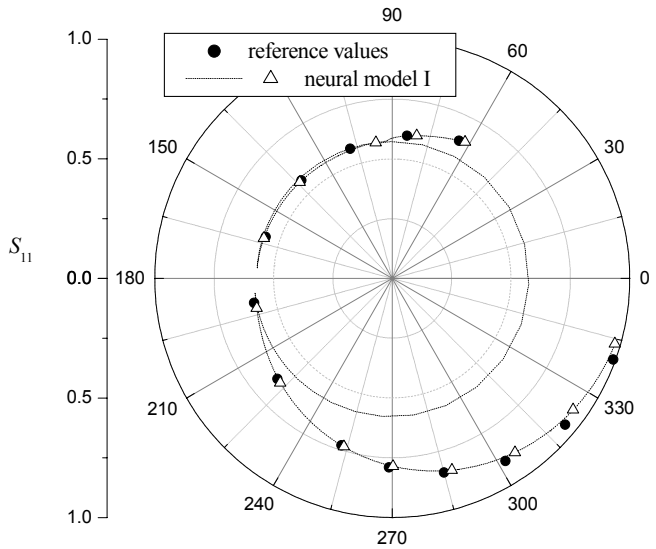


Fig. 7. S_{11} parameter – polar plot

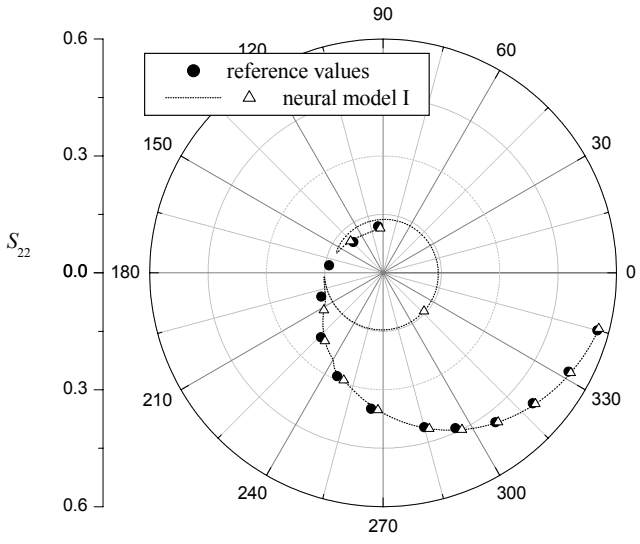


Fig. 8. S_{22} parameter – polar plot

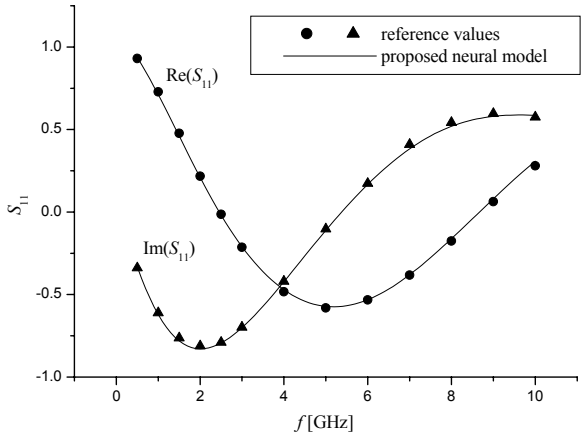


Fig. 10. S_{11} parameter – real and imaginary part

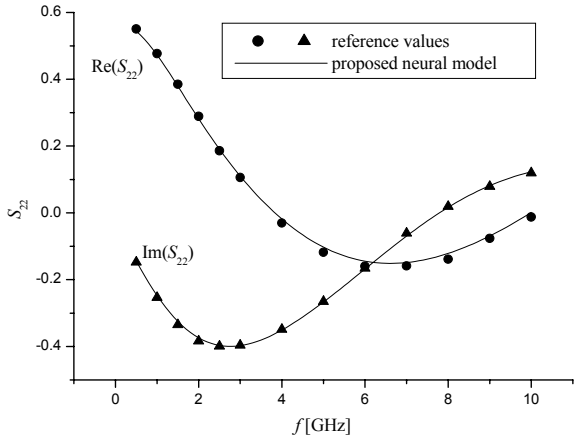


Fig. 11. S_{22} parameter – real and imaginary part

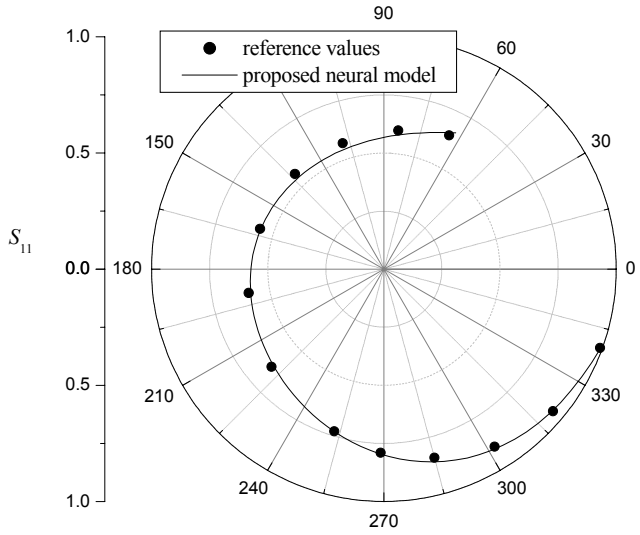


Fig.12. S_{11} parameter – polar plot

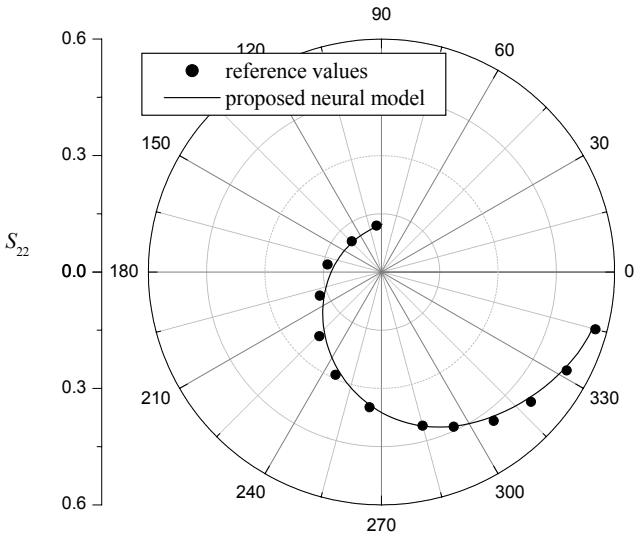


Fig. 13. S_{22} parameter – polar plot

Further, the newly proposed model was developed. S-parameters' data aimed for training purposes were preprocessed, i.e. represented in the form Real+j*Imag, and ANNs were trained. One of the trained ANNs with the best modeling results is a network with two hidden layers with ten neurons in each layer. Results for a bias point not used for the training are given in Figs. 10-13. Figs 10 and 11 refer to the real and imaginary part modeling and Figs 12 and 13 to polar plots of S_{11} and S_{22} parameters. It can be observed that the continuous ANN response (solid line) behaves smoothly and is very close to the reference values (symbols).

VI. Conclusion

In this paper we present results of investigation how different representations of the angles of the scattering parameters used as ANN training data affect modeling accuracy of ANN based bias dependent models of S-parameters' of microwave transistor. Earlier it was found that sharp changes in the frequency dependence might make modeling bad. Here we show that representation of S-parameters in the form Real+j*Imag, and development of the ANNs that model real and imaginary parts of the S-parameters may be a convenient way of overcoming the above mentioned problem. Beside the improvement of modeling accuracy, this modeling approach does not require previous detailed analyses of training data frequency dependence but only preprocessing of the complex training data by representing them by its real and imaginary parts instead by magnitudes and angles as usual.

References

[1] Q. J. Zhang, K. C. Gupta, *Neural Networks for RF and Microwave Design,* Artech House, 2000.

[2] F.Gunes, H.Torpi, F.Gurgen, "Multidimensional signal-noise neural network model", *Circuits, Devices and Systems, IEE Proceedings,* Vol.145, Iss.2, Apr 1998, pp. 111-117.

[3] Z. Marinković, A. Stošić, V. Marković, O. Pronić, "ANNs in Bias-Dependent Modeling of S-parameters of Microwave FETs and HBTs", *Microwave Review*, No.1, Vol. 12, June 2006, pp. 21-30

[4] D. Pozar, *Microwave Engineering*, John Wiley & Sons, Inc., New York, 1998

[5] http://www.semiconductor.agilent.com.

Synthesis and Analysis of Equal Power Division Reduced-Size Branch-line Hybrids

Marin V. Nedelchev, Ilia G.Iliev

Abstract: This paper proposes a method for synthesis and analysis of equal power division reduced size branch-line hybrids. Based on dispersion equation of the capacitively loaded transmission line, closed form synthesis formulae are derived. Using the fourfold symmetry of the reduced size branch-line hybrid, analysis formulae are derived. The output ports' phase difference is analyzed for different synthesis configurations. The fractional bandwidth of the reduced size hybrid is about 10% narrower than the conventional branch-line hybrid. The results presented in the paper are applicable to uniplanar structure designs.

Keywords: Reduced size, branch line hybrid, capacitively loaded transmission line.

I. INTRODUCTION

Quadrature branch-line hybrids play important role as power divider, or power combiner, image rejection mixers, and balanced amplifiers. The classic form of the quadrature branch-line hybrid is shown on Fig.1.

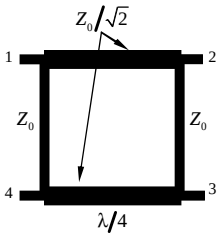


Fig.1 Conventional branch-line coupler

For a given input power at port 1, there will be 90° phase difference between the waves at ports 2 and 3 for the centre frequency. This phase difference varies over $\pm 5^{\circ}$ for a 10% fractional bandwidth. The usable bandwidth of the branch-line hybrid is constrained by the change in the isolation. Such branchline couplers can be designed for 3-10dB coupling.

Main arms of the coupler are of quarterwave length. This size is unacceptable large especially for the lower microwave bands and MMIC applications. The branch-line hybrid occupies large area on the printing circuit board or on the chip.

The use of lumped-element hybrid, which uses spiral inductors and lumped capacitors, is one possible solution to this problem [1,2]. However, the design of spiral inductors is requires precise empirical models based on measurement of test elements including parasitics.

Marin Nedelchev – Assist. Prof., PhD in Dept. of Radiocommunication and Videotechnologies in Faculty of Communications and Communication Technologies in TU –Sofia E-mail mnedelchev@tu-sofia.bg

Ilia G. Iliev – Assoc. Prof., PhD in Dept. of Radiocommunication and Videotechnologies in Faculty of Communications and Communication Technologies in TU –Sofia E-mail igiliev@tu-sofia.bg

This design approach becomes difficult for frequencies above 20GHz [3]. Another method for the hybrid coupler size reduction is proposed in [4]. The branch line is realized as lumped element Π or T equivalent networks. Each lumped element value is determined by equating the ABCD matrices of both structures – transmission line and lumped element. The coupler involves shunt inductors, which are inconvenient for fabrication. However, this method eliminates the uncertainty caused by the lumped inductors.

The method presented in [1,4], the branchline hybrid utilizes high impedance lines loaded by shunt lumped capacitors. The authors present a special case for the theory of this coupler.

In this paper, we present a general theory of reduced size branch-line hybrid, composed of capacitively loaded transmission lined. With the proposed method, a 3dB branch-line coupler can be designed for various impedance and capacitor values. The reduction of the size is connected with increase of the characteristic impedance of the transmission line. The phase difference between the output ports for various characteristic impedances and capacitors is studied. Because of the slow-wave effect of the capacitive loaded line, the phase difference is not symmetrical along the 90^0 value.

II. REDUCED SIZE HYBRID

The main element in the conventional hybrid coupler is the quarter-wavelength transmission line. In order to obtain reduced-size coupler, shown on Fig.2, in [1,4] is proposed to use capacitive loaded transmission line instead of quarter-wavelength line. To assure equivalent electrical parameters, it is examined both circuits shown on Fig.3a and Fig.3b. The circuit on Fig.3b may be considered as a unit element of periodic slow wave structure [4-7].

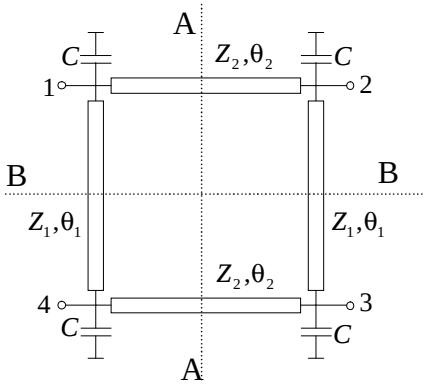


Fig.2 Reduced size branch-line coupler

The slow-wave line consists of a transmission line loaded on both sides by lumped capacitors C. The main parameters

of the transmission line are the characteristic impedance Z_c, length of the line l, and the propagation constant k.

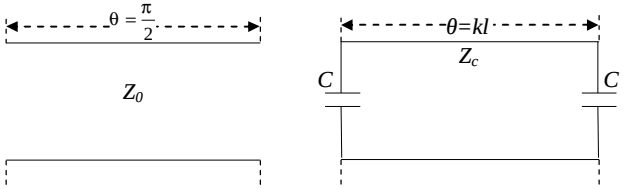


Fig.3 (a) Quarter wavelength line, (b) capacitive loaded line

The electric length of the unloaded line is $\theta = kl$. The electrical characteristics of the line are described by the ABCD matrix. The overall matrix of the resonator is a product of the three ABCD matrices – the lumped capacitors C and the unloaded transmission line.

$$[ABCD] = [ABCD]_1 [ABCD]_2 [ABCD]_3 \quad (1),$$

where

$$[ABCD]_1 = \begin{bmatrix} 1 & 0 \\ j\omega C & 1 \end{bmatrix}, \; [ABCD]_3 = \begin{bmatrix} \cos\theta & jZ_c \sin\theta \\ j\frac{1}{Z_c}\sin\theta & \cos\theta \end{bmatrix},$$

$$[ABCD]_3 = \begin{bmatrix} 1 & 0 \\ j\omega C & 1 \end{bmatrix}.$$

After the multiplication, the elements of the ABCD matrix are derived as:

$$A = \cos\theta - \omega C Z_c \sin\theta \quad (2a)$$

$$B = jZ_c \sin\theta \quad (2b)$$

$$C = 2j\omega C \cos\theta + j\left(\frac{1}{Z_c} - (\omega C)^2 Z_c\right)\sin\theta \quad (2c)$$

$$D = \cos\theta - \omega C Z_c \sin\theta \quad (2d)$$

It is easily checked out the main property of the ABCD matrix: $AD - BC = 1$.

The ABCD matrix of the quarter wavelength line is:

$$[ABCD] = \begin{bmatrix} 0 & jZ_0 \\ j\frac{1}{Z_0} & 0 \end{bmatrix} \quad (3)$$

Comparing the ABCD matrices of both circuits (2) and (3) it is obtained the main design equations:

$$\cos\theta - \omega C Z_c \sin\theta = 0$$
$$Z_c \sin\theta = Z_0 \quad (4)$$

It is obvious from Eq.4 that the characteristic impedance of the slow-wave line Z_c is greater than the characteristic impedance Z_0 of the conventional hybrid coupler.

There are two main approaches to design a reduced-size branch-line coupler. The first approach starts with the choice of the characteristic impedances Z_1 and Z_2 of the transmission lines (Fig.2) and calculation of their electrical lengths θ_1 and θ_2. On the final stage the value of the lumped element capacitor is calculated. The design equations become:

$$\theta_1 = \arcsin\frac{Z_0}{Z_1}$$

$$\theta_2 = \arcsin\frac{Z_0}{\sqrt{2}Z_2} \quad (5).$$

$$\omega C Z_0 = \cos\theta_1 + \sqrt{2}\cos\theta_2$$

The second approach starts with the choice of the electrical lengths of the transmission lines θ_1, θ_2 and calculation of the characteristic impedances Z_1 Z_2 and the value of the lumped capacitor. The design equations are as follows:

$$Z_1 = \frac{Z_0}{\sin\theta_1}$$

$$Z_2 = \frac{Z_0}{\sqrt{2}\sin\theta_2} \quad (6)$$

$$\omega C Z_0 = \cos\theta_1 + \sqrt{2}\cos\theta_2$$

The analysis of the reduced size hybrid is carried out using the fourfold symmetry of the schematic shown on Fig.2. Using proper excitation of the structure both symmetry planes AA and BB can be either magnetic (open circuit) or electric walls (short circuit). The equivalent schematics for the four different excitations are shown on Fig.4.

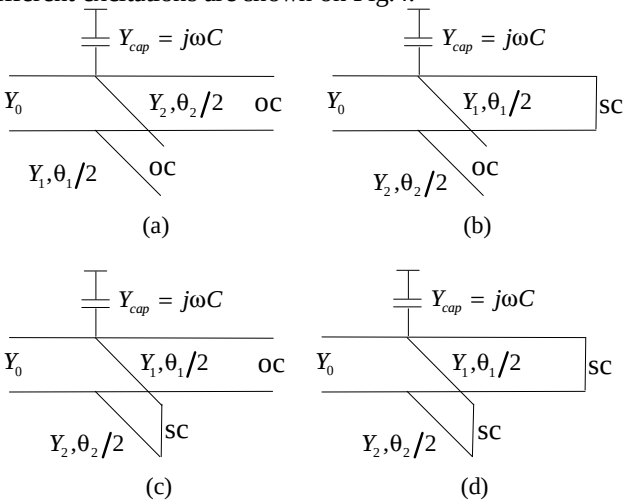


Fig.4. Equivalent schematics for one quarter section when (a) AA and BB are magnetic walls, (b) AA is electric wall, BB is magnetic wall, (c) AA is magnetic wall, BB is electric wall, (d) AA and BB are electric walls.

The reflection coefficients of each equivalent schematic are obtained as follows:

$$\Gamma_a = \frac{Y_c - j\omega C - jY_1 tg\frac{\theta_1}{2} - jY_2 tg\frac{\theta_2}{2}}{Y_c + j\omega C + jY_1 tg\frac{\theta_1}{2} + jY_2 tg\frac{\theta_2}{2}} \quad (7a)$$

$$\Gamma_b = \frac{Y_c - j\omega C - jY_1 tg\frac{\theta_1}{2} + jY_2 cotg\frac{\theta_2}{2}}{Y_c + j\omega C + jY_1 tg\frac{\theta_1}{2} - jY_2 cotg\frac{\theta_2}{2}} \quad (7b)$$

$$\Gamma_c = \frac{Y_c - j\omega C + jY_1 cotg\frac{\theta_1}{2} - jY_2 tg\frac{\theta_2}{2}}{Y_c + j\omega C - jY_1 cotg\frac{\theta_1}{2} + jY_2 tg\frac{\theta_2}{2}} \quad (7c)$$

$$\Gamma_d = \frac{Y_c - j\omega C + jY_1 cotg\frac{\theta_1}{2} + jY_2 cotg\frac{\theta_2}{2}}{Y_c + j\omega C - jY_1 cotg\frac{\theta_1}{2} - jY_2 cotg\frac{\theta_2}{2}} \quad (7d).$$

The lengths of the branches are shorter than quarter wavelength. Consequently the electrical lengths are less than $\pi/8\ rad$. The following approximations are assumed $tg\theta \approx \theta$ and $cotg\theta \approx 1/\theta$. The previously derived formulae for the reflection coefficients are reduced to:

$$\Gamma_a = \frac{Y_c - j\omega C - jY_1\frac{\theta_1}{2} - jY_2\frac{\theta_2}{2}}{Y_c + j\omega C + jY_1\frac{\theta_1}{2} + jY_2\frac{\theta_2}{2}} \quad (8a)$$

$$\Gamma_b = \frac{Y_c - j\omega C - jY_1\frac{\theta_1}{2} + jY_2\frac{2}{\theta_2}}{Y_c + j\omega C + jY_1\frac{\theta_1}{2} - jY_2\frac{2}{\theta_2}} \quad (8b)$$

$$\Gamma_c = \frac{Y_c - j\omega C + jY_1\frac{2}{\theta_1} - jY_2\frac{\theta_2}{2}}{Y_c + j\omega C - jY_1\frac{2}{\theta_1} + jY_2\frac{\theta_2}{2}} \quad (8c)$$

$$\Gamma_d = \frac{Y_c - j\omega C + jY_1\frac{2}{\theta_1} + jY_2\frac{2}{\theta_2}}{Y_c + j\omega C - jY_1\frac{2}{\theta_1} - jY_2\frac{2}{\theta_2}} \quad (8d).$$

The scattering matrix elements are derived as follows [6,7]:

$$\begin{aligned} S_{11} &= \frac{1}{4}(\Gamma_a + \Gamma_b + \Gamma_c + \Gamma_d) \\ S_{12} = S_{21} &= \frac{1}{4}(\Gamma_a - \Gamma_b + \Gamma_c - \Gamma_d) \\ S_{13} = S_{31} &= \frac{1}{4}(\Gamma_a - \Gamma_b - \Gamma_c + \Gamma_d) \\ S_{14} = S_{41} &= \frac{1}{4}(\Gamma_a + \Gamma_b - \Gamma_c - \Gamma_d) \end{aligned} \quad (9)$$

If the reduced size hybrid is designed according to Eq.5 or Eq.6, therefore the scattering matrix elements (Eq.9) will reduce to $S_{11} = S_{14} = 0$, $S_{13} = -1/\sqrt{2}$, $S_{12} = -j/\sqrt{2}$.

III. NUMERICAL RESULTS

In order to verify the applicability of the derived formulas (5,6) for design of reduced-size hybrid, it is examined the phase difference between the output ports for different electrical lengths and characteristic impedances of the branches.

It is convenient in some design cases to fix the characteristic impedances of the reduced-size hybrid branches. Utilizing Eq.(5) there are calculated the electrical lengths of the branches. The results are summarized in Table 1.

It is obvious that the increase in the characteristic impedance leads to shorter lines. The value of the lumped capacitance *C* is big enough and it could be realized in SMD as a trimmer for the low GHz range.

TABLE 1 EQUAL CHARACTERISTIC IMPEDANCE OF THE BRANCHES

Number	Z_1,Ω	Z_2,Ω	θ_1, deg	θ_2, deg	$C,[pF]$
1	55	55	65.38	40	4.77
2	70.7	70.7	45	30	6.15
3	90	90	33.75	23.13	6.78

Fig.4 shows the phase difference between the output ports for the design examples summarized in Table 1. It is drawn the phase difference for conventional branch-line coupler for reference.

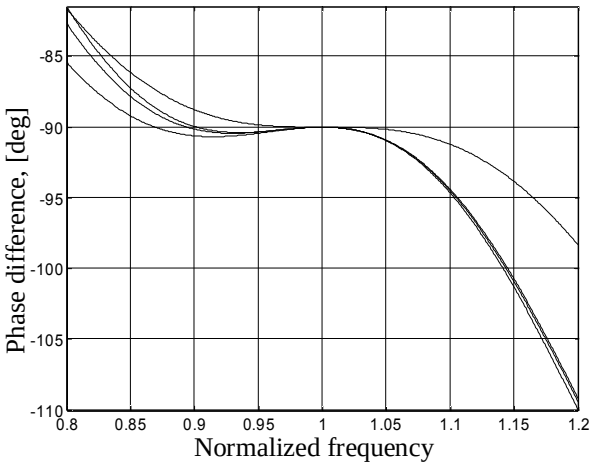


Fig.5 Phase difference between the output ports according Table1. (№1-dashed, №2- dotted, №3-dash-dotted). Conventional branch-line coupler-solid line

It is clearly seen from Fig.4 that the bandwidth of the reduced-size hybrid is narrower than the conventional branch-line coupler. The phase difference of reduced size hybrid is asymmetric against the central frequency. For the frequencies upper than the center, the phase difference increases very rapidly from the reference point of -90^o. For the frequencies lower than the center, the phase difference varies around -90^o. In contrast to the conventional branch-line coupler the output phase difference of the reduced size hybrid equals twice -90^o, while alternates around this value. This is due to the slow-wave effect observed in capacitively loaded transmission line. For lower frequencies the phase difference rises more rapidly than the conventional branch line coupler. Consequently as well-expressed is the slow-wave effect as narrower is the bandwidth.

In some design cases it is desirable to fix the electrical lengths of the branch lines. For the chosen lengths, it is computed the characteristic impedances of the lines and the values of the lumped capacitors (Eq.6). In Table2 are summarized the results for 6 different electrical lengths starting from *15 deg* to *75 deg*.

TABLE 2 EQUAL ELECTRICAL LENGTHS OF THE BRANCHES

Number	Z_1,Ω	Z_2,Ω	θ_1, deg	θ_2, deg	$C,[pF]$
1	193.18	136.6	15	15	7.42
2	118.31	83.66	25	25	6.96
3	87.17	61.64	35	35	6.29
4	70.71	50	45	45	5.43
5	57.73	40.82	60	60	3.84
6	51.76	36.6	75	75	2

Fig.5 shows the phase difference between the output ports for the design examples summarized in the first three rows in Table 2. For very small sized hybrids, it is impractical to realize in standard microstrip technology, because of the very high characteristic impedance of the lines and the necessity of high precision etching technology.

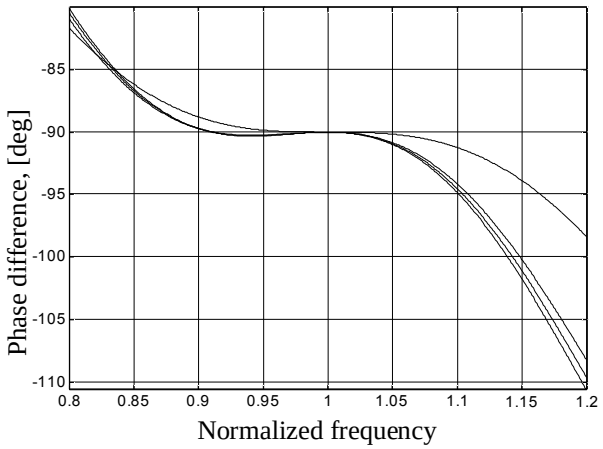


Fig.6 Phase difference between the output ports according Table2. (№1-dashed, №2- dotted, №3-dash-dotted). Conventional branch-line coupler-solid line

It is seen from Fig.5 that for highly dispersive structures the fractional bandwidth is much narrower than the conventional branch-line coupler. The ripple of the phase difference around the value of -90^{o} is under 0.8^{o} . For the frequencies upper the center frequency the phase difference decreases very rapidly and the output ports are not longer in quadrature.

Increasing of the branches electrical length leads to decreasing of their characteristic impedances. Therefore the lumped capacitance has lower value. Because of the very low value of the lumped capacitance a sensitivity analysis of the hybrid response is needed.

Fig.6 shows the phase difference for the design examples in the last three rows in Table2. The increasing of the electrical length of the branch lines causes their characteristic impedance to decrease.

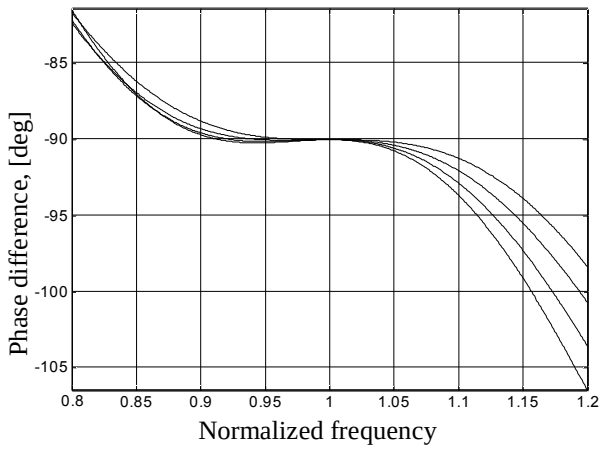


Fig.7 Phase difference between the output ports according Table2. (№4-dashed, №5- dotted, №6-dash-dotted). Conventional branch-line coupler-solid line

Comparing the results, shown on Fig.5 and Fig.6, the bandwidth of the reduced size hybrid is wider for longer branch lines. For electrical length $\theta_1 = \theta_2 = 75\text{deg}$ and lumped capacitor 2pF, the phase difference is close to the conventional branch-line hybrid for lower frequencies than the center frequency. Due to the slow-wave effect the phase difference for higher frequencies lowers more rapid than the conventional branch-line hybrid.

V. CONCLUSION

The paper proposes a method for synthesis and analysis of equal power division reduced size branch line hybrids utilizing slow wave effect of capacitively loaded line. Using the fourfold symmetry of the hybrid, analysis formulae are obtained. Closed form synthesis formulae are derived for fixed characteristic impedance or fixed electrical length of the branch lines. The numerical results show that the fractional bandwidth of the reduced size hybrid is narrower than the conventional hybrid. The phase difference between the output ports is no longer symmetrical around -90^{o} due to the slow wave effect. The proposed synthesis method is suitable for low GHz range hybrids, balanced amplifiers on a microstrip technology.

REFERENCES

[1] Hirota, Minakawa, Muraguchi, Reduced-Size Branch-Line and Rat-Race Hybrids for Uniplanar MMIC's, IEEE Trans on MTT March, 1990, pp.270-275

[2] Mongia, Bahl, Bhartia, RF and Microwave Coupled-Line Circuits, Artech House, 1999

[3] Bahl, Bhartia, Microwave Solid State Circuit Design, JohnWiley&Sons, 2003

[4] J.S.Hong, Lancaster, M, Theory and Experiment of Novel Microstrip Slow-Wave Open Loop Resonator Filters, IEEE Trans. On MTT-45, Dec.1997, pp.2358-2365

[5] Nedelchev, Iliev, Resonance and Dispersion Characteristics of Microstrip Slow-Wave Open-Loop Resonator, ICEST 2007, Bitola Macedonia

[6] Collin, Foundation for Microwave Engineering, McGraw-Hill, 1999

[7] Pozar, Microwave Engineering, JohnWiley&Sons, 1998

Design of Hexagonal Open Loop Filters on FR-4 Substrate

Marin V. Nedelchev, Ilia G.Iliev

Abstract - **This paper presents a design of microstrip hexagonal open loop resonator filter. The hexagonal form of the open loop resonator reduces the area on the PCB. Because of their structure, the hexagonal open loop coupled resonator offers wide variety of coupling schemes and different in value and sign coupling coefficient. The coupling schemes are electrical, magnetic and mixed. The paper presents a Chebyshev filter design of third order with tapped input/output lines. There is a good agreement between the theoretical and EM simulation results.**

Keywords - **Microstrip bandpass filter, hexagonal resonator, External quality factor, coupling coefficients.**

I. INTRODUCTION

In the modern communication systems, high selectivity and low passband loss are the main requirements for the microstrip filters. Low passband loss increases the system sensitivity and the high selectivity decrease the guard interval between two channels in a communication system. On the other hand it is important to reduce the filter size and weight in order to integrate them in MIC or MMIC. The problem for miniaturization is very important for the lower microwave range.

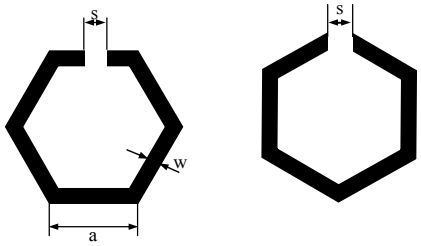
In order to reduce the size of the half wavelength resonator the authors of [1] fold back the ends of the resonator into a "U" shape. The hairpin resonator filter is one of the most popular microstrip filter configurations used in the lower microwave frequencies. It is easy to manufacture and adjust, because it has opencircuited ends that require no grounding. The hairpin resonator filter has the same design procedure as the parallel resonator filter.

The further miniaturization of half wavelength is achieved by the square open loop filters [2]. The resonator is bent in square form. This resonator configuration allows additional couplings and design of cross coupled filters [2-4]. These filters have quasi-elliptic response. The formulae for the coupling coefficient for different topologies are given in [2] using curve approximation of simulation results. Theoretical formulae for the coupling coefficients are derived in [5,6].

In order to achieve more flexible filter design, the authors of [7] proposed hexagonal structure of the resonator. It inherits the features of the halfwavelength resonator-position of the spur frequency, input impedance and slope parameter. There are two main topologies of hexagonal resonators shown on Fig.1.

Marin Veselinov Nedelchev and Ilia Georgiev Iliev –are with Dept. of Radiocommunication and Videotechnologies in Faculty of Telecommunication in TU –Sofia, N8, Kliment Ohridski bul., 1700 Sofia, Bulgaria. E-mail: mnedelchev@tu-sofia.bg, igiliev@tu-sofia.bg.

Hexagonal resonators offer different coupling topologies with different in value and sign coupling coefficients (Fig.2). Both configurations allow filter topologies with non-adjacent resonator couplings, performing quasi-elliptical response. More over, the hexagonal resonator makes easier the coupling for triplet and cascaded triplet filters than using square open loop resonators.



(a) (b)

Fig.1. Topologies of hexagonal open loop resonators.

Due to the short length of the coupled lines, the realization of loose couplings is possible with reasonable gap between them. This makes suitable the hexagonal resonator applicable in narrow and very narrow bandwidth microstrip filters.

The microstrip filter synthesis includes calculating the coupling coefficients for a given approximation (a method for Chebyshev approximation is presented in [2,7] and their realization. The theory of coupling for tuned resonators is given in [2].

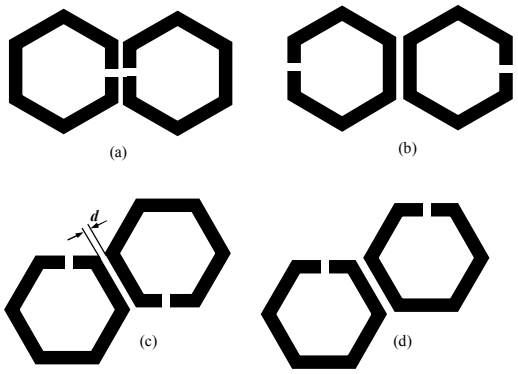


Fig.2 Coupled resonator topologies. (a) electrical (b) magnetic (c) and (d) mixed.

A full-wave electromagnetic (EM) simulator is used to calculate the couplings in the papers [2-4,7], which is an time-consuming, but reasonably accurate method. Closed form formulas are derived for the basic couplings in the CQ filters in [5,6].

In this paper is demonstrated a design of third order hexagonal open loop resonator filter. The filter is of Chebyshev type with tapped input/output lines with center

frequency 1080MHz and fractional bandwidth 4.7%. The size of the designed filter is two times shorter than the parallel edge coupled resonator filter, but remains the same selectivity. The coupling coefficient for mixed coupling and the external quality factor are extracted by EM simulation results.

II. HEXAGONAL RESONATOR AND COUPLING COEFFICIENT

Hexagonal open loop resonator is half wavelength long microstrip line with open ends (Fig.1). When the form of the resonator is symmetrical, there are two possible topologies. The hexagonal form of the resonator is derived from the square open loop resonator. Two adjacent faces join to an angle of 120°. In order to avoid the coupling between the open ends of the resonator the distance ***s*** between them should be bigger than the width oh the line ***w.*** The perimeter of the resonator is *(6a+s)* and should be equal to 180 degrees for the center frequency of the filter. The open ends are supposed to be shortened, because of the fringe capacitance by [1]:

$$\frac{\Delta l}{h} = 0.412\left[\frac{\varepsilon_{reff}+0.3}{\varepsilon_{reff}-0.258}\right]\left[\frac{w/h+0.264}{w/h+0.8}\right], \quad (1)$$

where ε_{reff} the effective dielectric permittivity and h is the substrate height.

The different orientations of the resonators on the substrate lead to different kinds of coupling structures (Fig2). Obviously the coupling is achieved by the fringe fields, when the resonators are close one another. Taking into account the field distribution along the hexagonal resonator it is possible to distinguish four main couplings. The electrical filed is stronger than the magnetic near the open end of the resonator. Consequently the coupled structure on Fig2a is dominantly electrical in nature and negative in sign. The maximum of the magnetic field is in the center of the resonator. The coupling is magnetic (Fig.2b) and positive in sign. The strength of the electrical field and magnetic field decays rapidly with the distance from the open end and the center of the resonator respectively. Then the coupling structures on Fig2c and Fig.2d perform mixed coupling. It is not possible to determine which field is dominant. The resonators on Fig.2c exhibit stronger coupling, because the currents in the coupled lines are equal and in-phase. The value of the coupling coefficient of the coupled resonators on Fig.2d is much lower, because the currents are out-of-phase. This topology is applicable in narrow bandwidth filters.

The filter demonstrated in the paper is fed by a tapped line.

III. THIRD ORDER CHEBYSHEV BANDPASS FILTER

Since all six arms of the hexagonal resonator are available for coupling, the filter design is very flexible. The topology of third order Chebyshev type bandpass filter is shown on Fig.3.

The filter design methodology is described in [2]. Following this procedure, a third order filter is designed. The filter specification is:

Order: *3*;
Approximation: *Chebyshev type*;
Center frequency: $f_0 = 1080MHz$ **;**
Bandwidth: $\Delta f = 50MHz$ **;**
Ripple: $0.1dB$

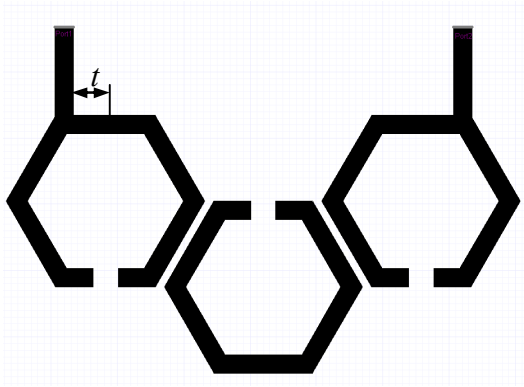


Fig.3 Filter topology

The coupling coefficients and the external quality factor of the bandpass filter are found to be:

$$k_{12} = k_{23} = 0.04 \quad (2)$$

$$Q_e = 20.79 .$$

Ansoft Designer SV is used for extraction the coupling lines parameters and the parameters of the hexagonal resonators.

Using FR-4 dielectric substrate with dielectric permittivity $\varepsilon_r = 4.5$, height $h = 1.5mm$ and tangent loss $tg\delta = 0.02$, the 50Ω hexagonal resonator dimensions are found to be $a = 13mm$, $s = 3.8mm$ and $w = 2.8mm$ for center frequency $f_0 = 1080MHz$. In comparison with it, the 50Ω halfwavelength resonator is $73.78mm$.

There are two resonant peaks in the response are observed if the coupled resonators are over-coupled, which occurs when the coupling coefficient is bigger than the value 1/Q, where Q is the quality factor of the resonator circuit. It is convenient to use full wave EM simulator to find the mode splitting resonance frequencies.

The coupling coefficient may be calculated by the resonance frequencies f_1 and f_2 as:

$$k = \frac{f_2^2 - f_1^2}{f_2^2 + f_1^2} \quad (3).$$

Fig.4 shows the dependence of the coupling coefficient against the distance between the coupled resonator topology shown on Fig2c. The coupling coefficient decreases with the increase of the distance between the resonators. This is due to the exponential decay of the fringe field against the distance. The coupling coefficient has a low value and decreases faster than the coupling coefficient between square open loop

resonators [2]. This makes the hexagonal open loop resonators more suitable for narrowband filters than the square open loop resonators.

For the design example, the distance between the resonators should be $d = 1.05mm$ in order to realize the coupling coefficient.

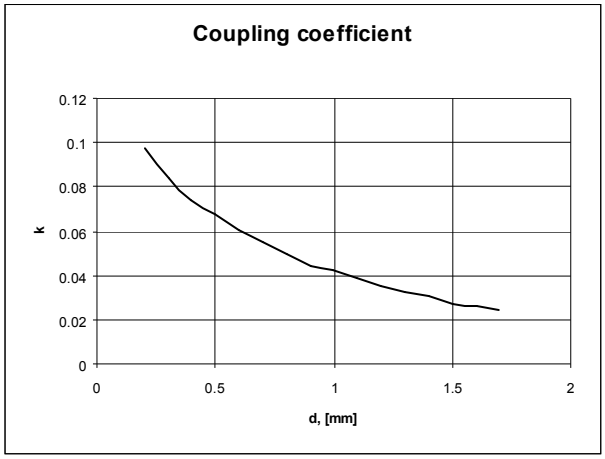


Fig.4 Coupling coefficient versus the distance between the resonators.

In order to determine the exact tapping position of the input/output lines, it is necessary to derive the external quality factor Q_e. The mathematical expression for it is [2]:

$$Q_e = \frac{f_0}{\Delta f_{\pm\frac{\pi}{2}}} \quad (4),$$

where f_0 is the resonance frequency of the resonator, and $\Delta f_{\pm\frac{\pi}{2}}$ is the bandwidth at which the phase of the s_{11} shifts to $\pm\frac{\pi}{2}, rad$ with respect to the phase at the resonance frequency.

Using Ansoft Designer SV, it is extracted the external quality factor. The results are shown on Fig.5.

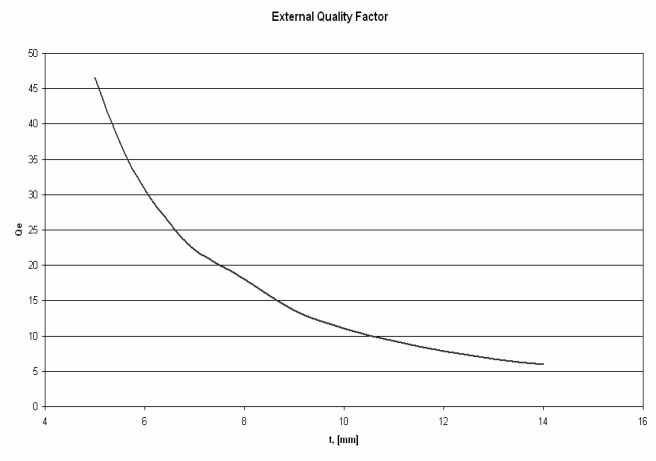


Fig.5 Coupling coefficient versus the distance between the resonators.

The external quality factor decreases with the distance from the center of the resonator. For narrow bandwidth filters Q_e is bigger and he tapping position is closer to the center of the resonator. This leads to bigger sensitivity of the filter's response against the tapping position. Consequently the passband reflection coefficient will degrade. The tapped position is found to be t=7mm.

A full wave EM simulation of the designed is performed in Ansoft Designer. The filter's passband response is shown on Fig.6.

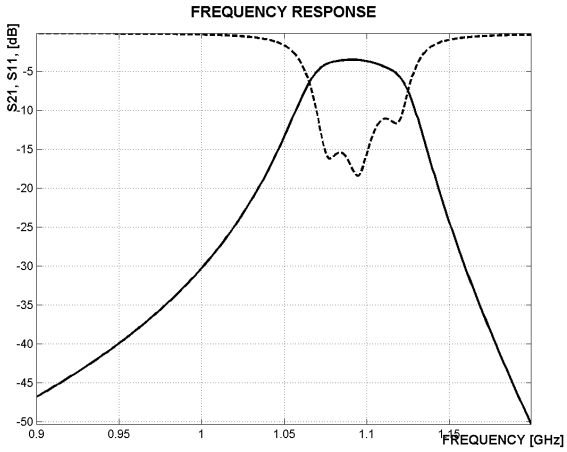


Fig.6 EM simulated filter's passband response. Solid line-S_{21}, dashed line- S_{11}

The designed filter size is reduced to the dimensions 0.497λ_g by 0.352 λ_g. The simulated 3dB passband is 47MHz at center frequency 1090MHz. It is clearly seen the three reflection zeroes in the passband to prove the Chebyshev response. The reflection coefficient in the passband is better than -11.4dB. The asymmetrical transmission coefficient is due to two factors-bad characteristics of the substrate FR-4 and the frequency dependent coupling coefficient. The minimum passband insertion loss is around 2.7dB. This is mainly due to the high dielectric loss of the substrate and the conductor loss.

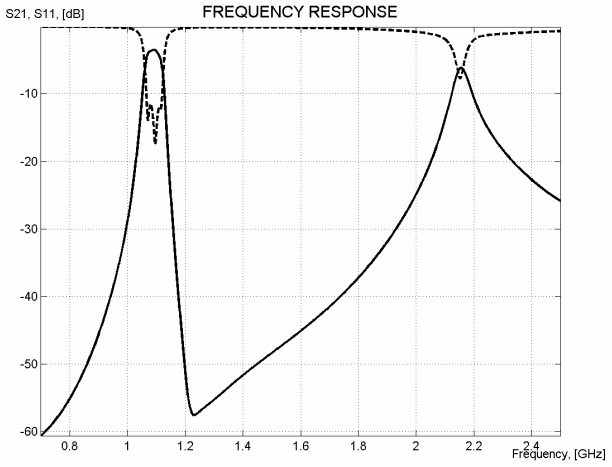


Fig.7 Wideband EM simulated filter's passband response. Solid line-S_{21}, dashed line- S_{11}

Fig.7 shows the wideband frequency response of the designed and EM simulated filter. As it is expected, the first spurious passband of the filter is around $2f_0$ as the hexagonal resonator length is half of the wavelength.

Further improvement of the reflection loss of the filter could be done by replacement of the tapped lines with coupled lines. Thus, a very loose coupling could be realized and the designed filters are narrow band.

IV. CONCLUSION

The paper demonstrates a design of third order hexagonal open loop resonator filter. The filter is of Chebyshev type with tapped input/output lines with center frequency 1080MHz and fractional bandwidth 4.7%. The size of the designed filter is two times shorter than the parallel edge coupled resonator filter, but remains the same selectivity. The coupling coefficient for mixed coupling and the external quality factor are extracted by EM simulation results. The paper presents the EM simulated responses of the designed filter. There is a very good agreement between the filter specification and the simulation results.

REFERENCES

[1] Morikazu Sagawa, Kenichi Takahashi, and Mitsuo Makimoto, "Miniaturized Hairpin Resonator Filters and Their ApplicationsTo Receiver Front-End MIC's," IEEE Transactions on Microwave Theory & Techniques, Vol. 37, No. 12, pp. 667-670, December 1989.

[2] J.S.Hong, Couplings of Asynchronously Tuned Coupled Microwave Resonators, IEE Proc. Microwave, Antennas, Propagation 147, Oct. 2000, pp.354-358

[3] J.S.Hong, Lancaster, M, Microstrip Cross-Coupled Trisection Bandpass Filters with Asymmetric Frequency Characteristics, IEE Proc.-Microwave, Antennas, Propagation 146, Feb, 1999, pp.84-90

[4] J.S.Hong, Lancaster, M, Theory and Experiment of Novel Microstrip Slow-Wave Open Loop Resonator Filters, IEEE Trans. On MTT-45, Dec.1997, pp.2358-2365

[5] J.S.Hong, Lancaster, M, Cross-Coupled Microstrip Hairpin-Resonator Filters, IEEE Trans on MTT-46, Jan.1998, pp.118-122

[6] Iliev, I, Nedelchev, M, CAD of Cross-Coupled Miniaturized Hairpin Bandpass Filters, ICEST, Nis, Yugoslavia, 2002

[7] Chang, Tam, Choi, Martins, Novel Quasi-Elliptic Microstrip Filter Configuration Using Hexagonal Open Loop Resonators, www.fst.umac.mo/en/staff/documents/rtorpm_pdf/43-SCAS2002.pdf

Microwave Absorbing Materials for Protection from Electromagnetic Radiation

R.I. Shtarkova[1], N.T. Dishovsky[2]

Abstract – **Elastomer-based microwave absorbers, intended for varied applications, especially for human protection from adverse influence of electromagnetic radiation, were developed. The influence of chemical character of the polymer matrix and chemical nature and concentration of some fillers with high values of the imaginary part of the complex dielectric permittivity and magnetic permeability on the microwave properties of the absorbers is investigated. Some more important microwave parameters of the absorbers as function of frequency and composition (mass ratio filler/rubber) are measured.**

Keywords - **Microwave absorber, Electromagnetic radiation Absorption active filler**

I. Introduction

As is well known radiofrequency (RF) and microwave (MW) radiation are electromagnetic radiation in the frequency ranges 3 kHz - 300 MHz, and 300 MHz - 300 GHz, respectively.

Electric and magnetic fields are complex physical agents whose potential health effects are the subject of much research [1-4]. Particularly controversial are the biophysical mechanisms by which these RF and MW fields may affect biological systems. General health effects reviews explore possible carcinogenic, reproductive and neurological effects. Health effects by exposure source are noted in radar traffic devices, wireless communications, broadcast communications, and industrial processes.

Spreading industrialization and increasingly powerful equipment raise issues about the health risks firstly to workers, then to the general public. At the same time, rapid technological advances in electronics, electro-optics and computer science have set the stage for an unprecedented drive towards improving existing medical devices and developing new ones. In particular, advances in RF/MW technology and computation techniques have paved the way for new treatments and diagnostic methods.

Today, more and more manufacturers are using microwave absorbing materials to enhance shielding performance at higher frequencies.

Microwave absorbers are specifically designed to attenuate or absorb microwave energy, and act as coatings with modified electrical or magnetic properties that allow absorption of microwave energy at discrete or broadband frequencies. Absorbers are used for radar cross section (RCS) reduction, electromagnetic interference (EMI) reduction, suppression of surface waves, and transmission line applications, as well as for radome lining and antenna side lobe reduction. On the other hand microwave absorbers are used or under study for protection of the environment and people from adverse influence of high frequency electromagnetic radiation.

Most of the contemporary microwave absorbers are produced of dielectric polymer matrix and specific functional fillers with high values of the imaginary part of the complex dielectric permittivity and magnetic permeability that absorb high frequency energy.

Since the single-layer absorber cannot own simultaneously a combination of broad-band and strong-absorption in the gigahertz (GHz) frequency band, multi-layer absorbing structures are employed to obtain effective microwave absorbers [5-7].

The aim of the paper is to present the results of our investigation on elastomer-based composites with different absorbing fillers developed for production of microwave absorbers, intended mainly for human protection from adverse influence of electromagnetic radiation.

II. Sample Preparation and Testing

Nitrile butadiene rubber (**NBR**) and chloroprene rubber (**CR**) were used as elastomeric matrix. Acetylene carbon black, carbonyl iron and natural magnetite were used as absorption active fillers. These fillers are accessible, inexpensive and very effective in a broad frequency range, as our investigations showed. The rubber-fillers mass ratio was shown in Table I.

TABLE I
BASIC COMPONENTS OF THE SAMPLES IN PHR

Basic components	Polysorb Fe	Polysorb Fe - C	Polysorb MB/S-MB
NBR	100	100	50
CR	-	-	50
Carbonyl iron	100; 300; 500	50; 100; 150; 300; 500	-
Acetylene carbon black	-	40	15; 55; 90
Natural magnetite	-	-	15; 110; 170

The fillers were compounded with rubber and the other obligatory ingredients in various proportions in a two-roll mill

[1]R.I. Shtarkova is with Technical University, 8 Kl. Ohridsky blvd, 1000 Sofia, E-mail: rchtarko@tu-sofia.bg
[2]N.T. Dishovsky is with University of Chemical Technology and Metallurgy, 8 Kl. Ohridsky blvd, 1756 Sofia, E-mail: dishov@uctm.edu

for samples **Polysorb Fe**, **Polysorb Fe-C** and **Polysorb MB** (Table I). Vulcanization was carried out by heating at 160°C for 10 min, using an electrically heated hydraulic press at a pressure of 10 MPa.

The material **Polysorb S-MB** is a liquid microwave absorber for elastic coatings. Acetone was used as dissolvent and the ratio dry mixture : dissolvent was 1:5. The vulcanization was going on at room temperature. These coatings were spreaded with a spraying pistol or with a brush, according to the circumstances.

The materials **Polysorb Fe - O** and **Polysorb Fe - W** are liquid microwave absorbers. **Polysorb Fe – W** is 40 % water dispersion of styrene acrylate copolymer and **Polysorb Fe – O** is based on an acrylate copolymer dissolved in a mixture of organic dissolvents (toluene and xylene). Carbonyl iron was used as an absorption active filler. These absorbers were prepared in two varieties:

Polysorb Fe – O – 15 **Polysorb Fe – W – 15**
Polysorb Fe – O – 30 **Polysorb Fe – W – 30,**

where the filler-polymer mass ratio was 15 phr and 30 phr respectively. It is possible to obtain coatings on the concentration gradient principle using these varieties in common.

Test specimens in different sizes according to the frequency of measurements were obtained. Some more important microwave parameters – absorption or attenuation, voltage standing wave ratio (VSWR) of the obtained rubber composites as function of frequency were measured using standard methods.

The waveguide measurements and the free-field measurements are very important for a comparative evaluation of the different absorbing materials. The measurement's methods were discussed in detail in our previous paper [8].

III. Results and Discussion

Our investigations indicated that the chemical nature of the elastomeric matrix had significant influence on the interaction between the electromagnetic waves and the absorbing composite (Fig. 1). Natural magnetite in concentration 100 phr was used as filler in all composites.

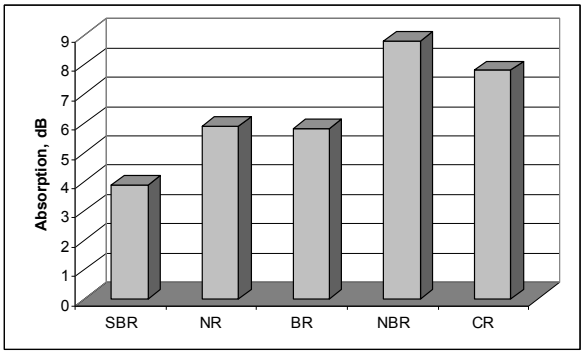


Fig. 1. Influence of the chemical nature of the elastomeric matrix on the absorption of electromagnetic waves at frequency 9, 4 GHz.

According to the obtained results the elastomers with polar functional groups or bonds (NBR, CR) determine better microwave properties of the absorbers.

Microwave absorbers intended for human protection from adverse influence of high frequency electromagnetic radiation have to absorb entirely electromagnetic energy. This is almost impossible to achieve with single-layer composite material. For that reason our investigations were directed mainly to develop multi-layer protective elastomeric coatings. In most cases the measurements were carried out in the frequency band 1-4 GHz, that is biologically significant.

Polysorb Fe and **Polysorb Fe – C** are flat, one-, two- and tree-layered materials made on gradient principle in respect to filler's concentration. The results of reflection coefficient measuring for tree different frequencies – 2, 3 and 4 GHz, are shown on Fig. 2.

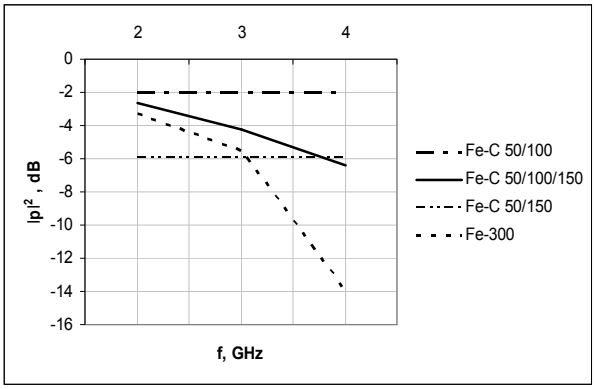


Fig. 2. Reflection coefficient frequency dependence for Polysorb Fe and Polysorb Fe – C

It is evident, that the best results for reflection coefficient are obtained for the material **Polysorb Fe – C 50 / 150** (constant value -5, 9 dB). Also good results showed **Polysorb Fe – 300 –** reflection coefficient -14 dB for the frequency 4 GHz, where as absorption active filler carbonyl iron was used. The filler – polymer mass ratio was 300 phr, but this filler is Bulgarian production and it did not raise the cost of material as a whole.

This group microwave absorbers are convenient to cover metal or nonmetal surfaces and rooms in order to decrease reflection of electromagnetic waves.

Microwave absorbers **Polysorb MB** are two- and multi-layered. An advantage for these absorbers is that the fillers (natural magnetite and acetylene carbon black) are accessible and inexpensive.

It is observed frequency dependence of the reflection coefficient for the 2-4 GHz band (Fig. 3). At 9,4 GHz the samples show reflection 88%, that means they are suitable for protective shields on metal and nonmetal surfaces.

Microwave absorbers **Polysorb MB** are fire safety (self-extinguishing) and they have good compatibility with decorative paints. The shield effect improves when using

combinations of coatings (**Polysorb MB 170/90** as base and **Polysorb MB 110/55** on it).

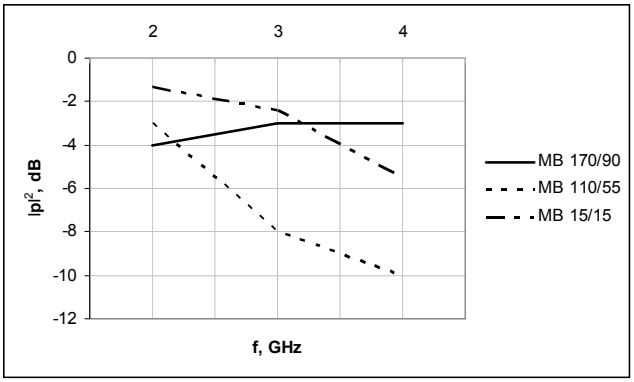


Fig. 3. Reflection coefficient frequency dependence for Polysorb MB

Rather effective broadband microwave absorber for frequency range 8-16 GHz is a multi-layered optimized structure on the base of **Polysorb MB** with concentration gradient of fillers (Fig. 4).

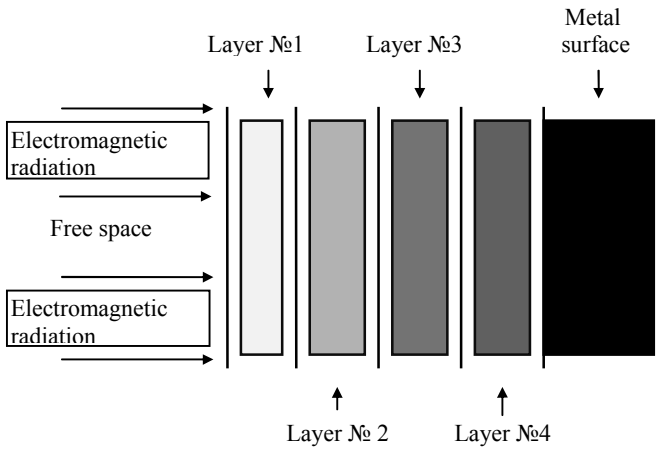


Fig.4. Multi-layered optimized structure on the base of **Polysorb MB**

Layer № 1 – 1,5 mm thick, **Polysorb MB** 15/15
Layer № 2 – 1,5 mm thick, **Polysorb MB** 110/55
Layer № 3 – 1,0 mm thick, **Polysorb MB** 170/90
Layer № 4 – 3,0-6,0 mm thick, **Polysorb MB** 15/15

This multi-layered optimized microwave absorber was measured at 7, 8 and 9,4 GHz. The total thickness of the structure was 7,35 mm. In Table II the results of voltage standing wave ratio (VSWR) and the reflection coefficient are given.

TABLE II
VSWR AND REFLECTION COEFFICIENT OF A MULTI-LAYERED STRUCTURE

f ,GHz	**VSWR**	**$\|p\|^2$, %**	**$\|p\|^2$, dB**	**$\|p\|^2$, dB/mm**
7	2,5	18	-7,36	0,5
8	1,8	8	-10,88	0,74
9,4	3,5	31	-5,11	0,35

The coatings **Polysorb S-MB** possess high adhesion to metal surfaces, they have water-proofing and anticorrosive effect. These coatings are elastic, weather-proof and their working temperature is from -45 ^{o}C to 110 ^{o}C.

To achieve the necessary thickness (1,5-2,0 mm) of the coatings **Polysorb S-MB,** as well as a gradient of filler's concentration, 3 – 7 layers from developed varieties of composites were used.

Free–field measurement at frequency 9,4 GHz of 0,5 mm thick sample was carried out. In Table III the results of this measurement are given.

TABLE III
RESULTS FROM FREE-FIELD MEASUREMENT OF POLYSORB S-MB SAMPLE

θ, deg	10	20	30	40	50	60
P_1, mW	22	22,8	23	11,7	10	10,8
P_2, mW	25	26,6	27,3	15,5	16	17,6
$R_{\varphi,\theta}$ dB	-0,56	-0,67	-0,74	-1,22	-2,04	-2,12

Very important distinctive feature of the liquid absorbing materials **Polysorb S – MB** is the possibility to soak-up fabrics (polyamide, polyester) and then to use them for individual protective clothing to shield workers from dangerous levels of electromagnetic radiation.

The results from measurement of **Polysorb Fe - O** and **Polysorb Fe – W** samples at frequencies 9 and 9,4 GHz are given in Table IV.

Advantage of microwave absorber **Polysorb Fe – O 30** is its low reflection of electromagnetic waves (at 9,4 GHz), when it is coated on metal surface. This means that it is convenient for antiradar camouflage coatings. Absorption coefficient is better for **Polysorb Fe – O 15/30** sample (Table IV), that is very important when the materials are coated on nonmetal surfaces.

TABLE IV
RESULTS FROM MEASUREMENT OF
POLYSORB FE-O AND FE-W SAMPLES

	Polysorb Fe – O 30	**Polysorb Fe – W 30**	**Polysorb Fe – O 15/30**	**Polysorb Fe – W 15/30**
f , GHz	9	9	9,4	9,4
VSWR	3,5	10	9	8
$\|p\|^2$, %	31	67	64	60
$\|p\|^2$, dB	-5,11	-1,74	-1,94	-2,18
$\|p\|^2$, dB/mm	1,28	0,44	0,28	0,27
α , dB	7	7	7,5	5,5

Microwave absorbers **Polysorb Fe – W** are convenient to use as protective shields when they are coated on metals. The compositions **Polysorb Fe – O** are suitable to coat on every solid surfaces, while **Polysorb Fe – W** are preferred to coat on walls and ceilings indoors.

V. CONCLUSION

The nitrile butadiene rubber would be the most suitable as dielectric matrix in preparation of elastomer based microwave absorbers, due to the highly polar nitrile group – C ≡ N in its macromolecules.

The effectiveness of some fillers with high dielectric or magnetic losses and their combinations was studied at different frequencies.

The two- and multi-layered microwave absorbers that contain combination of carbonyl iron and acetylene carbon black and combination of natural magnetite and acetylene carbon black are broadband and they ensure more effective protection of the human beings and environment from adverse influence of electromagnetic radiation.

REFERENCES

[1] "Health effects from radiofrequency electromagnetic fields", AGNR report, Doc. NRPB, vol. 14, no. 2, pp.1-177, 2003

[2] Review of the scientific evidence for limiting exposure to electromagnetic fields (0-300 GHz), Doc. NRPB, vol. 15, no. 3, pp.1-224, 2004

[3] J.C.Lin, "Human and environmental protection from electromagnetic emissions", IEEE Microwave Magazine, vol. 6, no.1, pp. 30-34, 2005

[4] *Radio-Frequency and Microwave Radiation*, Third Edition, American Industrial Hygiene Association (AIHA), 2004

[5] Y.He, R.Gong, X.Li, X.Wang, Q.Hu, "Design, fabrication and characteristic of two-layer microwave absorbers composed of magnetic micropowder – rubber composites in X-band frequency range", Europhysics Letters, vol. 77, no. 6, 2007

[6] Y.B.Feng, T.Qiu, C.Y.Shen, "Microwave Absorbing Materials", Journal of magnetism and magnetic materials", vol. 318, no. 1-2, pp. 8-13, 2007

[7] A.W.Guy, C.Chung-Kwang, J.A.McDougall, C.Sorensen, Measurement of shielding Effectiveness of Microwave Protective Suits", IEEE Trans., Microwave Theory and Techniques, vol. 35, no. 11, pp. 984-994, 1987

[8] R.I.Shtarkova, I.V.Petrova, N.T.Dishovsky, G.I.Krunev, "Magnetic Field Influence on Some Microwave Characteristics of Rubber Based Thin Films", ICEST'04, Conference Proceedings, pp. 225-228, Bitola, Macedonia, 2004

Block-based Analysis of Microstrip Structures with Stubs by use of 1D Wave Digital Approach

Biljana P. Stošić, Miodrag V. Gmitrović

Abstract – **The purpose of this paper is to show that the one-dimensional (1D) wave digital approach can be used in analysis of stub-line structures. A stub-line structure, divided into uniform sections, can be efficiently modeled by wave digital network. Frequency response is obtained by direct analysis of formed block-based network. Two application examples, proving the response accuracy of the new technique, are given.**

Keywords – **Wave digital approach, wave digital networks, microstrip circuits, stub-line structures, uniform segments.**

I. Introduction

Modeling of the planar structures by wave digital elements is based on well known theory of wave digital filters (WDF). A detailed review of WDF theory is given in references [1-4]. A large variety of WDF-based techniques has been developed for a wide range of applications [2]. Recently, microwave planar structures are modeled by one-dimensional (1D) [5-8], [10-15], and by two-dimensional (2D) [5], [9] wave digital elements (WDE). Standard WDE, such as delay, adder, multiplier and adaptor, are used in design of wave digital networks (WDN) which represent one signal flow diagram. 1D and 2D wave digital approaches can be efficiently used for analysis of these structures in both the time and the frequency domains.

There does not essentially exist just one type of microstrip structures, but a whole variety of quite distinct subclasses, each of which can again be divided into varies families, etc. For example, there exist stepped-impedance filters, linearly tapered nonuniform transmission lines, stub-line structures and elliptic filters. This reflects the richness of WDN, and the most appropriate one have to be chosen for structure at hand.

Till now, stepped-impedance filters and nonuniform structures with linearly tapered lines are analyzed by use of suggested 1D approach [7], [10-15]. Section II is devoted to some of the basic ideas of the 1D wave digital approach.

In the previously published paper [9] microstrip structure with stub is analyzed by use of 2D wave digital approach. In this paper, the authors will mainly be dealing with 1D wave digital approach. In fact, in view of their geometry, structures with stubs are also natural candidates for 1D approach. A new way for connecting segments in the structure with stubs is developed and represented in the Section III of this paper.

Biljana P. Stošić is with the University of Niš, Faculty of Electronic Engineering, Department of Telecommunications, Aleksandra Medvedeva 14, Niš, Serbia, (e-mail: biljana.stosic@elfak.ni.ac.yu)

Miodrag V. Gmitrović is with IMTEL-Communications Institute, Mihajlo Pupin Boulevard, 165b, Belgrade, Serbia, (e-mail: gmitrovic@insimtel.com, miodrag.gmitrovic@elfak.ni.ac.yu)

II. The Basic Ideas of the 1D Wave Digital Approach

A theoretical model for the modeling of the microstrip stepped-impedance structure as well as one type of regular discontinuity (step) is described in the papers [5-7], [10-15]. A nonuniform structure has to be divided into cascade-connected uniform transmission lines (UTL segments). A lossless uniform transmission line is modeled by a two-port digital element with a delay appears in forward path. This wave digital two-port is called the unit element (UE) [2]. The port resistances of the UE are equal and correspond to the characteristic impedance of UTL segment. The connection of two UE with different port resistances is achieved by two-port series adaptors (TA), [3]. A microstrip structure, divided into cascade connection of uniform sections, can be efficiently modelled by wave digital networks. The WDN is a model of the microstrip structure modeled by wave digital elements, and it is composed of two types of building blocks UE and TA.

In the complex microstrip structures, delays of the transmission lines vary from one another and because of this, each transmission line has to be represented as a cascade connection of a certain number of UE. A way of determination a minimal section numbers in WDN of complex microstrip structure is based on the given relative error [10], [14], [15]. Appropriate choice of a minimal section number in that model is very important because of the direct influence on the sampling frequency of that digital model, and on accuracy of the desired response. Also, it is very important to achieve a good compensation of the effects of identified step discontinuities.

Response in WDN of stepped-impedance structure can be found in two different ways. In the first way, efficient and very simple algorithms for calculating transmission and input reflection coefficients of WDN are described in the papers [7], [8]. The algorithms are very easily implemented in the MATLAB environment. The analysis of the wave digital network is efficiently automated, which is inevitable when structures with large numbers of building blocks are to be dealt with. Response can be counted in frequency or in time domain directly from known network function in z-domain.

In the second way, direct analysis of WDN is used. A very simple method of analysis of the WDN is a block-diagram method. WDN is formed directly in the Simulink toolbox of the MATLAB environment [16], [17]. Signal flow diagrams are the basis for block-oriented simulation programs such as Simulink. Response is obtained directly in the time domain, and Fourier transformation is used for frequency response calculation.

III. MODELING OF THE STRUCTURE WITH STUB

If 1D approach is used for analysis of structures with stubs, a new way for connecting UTL segments in that structures have to be developed. In order to better explain the approach, one microstrip structure with one stub, depicted in Figure 1*a*, is used. This structure can not be observed as simple cascade connection of UTL segments, these segments have to be connected as depicted in Figure 1*b*.

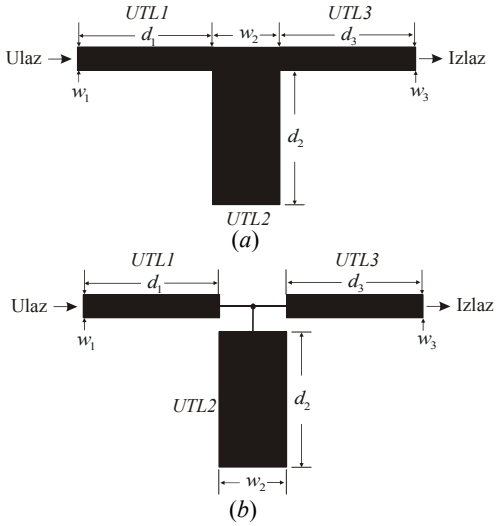


Fig.1. Stub-line microstrip structure:
(a) layout and (b) segment connections

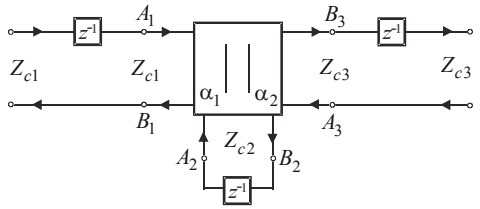


Fig.2. Network of the stub-line structure

A wave digital nertwork which corresponds to the segment connection shown in Figure 1*b* is depicted in Figure 2. In this network, the effects of discontinuity are not considered. Each uniform segment in the Figure 1*b* can be approximated with one transmission line of characteristic impedance Z_{ck}, $k = 1,2,3$. Models of UTL segments are connected by use of one three-port parallel adaptor with port 2 being dependent. Adaptors are memoryless devices whose task is to perform transformations between pairs of wave variables that are referred to different levels of port resistance. In the symbolic representation of the three-port parallel adaptor [4] given in Figure 3, they are shown explicitly the adaptor coefficients α_1 and α_2 next to the ports 1 and 3, respectively. These adaptor coefficients are

$$\alpha_j = \frac{2G_j}{G_1 + G_2 + G_3}, \qquad j = 1,2, \tag{1}$$

where port conductances are

$$G_k = 1/Z_{ck}, \qquad k = 1,2,3. \tag{2}$$

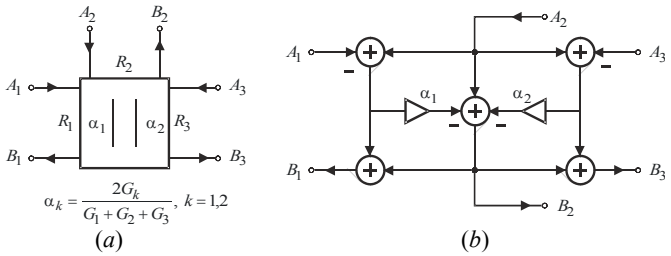


Fig.3. Three-port parallel adaptor with port 2 being dependent:
(a) symbol and (b) WDE

Segment UTL2 which represent open stub is connected to adaptor's dependent port 2. All UTL segments are modeled by several cascaded UE. In case of equal delays of UTL segments, each segment is modeled by one UE as shown in Figure 2.

The formed Simulink model of WDN is depicted in Figure 4. The blocks **TLine_1**, **Stub_2** and **TLine_3** represent UTL segments. The blocks **ADP-S** and **ADP-L** represent two-port series adaptors, and block **ADP_T1S2T3** three-port adaptor. The two-port adaptors at the ends are used for matching source and load resistances to the rest of the WDN. In the symbolic representation of a two-port series adaptor [1-4] given in Figure 5, it is shown explicitly the parameter α next to the port 1. The adaptor coefficients for the blocks **ADP-S** and **ADP-L** in the WDN are

$$\alpha_S = (R_S - Z_{c1})/(R_S + Z_{c1}), \tag{3}$$
$$\alpha_L = (Z_{c2} - R_L)/(Z_{c2} + R_L).$$

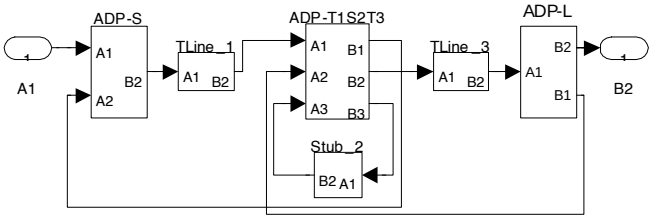


Fig.4. Simulink model of the structure

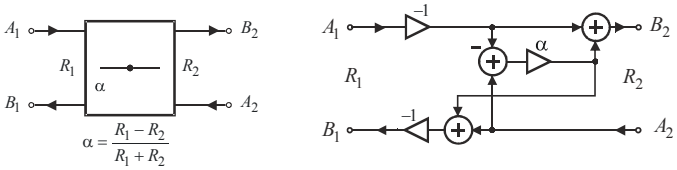


Fig.5. Two-port series adaptor: (a) symbol and (b) WDE

IV. EXAMPLES OF APPLICATION

In this section, some simple examples of stub-line microstrip structures are analyzed by use of suggested 1D approach.

A. Microstrip Structure with a Single Stub

A microstrip structure with a single stub, depicted in Figure 1, is used for verification of the proposed method. The substrate dielectric constant is $\varepsilon_r = 2.32$ and the board thickness $h = 1.58\,mm$.

TABLE I
TRANSMISSION LINE PARAMETERS

nv	d [mm]	w [mm]	Zc [Ohm]	Tv [ps]
1	30.0000	4.7100	50.2540	139.8003
2	30.0000	15.7600	20.0016	145.0746
3	30.0000	4.7100	50.2540	139.8003

Microstrip structure given in Figure 1 is approximated by connection of 3 transmission lines with parameters given in Table I.

For given error of $n\_er = 0.01\%$, a total minimal number of sections in *WDN* is $n_t = \sum_{k=1}^{3} n_k = 161$, [10], [14]. For segments UTL1 and UTL3, a number of sections n_k is 53, and for segment UTL2 is 55. A total delay for the digital model of the structure is $T_t = n_t \cdot T_{\min} / q = 424.6763\ ps$ where $q = 53$ is a multiple factor and $T_{\min} = \min\{T_1, T_2, T_3\} = 139.8003\ ps$ is a minimum delay. A total real delay of the structure is $T_\Sigma = \sum_{k=1}^{3} T_k = 424.6752\ ps$. A sampling frequency of the digital model of the planar structure for the chosen minimal number of sections is $F_s = n_t / T_t = 379.1123\ GHz$. In this case, a relative error of delay is $er = \frac{T_\Sigma - T_t}{T_\Sigma} \cdot 100 = -0.00027\ \%$. According to the relations (1)-(3), adaptor coefficients are $\alpha_1 = \alpha_2 = 0.4432$ and $\alpha_S = -\alpha_L = -0.0025$.

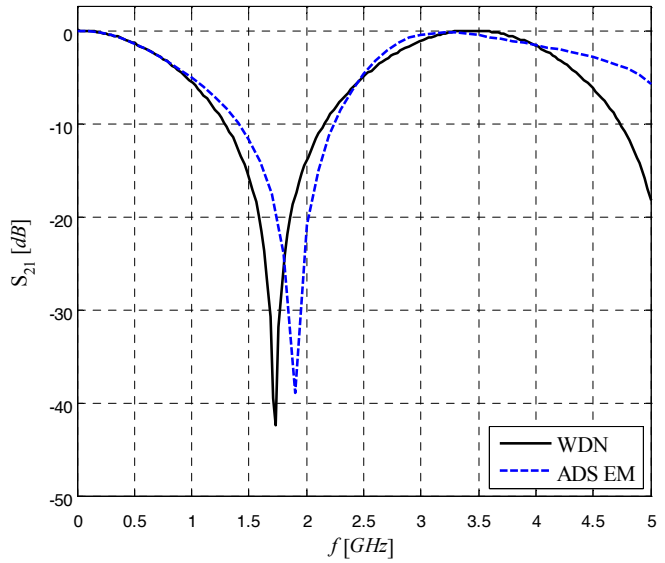


Fig.6. Frequency response

Response comparison for the proposed way of modeling structrure with stub by use of 1D approach is shown in Figure 6. It is clear that the curve corresponding to the WDN result is slightly shifted to the left comparing to the one of electromagnetic simulation done in ADS [18]. If the effects of the T-junction discontinuty are treated, the agreement between the results is going to be better. That will be considered in the future.

B. Microstrip Structure with Two Stubs

Consider a microstrip structure with two stubs depicted in Figure 7. The substrate dielectric constant is $\varepsilon_r = 9.9$ and the board thickness $h = 127\ \mu m$. Metalization is cooper and the strip thickness is $t = 36.068\ \mu m$. The observed structure is treated as a specific connection of 5 transmission lines with parameters given in Table II.

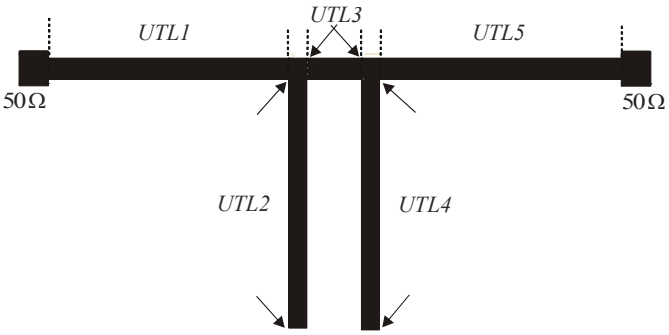


Fig.7. Layout

TABLE II
TRANSMISSION LINE PARAMETERS

nv	d [mm]	w [mm]	Zc [Ohm]	Tv [ps]
1	2.8778	0.1308	41.7076	25.1683
2	2.9210	0.1308	41.7076	25.5459
3	0.3810	0.1308	41.7076	3.3321
4	2.9210	0.1308	41.7076	25.5459
5	2.8778	0.1308	41.7076	25.1683

For given error of $n\_er = 0.01\%$, a total minimal number of sections in *WDN* is $n_t = 1006$, [10], [14]. For UTL segments, a number of sections n_k is 242, 245, 32, 245 and 242, respectively. A total delay for the digital model of the structure is $T_t = 104.7521\ ps$ where $T_{\min} = 3.3321\ ps$ and $q = 32$. A total real delay of the structure is $T_\Sigma = 104.7605\ ps$. A sampling frequency of the digital model of the planar structure for the chosen minimal number of sections is $F_s = 9603.6215\ GHz$. In this case, a relative error of delay is $er = 0.00795\ \%$. According to the relations (1)-(3), adaptor coefficients are $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.6667$ and $\alpha_S = -\alpha_L = 0.0904$.

The result obtained by direct analysis of WDN and the ones obtained by linear simulations in the programs GENESYS [19] and ADS [18] are depicted in Figure 8. From the given responses, it can be inferred that the curve corresponding to WDN is slightly shifted to the left, and the curve corresponding to ADS result to the right, comparing to GENESYS curve.

In this example there are two types of discontinuities: T-junction and open-end. Here, the effects have not been taken under consideration. But, when the discontinuities are identified in the structure, they have to be treated. Only in that case, WDN curve is going to have better agreement with other results.

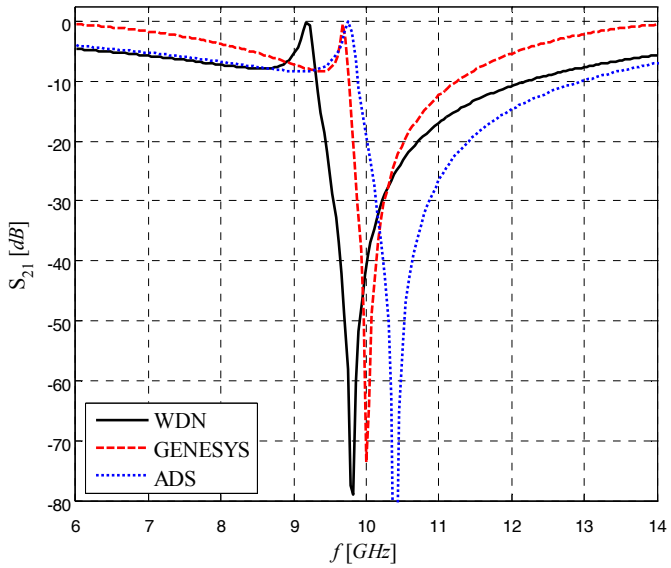


Fig.8. Frequency response

V. Conclusion

Approaches for frequency and time analyses of stepped-impedance microwave structure based on wave digital filter theory are described in references [5-8], [10-15]. The microwave structure is treated there as a cascade connection of uniform transmission lines with various lengths and widths.

Here, it is shown that the 1D wave digital approach used for analysis of microstrip stepped-impedance structures can be also used in analysis of stub-line structures. A new way for connection of identified UTL segments is described. That leads to a universal and effective procedure capable of solving a wide range of practical problems.

The efficiency and accuracy of the suggested procedure is shown in two examples realized in the microstrip line technique, such as structure with a single stub and with two stubs. The results of the analysis obtained by the WDN have shown good agreement with those obtained by other programs mentioned above. A much better agreement can be achieved by modeling discontinuities in the observed structure.

A very simple block-diagram method of analysis of the WDN is used here. WDN is formed directly in the Simulink toolbox of the MATLAB environment. Frequency response is obtained by direct analysis of formed block-based network.

References

[1] A. Fettweis, "Digital Circuits and Systems", *IEEE Transactions on Circuits and Systems*, Vol. CAS-31, No. 1, January, 1984, pp. 31-48.

[2] A. Fettweis, "Wave Digital Filters: Theory and Practice", *Proc. IEEE*, Vol. 74, 1986, pp. 270-327.

[3] W. K. Chen, *The Circuits and Filters Handbook*, CRC Press, 1995 (Wave Digital Filters, pp. 2634-2661).

[4] M. V. Gmitrović, *Microwave and Wave Digital Filters*, Faculty of Electronic Engineering, Niš, 2007 (in Serbian).

[5] B. P. Stošić, *Analysis of Planar Microwave Structures Modeled by Wave Digital Elements*, Doctoral thesis, Faculty of Electronic Engineering, University of Niš, Niš, September 2008 (in Serbian).

[6] M. V. Gmitrović and B. P. Stošić, "Analysis of Planar Structures Modeled by Wave 1D Digital Elements", *14th Telecomm. forum - TELFOR 2006*, Serbia, Belgrade, November 21-23, 2006, pp. 418-421.

[7] B. P. Stošić and M. V. Gmitrović, "Implementation of Wave Digital Model in Analysis of Arbitrary Nonuniform Transmission Lines", *Microwave and Optical Technology Letters*, Vol. 49, No. 9, September 2007, pp. 2150-2153.

[8] B. P. Stošić and M. V. Gmitrović, "Equivalent Thevenin Source Method as Tool for Response Computation of Wave Digital Structures", *8th Inter. Conference on Telecommunications in Modern Cable, Satellite and Broadcasting Services - TELSIKS 2007*, Serbia, Niš, September 26-28, 2007, Vol. 1, pp. 203-206.

[9] B. P. Stošić and M. V. Gmitrović, "Generating of Basic Wave Digital Elements for Modeling of Two-dimensional Planar structures", *XLII International Scientific Conference - ICEST 2007*, Macedonia, Ohrid, June 24-27, 2007, pp. 309-312.

[10] B. P. Stošić and M. V. Gmitrović, "Choice of Section Number in Wave Digital Model of Microstrip Structure",*52nd Conference ETRAN*, Serbia, Palić, June 8-12, 2008, MT1.1-1-4 (in Serbian).

[11] M. V. Gmitrović and B. P. Stošić, "Analysis of Wave Digital Model of Microstrip Structure with Step Discontinuity", *52nd Conference ETRAN*, Serbia, Palić, June 8-12, 2008, MT1.2-1-4 (in Serbian).

[12] B. P. Stošić and M. V. Gmitrović, "Modeling of Step Discontinuity in Microstrip Structures by using Wave Digital Approach", *XLIII Intern. Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST 2008*, Serbia, Niš, June 25-27, 2008, pp. 347-350.

[13] B. P. Stošić and M. V. Gmitrović, "Wave Digital Approach – A Theoretical Model of Step Discontinuity", *16th Telecomm. forum - TELFOR 2008*, Serbia, Belgrade, November 25-27, 2008, pp. 539-542.

[14] B. P. Stošić and M. V. Gmitrović, "Direct Analysis of Wave Digital Network of Microstrip Structure with Step Discontinuities", *The 7th WSEAS International Conference on System Science and Simulation in Engineering – ICOSSSE'08*, Italy, Venice, November 21-23, 2008, pp. 25-29.

[15] B. P. Stošić and M. V. Gmitrović, "Wave Digital Approach - A Different Procedures for Modeling of Microstrip Step Discontinuities", *International Journal of Circuits, Systems and Signal Processing*, Issue 3, Volume 2, 2008, pp. 209-218, http://www.naun.org/journals/circuitssystemssignal/

[16] MATLAB - The Language of Technical Computing, Version 7.6.0.324 (R2008a), February 10, 2008, The MathWorks Inc. 1984-2008.

[17] A. Gilat, *MATLAB – An Introduction with Applications*, John Wiley & Sons, 2005.

[18] Advanced Design Software 2002, Agilent Technologies 1983-2002, 395 Page Mill Road, Palo Alto, CA 94304, USA.

[19] RF and Microwave Design Software *GENESYS*, Eagleware Corporation, 635 Pinnacle Court, Norcross, GA 30071, 2001.

Session

ANTENNAS AND PROPAGATION

Theory Approach and Method for Linear Antenna Array Design with Improved Selectivity

Peter Apostolov[1]

Abstract – **In the article a theory for design of linear antenna array, based on approximation of "shifted" delta function, is shown. An appropriate basis function, which leads to obtaining narrow beam pattern, is offered. An analysis of antenna array's parameters and comparison with such of Dolph-Chebyshev's and Riblet's is done. An idea for practical realization of antenna array is given. With offered method, antenna array with better selectivity than those of Dolph-Chebyshev's and Riblet's, can be obtained.**

Keywords – **Approximation, Polynomial, Antenna array, Pattern, Phase response.**

I. INTRODUCTION

Antenna array are used for direct transmitting or receiving of electromagnetic power to definite direction. They consist of n equal uniform set elements. Basic parameters of antenna array are:

- general array shape (linear, circular, planar, etc.);
- element spacing d normalized against wave length λ ;
- element excitation amplitude I_n ;
- element excitation phase φ_n ;
- pattern of array element.

The radiation of antenna array is defined by the pattern multiplication theorem [1]:

Array pattern = Array element pattern x Array Factor.

The array Factor depends of general array shape, element excitation amplitude and element excitation phase. For linear antenna array it has the form of the function $\sin(x)/x$. The synthesis procedure reduces to obtaining the array factor.

Object of investigation in this article are linear array antennas with narrow-beam, low-sidelobe pattern. The most popular methods [2] are Schelkunoff's, Dolph–Chebyshev's Taylor's, Villeneuve's, Orchard's etc. Dolph–Chebyshev's method and its Riblet's modification [3] are with best properties.

In this article a method for linear antenna array design with odd number elements, which has better parameters then above mentioned is offered.

II. THEORETICAL ESTABLISHMENT

The theoretical base of the method is the Alternation Theorem:

- If a function $f(x)$ is defined and continuous in the closed definition area it can be approximated by trigonometric polynomial $P_n(x)$ by power n, with basis function $\cos(x)$.
- The polynomial is the unique and best approximation, if the error function $w(x) = f(x) - P_n(x)$ has least $n+2$ extremes in the definition area.
- All extremes are alternatively and theirs modules are equal at positive number ε .

The idea of the method is an approximation of "shifted" delta function to be done

$$\delta(\theta) = \begin{cases} 1, & \theta = \pi/2 \\ 0, & \theta \neq \pi/2 \end{cases}; \theta \in [0, \pi] \quad (1)$$

in the interval $[0, \pi]$ with polynomial, containing basis function

$$y(\theta) = \cos\left[nC \arctan(kd\cos\theta) - \frac{n\pi}{2} \right]. \quad (2)$$

In the above equation $k = 2\pi/\lambda$, θ is elevation angle between axis of antenna array and the direction of transmitted/received signal. The factor C is defined by

$$C = \frac{\pi}{2}\arctan(kd). \quad (3)$$

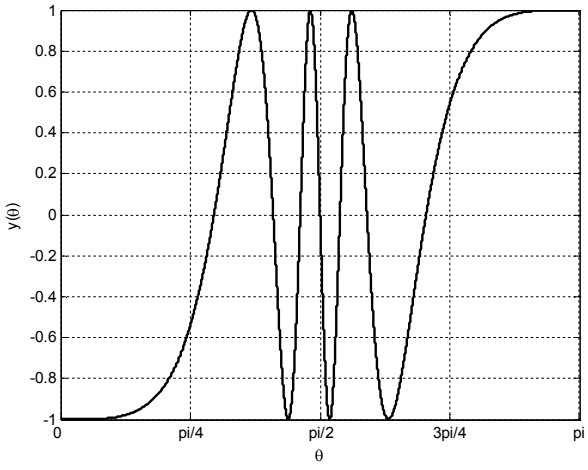


Fig.1. Basis function $n = 7$, $d = \lambda/2$

On Fig.1 is shown the function graphics by $n = 7$ and $d = \lambda/2$. From the figure it is seen that the function thickens its oscillations in the middle of the definition area, where the mainlobe lies. This property makes it appropriate for obtaining a narrow beam patterns. The polynomial is obtained

[1] Assoc. Prof. Peter Stoyanov Apostolov, Phd, Institute for Special Technical Equipment-MI, Sofia, e-mail: p_apostolov@abv.bg

by Remes' algorithm. It comprises an iterative computing of a system of $n+2$ linear equations. The obtained $n+2$ solutions are $n+1$ coefficients of the desired polynomial of power n, and the approximation error ε. The polynomial looks like this

$$(4)\quad P_n(\theta)=\sum_{m=1}^{n+1} b_m \cos\left[mC\arctan\left(kd\cos\theta\right)-\frac{m\pi}{2}\right].$$

On Fig. 2 the approximation of "shifted" delta function with polynomial of power 7 is shown.

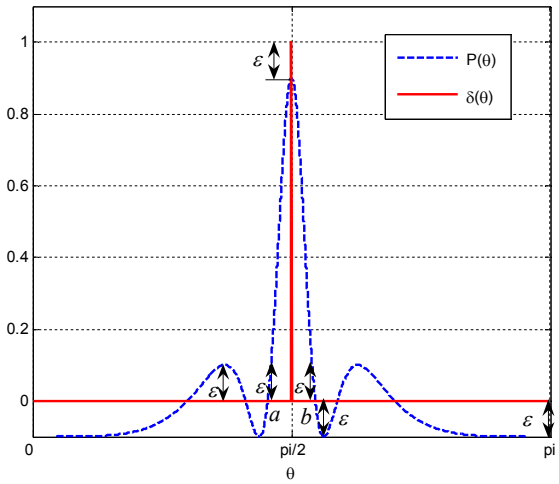


Fig.2. Approximation of "shifted" delta function with $P_7(\theta)$

The approximation is optimal, because is equiripple outside the mainlobe.

Excitation amplitude currents of the antenna array are obtained by nonzero coefficients of polynomial in following sequence

$$(5)\quad I_n=\frac{|b_{n+1}|}{2},\frac{|b_n|}{2},\ldots,\frac{|b_2|}{2},|b_1|,\frac{|b_2|}{2},\ldots,\frac{|b_n|}{2},\frac{|b_{n+1}|}{2}.$$

The final form of array factor has complex function and looks like this

$$(6)\ A_n(\theta)=\sum_{m=1}^{n} I_m \exp\left[j(n-m)C\arctan\left(kd\cos\theta\right)-j(n-m)\frac{\pi}{2}\right].$$

Its module defines the antenna array's pattern and the argument – its phase response.

III. Analisis Antenna Array's Parameters. Comparison with Dolph-Chebyshev's and Riblet's Antenna Arrays

On Fig.3 antenna array's pattern with following initial data: number of elements $n=5$, element spacing $d=\lambda/2$; sidelobe attenuation $R=-20\text{dB}$ is shown. Normalized excitation amplitude currents are: $I_1=1$; $I_2=3.0087$; $I_3=4.1402$; $I_4=3.0087$; $I_5=1$. The circle with dashed line defines the angle $\Delta\theta_{3\text{dB}}$ of the pattern at the level -3dB. In the particular case $\Delta\theta_{3\text{dB}}=16.431^0$. In the Remes' algorithm, for every iteration are assigned $n+2$ points from the definition area of the argument θ, for which the graphics of approximation function must be at ε distance from approximated. This means, that $\Delta\theta_{3\text{dB}}$ can be varied by appropriate selection of the two closest points near the middle of the definition area $\pi/2$. On Fig.2 they are signed by a and b. The shrinking of the angle leads to increasing of the sidelobe level.

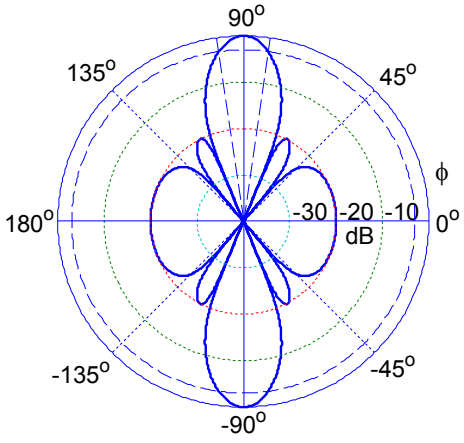


Fig.3. Offered method: $n=5$; $d=\lambda/2$; $R=-20\text{dB}$

Increasing the array element spacing d leads to significant improvement of its selectivity – shrinks the mainlobe and decreases the level of the sidelobes. On Fig.4 the pattern of array when $d=\lambda$ is shown.

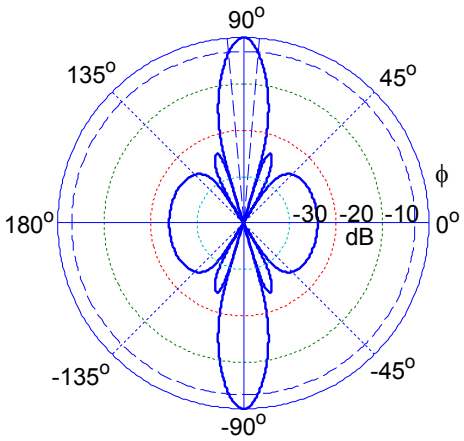


Fig.4. Offered method: $n=5$; $d=\lambda$; $R=-23.98\text{dB}$

In this case $R=-23.98\text{dB}$, and $\Delta\theta_{3\text{dB}}=9.648^0$. Normalized excitation amplitude currents are: $I_1=1$; $I_2=3.3463$; $I_3=4.746$; $I_4=3.3463$; $I_5=1$. The decreasing d leads to opposite effect.

On Fig.5 the pattern of Dolph-Chebyshev's antenna array with the same initial data like this on Fig.3 is shown.

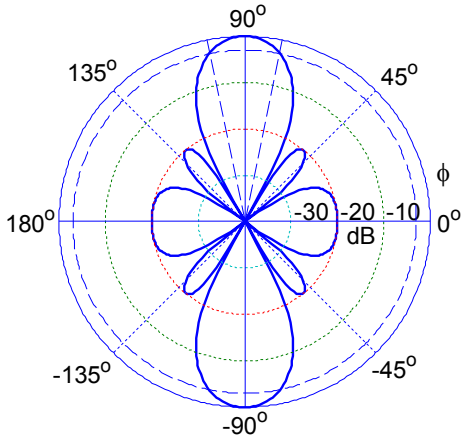


Fig.5. Dolph-Chebyshev: $n = 5$; $d = \lambda/2$; $R = -20\text{dB}$

The angle $\Delta\theta_{3\text{dB}} = 23.538^0$ is with 7.1^0 wider from this on the pattern from Fig.3. In this case with the offered method a narrower beam with 30.2% is obtained.

The Riblet's antenna arrays are with odd number of elements and have improved selectivity for $d < \lambda/2$. For higher values of d they have Dolph-Chebyshev's pattern. On the following two figures the patterns of antenna arrays, designed by the offered method and the Riblet's with initial data: $n = 5$; $d = \lambda/4$; $R = -20\text{dB}$ are shown.

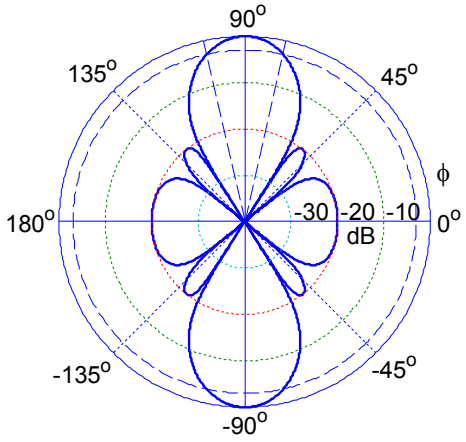


Fig.6. Offered method: $n = 5$; $d = \lambda/4$; $\Delta\theta_{3\text{dB}} = 25.632^0$

From the comparison can be seen that the pattern of Fig.6 is with better selectivity. The difference between the angles $\Delta\theta_{3\text{dB}}$ is 7.362^0. In this case with the offered method a narrower beam with 22.3% is obtained.

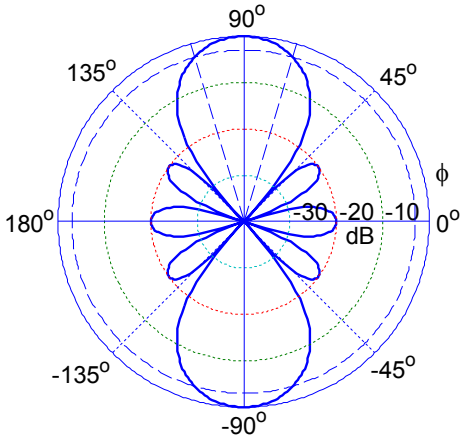


Fig.7. Riblet's method: $n = 5$; $d = \lambda/4$; $\Delta\theta_{3\text{dB}} = 32.994^0$

IV. Practical Realisation

On Fig. 8 an equivalent configuration for the determining of array factor is shown.

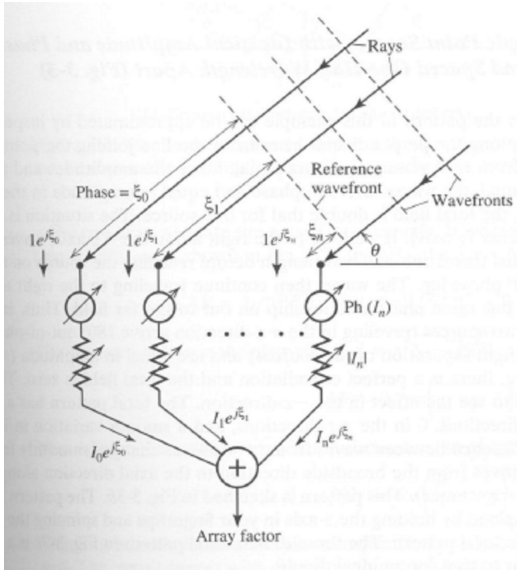


Fig.8

In previous part it was shown that the excitation amplitude currents $|I_n|$ are given by Remes' algorithm. The currents distribution realization is obtained with variable attenuators. Their signal attenuation will be inversely proportional to the excitation currents relation. From the equation (6) can be seen, that the function presenting phase's change takes the form

$$(7)\quad \varphi(\theta) = C\arctan(kd\cos\theta) - \frac{\pi}{2} = C\arctan(\psi) - \frac{\pi}{2},$$

where $\psi = kd\cos\theta$. On Fig. 9 a function graphic's is shown.

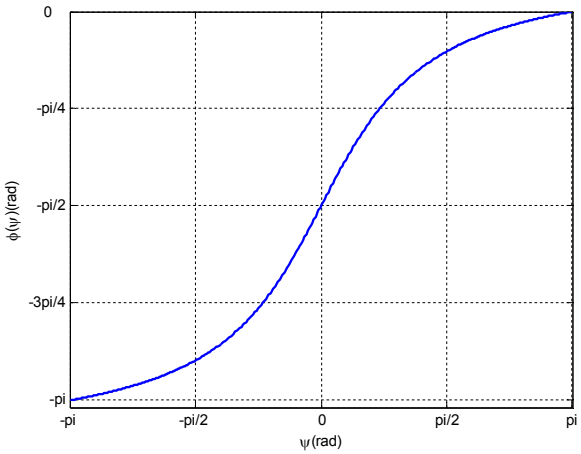


Fig. 9. Phase response

The implementation of the phase function can be done with phase modulators. It must be applied to modulation inputs simultaneously, with scanned frequency, a signal proportional to the change of angle $\psi = kd\cos\theta$.

V. CONCLUSION

With the offered theory and method for linear array antenna design optimal patterns with high selectivity are obtained. That is due to the specific properties of the offered basis function and the use of Remes' algorithm. Methods, with better properties than those based on Dolph-Chebyshev and Riblet can not be found in literature and that is why the offered theory is original. The theory and the method enrich the knowledge in the domain of antenna arrays synthesis and can find practical application.

REFERENCES

[1] Hubregt, J. Visser. Array and Phased Array Antenna Basics. John Wiley & Sons, Inc., 2005.
[2] Milligan, Thomas A. Modern Antenna Design. Hoboken, NJ, John Wiley & Sons, Inc., 2005.
[3] Orphanidis, Sophocles. Electromagnetic Waves and Antennas. Rutger University, 2008.

Investigation on Millimeter Waves Usage in Hybrid FSO/RF Communication Systems

Boncho Bonev[1] and Kliment Angelov[2]

Abstract **In this paper the comparative analysis of radio wave attenuation at frequencies of 60 GHz, 93 GHz and 77 GHz has been taken into consideration. The maximum rain range admitting radio connection has been estimated. The possibility of using the radio waves with 77 GHz frequency in hybrid FSO/RF communication systems has been investigated.**

Keywords – **Hybrid FSO/RF, Rain attenuation, Atmospheric gaseous attenuation.**

I. Introduction

Free-Space Laser Communications (Free-Space Optics – FSO) are subject of increasing interest in last ten years. This is caused to a great extent by their technical and economical advantages over cables and microwave networks – quick and easy installation, portability, relatively low price and high speed information flows. Furthermore through FSO the "last mile" problem is resolved.

Despite these advantages, the utilization of FSO is related to some important problems, which decision is subject to a number of researches [1–5]. First of all these problems are due to influence of atmospheric extinction especially in fog and clouds [6–8] and influence of atmospheric turbulence and mechanic vibration of optical antennas [9–12] on link quality of FSO communication systems. Investigation on effect of turbulent medium on laser beam propagation is subject of many publications, where authors suggested some methods and techniques for reducing of this effect – using of partially coherent laser beam [13, 14] or different methods of coding, modulation and detection [15–19]. Decreasing of antennas vibrations influence can be achieved with tracking systems [20]. That leads to significant improvement of link quality on FSO even in case of long distance communication lines. Therefore the main limiting factor on link range of FSO remains atmospheric transmittance [2, 6].

This problem can be resolved with backing up the FSO with RF communication line, which will take up communication when optical connection is down. That is possible because of insignificant attenuation of radio waves in fog and clouds.

The operating frequency of 2,4 GHz for RF part of hybrid FSO/RF communication system is frequently discussed in free-space laser communications literature [21, 22]. The Earth atmosphere is actually transparent for radio waves with this frequency and RF part of hybrid system can work in all kinds of weather conditions. However in this case the communication speed is significantly lower than on FSO. In [23] is examined hybrid FSO/RF system with radio line working on 60 GHz. This frequency is unlicensed in some countries (USA, Canada), because of maximum of oxygen absorption. Some authors are investigated the possibility of using 93 GHz RF line for hybrid systems [24]. Where using these frequencies of MMW band the speed of information flows can be much greater than in 2,4 GHz radio line, but radio waves suffer on significant rain and atmospheric gases attenuation. That cause decreasing of communication links range.

In this paper the comparative analysis of radio wave attenuation at frequencies of 60 GHz, 93 GHz and 77 GHz has been taken into consideration. The possibility of using the radio waves with 77 GHz frequency in hybrid FSO/RF communication systems has been investigated.

II. Theoretical Analysis

Let RF line operates on frequency f, respectively on wavelength λ and has a link range of d. For received power P_r in dBm can be written

$$P_r = P_t + G_t - L + G_r , \quad (1)$$

where P_t is transmitters power in dBm, G_t is transmitters antenna gain in dB with transmission line losses at transmitting terminal included, G_r is receivers antenna gain in dB with transmission line losses at receiving terminal included and L are total propagation losses in dB.

$$L = L_{FS} + L_{rain} + L_{atm} , \quad (2)$$

where

$$L_{FS} = 20\lg\frac{4\pi d}{\lambda} = 20\lg\frac{4\pi df}{c} \quad (3)$$

is free space loss, L_{rain} are rain attenuation losses and L_{atm} is atmospheric gaseous absorption loss. In Eq. (3) $c = 3.10^8\ m/s$ is light velocity.

Rain attenuation losses are given by relation

$$L_{rain} = L_{sp\_r}.d/1000 , \quad (4)$$

where L_{sp_r} is specific rain attenuation in dB/km and d is distance in m. L_{sp_r} according [25] can be calculates as follow

$$L_{sp\_r} = a(f).I^{b(f)} , \quad (5)$$

[1]Boncho Bonev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: bbonev@tu-sofia.bg.

[2]Kliment Angelov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kliment.angelov@gmail.com.

where I is rain rate in mm/h, $a(f)$ and $b(f)$ are frequency and polarization depending constants.

Atmospheric gaseous absorption losses are due to oxygen and water vapor absorption and are specified as [26]

$$L_{atm} = L_{sp\_atm}.d/1000, \quad (6)$$

where L_{sp_atm} is specific atmospheric gaseous attenuation in dB/km and d is distance in m.

For the total propagation losses after substitution of Eq. (3-6) in Eq. (2) are obtained the expression

$$L = 20\lg\frac{4\pi df}{c} + a(f).I^{b(f)}.\frac{d}{1000} + L_{sp\_atm}.\frac{d}{1000} = \\ = L(f,d,I), \quad (7)$$

or they depend of frequency, distance and rain rate.

The maximum value of rain rate that has been admitted connection can be given by

$$I_{\max} = \left\{\frac{1}{a(f)}\left[\frac{1000}{d}\left(L_{\max} - 20\lg\frac{4\pi df}{c}\right) - L_{sp\_atm}\right]\right\}^{1/b(f)} \quad (8)$$

In Eq.(8) maximum admissible total losses L_{max} can be expressed from Eq. (1) as

$$L_{\max} = P_t + G_t + G_r - P_{r\_\min} \quad (9)$$

where $P_{r\_\min}$ is receivers sensibility in dBm.

III. NUMERICAL RESULTS

Equations (7) – (9) are used for an example investigation on total losses and maximum admissible rain rate for frequencies of 60 GHz, 77 GHz and 93 GHz. The following input data (the same for all frequencies) has been assumed: P_t=10 dBm; G_t=G_r=43 dB; $P_{r\_\min}$= – 60 dBm; d=var.

In that case the maximum admissible total losses from Eq. (9) are calculated as L_{max}=156 dB.

The constants $a(f)$ and $b(f)$ for examined frequencies has been taken from [25] and are given in Table 1 and Table 2. The specific atmospheric attenuation [26] for examined frequencies is given in Table 3.

TABLE 1
SPECIFIC RAIN ATTENUATION CONSTANTS FOR HORIZONTAL POLARIZATION

Frequency, GHz	$a(f)$	$b(f)$
60	0,8606	0,7656
77	1,132	0,7177
93	1,3089	0,6901

TABLE 2
SPECIFIC RAIN ATTENUATION CONSTANTS FOR VERTICAL POLARIZATION

Frequency, GHz	$a(f)$	$b(f)$
60	0,8515	0,7486
77	1,1276	0,7073
93	1,3083	0,684

TABLE 3
SPECIFIC ATMOSPHERIC GASEOUS ATTENUATION

Frequency, GHz	L_{sp_atm}, dB/km
60	16
77	0,26
93	0,37

The results for maximum admissible rain fate for horizontal and vertical polarization are illustrated on Fig. 1 and Fig. 2 respectively. The total losses for 77 GHz and 90 GHz for horizontal polarization are given on Fig. 3.

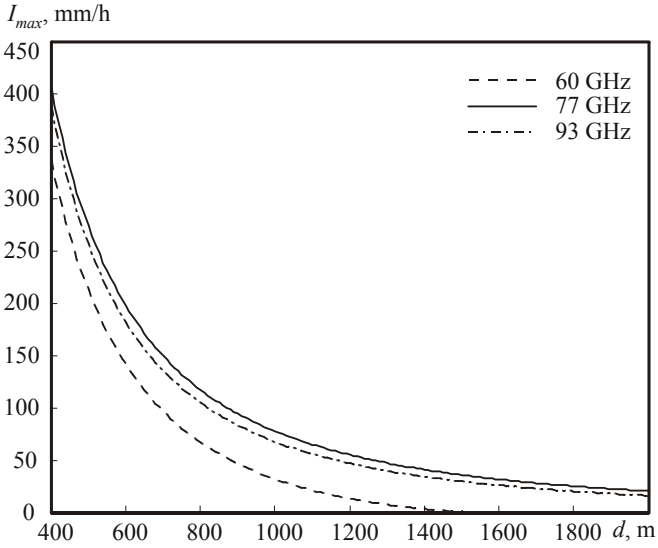


Fig. 1. Maximum admissible rain rate I_{max} as function of distance d for horizontal polarization

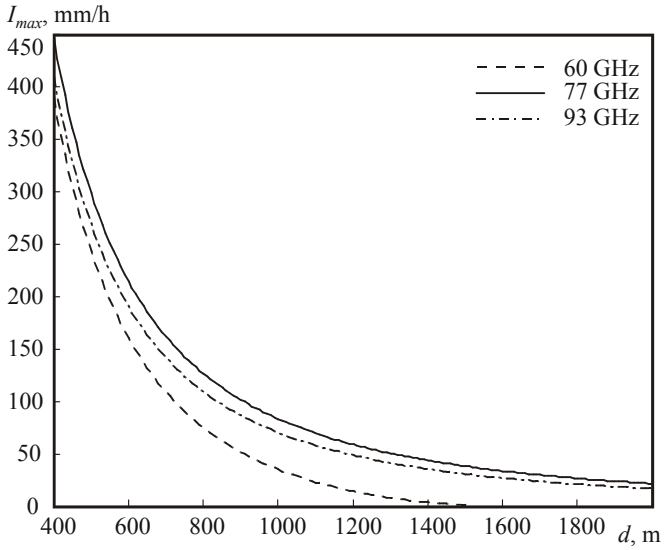


Fig. 2. Maximum admissible rain rate I_{max} as function of distance d for vertical polarization

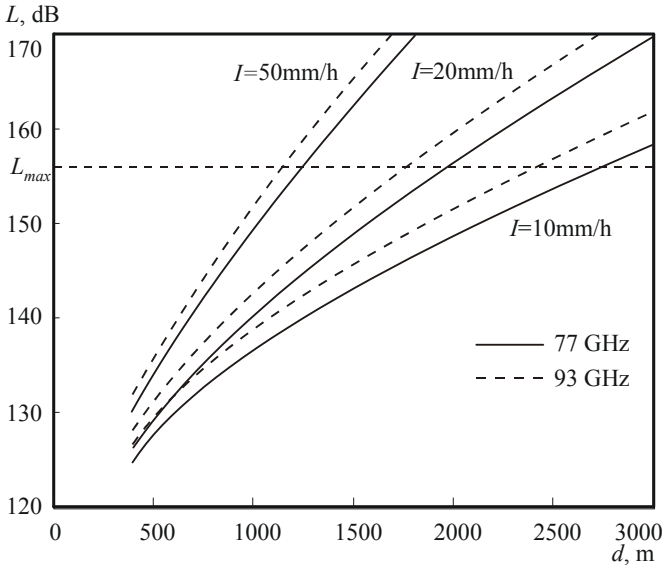


Fig. 3. Total losses L for 77 GHz and 93 GHz as function of distance d for horizontal polarization

IV. Conclusion

Fig. 1 and Fig. 2 show that the connection on RF line working on 60 GHz will interrupt at considerably lower values of rain rates than in 77 GHz and 93 GHz lines, because of very high oxygen absorption for that frequency. Nevertheless the radio line on frequency of 60 GHz can be used in short range (up to 800 – 1200 m) hybrid FSO/RF systems.

On the other hand the 77 GHz RF line will be available for rain rate 10-20 mm/h greater than 93 GHz one (Fig. 1 and Fig. 2) and can be used for back-up connection for FSO systems with range of about 2000 – 2500 m, while 93 GHz RF line for distances to 1800 – 2200 m. That is clearly shown on Fig. 3 with drawings for rain rates of 10 mm/h and 20 mm/h. The total losses for 77 GHz are 3-5 dB lower than on 93 GHz. For rain rate of 50 mm/h link range is shorter, but in that case hybrid system can connect on FSO, because optical waves attenuate insignificantly in rain, and heavy rain and fog don't appear simultaneously.

In addition Eq. (8) can be used in future works for estimation of link availability of hybrid FSO/FR networks for some regions if probability density function of rain rate *pdf(I)* is known.

Acknowledgement

Scientific researches whose results are presented in this paper are financed by Internal competition of Technical University of Sofia.

References

[1] Q. Liu, C. Qiao, G. Mitchell, S. Stanton, Optical wireless communication networks for first- and last-mile broadband access, Journal of Optical Networking, Vol. 4, Issue 12, December 2005, pp. 807-828.

[2] D. Killinger, Free Space Optics for Laser Communication Through the Air, Optics & Photonics News, p. 36-42, October 2002.

[3] V. Sidorovich , Solar background effects in wireless optical communications, SPIE, Volume 4873, Optical Wireless Communications V, p. 133-142, October 2002.

[4] M. Zaatari, Wireless Optical Communications Systems in Enterprice Networks, The Telecommunications Review, p. 49-57, 2003.

[5] E. Ferdinandov, Ts. Mitsev, Link Range of Free Space Communication System, ICEST 2003, Sofia, 2003.

[6] M. Al Naboulsi, H. Suzin, F. de Fornel, Propagation of optical and infrared waves in the atmosphere, XXVIII URSI General Assembly, New Delhi, India, October 2005.

[7] B. Strickland, M. Lavan, E. Woodbridge, V. Chan, Effects of fog on bit-error rate of a free-space laser communication system, Applied Optics, Vol. 38, No 3, p. 424-431, January 1999.

[8] I. Kim, B. McArthur, E. Korevaar, Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications, SPIE - Volume 4214, Optical Wireless Communications III, p. 26-37, February 2001.

[9] X. Zhu, J. Kahn, Free-Space Optical Communication Through Atmospheric Turbulence Channel, IEEE Transactions on Communications, Vol. 50, No 8, p. 1293-1300, August 2002.

[10] X. Guoliang, Zh. Xuping, W. Junwei, F. Xiaoyong, Influence of atmospheric turbulence on FSO link performance, Conference – Optical transmission, switching and subsystems, Wuhan, China, 4-6 November 2003, vol. 5281, pp. 816-823.

[11] Sh. Arnon, Effect of atmospheric turbulence and building sway on optical wireless-communication systems, Optics Letters, Vol. 28, No 2, p. 129-131, January 2003.

[12] M. Toyoshima, T. Jono, K. Nakagawa, A. Yamamoto, Optimum divergence angle of a Gaussian beam wave in the presence of random jitter in free-space laser communication systems, Journal of the Optical Society of America, Vol. 19, No3, p. 567-571, March 2002.

[13] O. Korotkova, L. C. Andrews, R.L. Phillips, "A Model for a Partially Coherent Gaussian Beam in Atmospheric Turbulence with Application in Lasercom" Opt. Eng. 43(2), pp.330-341, 2004.
[14] Y. Baykal, Average transmittance in turbulence for partially coherent sources, Optics Communications, Vol. 231, Issue 1-6, pp. 129-136, February 2004.
[15] J. Xing, G. Xu, X. Zhang, G. Wang, T. Ding, Effect of interlaced code for free-space optical communication, SPIE, Vol. 6021, December, 2005.
[16] I. Djordjevic, B. Vasic, M. Neifeld, Power Efficient LDPC-Coded Modulation for Free-Space Optical Communication over Atmospheric Turbulence Channel, Optical Fiber Communication and the National Fiber Optic Engineers Conference 2007, pp. 1-3, March 2007, Anaheim, CA, USA.
[17] J. Wang, D. Huang, Y. Xiuhua, Adaptive detection technique for optical wireless communication over strong turbulence channels, Optik – International Journal for Light and Electron Optics, Vo. 118, Iss. 11, November 2007, pp. 515-520.
[18] S. Hitam, M. Kh. Abdullah, M. A. Mahdi, K. Dimyati, Improvement of BER by Increasing Decision Threshold Level for Multi-Gigabits-per-second Free Space Optical Communications, International Conference on Next Generation Communication Systems – ICONGENCOM-06, pp. 158-163.
[19] W. O. Popoola, Z. hassemlooy, J. I. H. Allen,: "Performance of subcarrier modulated free-space optical communications", 8th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting (PGNet 2007), Liverpool, UK. pp. 75-80, June 2007.
[20] Sh. Arnon, Optimization of urban optical wireless communication systems, IEEE Transactions of Wireless Communications, Vol. 2, Issue 4, pp. 626 – 629, July 2003.
[21] I. Kim, E. Korevaar, Availability of free-space optics (FSO) and hybrid FSO/RF systems, SPIE Vol. 4530, p. 84-95, Optical Wireless Communications IV, Eric J. Korevaar; Ed., November 2001.
[22] A. Akbulut, M. Efe, A. Ceylan, F. Ari, Z. Telatar, H. Ilk, S. Tugas, An Experimental Hybrid FSO/RF Communication System, Communication Systems and Networks – 2003, Benalmadena, Spain, 2003.
[23] S. Bloom, W. Hartley, The Last Mile Solution: Hybrid FSO Radio, AirFiber Inc., May 2002.
[24] V. Kvicera, M. Grabner, O. Fiser, Availability Performance of a Simulated 850 nm FSO/93 GHz MMW Hybrid System, Wireless and Optical Communications – 2007, Montreal, Canada, 2007.
[25] Specific attenuation model for rain for use in prediction methods, Rec. ITU-R P.838-3.
[26] R. L. Freeman, Radio System Design for Telecomunications, John Wiley & Sons Inc., Hoboken, New Jersey, 2007.

Neural Network Approach in Modeling Microwave Slotted Patch Antennas

Bratislav Milovanović, Marija Milijić, Zoran Stanković

Abstract – **This paper presents a neural model based on multi-layer perceptrons (MLP) network for modeling slotted patch antennas. Neural model were trained to calculate accurately the resonant frequency *f* and the minimum value of S_{11} parameter (S_{11min}). Besides dimension parameters: antenna patch length *L*, antenna patch width *W*, deep of antenna patch slot *s* and length of antenna patch slot *l*, neural network relates and substrate layer height *h*. The developed neural model speeds up the process of collecting results by replacing repeated EM simulations and retains good accuracy compared with the MoM modeling.**

Keywords - **slotted patch antenna, neural network**

I. INTRODUCTION

Modern communication systems, from wireless to space communications, require compact and light weight components. One important component is the antenna, which, in addition, is required to have sufficiently high gain, wide bandwidth, high efficiency and to be easy to fabricate. Printed antennas are commonly used in these applications, since they inherently have low profile and low weight apart from having a low fabrication cost. They, thanks to their planar geometry, are usually integrated in devices mounted on a platform (for example in a mobile telephone).

Microstrip patch antenna is the most common form of printed antennas [1]. Due to its planar configuration and ease of integration with microstrip technology, it has been heavily studied and is often used as elements for an array. Consider the microstrip antenna shown in Figure 1, fed by a microstrip transmission line. The patch is of length *L*, width *W* with slot of deep *s* and length *l*. It sits on top of a substrate of thickness *h* with permittivity ε_r. Typically the height *h* is much smaller than the wavelength of operation. A substrate with a low dielectric constant (ε_r) is used (typically ~ 2.5), but in loss critical applications, greater dielectric constant (ε_r =10) must be used.

The most known method used for modeling patch antennas is the electromagnetic simulation. Although it is very correct process, it has some disadvantages which can not satisfy requirements of communication systems designing under some circumstances. Its basic disadvantage is that electromagnetic simulation has high demands concerning the hardware resources necessary for its software implementation. The software implementation itself might be very complicated and faced with many difficulties. Also the time needed for numerical calculation by electromagnetic simulation could be unacceptably long.

The authors are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, Niš, Serbia E-mail:batam@pogled.net, marija.milijic@elfak.ni.ac.rs,zoran.stankovic@elfak.ni.ac.rs

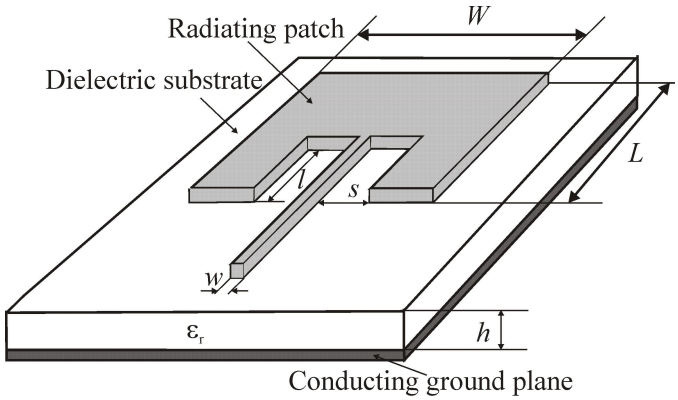


Fig. 1 Slotted patch antenna

The method of moments (MoM) is a very popular algorithm of computer electromagnetic calculations. It is widely used for antenna simulations and electromagnetic wave scattering analysis due to its good accuracy. The main disadvantages of the MoM is demand to solve a large number of time-consuming complex electromagnetic equations.

In this paper, the modeling patch antenna using neural networks is studied as an alternative method to detailed EM modeling. The previous researching, concerning the modeling of patch antennas have showed that the neural network models can have satisfactory accuracy similar as MoM but also can have higher simulation speed then EM simulations [2-5]. In this researching, it is shown that neural network could be very successfully for slotted patch antenna modeling considering dimension parameters: antenna patch length *L*, antenna patch width *W*, deep of antenna patch slot *s* and length of antenna patch slot *l*. As further study of modeling patch antennas, the authors developed in this paper neural network which considers and substrate layer height *h* influences on the resonant frequency f_r and the minimum value of S_{11} parameter (S_{11min}).

II. MAIN CONCEPT OF NEURAL NETWORKS

Owing to the capability of a functional dependence's modeling exclusively on the basis of input data, Multilayer Perceptron Network (MLP) is a type of neural network that can be successfully applied in the modeling of slotted patch antennas [6].

Neural networks, created in imitation of biological nervous systems, consist of connected cells (neurons) which parallel process data [7]. Connections between neural networks cells feature with weights. During process of training neural network for special problem, values of weights between neural network neurons change to give expecting output for

specified input. Such type of training is called supervised training and it is used for training neural network for modeling slotted patch antennas. Proposed neural multi-layers network for modeling patch antennas has four layers (Fig. 2). Outputs from all neurons of previous layer are input in every neuron in next layer whereas neurons in same layer are not connected together. Input data vector is run into input layer whose number of neurons is equal to input vector dimension (the number of input data). The latest layer of neural network is output layer and its neurons give output of neural network. Other neurons, which are not connected with input or output links, are hidden neurons. Signals through neural network runs in one direction, from input to output, and this network is example of feed forward network [7].

MLP neural network for modeling slotted patch antennas consists of input, two hidden and output layers. The number of input layer neuron is equal to number of antenna parameters that determine modeling. In this application, there are five input parameters: antenna patch length L, antenna patch width W, deep of antenna patch slot s, length of antenna patch slot l and substrate layer height h. The number of hidden layer neurons is variable during training process and output layer has two neurons that give resonant frequency f_r and S_{11min} parameter. The MLP network models the function:

$$[f_r, S_{11\min}] = f\ (L, W, l, s, h) \qquad (1)$$

The activation functions of the hidden layers are sigmoid, while the neurons of the output layers have linear activation functions. The neural networks were trained using Levenberg-Marquardt method with 10^{-3} performance goal. The notation of MLP models is MLPn-l_1-l_2-…-l_{n-2} where n represents layer number and l_1-l_2-…-l_{n-2} are the numbers of neurons of its each hidden layer.

The values of resonant frequency fr and S_{11min} parameter, necessary for the training and the test MLP neural networks, were obtained by ADS 2006 software. This software uses method of moments to calculate S_{11min} parameter for certain frequency f of slotted patch antenna with specific dimension. Training and test sets consist of only resonant frequency f_r defined by minimum value of S_{11} parameters for specific slotted patch antenna.

Patch antenna, modeled in this paper, has a substrate with dielectric constant ε_r=2.17. The width of fed line w depends on substrate layer height h and its range is [0.5 mm, 3.5 mm][8]. The other antenna parameters are changeable. In the training set with 3456 samples five input parameters have following range: 45 mm≤L≤120 mm, 70.5 mm≤W≤133.5 mm, 1.5 mm≤l≤31.5 mm, 1.5 mm≤s≤31.5 mm and 0.208 mm≤h≤1.108 mm.

III. TEST RESULTS

The test set contained 405 samples that have not been used in training process. Test results of successfully trained neural networks are presented in the Table I together with the average test error (*ATE*), the worst case error (*WCE)* and the

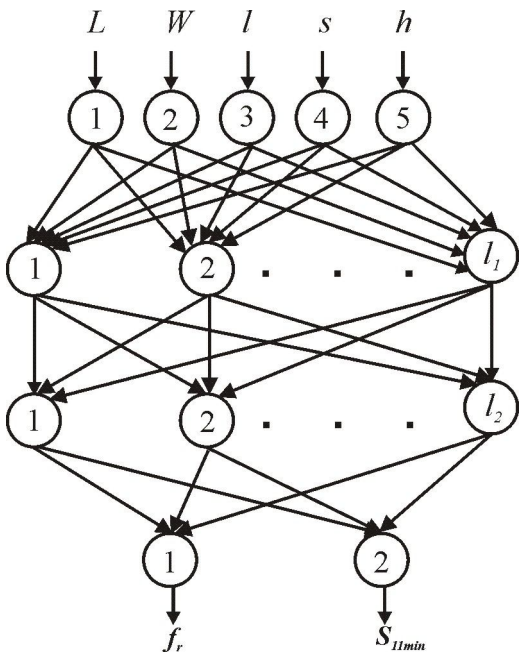


Fig. 2 MLP neural network for modeling slotted patch antenna

TABLE I TEST RESULTS

MLP model	*WCE*[%]	*ACE*[%]	r^{ppm}
MLP4-12-6	43.09	2.59	0.93
MLP4-15-10	45.68	2.54	0.93
MLP4-20-10	43.12	2.84	0.92
MLP4-16-8	46.74	2.90	0.92
MLP4-9-9	48.60	2.68	0.92
MLP4-20-15	49.27	2.80	0.91
MLP4-18-16	48.01	2.87	0.91
MLP4-14-12	46.09	2.93	0.90

Pearson Product-Moment correlation coefficient (r^{PPM}). The minimum of worst test error and the maximum value of Pearson Product-Moment correlation coefficient represent the basic criterion for selection the best MLP network. Selected neural model is MLP4-12-6.

IV. SIMULATION RESULTS

Beside its best test results, generalization level of MLP4-12-6 model should be checked as its advantages over previous methods for modeling slotted patch antenna. First, dependence of antenna parameter S_{11min} on patch length L for different values of substrate layer height h obtained by MLP4-12-6 is compared with MoM simulation results in Fig. 3. This figure shows the satisfying accuracy of MLP model compared with MoM simulation results. The simular conclusion can be done in Fig. 4. that shows how resonant frequency f_r depends on antenna patch parameters L for different values of substrate layer height h. Results obtained by MLP model very good agree with results of MoM simulation.

Further, preposed MLP model is used for representing both parameter S_{11min} and resonant frequency f_r as function of

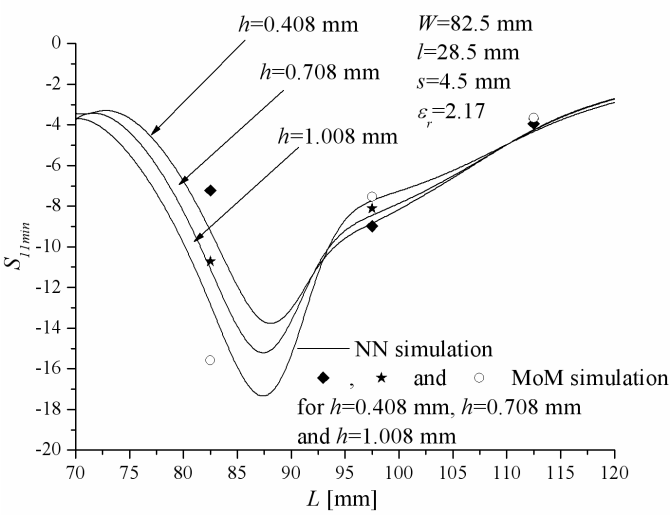


Fig. 3 S_{11min} parameter vs. antenna patch length L for different values of substrate layer height h

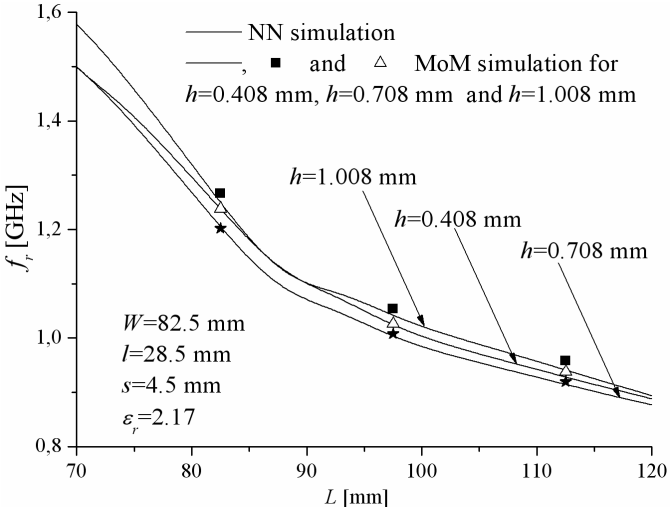


Fig. 4 Resonant frequency f_r vs. antenna patch length L for different values of substrate layer height h

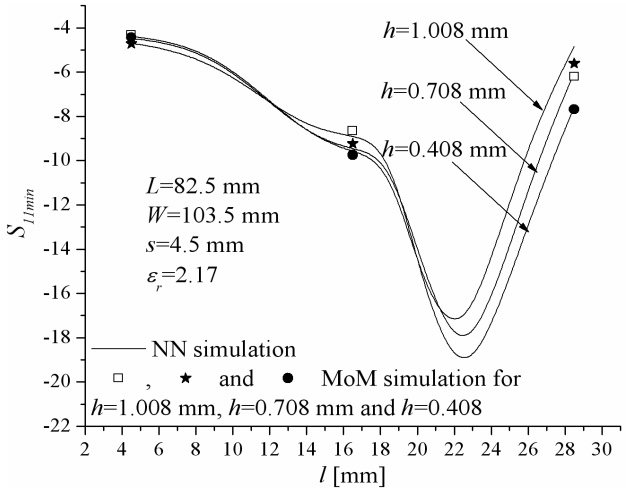


Fig. 5 S_{11min} parameter vs. antenna patch slot length l for different values of substrate layer height h

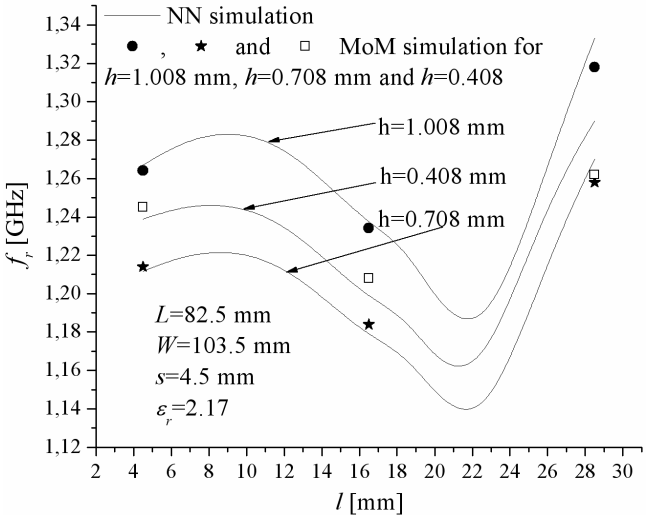


Fig. 6 Resonant frequency f_r vs. antenna patch slot length l for different values of substrate layer height h

antenna parameters l for different values of substrate layer height h and these results have been compared with values obtained by MoM simulation (Fig. 5. and Fig. 6.).

As all these figures show, MLP4-12-6 model is correct and it can be used for modeling patch antenna S_{11min} parameter and resonant frequency f_r with acceptable accuracy compared with MoM simulation used in ADS 2006 software.

Added to its accuracy, MLP model improves patch antenna modeling with very great speed of work. Fig. 7. shows the dependence f_r on parameter L and W, when h, l and s parameters are constant. This dependence is presented using 352 values of f_r obtained by NN simulation for 2 seconds. If we use MoM simulation in ADS 2006 software to obtain the same number of f_r values, we will do it for one day. Also, Fig. 8. represents S_{11min} as function of parameters s and l when h, W and L are constant values. MoM simulation from ADS 2006 sofware, obtains 256 values of S_{11min} parameter for 20 hours. NN simulation obtains needed 256 values of S_{11min} parameter for 1 second. For these reasons, NN simulation is better alternative in applications where simulation has to be finished in certain period of time.

V. CONCLUSION

Patch antennas are increasing in popularity for use in different communication systems due their numerous advantages. Therefore, requests concerning quality, performance and realization quickness of wireless communication systems dictate that tools for patch antenna modeling must be fast and failsafe. Detailed EM modeling based on numerical techniques is the most often used method for modeling patch antennas. It features with great accuracy. But, it has many limitations. First limitation of

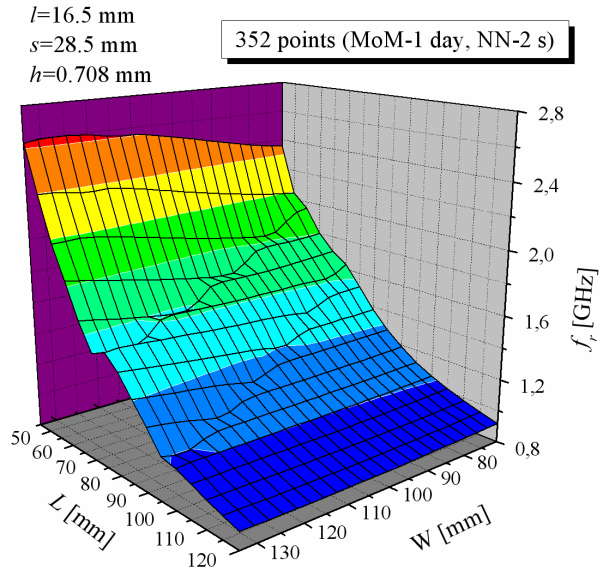


Figure 7 f_r vs. parameters L and W with constant h, l and s parameters

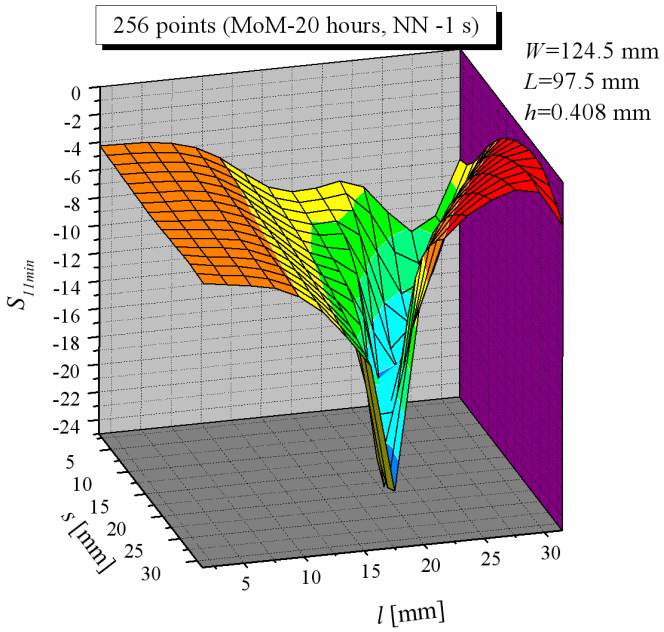


Figure 8 S_{11min} vs. parameters l and s with constant h, L and W parameters

EM methods is very strong hardware platform request. Also, its numerical calculations are done for a long period of time.

Neural model presented in this paper, thanks to its features, can surpass these problems. After successfully finished training, its accuracy is not significantly less then accuracy of detailed EM modeling, although it is much faster. It considers five parameters of patch antenna, four dimensional parameter - antenna patch length L, antenna patch width W, deep of antenna patch slot s and length of antenna patch slot l and one which features substrate layer - its height h. The great speed of neural models can give good base for introducing active models in the optimization process of planar structures.

REFERENCES

[1] Lal Chand Godara "Handbook of Antennas in Wireless Communications", Chapter 3, CRC Press, 2002

[2] Bratislav Milovanović, Marija Milijić, Aleksandar Atanasković, Zoran Stanković: "Modeling of Patch Antennas Using Neural Networks", 7th International Conference TELSIKS 2005, Niš, Serbia and Montenegro, September 28-30 2005, Vol.2, pp.385-388.

[3] Zoran Stanković, Bratislav Milovanović, Marija Milijić, "Efficient Neural Model of Microwave Patch Antennas" Proceedings of XLI International Scientific Conference ICEST2006, Sofia, Bulgaria, 29 June – July 1 2006, pp.49-52.

[4] Bratislav Milovanović, Zoran Stanković, Marija Milijić, "Hybrid Empirical-Neural Model of Microvawe Slotted Patch Antennas", Eight Seminar of- NEUREL 2006, Belgrade, September 25-27 2006, pp.181-184.

[5] Bratislav Milovanović, Zoran Stanković, Marija Milijić, Maja Sarevska, "Software for Microwave Slotted Patch Antenna Design Based on Hybrid Empirical-Neural Model ", XLII International Scientific Conference ICEST 2007, Bitola, Macedonia, June 24-27 2007, Vol.1, pp.235-238.

[6] Q. J. Zhang, K. C. Gupta, Neural Networks for RF and Microwave Design, Artech House,2000.

[7] S. Haykin, Neural Networks, New York, IEEE, 1994.

[8] Advanced Design System Documentation 2006A.

Reliability Testing of Free Space Optical Systems in Laboratory Conditions

Tsvetan Mitsev[1], Kalin Dimitrov[2] , Nikolay Kolev[3]

Abstract – **This paper deals with principle schemes of transmitter and receiver of a Free Space Optic (FSO) system and the organisation of experiments with it. We have shown measurements for the verification of its working ability. We have made comparisons with theoretical results.**

Keywords – Free Space Optical Systems, Experimental Setup, Reliability Testing.

I. INTRODUCTION

In the literature there are many theoretical results from research on the working reliability of FSO systems [1-6]. Most papers usually deal with the dependence of FSO on the length of the communication channel, the working wavelength [7,8] and the concrete meteorological conditions [9]. The losses caused by the atmosphere are usually expressed by the visibility S_M [3,10]. Some papers also discuss the fluctuations in the direction of laser ray propagation, caused by other various non-atmospheric factors (building sway [11]).

According to the authors, the experimental verifications and checks of these results present in the literature are very scarce. Probably the reason is basically the labour consumption during the setting up of such experiments, the impossibility for some of the parameters of the experiment to be easily changed or even predicted. We have to add to the abovementioned facts the necessary long overall time duration of possible complex research.

These reasons have resulted in the setting up of a laboratory setup with a controlled simulation of conditions maximally close to the real working ones, for a research on the working reliability of FSO systems depending on the characteristics of their transmitter and receiver systems and also on the parameter of the atmospheric channel.

II. THEORETICAL BACKGROUND

Before we discuss the practical research, we will derive the theoretical expressions necessary for the comparison.

We will start by discussing a laser source [10,12,13]. During a Gaussian amplitude distribution of the light intensity in the transmitting aperture, the intensity at point (ρ, z), in a cylindrical coordinate system and during assumption of azimuthal symmetry, is shown [4] by the expression

$$I(\rho,z) = I_m \frac{\rho_0^2}{\rho_z^2(z)} \exp\left(-2\frac{\rho^2}{\rho_z^2(z)}\right)\tau_a(z) \quad , \qquad (1)$$

where

$$I_m = \frac{2\Phi_L}{\pi\rho_0^2\left(1-e^{-2}\right)} \qquad (2)$$

is the intensity of the light radiation in the centre of the transmitting aperture (maximal intensity), when the power of the source of optical radiation is Φ_L and the losses in the transmitter optics are not taken into account, and ρ_0 is the radius of the Gaussian laser beam in the transmitting aperture (that is, when $z = 0$). The radius of the Gaussian laser beam at a distance z when the wavelength of the optical radiation λ is defined by

$$\rho_z(z) = \frac{\lambda}{\pi\rho_0} z \, . \qquad (3)$$

The expression (3) is valid when the condition $z \geq 10\left(\rho_0^2/\lambda\right)$ is fulfilled. To measure the losses during the propagation of the optical radiation through the communication channel we will use the quantity transmittance

$$\tau_a(z) = \exp\left(-0{,}23.b_f[dB/km].z[km]\right), \qquad (4)$$

where $z \in [0, z]$, and b_f are the distributed losses of the channel. For the atmosphere Eq. (4) is defined by the extinction coefficient

$$b_f[dB/km] = 4{,}343.\alpha_e[km^{-1}] \, . \qquad (5)$$

To define a concrete numerical value Eq. (5) we usually use the half-empirical relation with the meteorological distance of visibility

$$\alpha_e[km^{-1}] = \frac{3{,}92}{S_M[km]}\left(\frac{\lambda[\mu m]}{0{,}55}\right)^{-0{,}585\sqrt[3]{S_M[km]}} \, . \qquad (6)$$

It is evident from expressions (1) and (2) that the change in the intensity of the optical radiation in the point (ρ, z) with constant transmitter parameters is due to two factors: lowering

[1] Tsvetan Mitsev is with the Faculty of Telecommunications at Technical University Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: mitzev@tu-sofia.bg.

[2] Kalin Dimitrov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kld@tu-sofia.bg.

[3]Nikolay Kolev is with the Faculty of Telecommunications at Technical University Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria

of the intensity $I(0,z)$ along the axis of the ray due to its divergence and the losses in the transmission channel.

If the radius of the transmitter aperture R_r meets the condition $R_r << \rho_z << z$, then the power $\Phi_{r,s}$ coming form the transmitter into the receiver, when the losses are not taken into account, is

$$\Phi_{r,s} = \pi R_r^2 I(\rho,z). \quad (7)$$

During background radiation with receiver power $\Phi_{r,B}$ the total received power can be calculated by the formula

$$\Phi_r = \Phi_{r,s} + \Phi_{r,B}. \quad (8)$$

We substitute in Eq. (8) the expressions (1), (4), (5), (6) and (7) and we derive

$$\Phi_r = 10\lg\left\{\frac{2}{1-e^{-2}}\cdot\left(\frac{R_r}{\theta_t z}\right)^2 .10^{\frac{\Phi_L}{10}} \times \right.$$
$$\left. \times \exp\left[-2\left(\frac{\theta}{\theta_t}\right)^2 - 2{,}3.10^{-4}.b_f.z\right] + 10^{\frac{\Phi_{r,B}}{10}}\right\}, \quad (9)$$

where θ_t is the angle divergence of the ray after the transmitter aperture, and θ is the angle displacement of the axis of the laser ray according to the direction of the receiver aperture centre. The powers participating in the equation are in [dBm], the angles in [rad], and the distributed losses in [dB/km].

When there are very low values of the background optical flux $\Phi_{r,B}$ and very low distributed losses along the transmission channel Eq. (9) is defined by

$$\Phi_r[dBm] =$$
$$= \Phi_L[dBm] + 20.\lg\left(\frac{R_r}{\theta_t z}\right) - 8{,}686.\left(\frac{\theta}{\theta_t}\right)^2 + 3{,}642 . \quad (10)$$

III. Experimental Setup

The structural scheme of the laboratory test-bed is displayed on Fig. 1.

On the immobile rail (Fig.1, pos.1) we have placed the transmitter (Fig.1, pos.2) and the receiver (Fig.1, pos.3). The distance L between them meets the condition according to which the receiver has to be in the Fresnel zone, and in separate cases only in the Fresnel zone.

The transmitter has been placed on a horizontally rotating (by azimuth) stage (Fig.1, pos.4). With a view to the possibility of a fine adjustment, the receiver has been placed on a x-y-z stage {3D} (Fig.1, pos.5) and it can also be rotated by azimuth. The elevation angle of both blocks is constant and it equals 0. The distribution of the optical radiation of the source in a plane which is cross to the distribution is synphase

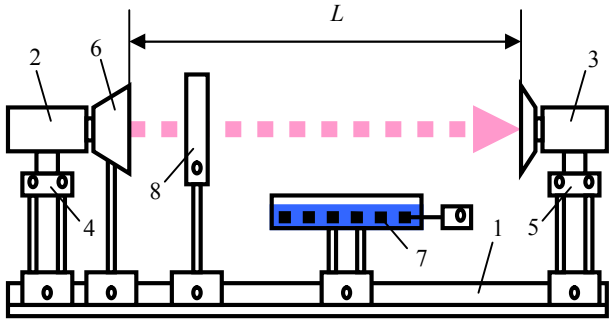


Fig. 1. Laboratory test-bed

and Gaussian amplitude. By means of a readjustable two lens collimator, Kepler's type (Fig.1, pos.6), the necessary beam changes are provided for the simulations of real conditions. For the fixed distance $z = const$ the Gaussian amplitude distribution is preserved, but the current beam radius of $\rho_z(z)$ changes, which equals a change in the divergence of the beam θ_t.

The increasing of θ_t brings about a change in the character of the interaction between the laser radiation and the heterogenities in the field of distribution. At a short distance there is a simulation of an influence over the distribution of the radiation of random heterogenities, turbulence vortices, with larger sizes, which is characteristic of longer real fields.

The collimator makes possible the change of θ_t within wide limits – $\theta_t \in [1mrad; 0{,}3rad]$.

The vaporizer (Fig.1, pos.7) is used for the creation of a randomly heterogeneity medium in the connection channel, which stimulates thoroughly the real turbulence heterogenities during the distribution of the optical radiation in the atmosphere. Its characteristics can change by means of altering the working intensity of the vaporizer. The level of turbulence or vortices distribution along the field changes by means of a fan speed.

The change of the attenuation along the connection channel resulting from the alterations in its transparency, that is the integral changes of the volume extinction coefficient α_e, respectively the meteorological visibility distance S_M, is simulated by placing neutral filters along the laser radiation distribution field. (Fig.1, pos.8). Their location is in conformity with the condition according to which they must not influence the value of the background radiation during the measurements. The filters used have band pass values within the interval $\tau_F \in [0{,}4; 0{,}97]$, which provide values of the simulated meteorological visibility distance along the connection channel within the interval $S_M \in [5{,}5km; 60km]$.

In order to avoid an unnecessary drain of money and time, the electronic blocks of the system have also been programme-tested as early as the design stage, before their physical implementation [14].

The structural scheme of the electronic blocks of the measurement test-bed for BER in Fig.1 consists of a transmitter TX, receiver RX (Fig. 2).

The transmitter TX has 3 outputs – B1, B2 and SYNC. Signals are transmitted to the receiver, which are subject to

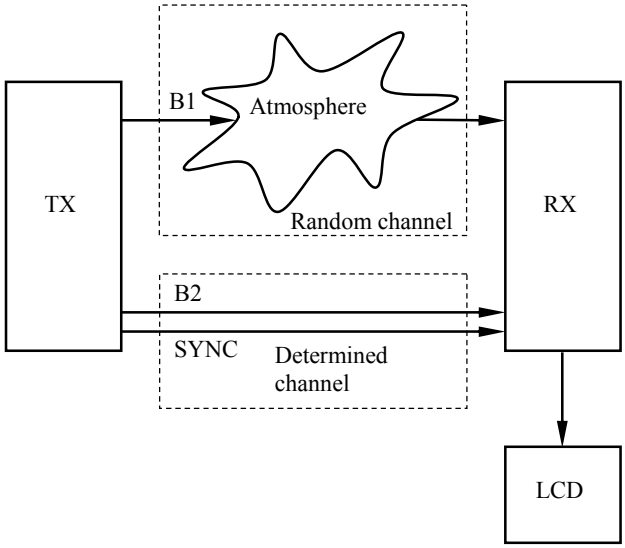


Fig. 2. Structure of BER calculating blocks

strong distortions through the atmosphere connection channel and there are also signals transmitted through a fixed-parameter channel. If the information has been transmitted without an error through the atmosphere channel, then at the output of the logical element in the receiver block at the moment of comparison there is a logical zero and there is no error aggregated. In the opposite case at the output of the logical element there is a logical one and respectively there is error aggregated. The results are visualized on a liquid crystal display.

The research stand of BER is based on microcontrollers PIC18F452. As the rest of the PIC microcontrollers, the microcontroller PIC18F452 uses HARVARD architecture, characteristic of which are the separate buses for access to the data memory and the programme memory Fig.3.

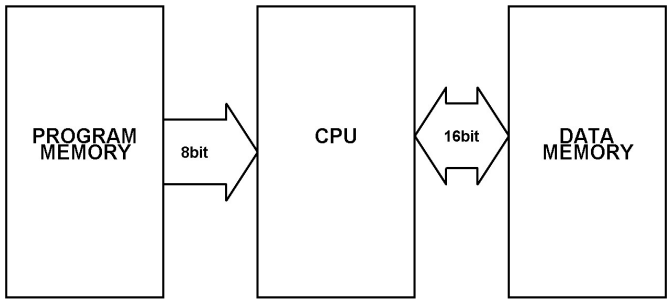


Fig. 3. Structure of chosen processor

The advantages of this are many but the most important is that the two buses can have different length (number of bits). This allows the 16-bit instructions of the microcontroller PIC18F452 to be extracted from a 16-bit data path for instructions, and the 8-bit data, respectively from an 8-bit data path. In these conditions the extracting and implementing of a 16-bit instruction takes one cycle, that is we have a greater productivity of the central processor.

The maximum clock frequency of the microcontroller PIC18F452 is 40 MHz. Before entering the central processor the clock signal is divided into 4 non-overlapping series which form an instruction cycle with a frequency of 10 MHz. Having in mind that the implementation of one instruction takes one cycle, when working with a maximum clock frequency, the productivity of the central processor will be 10MIPS (MIPS – million instructions per second). An exception to this are the transition instructions and the 32-bit instructions, the implementation of which takes two cycles.

A detailed description of the numerical part of the system as well as the consideration on the signal and subsystem level [15] is to be done in some of our further works.

IV. EXPERIMENTS

The initial check of the productivity of the system described in section III includes the power measurement of the receiver signal at random points and its comparison with the respective values calculated according to the theoretical model given in section II. The main experiments are two: check of the axial dependency of the optical field on the source within the distance of the test-bed; check of the radial dependency of the field by means of changing the azimuthally angle θ of the distribution direction of the beam.

For this purpose the receiver is replaced by a highly sensitive fiber-optic power meter with a low receiver aperture $A_r = 0{,}1mm^2$, sensitive within a wide wavelength interval, also including the central length of the optical radiation of the source $\lambda_0 = 850nm$.

The values of the parameters of the system at which the measurements are done and which are necessary for the numerical calculation, are: average optical source power $\Phi_L = 2mW\,(3dBm)$; angle divergence of the radiation after the transmitter antenna (the collimator) $\theta_t = 0{,}3rad$; distributed losses along the connection channel – $b_f = 0{,}1dB/km$ have been accepted, which is equals absolutely clean atmosphere ($S_M = 60km$).

The background optical radiation measured during the experiment by using measurements and wavelengths $1{,}31\mu m$ and $1{,}55\mu m$ (to increase the sensitivity and decrease the measurement error) is $\Phi_{r,B} = -69dBm$.

In Fig.4 we have shown the experimental (with dots) and theoretical (flat line) results for the change in the receiver power along the axis of the ray, that is the dependency $I(0,z)$, resp. $\Phi_r(0,z)$, for the distance between the transmitter and receiver $L \in [0{,}2m;1{,}7m]$ and for the movement of the ray away from the direction of the receiver antenna $\theta_t = 0$. There is a very good matching of the two results.

In Fig.5 we have presented the results from the comparison between the experimental and theoretical values for the

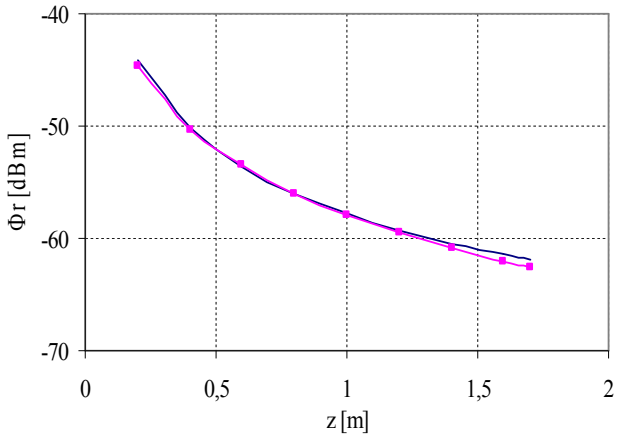


Fig. 4. Measured vs Calculated Data for Received Power

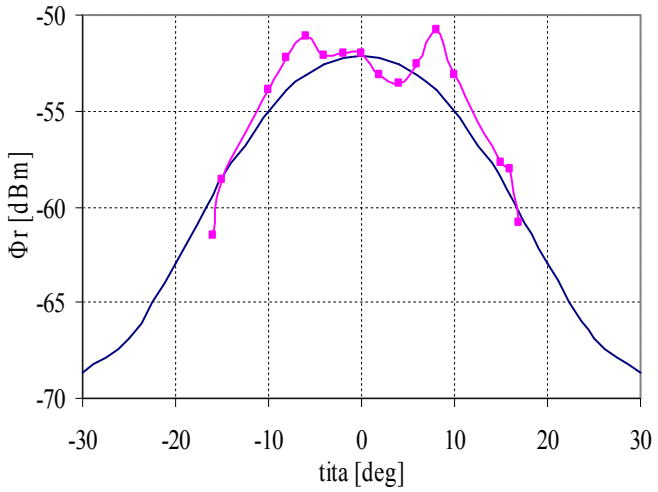


Fig. 5. Measured vs Calculated Data for axial Gaussian Distribution

checking of the axial Gaussian amplitude distribution of the optical radiation. The comparison has been conducted with $L = 0{,}5m$, $|\theta_t| \in [0;0{,}52rad]$, respectively $\rho \in [0;0{,}25m]$. The visible differences are mostly due to deviations in the amplitude distribution of the source of optical radiation from the ideal Gaussian one, which is observed with a great number of semiconductor sources, and even a very small azimuthally dependency of the background optical radiation $\Phi_{r,B}$, inevitable for the particular experimental setting.

V. Conclusion

We have conducted a theoretical analysis which has been verified by means of experimental results. This paper is part of a work which will include gradual detailed research on the FSO technology in laboratory and real environment.

References

[1] H. Willebrand, B. Ghuman, *Free Space Optics: Enabling Optical Connectivity in Today's Networks*, Indianapolis, SAMS, 2001.

[2] S. Hranilovic, *Wireless Optical Communication Systems*, Springer, 2004.

[3] S. Bloom, E. Korevaar,J. Schuster,H. Willebrand, "Understanding the performance of free-space optics", Optical Society of America, JON 2330 June 2003 / Vol. 2, No. 6 / JOURNAL OF OPTICAL NETWORKING, pp.178-200,2003.

[4] E. Ferdinandov, B. Pachedjieva, K. Dimitrov, *Optical Communication Systems*, Sofia, Technika, 2007.(in bulgarian)

[5] X. Zhu, J. M. Kahn, "Free-Space Optical Communication through Atmospheric Turbulence Channels", IEEE Trans. on Commun., vol. 50, no. 8, pp. 1293-1300, 2002.

[6] X. Zhu and J. M. Kahn, "Performance Bounds for Coded Free-Space Optical Communications through Atmospheric Turbulence Channels", IEEE Trans. on Commun., vol. 51, no. 8, pp.1233-1239, 2003.

[7] E Korevaar; I. Kim, B. McArthur, Atmospheric propagation characteristics of highest importance to commercial free space optics Proc. SPIE, Vol. 4976, Atmospheric Propagation, C. Young; J. Stryjewski, Editors, pp.1-12, 2003.

[8] A. Tunick, Optical turbulence parameters characterized via optical measurements over a 2.33 km free-space laser path, Optics Express, Vol. 16, Issue 19, pp.14645-14654, 2008.

[9] Al. Naboulsi1, M. Sizun H. de Fornel, "Propagation of optical and infrared waves in the atmosphere", XXVIIIth UNION RADIO-SCIENTIFIQUE INTERNATIONALE General Assembly, New Delhi, India, 2005.

[10] E. Ferdinandov, *Laser Radiation in Radiotechnique*, Sofia, Technika, 1981.(in bulgarian)

[11] X.Liu, Free-space optics optimization models for building sway and atmospheric interference using variable wavelength, IEEE Transactions on Communications, Vol.57, Issue: 2, pp.492-498, 2009.

[12] P. Corrigan, R. Martini, E. Whittaker, C. Bethea, Quantum cascade lasers and the Kruse model in free space optical communication, Optics Express, Vol. 17, Issue 6, pp. 4355-4359, 2009.

[13] K. Veselinov, F. Grillot, Al.Bekiarski, J. Even, S. Loualiche, "Numerical modelling of the two-state lasing in 1.55um (113)B InAs/InP quantum dot lasers for optical telecommunications", Sofia, ICEST, pp.260-261, 2006.

[14] K. Kazaura, K. Omae, T. Suzuki, M. Matsumoto, E. Mutafungwa, T. Korhonen, T. Murakami, K. Takahashi, H. Matsumoto, K. Wakamori, Y. Arimoto, Enhancing performance of next generation FSO communication systems using soft computing-based predictions, Optics Express, Vol. 14 Issue 12, pp.4958-4968, 2006.

[15] V. Georgieva, *Signals and Systems (laboratory manual)*, Sofia, Novi Znania, 2009.

60 GHz Range High Gain Millimeter Wave Antenna Array With Cylindrical-Parabolic Reflector

Aleksandar Nešić[1], Ivana Radnović and Milan Šunjevarić[1]

Abstract - **The paper presents concept, simulation, design and realization of the printed antenna array with a subreflector and a cylindrical-parabolic reflector. This antenna is intended for higher millimeter wave ranges and is characterized by many advantages such as efficacy, compactness, broad bandwidth and possibility of integration with other passive and active microwave circuits. Antenna gain is nearly 34 dBi and aperture efficiency is almost 50%. The antenna operates in the frequency range around 60 GHz. Measured results are in good accordance with those obtained by a simulation.**

Keywords – **Printed antenna arrays, Millimeter wave antennas, Cylindrical-parabolic reflector, Printed subreflector**

I. Introduction

Conventional high gain antennas that are used in microwave and millimeter-wave telecommunication, radar and navigation systems usually are: (1) parabolic antennas with central symmetry, (2) antennas with off-set parabolic reflector and (3) two-dimensional antenna arrays.

Due to relatively high losses in feed lines, conventional high gain printed antenna arrays are rarely used. In the suggested solution we use a cylindrical-parabolic reflector which is mainly applied in VHF and UHF ranges [1]. In millimeter wave ranges this type of antenna, to the author's knowledge, has not been used, except "pillbox" (or "cheese") antennas which have very wide E-plane beamwidth [2,3].

We will consider a suitable solution for high gain (around 34 dBi) antenna operating in millimetre wave range about 60 GHz according to ECC Recommendation (09)01 concerning the use of (57-64)GHz frequency band for point-to-point fixed wireless systems [4]. As a primary feed, in the focal line of the cylindrical-parabolic reflector, there is a printed antenna array consisting of 16 axial dipoles integrated with a subreflector, feed network and a transition from symmetrical microstrip to a waveguide. Obtained measured results of this antenna array are completely in accordance with ECC Recommendation [4] and show many advantages such as low cost, reproducibility and simple realization.

Although the simulation was carried out in the range (57-64) GHz, only the radiation patterns at 60 GHz are presented while the measurements at other frequencies are in progress at the moment (measuring procedure is rather slow due to lack of adequate equipment).

[1] Aleksandar Nešić, Ivana Radnović and Milan Šunjevarić are with the IMTEL Institute, Bulevar M. Pupina 165b, 11070 Beograd, Serbia,
e-mail: aca@insimtel.com, ivana@insimtel.com

II. Concept

The antenna consists of a cylindrical-parabolic reflector and a linear tapered axial array of printed pentagonal dipoles with printed subreflector. One half of each dipole is placed on one side and another half on the opposite side of the dielectric substrate. The dipoles operate on the second resonance (antiresonance) and their impedances are about 100Ω. The strip printed ahead of the array plays a role of a subreflector. Dipoles are fed through the printed feed network realized with symmetrical microstrip lines, like in [4]. Transition from symmetrical microstrip structure to a waveguide is inserted between the antenna and the feed line. All the elements of this antenna system, except the cylindrical-parabolic reflector, are placed on the same dielectric substrate (Fig. 1).

III. Design and Realization

Linear Array: Radiating elements in the linear array are pentagonal dipoles axially placed in the focal line of the cylindrical-parabolic reflector and operating on the second resonance. The dipoles, feed network, and the subreflector (printed on both sides of the dielectric), are realized on the dielectric substrate of h=0.127mm, ε_r=2.1 and tgδ=4x10^{-4}. Distance between the subreflector's axis and dipole's axis is λ/4 at the central frequency, while width of the subreflector is optimized so to obtain as high as possible gain. Dimensions of the pentagonal dipoles are optimized so to obtain impedance of about (100+j0) Ω at the central frequency of 60 GHz.

In the next step, an array of 16 dipoles has been modeled. The distance between the dipoles is chosen in such a way as to obtain high array gain with relatively high suppression of grating lobs. In our case, the distance between axial dipoles is L=1.1λ. Feed network realized in symmetrical (balanced) microstip is like in [5]. Feeding lines for dipoles penetrate the junction of two reflector halves. In the place of this junction there are holes with diameter d=1mm through which symmetrical microstrip lines of the feed network pass. Software package WIPL-D [6] has been used in these analyses.

Cylindrical-parabolic reflector: Cylindrical-parabolic reflector is assembled of its two halves. Dielectric substrate with printed radiating elements, subreflector, feed network and transition to waveguide is positioned between the reflector's halves, Fig.2. Length of the cylindrical-parabolic reflector (D=100mm) is practically defined by the linear array's length. Reflector's aperture is obtained from the condition that H-plane beamwidth is close to the E-plane beamwidth.

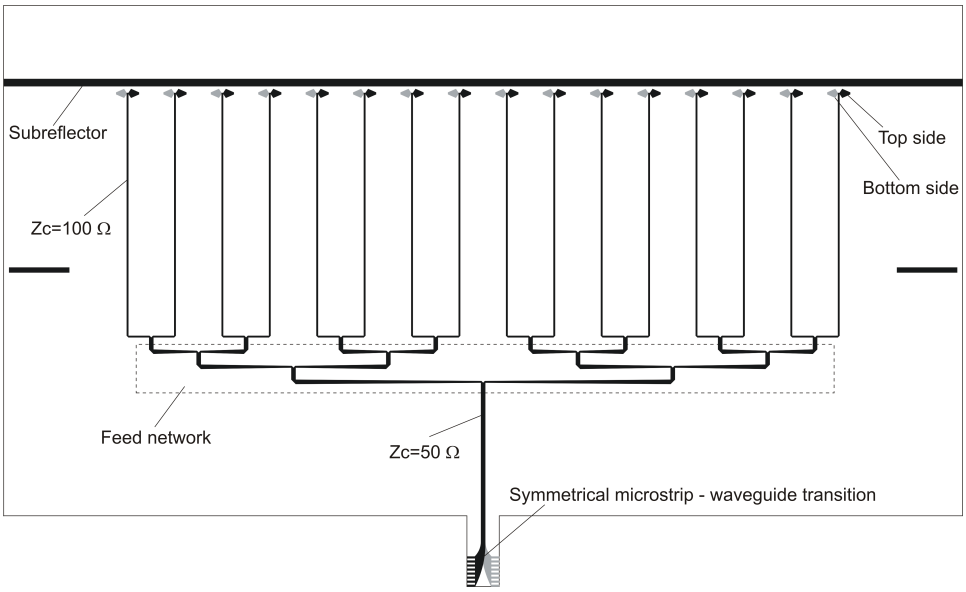


Fig. 1. Printed antenna array with feed network and symmetrical microstrip – waveguide transition integrated on the same dielectric substrate.

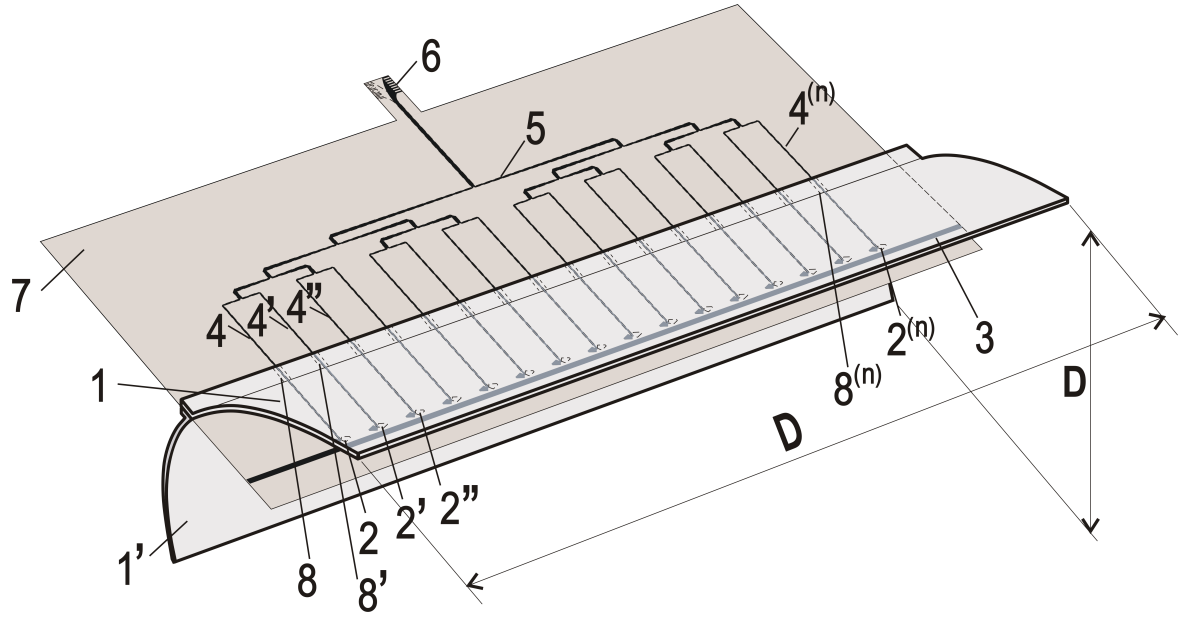


Fig. 2. Layout of the printed antenna array with subreflector placed in cylindrical-parabolic reflector: 1 and 1': two halves of the parabolic reflector; 2-2$^{(n)*}$: printed pentagonal dipoles; 3: subreflector; 4-4$^{(n)*}$: 100Ω-lines of the feed network; 5: feed network; 6: transition from symmetrical microstrip to a waveguide ; 7: dielectric substrate; 8-8$^{(n)*}$: holes in the cylindrical-parabolic reflector (* n= number of dipoles in the array).

Focal length (L_f) is 16λ/4 (20mm at f=60 GHz). Thus, L_f/D is around 0.2 that makes the overall size and depth of the antenna smaller. Owing to the subreflector, more suitable illumination distribution, i.e. higher gain and better side lobe suppression in H-plane has been achieved.

IV. OBTAINED RESULTS

E- and H-plane radiation patterns of the antenna with and without a subreflector, obtained by a simulation are shown in Fig. 3a and Fig 3b, respectively. Simulated gain of the antenna without a subreflector is nearly 34 dBi. There is an improvement in the antenna gain of about 2 dB in the case with subreflector. Figures 4a and 4b show simulated and measured E- and H-plane radiation patterns of the array. Measured gain is approximately 2 dB lesser than simulated due to losses in the feeding lines and the transition from balanced microstrip to waveguide, which were not taken into account. VSWR against frequency measured in the range from 57.5 GHz to 75 GHz is given in Fig. 5, while photograpf of the realized printed antenna array in the parabolic reflector is presented in Fig. 6.

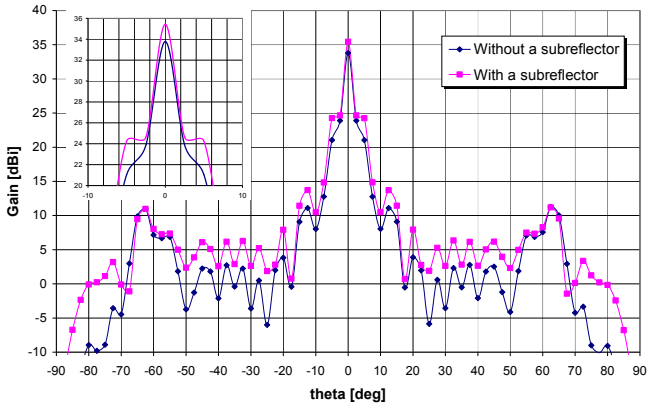


Fig. 3a. Simulated E-plane radiation patterns of the antenna with and without printed subreflector

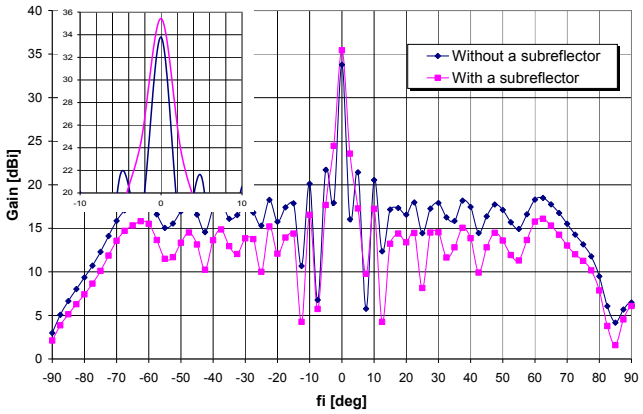


Fig. 3b. Simulated H-plane radiation patterns of the antenna with and without printed subreflector

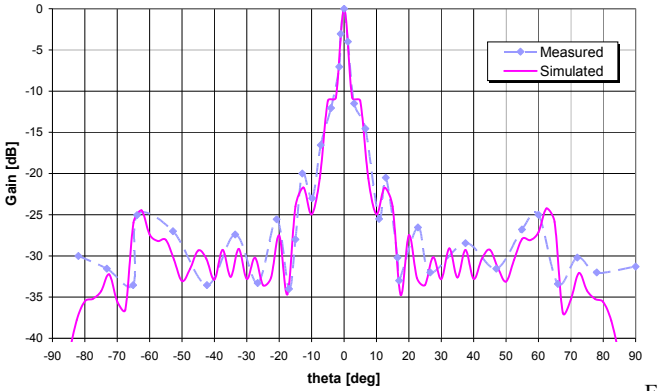


Fig. 4a. Simulated and measured E-plane radiation patterns of the antenna

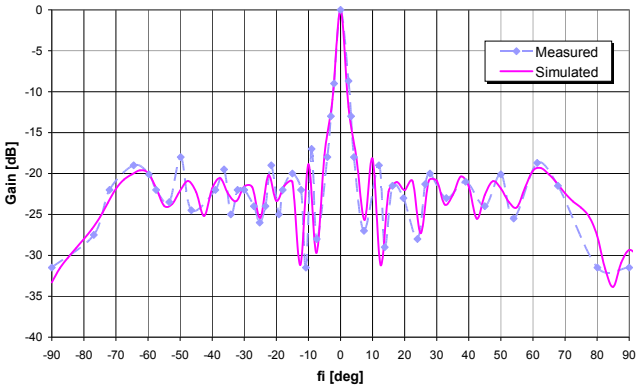


Fig. 4b. Simulated and measured H-plane radiation patterns of the antenna

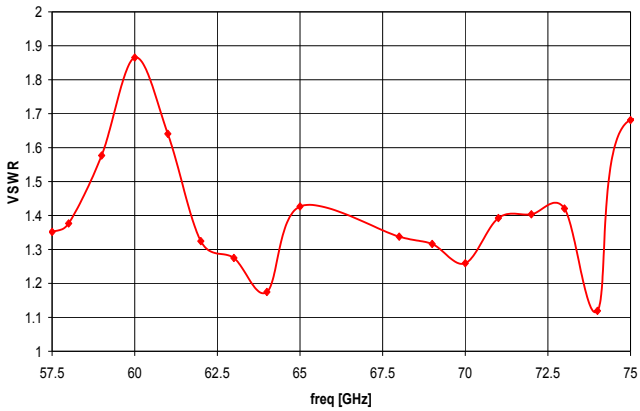


Fig. 5. Measured VSWR of the antenna against frequency

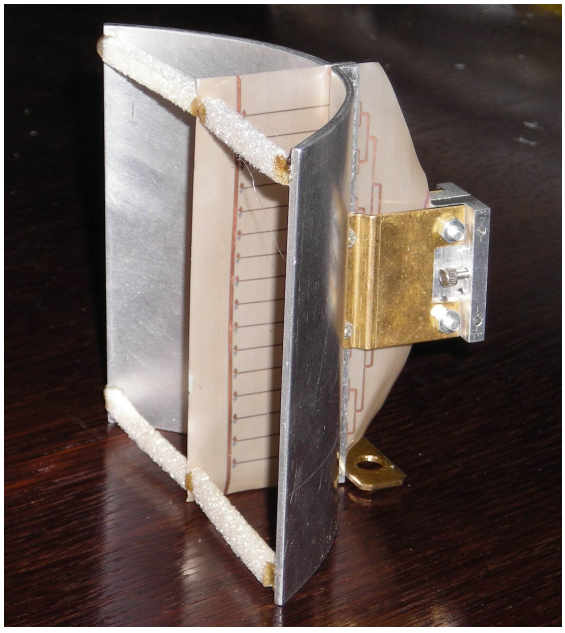


Fig. 6. Photograph of the realized antenna

V. CONCLUSION

The paper presents a new type of printed antenna array with a subreflector, placed in a cylindrical-parabolic reflector. Axial array of 16 pentagonal dipoles operating on the second resonance plays a role of a primary feed. In front of the array, there is a printed strip that acts as a subreflector. Symmetrical microstrip – waveguide transition, feed network, printed dipoles array and subreflector are all on the same dielectric substrate. Measured gain at 60 GHz is about 34dBi which is about 2 dB lesser than the gain obtained by a simulation where losses in the dielectric substrate and microstrip lines were neglected. There is a very good accordance between measured and simulated results. VSWR of the array is lesser than 2 in the range from 57.5 GHz to 75 GHz. Proposed concept of the cylindrical-parabolic antenna with a printed primary feed can be used in millimeter ranges up to 110 GHz.

ACKNOWLEDGMENTS

Authors would like to thank Ms. M. Marjanovic and Mr. M. Tasic for their help in realization of the experimental model. This work has been supported by Serbian Ministry of Science.

REFERENCES

[1] J. D. Kraus: "ANTENNAS" McGraw-Hill Book Company, 1988, pp. 572-573

[2] R. E. Collin, F.J. Zucker: "ANTENNA THEORY", McGraw-Hill Book Company, 1969, pp. 83-86

[3] Henry Jasik: “Antenna Engineering Handbook”, McGraw-Hill Book Company, 1961

[4] ECC Reccomendation (09)01 – Use of the 57-64 GHz Frequency Band for Point-to-point fixed Wireless Systems, Electronic Communications Committee (ECC) within the European Conference of Postal and Telecommunications Administrations (Recommendation approved by the Working Group “Spectrum Engineering” (SE)), pp.1-2

[5] A. Nesic, Z. Micic, S. Jovanovic, I. Radnovic,Dusan Nesic: “Millimeter Wave Printed Antenna Arrays for Covering Various Sector Widths”, IEEE Antennas and Propagation Magazine, Vol. 49, No 1, February 2007, pp. 113-118

[6] WIPL-D Pro v5.1, Branko Kolundzija, Jovan Ognjanovic, Tapan K. Sarkar

Smart Antenna Configurations – Analysis Based on ESPRIT Methods

Sava Savov[1], Viara Vasileva[2] and Miroslava Doneva[3]

Abstract – **In this paper, the simple direction of arrival (DOA) methods for uniform rectangular array (URA) and uniform circular array (UCA) based on narrowband radio-frequency signals are introduced. Arrays are composed of a number of uniformly distributed identical half-wavelength dipoles. Numerical examples are presented to illustrate these methods.**

Keywords – **Direction of arrival, Half-wavelength dipole, Smart antenna, Uniform circular array, Uniform rectangular array.**

I. Introduction

Smart antennas (adaptive arrays) can provide a substantial capacity improvement in the frequency-resource-limited radio-communication system. They are used in order to improve capacity per user channel and better quality [6].

Spatial processing is the central idea for smart antennas. This is a technology that would be implemented into the existing wireless communications infrastructures to provide broader channel bandwidth and new improved services.

The investigation of smart antennas suitable for wireless communication systems has involved primary uniform circular arrays (UCA) and uniform rectangular arrays (URA). They possess the ability to scan in 3-D space, and it is necessary for wireless devices to scan the main beam in any direction of elevation and azimuth.

Two DOA estimation methods are introduced for the significant improvement in smart antenna resolution [6]. This paper presents UCA-ESPRIT and the 2-D unitary ESPRIT for direction of arrival estimation analysis of UCA and URA, respectively. Limited numerical examples are depicted to illustrate these algorithms.

[1]Sava V. Savov is with Department of Electrical Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, Senior Member IEEE
E-mail: savovsv@yahoo.com
[2]Viara Y. Vasileva is with Department of Electrical Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria,
E-mail: via_vas@abv.bg
[3]Miroslava G. Doneva is with Department of Electrical Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria,
E-mail: m_grisheva@abv.bg

II. Uniform Circular and Rectangular Array Configurations

A. Circular antenna structure

The UCA with radius a consisting N equally distributed identical half-wavelength dipoles is located on x-y plane, as illustrated in Fig. 1.

An incoming plane wave (narrowband signal) with wavelength λ arrives at the array from elevation angle θ and azimuth angle ϕ. A spherical coordinate system is used to denote the arrival direction. The origin of coordinate system is located at the center of the array.

As presented in Fig. 1, the array factor *(AF)* of UCA is given by [4]

$$AF(\theta,\phi)=\sum_{n=1}^{N} w_n e^{-jka\sin\theta\ \cos(\phi-\phi_n)} \qquad (1)$$

where w_n, $\theta=\pi/2$ and ϕ_n are the estimated weights and angular positions of the nth element, respectively, a is the radius of the UCA, and k is the wave number ($k=2\pi/\lambda$).

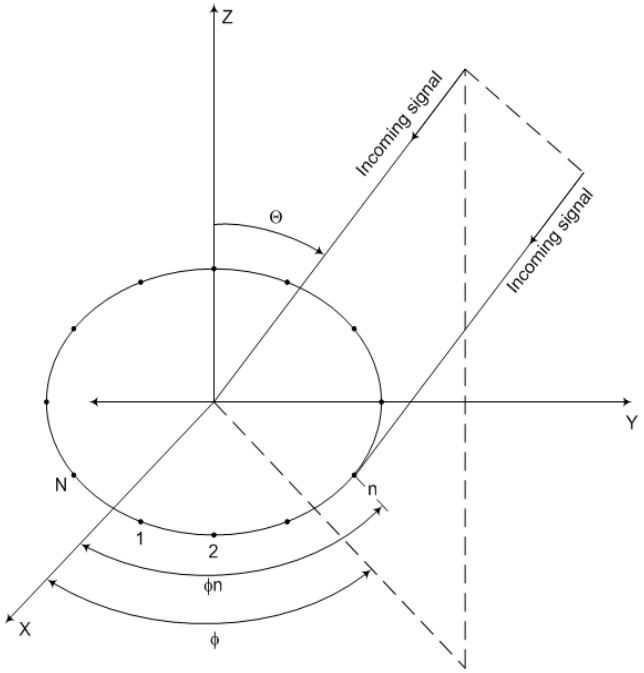


Fig. 1. UCA geometry, along with an incoming plane wave

B. Rectangular antenna structure

The URA consisting N x M equally distributed identical half-wavelength dipoles (M, N – even) is located symmetrical in x-y plane, as illustrated in Fig. 2.

The array factor *(AF)* of URA with its maximum along θ_0, ϕ_0 is given by [3]

$$[\mathrm{AF}(\theta,\phi)]_{\mathrm{MxN}} = 4\sum_{m=1}^{M/2}\sum_{n=1}^{N/2} A_{mn}\cos[(2m-1)u]\cos[(2n-1)v] \quad (2)$$

where

$$u = \frac{\pi d_x}{\lambda}(\sin\theta\cos\phi - \sin\theta_0\cos\phi_0) \quad (3)$$

$$v = \frac{\pi d_y}{\lambda}(\sin\theta\sin\phi - \sin\theta_0\sin\phi_0) \quad (4)$$

and A_{mn} is the amplitude excitation of the individual element, d_x, d_y are the interelement spacing along the x-axis and the y-axis, respectively.

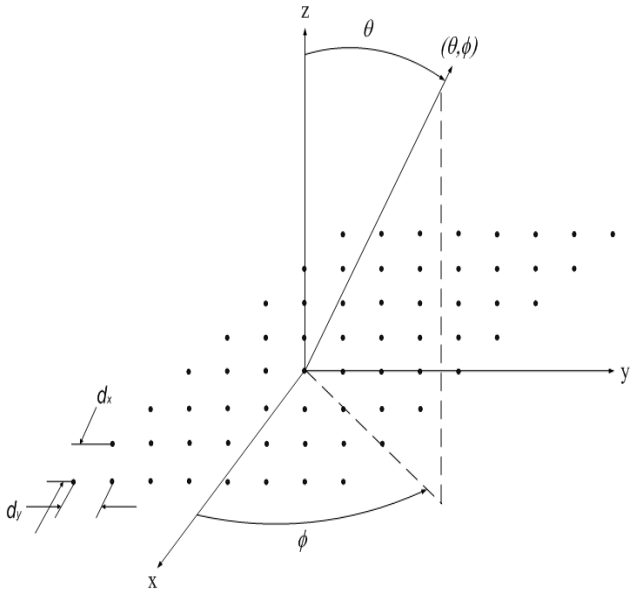


Fig. 2. URA geometry, along with an incoming plane wave

III. DIRECTION OF ARRIVAL ESTIMATION METHODS

A. UCA-ESPRIT method

The UCA-ESPRIT method is unique different from the classical ESPRIT. Applying this algorithm under the conditions of a UCA structure, the eigenvalues of each correlation array matrix have the form [4]

$$\lambda_i = \sin\theta_i e^{j\phi_i} \quad (5)$$

where (θ_i, ϕ_i) are respectively elevation and azimuth angles of incoming plane wave of *i*th signal source ($i = 1,2,...,M$), M is the number of the narrowband sources.

The three basic steps of real valued estimation are [1]:

1. The signal eigenvector estimation.
2. The equation system solution derived from eigenvectors computed in step 1.
3. The eigenvalues estimation of the solution to the system worked out in step 2.

This method gives several advantages in comparison with classical ESPRIT, such as: a) reduced computational complexity; b) very accurate finds simultaneously both the elevation and azimuth angles of arrival for incoming signals, and c) lower SNR (signal-to-noise ratio) resolution thresholds [5].

B. 2-D unitary ESPRIT method

The 2-D unitary ESPRIT algorithm is unique different from the classical ESPRIT, first of them provides closed-form automatically paired two dimensional estimation as long as the elevation and azimuth of each narrowband signal arrives at the URA [1]. Applying this method under the conditions of a URA structure (Fig. 2), the array manifold has the matrix form

$$\mathbf{A}(\mu,\nu) = \mathbf{a}_N(\mu)\mathbf{a}_M^T(\nu) \quad (6)$$

where the array manifold is

$$\mathbf{a}_N(\mu) = \left[e^{-j\left(\frac{N-1}{2}\right)\mu}, \ldots, e^{-j\mu}, 1, e^{j\mu}, \ldots, e^{j\left(\frac{N-1}{2}\right)\mu}\right]^T \quad (7)$$

$$\mu = \frac{2\pi}{\lambda} d_x p \quad (8)$$

λ is the wavelength, p is the direction cosine variable relative to the *x*-axis and

$$\mathbf{a}_M(\nu) = \left[e^{-j\left(\frac{M-1}{2}\right)\nu}, \ldots, e^{-j\nu}, 1, e^{j\nu}, \ldots, e^{j\left(\frac{M-1}{2}\right)\nu}\right]^T \quad (9)$$

is defined from $\mathbf{a}_N(\mu)$ with N, μ replaced by M, ν respectively

$$\nu = \frac{2\pi}{\lambda} d_y q \quad (10)$$

is the spatial frequency variable, q is the direction cosine variable relative to the *y*-axis.

Pre-multiplying $\mathbf{A}(\mu,\nu)$ by $\mathbf{Q}_N^H$ and post-multiplying by $\mathbf{Q}_M^*$, creates the real valued N *x* M array manifold

$$\begin{aligned}\mathbf{D}(\mu,\nu) &= \mathbf{Q}_N^H\mathbf{A}(\mu,\nu)\mathbf{Q}_M^* = \\ &= \mathbf{Q}_N^H\mathbf{a}_N(\mu)\mathbf{a}_M^T(\nu)\mathbf{Q}_M^* = \mathbf{d}_N(\mu)\mathbf{d}_M^T(\nu)\end{aligned} \quad (11)$$

where $\mathbf{Q}_N^H$, $\mathbf{Q}_M^H$ are a sparse unitary matrices that transforms $\mathbf{a}_N(\mu)$ into an N *x 1* real valued vector manifold

$$\mathbf{d}_N(\mu) = \mathbf{Q}_N^H\mathbf{a}_N(\mu) \quad (12)$$

$\mathbf{d}_M(\nu)$ is defined by $\mathbf{d}_N(\mu)$ with N, μ replaced by M, ν respectively.

The real valued N *x* M array satisfies

$$\tan\left(\frac{\mu}{2}\right)\mathbf{K}_1\mathbf{D}(\mu,\nu) = \mathbf{K}_2\mathbf{D}(\mu,\nu) \quad (13)$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are the real and imaginary parts of $\mathbf{Q}_{N-1}^H\mathbf{J}_2\mathbf{Q}_N^H$, $\mathbf{J}_2$ select the last *N-1* elements of an *N x 1* vector.

The *NM x 1* real valued manifold in vector form $\mathbf{d}(\mu,\nu) = vec[\mathbf{D}(\mu,\nu)]$ satisfies the requirement

$$\tan\left(\frac{\mu}{2}\right)\mathbf{K}_{\mu 1}\mathbf{D}(\mu,\nu)=\mathbf{K}_{\mu 2}\mathbf{D}(\mu,\nu) \quad (14)$$

where $\mathbf{K}_{\mu 1}$ and $\mathbf{K}_{\mu 2}$ are the *(N-1)M x NM* matrices

$$\mathbf{K}_{\mu 1}=\mathbf{I}_M\otimes\mathbf{K}_1,\ \mathbf{K}_{\mu 2}=\mathbf{I}_M\otimes\mathbf{K}_2 \quad (15)$$

where $\otimes$ denotes the Kronecker matrix product.

Similarly, $\mathbf{K}_3\mathbf{d}_M(\nu)=\mathbf{K}_4\mathbf{d}_M(\nu)$

where $\mathbf{K}_3=\mathrm{Re}\{\mathbf{Q}_{M-1}^H\mathbf{J}_2\mathbf{Q}_M\}$, $\mathbf{K}_4=\mathrm{Im}\{\mathbf{Q}_{M-1}^H\mathbf{J}_2\mathbf{Q}_M\}$

It follows that

$$\tan\left(\frac{\nu}{2}\right)\mathbf{D}(\mu,\nu)\mathbf{K}_3^T=\mathbf{D}(\mu,\nu)\mathbf{K}_4^T$$

$$\tan\left(\frac{\nu}{2}\right)\mathbf{K}_{\nu 1}\mathbf{d}(\mu,\nu)=\mathbf{K}_{\nu 2}\mathbf{d}(\mu,\nu) \quad (16)$$

where $\mathbf{K}_{\nu 1}$ and $\mathbf{K}_{\nu 2}$ are the *N(M-1) x NM* matrices

$$\mathbf{K}_{\nu 1}=\mathbf{K}_3\otimes\mathbf{I}_N,\ \mathbf{K}_{\nu 2}=\mathbf{K}_4\otimes\mathbf{I}_N \quad (17)$$

The real valued DOA matrix $\mathbf{D}=[\mathbf{d}(\mu_1,\nu_1),\ldots,\mathbf{d}(\mu_d,\nu_d)]$ satisfies

$$\mathbf{K}_{\mu 1}\mathbf{D}\boldsymbol{\Omega}_\mu=\mathbf{K}_{\mu 2}\mathbf{D} \quad (18)$$

where

$$\boldsymbol{\Omega}_\mu=diag\left\{\tan\left(\frac{\mu_1}{2}\right),\ldots,\tan\left(\frac{\mu_d}{2}\right)\right\},\ \mathbf{K}_{\nu 1}\mathbf{D}\boldsymbol{\Omega}_\nu=\mathbf{K}_{\nu 2}\mathbf{D} \quad (19)$$

where $\boldsymbol{\Omega}_\nu=diag\left\{\tan\left(\frac{\nu_1}{2}\right),\ldots,\tan\left(\frac{\nu_d}{2}\right)\right\}$

If $\mathbf{X}$ denotes the *NM x* N_s complex valued element space data matrix, the array output data matrix may be expressed as $\mathbf{Y}=\left(\mathbf{Q}_M^H\otimes\mathbf{Q}_N^H\right)\mathbf{X}$ and the appropriate *NM x d* signal eigenvector matrix $\mathbf{E}_s$ may be computed as the *d* "largest" left singular vectors of [Re{**Y**}, Im{**Y**}]. Viewing the signal relations as

$$\mathbf{K}_{\mu 1}\mathbf{E}_s\boldsymbol{\psi}_\mu=\mathbf{K}_{\mu 2}\mathbf{E}_s \text{ where } \boldsymbol{\psi}_\mu=\mathbf{T}^{-1}\boldsymbol{\Omega}_\mu\mathbf{T} \quad (20)$$

$$\mathbf{K}_{\nu 1}\mathbf{E}_s\boldsymbol{\psi}_\nu=\mathbf{K}_{\nu 2}\mathbf{E}_s \text{ where } \boldsymbol{\psi}_\nu=\mathbf{T}^{-1}\boldsymbol{\Omega}_\nu\mathbf{T} \quad (21)$$

Finally, from the eigenvalues of the matrix $\boldsymbol{\psi}_\mu+j\boldsymbol{\psi}_\nu=\mathbf{T}^{-1}\{\boldsymbol{\Omega}_\mu+j\boldsymbol{\Omega}_\nu\}\mathbf{T}$ compute spatial frequency estimates μ_i, ν_i.

This method provides closed form 2-D angle estimation in real time and has selfsame advantages as UCA-ESPRIT algorithm.

IV. Numerical examples for DOA estimation

The DOA estimation is investigated under the conditions of a UCA structure and URA structure with half-wavelength dipoles. Methods described above are utilized to perform the estimation [1]. The signal of interest (SOI) incomes from ($\theta=70^0,\phi=120^0$), while the three signals not of interest (SNOI) are directed from ($\theta=50^0,\phi=100^0$), ($\theta=45^0,\phi=95^0$), and ($\theta=90^0,\phi=140^0$) are given in Table I and Table III. The signal-to-noise ratio in all the considered cases is assumed to be (- 40) dB.

The results from UCA simulations are presented in Table II. The UCA with radius $a=0.6\lambda$ is examined about two scenarios: a) when array consists of *N=6* elements; b) when array consists of *N=8* elements.

The results from simulations for URA are described in Table IV. Two configurations of URA are examined: a) a (*N*=6, *M*=6) elements uniform rectangular array with $d_x=d_y=0.6\lambda$; b) a (*N*=8, *M*=8) elements uniform rectangular array with $d_x=d_y=0.6\lambda$.

TABLE I

THE UCA-ESPRIT DOA DATA

	Case 1	Case 2
Number of elements	M=6	M=8
Interelement spacing	0.6λ	0.6λ
Number of incoming signals	1	1
Number of data samples	2000	2000
Actual		
SOI	$\theta_1=70^0$, $\varphi_1=120^0$	$\theta_1=70^0$, $\varphi_1=120^0$
SNOI 1	$\theta_2=50^0$, $\varphi_2=100^0$	$\theta_2=50^0$, $\varphi_2=100^0$
SNOI 2	$\theta_3=45^0$, $\varphi_3=95^0$	$\theta_3=45^0$, $\varphi_3=95^0$
SNOI 3	$\theta_4=90^0$, $\varphi_4=140^0$	$\theta_4=90^0$, $\varphi_4=140^0$

TABLE II

THE DOA ESTIMATIONS OBTAINED UTILIZING UCA-ESPRIT

	Case 1	Case 2
Number of elements	M=6	M=8
Interelement spacing	0.6λ	0.6λ
Number of incoming signals	1	1
Number of data samples	2000	2000
DOA Estimations		
SOI	$\theta_1=70.020^0$, $\varphi_1=120.054^0$	$\theta_1=69.991^0$, $\varphi_1=119.994^0$
SNOI 1	$\theta_2=50.057^0$, $\varphi_2=100.051^0$	$\theta_2=49.998^0$, $\varphi_2=99.992^0$
SNOI 2	$\theta_3=44.971^0$, $\varphi_3=94.963^0$	$\theta_1=44.991^0$, $\varphi_1=94.992^0$
SNOI 3	$\theta_4=90.015^0$, $\varphi_4=139.969^0$	$\theta_2=89.991^0$, $\varphi_2=139.992^0$

TABLE III

THE 2-D UNITARY ESPRIT DOA DATA

	Case 1	Case 2
Number of elements	M=6 ,N=6	M=8, N=8
Interelement spacing	0.6λ	0.6λ
Number of incoming signals	1	1
Number of data samples	2000	2000
Actual		
SOI	$\theta_1=70^0$, $\varphi_1=120^0$	$\theta_1=70^0$, $\varphi_1=120^0$
SNOI 1	$\theta_2=50^0$, $\varphi_2=100^0$	$\theta_2=50^0$, $\varphi_2=100^0$
SNOI 2	$\theta_3=45^0$, $\varphi_3=95^0$	$\theta_3=45^0$, $\varphi_3=95^0$
SNOI 3	$\theta_4=90^0$, $\varphi_4=140^0$	$\theta_4=90^0$, $\varphi_4=140^0$

TABLE IV

THE DOA ESTIMATIONS OBTAINED USING 2-D UNITARY ESPRIT

	Case 1	Case 2
Number of elements	M=6 ,N=6	M=8, N=8
Interelement spacing	0.6λ	0.6λ
Number of incoming signals	1	1
Number of data samples	2000	2000
DOA Estimations		
SOI	$\theta_1=70.018^0$, $\varphi_1=120.051^0$	$\theta_1=69.994^0$, $\varphi_1=119.995^0$
SNOI 1	$\theta_2=50.055^0$, $\varphi_2=100.046^0$	$\theta_2=49.995^0$, $\varphi_2=99.995^0$
SNOI 2	$\theta_3=44.973^0$, $\varphi_3=94.966^0$	$\theta_3=44.983^0$, $\varphi_3=94.997^0$
SNOI 3	$\theta_4=90.018^0$, $\varphi_4=139.973^0$	$\theta_4=89.992^0$, $\varphi_4=139.998^0$

Both types of arrays are investigated in the presence of the Additive White Gaussian Noise (AWGN) with the zero mean, and variance 0.1. The results demonstrate theirs ability for accurate estimation, great performance, and robustness.

V. CONCLUSION

This paper investigated uniform circular and rectangular smart antennas with half-wavelength dipoles. A brief theory of two different antenna arrays to distinguish the direction of arrival by ESPRIT algorithm is considered. This theory was supported by suitable numerical data (see the Tables). The antennas were exploited in order to obtain more efficient method for a calculation of accurate DOA of impinging signals at the array.

Numerical examples have illustrated that the optimal scenario for the antenna geometry is UCA with *M*=6 elements, because of the symmetry UCA has almost the same performance as URA but with lower number of elements.

ACKNOWLEDGEMENT

Project no. 21 "Improvement of the research potential in the area of engineering and information technology" in the frame of Program "Development of human resources" and Ministry of Education of Bulgaria.

REFERENCES

[1] M. Zoltowski et al., “Closed-Form 2-D Angle Estimation with Rectangular Arrays in Element Space or Beamspace via Unitary ESPRIT”, IEEE Transactions on Signal Processing, vol. 44, no. 2, pp. 316-328, 1996.

[2] R. Roy, T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 7, pp. 984-995, 1989.

[3] L. C. Godara, “Smart Antennas”, CRC Press LLC, Florida, 2004.

[4] N. Herscovici, C. Christodoulou, "Uniform Circular Arrays for Smart Antennas", IEEE Antennas and Propagation Magazine, vol. 47, no. 4, pp. 192-206, 2005.

[5] C. Mathews, M. Zoltowski, “Eigenstructure for 2-D Angle Estimation with Uniform Circular Arrays”, IEEE Transactions on Signal Processing, vol. 42, no. 9, pp. 2395-2407, 1994.

[6] S. Bellofiore et al., “Smart-antenna systems for mobile communication networks, part 1: overview and antenna design”, IEEE Antennas and Propag. Magazine, vol. 44, pp. 106-114, 2002.

[7] K. Lonngren, S. Savov, “Fundamentals of Electromagnetics with Matlab”, SciTech, 2005.

Path loss calculation for a surface duct statistics

I. Sirkova[1]

Abstract – **This report applies recently published tropospheric ducts statistics for the region of Bosporus to the simulation of path loss for WCDMA FDD DL frequency band. The ducts parameters from the statistics are used as input to the parabolic wave equation method known to provide accurate path loss calculations in complex environments. Seasonal, monthly and daytime/nighttime variations of the path loss are presented. The results may be used for a preliminary assessment of coverage and possible interferences for the studied region and frequency band.**

Keywords – **Microwave propagation modelling, Duct statistics, Parabolic equation.**

I. INTRODUCTION

To account for the propagation conditions in coastal and maritime regions the propagation prediction models used need as input, besides other quantities, the meteorological parameters defining the tropospheric refractivity, especially when it differs from the standard troposphere case. The most severe deviation in refractivity from the standard conditions is the formation of tropospheric duct. The effect of ducted propagation on communications links has been studied in [1]-[4]. The difficulties in producing continuously *in situ* meteorological parameters measurements have lead to the use of global [5], [6] or local [7] climatology statistics in order to obtain the duct parameters. In [7], a two year statistics on surface duct formation over Istanbul has been reported. This area is rich in trapping layers forming different duct types: the annual percentage of surface duct occurrence is 31%, [7], the global seasonal statistics in [6] exhibit ducting occurrences for all type of ducts with averaged frequencies peaking at 60% in summer late afternoon. The availability of even short-time, but reliable, duct statistics has determined the interest in applying it to assess the microwave path loss variations due to anomalous tropospheric conditions for this region. In the present work, the duct parameters derived from [7] are used as input to the parabolic equation method known to provide accurate path loss calculations under complicated propagation conditions [8]-[10]. This report is related to [11] and follows the reference scenario relevant to coastal and maritime regions defined there.

The data reported in [7] are based on radiosonde measurements recorded at a meteorological station situated near Bosporus strait, Istanbul. The duct statistics provided refers particularly to the surface ducts for which the necessary parameters to reconstruct the modified refractivity M profiles using bi-linear model are the duct thickness Z_d and the duct strength ΔM (or M-deficit), [11]. Other types of ducts, surface-based or elevated, [11], [7], are not included in the statistics of [7]. Also, there is no differentiation of the evaporation ducts from the other surface ducts. Among the variety of data reported in [7] the present study makes use of the monthly variation of Z_d and ΔM (mean values extracted from Fig. 4, [7]) during the year, monthly variation of the nighttime and daytime mean values of the same parameters (Fig. 5 a) and c), [7]) and their seasonal variation for stable tropospheric conditions (extracted from Fig. 8 a) and b), [7]). The stable troposphere has been chosen because, following the reported statistics, the surface ducts occurrence has been higher under stable conditions than under unstable ones.

[1]Irina Sirkova is with the Institute of electronics, Bulgarian Academy of Sciences, 72 Tzarigradsko chaussee Blvd, Sofia 1784, Bulgaria, E-mail: irina@ie.bas.bg

II. PATH LOSS CALCULATIONS

The reconstructed through the bi-linear model M-profiles serve as input to the Advanced Propagation Model (APM) routines, used to compute the path loss. Those routines are based on a hybrid ray optics/PE method [12] and account simultaneously for microwave diffraction, refraction and scattering, thus providing accurate path loss computation. The initial field required to start the APM is provided by horizontally polarized Gaussian beam source with pattern factor given by (1) where θ_0 and θ_s are the half power beamwidth and the antenna elevation angle. Perfect conductivity of the ground/sea is assumed. The results are presented in the form of path loss (PL, in dB), equation (2), versus range for a fixed height.

$$F(\theta) = \exp\left[\frac{\ln(0.707)(\theta-\theta_s)^2}{\left(\theta_0/2\right)^2}\right], \qquad (1)$$

$$PL = 20\log\left(\frac{4\pi r}{\lambda}\right) - PF, \qquad (2)$$

In (2) λ is the free-space wavelength, r is the distance between the corresponding points and PF is the pattern propagation factor defined as the square of the ratio of the electric field amplitude E received at a given point under specific conditions to the amplitude of the electric field E_0 received under free-space conditions with the beam of the transmit antenna directed toward this given point. In this work, we used two carrier frequencies, f_{DLmin}=2112.4 MHz and f_{DLmax}=2167.4 MHz, situated at the two ends of the UMTS WCDMA-FDD downlink band. Ducting is the most important short-term interference mechanism over water and in flat coastal areas due to the increased propagation range. The possible interference on other, geographically close links

is more likely to come from the base station; that is why the downlink band has been chosen. The most samples are for f_{DLmin}, the higher the frequency the more affected it will be by the ducting. The reported results are for transmitter antenna height Z_t=30 m, θ_0=5^0, θ_s=0^0 and fixed height Z_r=20 m.

III. Results and Discussion

Fig. 1 shows comparison between path loss variations obtained for *M*-profiles based on monthly mean values of the surface duct parameters. It is to be noted that in this case Z_r is always "submerged" within the duct; the same is true for Z_t except for January, March and May. For distances greater than 6 km path loss under ducting differs significantly from month to month and may exceed or be lower than the standard troposphere case reaching a difference of more than 30 dB for September and r=7.8 km. September is characterized by moderate Z_d but it is with strongest ΔM. Due to the bi-linear profile of M, months with similar ratio $\Delta M/Z_d$ demonstrate similar path loss behavior. Fig. 2 depicts path loss variations for seasonal mean surface ducts parameters. For summer and autumn months path loss variations compared to standard tropospheric case exceed 15 dB for some ranges (r=9.8 km, r=12 km, respectively). Fig. 3 compares path loss variations for the worst summer case for which the pair Z_d-ΔM is available for f_{DLmin} and f_{DLmax}. The two frequencies are compared also for September, a month with strong values of ΔM. The behavior of path loss variations for this particular scenario for the two frequencies is similar. It is to be noted that for different scenario (different values of Z_t, Z_r and/or Z_d-ΔM) the differences in the two end of DL band may be grater. Figs. 4 and 5 compare path loss variations obtained for *M*-profiles for daytime and nighttime mean duct parameters for winter/spring months and summer/autumn months, respectively. Nighttime/daytime path loss differences reach significant values: 9 dB for May (r=9.8 km), 10 dB for September (r about 12 km), more than 12 dB for October (r=9.5 km).

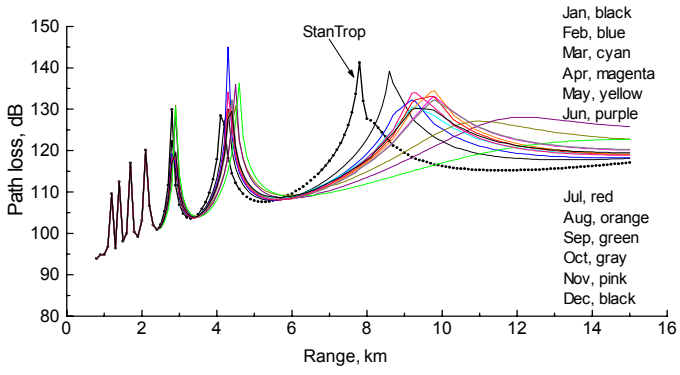


Fig. 1 Comparison between path losses obtained for M-profiles based on monthly mean values of duct parameters.

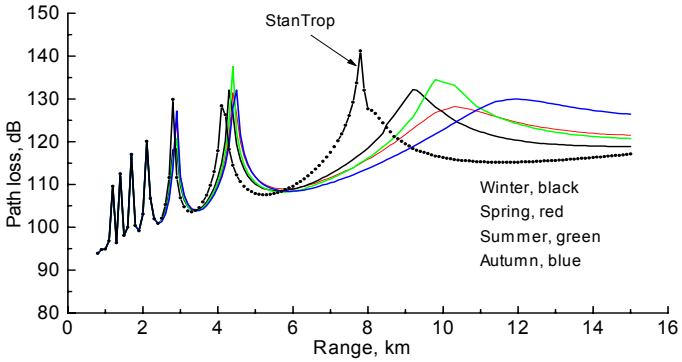


Fig. 2 Comparison between path losses obtained for M-profiles for seasonal mean duct parameters.

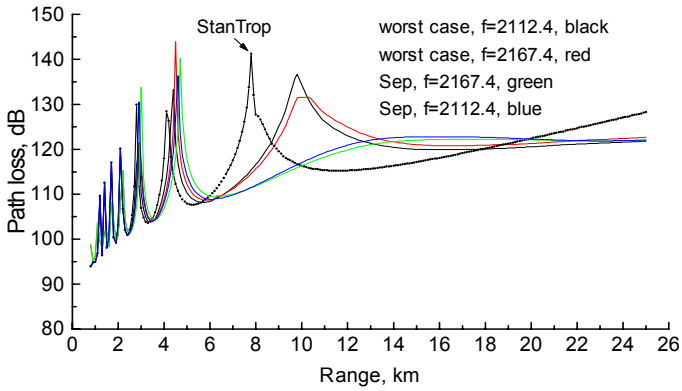


Fig. 3 Comparison between f_{DLmin} and f_{DLmax} path losses obtained for M-profiles for the worst summer case (Z_d=70 m, ΔM=12 M-units) and for September mean duct parameters.

IV. Conclusion

The presented study is limited to the effect due to surface ducts only. For more detailed study a long-term duct statistics differentiating between different duct types and accounting for range-dependent duct thickness and strength is needed. The last is required by the mixed land-sea-land path in the area. Finally, the duct statistics reported in [7] tends to underestimate the ducts occurrence due to the limitations in the resolution of the radiosonde data used. Nevertheless, even though tentative, the presented study shows: a) the path loss differences of scores of dB due to surface ducting will affect the link budget and increase the requirements to the WCDMA power control range; b) how the duct statistics should be used: when predicting the *PL*, use of monthly mean values for the duct parameters rather than annually averaged values is needed. A further differentiation between nighttime/daytime cases may also be required to account for specific climatic characteristics and influences.

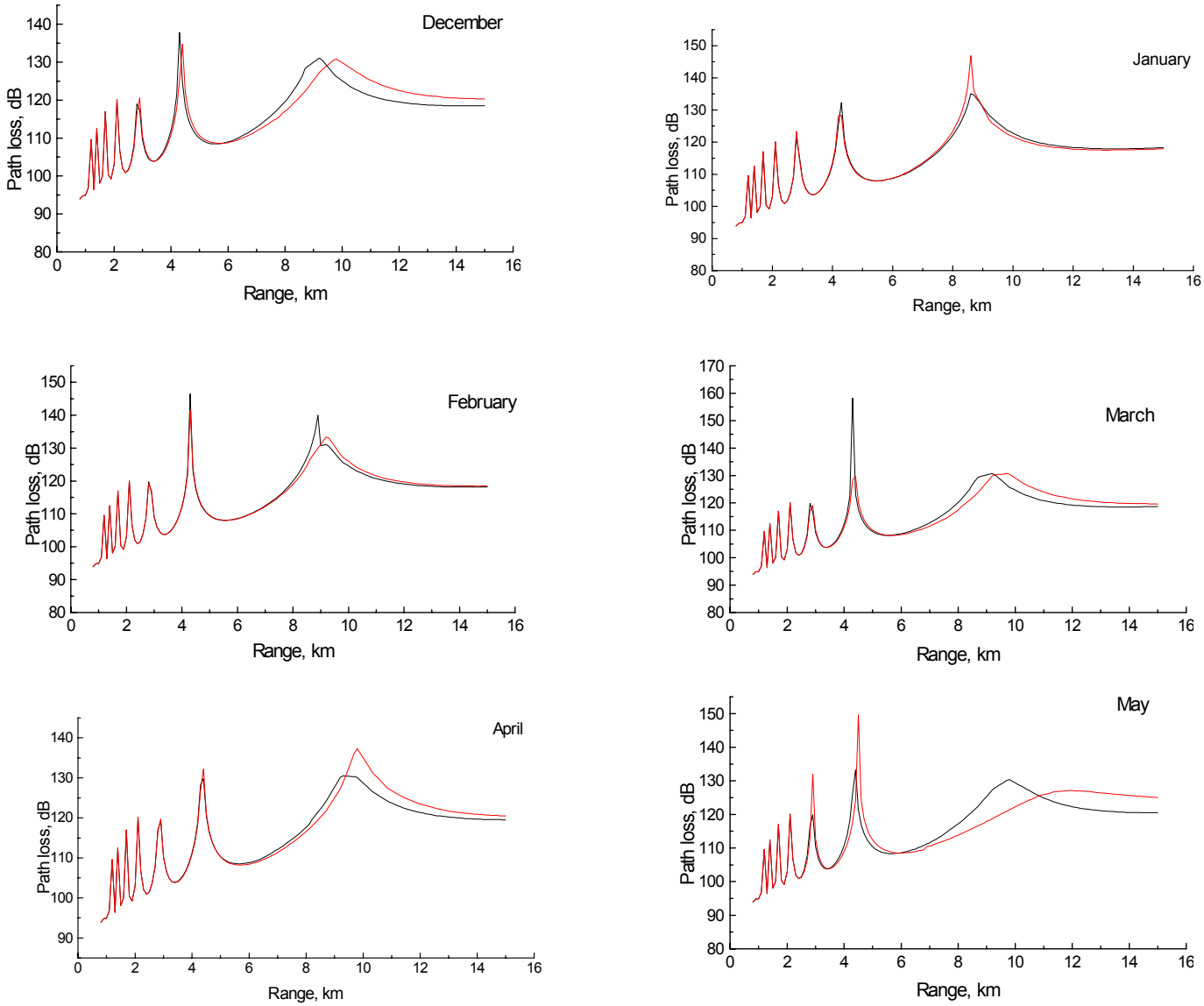


Fig. 4 Comparison between path losses obtained for *M*-profiles for daytime (red) and nighttime (black) mean duct parameters for winter and spring months.

ACKNOWLEDGEMENT

This work was partially supported by Action 2100 of the European program COST. The author is grateful to the SPAWAR Systems Centre, San Diego, USA, for making available the APM code.

REFERENCES

[1] A. Kerans, A. S. Kulessa, E. Lensson, G. French, G. S. Woods, "Implications of the evaporation duct for microwave radio path design over tropical oceans in Northern Australia", Proc. WARS (Workshop Applications of Radio Science), 4 p., Leura, Australia, 20-22 Feb 2002.

[2] M. J. Willis, K. H. Craig, "Results from a Long Term Propagation Measurement Campaign", 2nd EuCAP, pp. 1-8, Edinburgh, UK, 11-16 Nov 2007.

[3] I. Sirkova, M. Mikhalev, "Parabolic-equation-based study of ducting effects on microwave propagation", J. of Microw. Opt. Technol. Lett., vol. 42, no. 5, 2004, pp. 390-394.

[4] D. Siddle, E. M. Warrington, S. Gunashekar, "Transhorizon propagation at 2GHz", 4th COST 2100 MCM, 13 p., Wroclaw, Poland, 6-8 Feb 2008.

[5] A. von Engeln, J. Teixeira, "A Ducting Climatology derived from ECMWF Global Analysis Fields", J. Geophys. Res., vol. 109, no. D18, 2004.

[6] Ph. Lopez, "A 5-year 40-km Resolution Global Climatology of Super-refraction for Ground-based Weather Radars", J. of

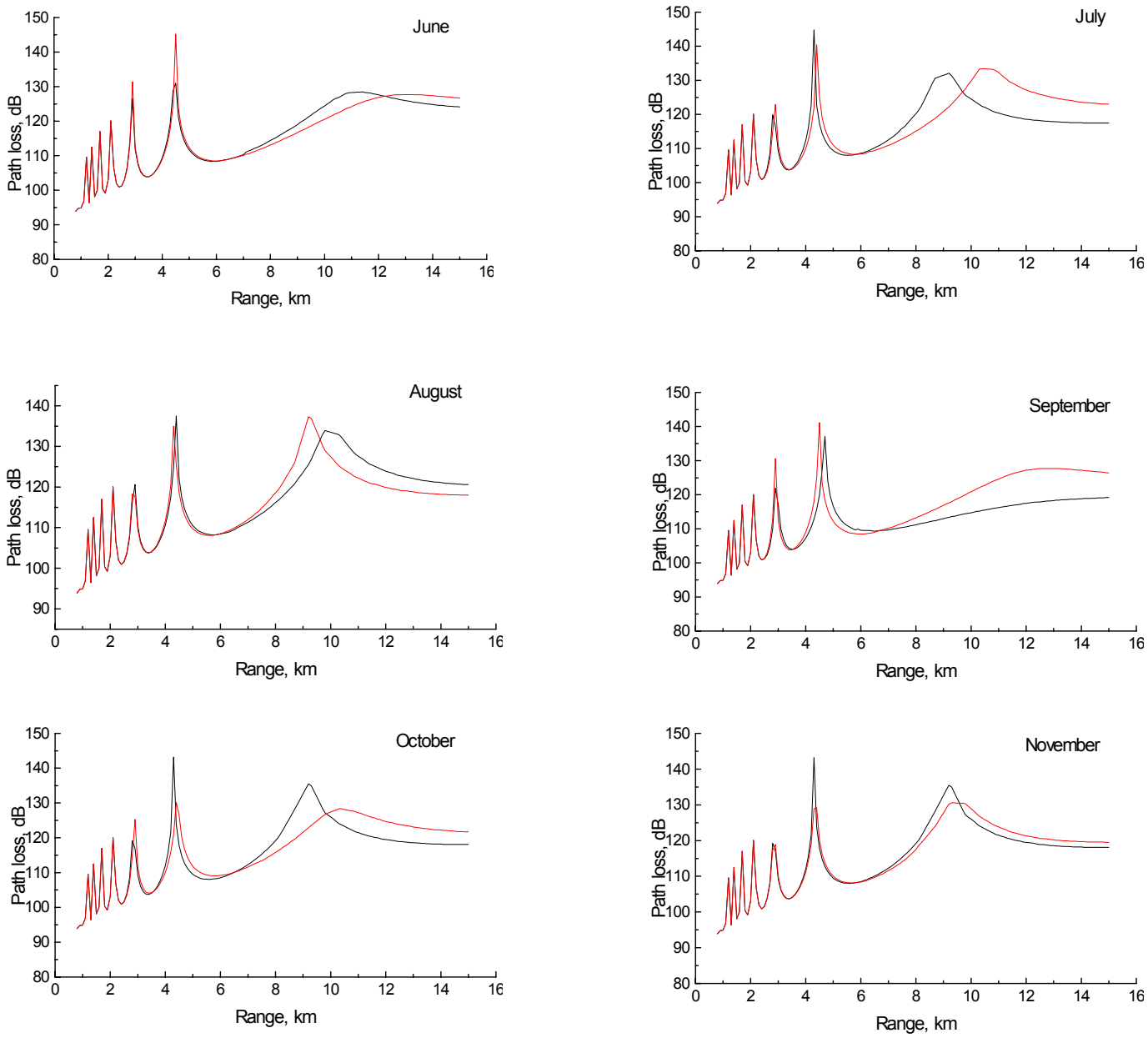


Fig. 5 Comparison between path losses obtained for *M*-profiles for daytime (red) and nighttime (black) mean duct parameters for summer and autumn months.

Applied Meteorology & Climatology, doi: 10.1175/2008JAMC1961.1, 2008.

[7] S. S. Mentes, Z. Kaymaz, "Investigation of Surface Duct Conditions over Istanbul, Turkey", J. of Applied Meteorology & Climatology, vol. 46, pp. 318-337, 2007.

[8] M. Levy, *Parabolic equation methods for electromagnetic wave propagation*, IEE electromagnetic waves series 45, UK, 2000.

[9] A. E. Barrios, K. Anderson, G. Lindem, "Low Altitude Propagation Effects - A Validation Study of the Advanced Propagation Model (APM) for Mobile Radio Applications", IEEE Trans., Antennas &. Prop., vol. 54, no.10, pp. 2869-2877, 2006,.

[10] I. Sirkova, M. Mikhalev, "Parabolic Wave Equation Method Applied to the Tropospheric Ducting Propagation Problem – a Survey", Electromagnetics, vol. 26, no. 2, pp. 155-173, 2006.

[11] I. Sirkova, "Proposal for a reference scenario relevant to coastal and maritime regions", 4th COST 2100 MCM, 6 p., Wroclaw, Poland, 6-8 Feb 2008.

[12] A. E. Barrios, W. L. Patterson, R. A. Sprague, Advanced Propagation Model (APM). Computer Software Configuration Item (CSCI) Documents. *SSC SD TD 3214*, 2007.

Suppression Mutual Coupling between the Resonators in Microstrip Antennas by using PBG Substrates

Nikolay M. Stoyanov[1], Tchavdar P. Levtchev[2], Georgi K. Georgiev[3]

Abstract – **In the paper investigations related to suppress the mutual couplings using 1D and 2D dielectric structures are presented. Investigations and analysis are made in the sequence: offering dependencies applicable method – moments in the case of one-and two-dimensional periodic dielectric structures; simulation study of periodic and non-periodic dielectric structures to reduce mutual influences between the planar microstrip antenna grids.**

Keywords – **Microstrip antenna, PBG, substrate, antenna array.**

I. Introduction

Reduction of mutual coupling between the resonators in microstrip antenna arrays can be achieved by change of permittivity in the area between resonators or by periodic dielectric structures. In essence, these techniques are further developing the theory of dielectric antennas, but the regimes in which they do not radiate or play the role of dielectric band-stop filters.

Periodic dielectric materials (PBG - photonic-band gap), or also known as photon crystals are periodic inhomogeneous materials that stops the propagation of electromagnetic waves in a given bandwidth. The stop-band of the frequency range depends on the lattice constant of *b* (the ratio of the size of individual elements to the distance between elements), which are built this structure [1, 2]. Exact frequency band of a PBG structure depends on many physical parameters such as parameters of the environment and type of the lattice geometry of the individual element. In the antenna arrays, applications the structures are mainly constructed in one and two-dimensional grids. In the most commonly used substrate, are made holes or slots with a fixed period of the lattice. In more complex lattice configurations can have any configuration and shape.

Besides these parameters influence has and thickness of the substrate. As a rule we can say that in order to obtain the effect of suppression of surface wave it is necessary to use relatively thick (compared to wavelength) substrate with a large relative permittivity. Of course, in a real dielectric substrate thickness affects the transmission losses and the relative coefficient of refraction of electromagnetic waves.

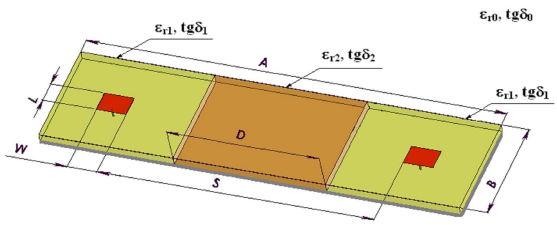


Fig. 1. Test setup for the study of the mutual influences between the two microstrip resonators by changing of permittivity in the area between them and the change in the width of this area.

II. One Dimensional Periodic Dielectric Structures

A. Mutual coupling between two resonators in the change of permittivity in the area between them

In this point has been made simulation investigation of the mutual coupling between two microstrip resonators as the parameters of the dielectric substrate does not modify the whole structure, but only in a specific area between them.

In other words interest represents, what impact has the introduction of the permittivity no uniformity with certain size and a relative permittivity in the area between resonators. The calculations were made with parameters: substrate thickness h = 1,5 mm; relative permittivity for the area of the resonators ε_{r1}=10,2; modification of the relative permittivity of the area between resonators ε_{r2}=1… 20; dielectric losses $tan\delta_{\varepsilon}$ = 0.003; thickness of the metallization t = 17.5 μ m; surface resistance of copper R_s = 1,82 μς/cm; baseline distance between the resonators $S[mm] = 0.5\lambda_0$; modification of the width D *[mm]* permittivity of non-uniformity $0 \le D \le S$, i.e. change in the ratio $D/S \in [0,1]$.

B. Numerical results

Electromagnetic simulation is made using RF electromagnetic simulator. Parameters as defined in the study are conducted by maintaining the best matching of the resonators. On Fig.2 is a Smith chart of the configuration of Fig.1. It clearly can be seen that the change of permittivity dielectric between the two does not submit additional parasitic resonance.

Fig. 3 shows the dependence of coefficients S_{12} (S_{21}) in the modification of the ratio D / S (Fig.1.) for three different values of permittivity of the intermediate dielectric. Clearly the highest value of the mutual coupling between the resonators is in the case when is used highest dielectric permittivity value, but at lowest value is at ε_{r2}=1. When

[1] Nikolay M. Stoyanov, phd, e-mail: nmstoyanov@gmail.com
[2]Chavdar P. Levchev, phd, e-mail: ch_levchev@ncc.mvr.bg
[3]Georgi K. Georgiev, e-mail: gogodreamer@yahoo.com

ε_{r2}=11 the amendment is minor, because the value of permittivity is very close to that of r_1 = 10.2 (patches area).

From the figure can be observed that for $D/S \cong 0{,}73$ in the value of ε_{r2}= 20 is seen maximum and down on the value of the S_{12} starts to decline slightly. Graph relating to ε_{r2} = 1 in value of the ratio $D/S \cong 0{,}64$ starts to decrease more quickly. This character of change of mutual influence as a function of the size of the second dielectric for different values of ε_{r2} has important practical significance. Because of this feature may be designed frequency-selective structures to suppress the propagation of surface wave. Given the above can be made the following conclusions:

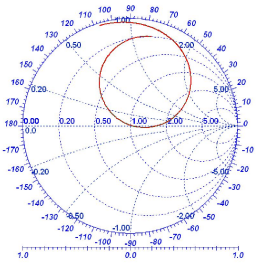


Fig. 2. Smith chart of S11 and S22 parameters of the two patches.

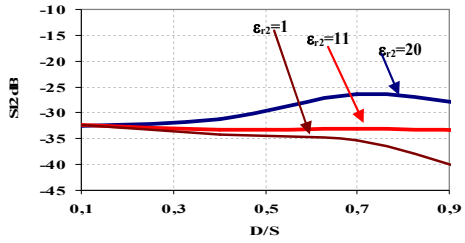


Fig. 3. Results from the study of mutual influence between two microstrip resonator with a change of permittivity in the field between them and the change in the breadth of this area

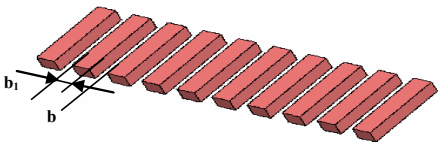


Fig. 4. One-dimensional dielectric periodic structure with a period b and width of the dielectric layer b_1.

Changing the relative permittivity in the area between resonators does not lead to significant suppressing of surface waves and the mutual influence between the resonators.

Weak resonance character, which is seen when comparing the graphs of Fig. 3 due to mish-matching of the resonators as result of changes in the permittivity adjacent to them;

Least mutual influence was observed in the case using air. This is because in this case have not propagation of surface waves and the mutual influence is mainly due to the impact of the metal ground plane. Reduce this influence can be further achieved if instead of a complete ground plane metallization using resonance structure.

A. Analytical Model for The Presentation Of The Field In One-Dimensional Periodic Dielectric Structures

In this section is made mathematical analysis and simulation of one-dimensional dielectric structures in order to reduce the mutual influences between the planar antenna grids.

For spreading modes in a periodic permittivity structure, field in the neighboring cells is associated with a complex constant [4]. This may be shown in the following ways:

$$E(x,y,z+b)=E(x,y,z)e^{-j\beta_0 b}, \quad (1)$$

where *E (x, y, z)* is a periodic function of *z* with a phase constant in a direction z β_0.

Distributed modes in the *z* direction can be represented as:

$$E(x,y,z)=E_p(x,y,z)e^{-j\beta_0 z}, \quad (2)$$

where $E_p(x,y,z)$ is a periodic function of *z* with period *b*, then

$$E(x,y,z+b)=E_p(x,y,z+b)e^{-j\beta_0(z+b)}, \quad (3)$$

Because $E_p(x,y,z)$ is a periodic function with period *a:*

$$E_p=(x,y,z+b)=E_p(x,y,z) \quad (4)$$

By substitute (3) in (4) can be obtained the following formula:

$$E(x,y,z+b)=E_p(x,y,z)e^{-j\beta_0(z+b)}. \quad (5)$$

But taking into account (2) is obtained exactly theorem of Floquet:

$$E(x,y,z+b)=E(x,y,z)e^{-j\beta_0 b} \quad (6)$$

$$E(x,y,z)=\sum_n E_n(x,z)e^{-j\frac{2\pi n}{b}z}e^{-j\beta_o z}=\sum_n E_n(x,y)e^{-j\beta_n z} \quad (7)$$

where

$$E_n(x,y)=\frac{1}{b}\int_0^b E_n(x,y,z)e^{j\frac{2\pi n}{b}z}dz \quad (8)$$

are coefficients which represent the dependencies on *x* and *y*

$$\beta_n=\beta_0+\frac{2\pi n}{b}$$

is the phase constant of these n-harmonics.

Thus, the field in the periodic structure may be decayed to the infinite number of harmonics by the theorem of Floquet, each with frequency f and propagation constant β_n.

Be considered a flat electromagnetic wave, which propagates across the periodic structure with a permittivity and a dielectric layer of width *b*.

The decision of the wave equation for one-dimensional dielectric structures of the type shown in Fig. 4 can be obtained by decomposition in a periodic series of Fourier.

Electrical field is a periodic function decay of flat waves in x with period b, (which determines the constant of distribution k_{x0}) and obtained in following form:

$$E(x,y)=\hat{z}E_z(x,y)=\hat{z}E_p(x)e^{-jk_{x0}x}e^{-jk_y y} \tag{9}$$

where $E_p(x)$ is a periodic electric field which is distributed only in the XY-plane, i.e. $k_z = 0$. Assuming that the parallel layers in endless directions y and z, wave equation can be simplified and presented as:

$$-\frac{d^2}{dx^2}E_z(x,y)+k_0^2\varepsilon_r(x)E_z(x,y)=k_0^2\varepsilon_r(x)E_z(x,y) \tag{10}$$

Periodic electric field to decay by a Fourier series along the axis x with unknown coefficient a_n, which serves to present dependence on the axis Y.

$$E_p(x)=\sum_n b^n e^{-j\frac{2\pi n}{b}x} \tag{11}$$

Permittivity is also periodic and is suitable for the decomposition in Fourier series with coefficient b_m

$$\varepsilon_r(x)=\sum_m b_m e^{-j\frac{2\pi m}{b}x} \tag{12}$$

Make a substitution of the Fourier decomposition of the field and in the dielectric (9) and obtained:

$$\sum_n\left[\left(\frac{2\pi n}{b}+k_{x0}\right)^2+k_y^2\right]a_n e^{-j\frac{2\pi n}{b}x}=$$

$$k_0^2\sum_n\sum_m a_n b_m e^{-j\frac{2\pi m}{b}x}e^{-j\frac{2\pi n}{b}x} \tag{13}$$

In accordance with the determination of the unknown coefficients a_n and b_m Eq. (12) are multiplied by orthogonal function and is integrated for a single cell with a specific index using Kroneker's delta function.

$$\sum_n\left[\left(\frac{2\pi n}{b}+k_{x0}\right)^2+k_y^2\right]a_n\delta\left(\frac{2\pi p}{b}-\frac{2\pi n}{b}\right)=$$

$$k_0^2\sum_n\sum_m a_n b_m\delta\left(\frac{2\pi p}{b}-\frac{2\pi m}{b}-\frac{2\pi n}{b}\right) \tag{14}$$

Likewise can be determined and propagating TM-modes. By using the derived relationships is relatively easy to determine the frequency and extent of distribution of TE modes in periodic-permittivity substrate. This is a convenient approach to applying the method of moments.

B. Evaluation of mutual influence between two microstrip resonators using one-dimensional periodic structure in permittivity between them

Dielectric periodic structures as explained can be constructed in such a way as to achieve effective filtration structure for suppression of certain modes. Given the complexity of this issue in such configurations must be taken into account the fact that the most influential has TM_0 mode of surface waves. Matching the appropriate values of thickness and relative permittivity of the dielectric substrate may reduce the propagation of the TE_1 mode. Thus, the periodic structure can be realized only suppression of TM_0 mode of surface waves. After defining the geometry of the periodic structure, the dispersion relationships for transverse field components can be calculated by the method described above. The main parameter to be calculated is the ratio b_1 / b to obtain the desired effect of reducing the respective mode.

Practical computing of band-stop ranges can be made using the microstrip line located on one particular made periodic permittivity substrate with a b_1 / b (Fig. 5). Graphs obtained from the calculations can be presented as Brillouin zones (BZ). They represent the dependence of normalized frequency, depending on the distribution constant for the area. In the case of $1D$ structure constant of the distribution depends only on the ratio b_1 / b.

Basic configuration, which may find application in the antenna arrays is shown in Fig. 5. Again as in previous models, the parameters of permittivity substrate in the resonators are fixed and they not changed during the study but the relative permittivity of the area between resonators is changing [3]. This is a modular approach to build antenna arrays, i.e. have not used a single permittivity substrate. Thus it is possible to optimize the antenna further when in the area between radiating elements are placed specific filtering material for the wished bandwidth properties. Although seemingly more complex structure such antenna array is not a problem in technological terms, and in some cases can be used to reduce cost, weight and dimensions.

Parameters of permittivity substrate and microstrip line are:

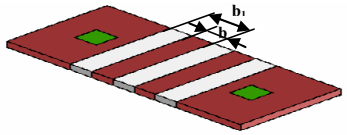


Fig. 5. Setup for determining the level of reduction of the mutual influences through $1D$ periodic structure with permittivity b time and width of dielectric layer $b1$

input impedance Zin =50Ω; filling factor of the intermediate dielectric $b1/b$ = 0,5; permittivity of the substrate ε_r= 10,2; width of the dielectric substrate h=1,27mm;

Graphic results of calculations of the mutual influence between two adjacent microstrip resonators $1D$ dielectric structure with between them are presented in Fig. 6. Can be seen that the main focus the analysis on the impact, which has the value of relative permittivity on the coefficient of correlation of two adjacent microstrip resonator in E-plane to one another. It should be noted that in the entire sequence of tests for different values of ε_r resonance frequency of microstrip resonators is constant and is therefore presented a picture for the S_{11}.

Completely logically equivalent when increasing the relative permittivity in the area between the two resonators increases the mutual influence between the frequency of resonance. The latter is indeed true and observed values in the relative permittivity of ε_r = 1 to ε_r = 31. For large value of relative permittivity ε_r = 50 (eg using special ceramics) the value of the S_{12} is considerably smaller than even the case when using air. This fact can be explained by the fact that in a

certain ratio between the permittivity of air and the second dielectric and, in certain size resonance occurs.

Resonance character can be observed with other graphics, but for other frequencies. Graph of these areas are marked with circles. Watching them more closely you can see that in the linear change of relative permittivity is observed slightly exponentially change the frequency of such resonance. However with increasing frequency, ε_r is reduced. This dependence is illustrated in Fig.7.

It is observed that a large change in relative permittivity does not lead to a drastic change of resonance frequency. However, the study shows that in the optimal selection of sizes and permittivity can achieve significant reduction of the mutual influences between the elements.

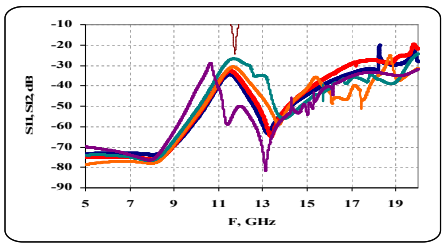


Fig. 6. Numerical results of the investigation of mutual influence between two microstrip resonators with permittivity placing *1D* structure

III. Output Data and Method of Analysis to Determine Areas of Surface Wave Propagation in the Case of 2D Structures

The simplest way to obtain the permittivity of two-dimensional structure is in plain permittivity substrate with certain values of relative permittivity and thickness to drill holes with defined diameter and period (Fig. 8).

Parameters of permittivity substrate and microstrip line are: input impedance of microstrip line Z_{in} = 50Ω; b_1 / b = 0,5; permittivity ε_r = 10,2; substrate thickness h = 1,27 mm.

Calculation of microstrip line is made using the electromagnetic simulator *SV Ansoft Designer*, whose width is W = 1.3314mm. To calculate the Brillouin zones is necessary to determine the configuration of the structure (rectangular or triangular lattice of holes).

Fig. 9 presents the results of the simulation study of the structure of Fig.8. For comparison, the graph shows the coefficient of correlation in the case when no holes. Can be seen that is possible to receive frequency band, in which the structure has high impedance (11 GHz).

Survey shows that in an appropriate selection of the relative

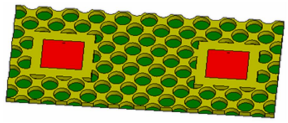


Fig. 8. Setup for investigation of mutual coupling in *2D* periodic structure (PBG) patch antenna.

permittivity and thickness of the permittivity substrate can reduce the degree of mutual influence between the planar antenna grids. From the graph in Fig. 9 shows that the suppression of mutual influences is very narrow bandwidth.

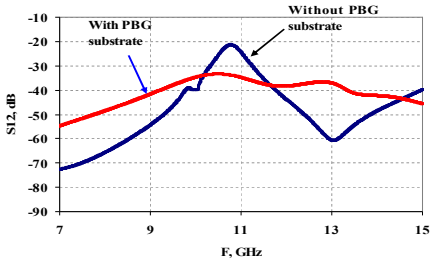


Fig.9. Comparison of the degree of mutual influence in the cases with and without the use of PBG structures

IV. Conclusions

Based on the theoretical and simulation studies on the possibility of suppression of mutual influence between the planar antenna arrays can be used fully periodic dielectric structures (one or two-dimensional). As a disadvantage may be stated that to obtain a strong effect is necessary to use materials with high permittivity (ε_r> 10) and substrates with high thickness. In addition, it is necessary to note that in this way is difficult to reduce the size of structures, because to obtain a desired field distribution of electromagnetic field (to achieve high impedance) is necessary to ensure a minimum volume. Another feature is that when using periodic dielectric structures can reduce the weight of the antenna, but at the expense of her physical strength.

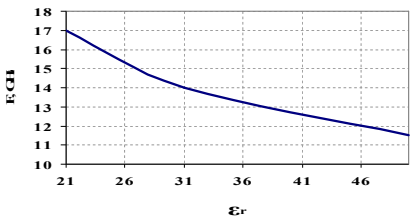


Fig. 7. Graphical representation of the variation of resonance frequency change of the value of the relative permittivity of the periodic structure

References

[1] T. Itoh, C. Caloz, "Electromagnetic metamaterials: transmission line theory and microwave applications: the engineering approach", John Wiley & Sons, Inc, New Jersey, 2006.

[2] J. D. Shumpert, "Modeling of Periodic Dielectric Structures, PHD Dissertation, University of Michigan, 2001.

[3] R. Gonzalo, P. De Maagt, M.Sorolla, "Enhanced patch-antenna performance by suppressing surface waves using photonic-bandgap substrates", IEEE Transactions on HMicrowave Theory and Techniques,vol.47, No.11, pp.2131-2138, November 1999.

Session

RADIO COMMUNICATION SYSTEMS – 1

Statistics of the SSC Output SIR in the Presence of Correlated Weibull Fading and Interference

Aleksandra M. Cvetkovic[1], Jelena A. Anastasov[2], Goran T. Đorđević[3], Mihajlo Č. Stefanović[4]

Abstract - **In wireless communication systems, effects of fading and cochannel interference can be diminished using diversity reception. In those environments, the level of the cochannel interference is sufficiently high as compared to the thermal noise so the effect of the thermal noise can be ignored. In this paper SSC (switch and stay combining) diversity system, based on signal-to-interference ratio (SIR) in the presence of correlated Weibull fading and interference, is analyzed. The probability density function (pdf), cumulative distribution function (cdf) and moments of the SSC output SIR are analytically defined. The influences of threshold, parameters of fading and correlation coefficient to the first order statistics are also considered. Numerical results, based on proposed analytical analysis are confirmed by Monte Carlo simulations.**

Keywords – **diversity receivers, moments, Monte Carlo simulation, Weibull fading.**

I. INTRODUCTION

There is great need for alleviating bad influences of fading and co-channel interference in wireless communication systems. Diversity reception is one of commonly used technique to combat all those bad influences and to upraise system performance gain [1-2]. There are several combining methods, which division can be performed by their dependence on complexity of communication system and amount of channel state information available at the receiver. Equal-Gain Combining (EGC) and Maximal-Ratio Combining (MRC) require all or some of the amount of the channel state information of received signal. So, there is need of one receiver for every branch, which means higher complexity for practical realization. In opposition to these techniques (MRC and EGC), there is Selection Combining (SC) and Switch and Stay Combining (SSC), which process only one of the diversity branches. SC receiver processes all branches continuously and chooses the best one, while SSC receiver processes selected branch and proceed to the other one, when the previous becomes unacceptable [3]. When a SSC diversity combiner is subject to co-channel interference, the combining decision algorithm is not unique. There are three different algorithms of determining: based on total (desired plus interference) power; based on desired signal power and based on signal-to-interference ratio (SIR) [2], [4]. The algorithm of determining based on SIR provides the best performance gain of system with cochannel interference [4-5]. In interference-limited systems, the level of cochannel is sufficiently high as compared to thermal noise so the effects of thermal noise can be ignored.

In some papers [5-8], the outage probability over uncorrelated fading channels in the presence of interference has been analyticaly determinated. However, uncorrelated fading assumes antennas to be placed sufficiently apart, which is not always realized in practice due to insufficient antenna spacing when diversity is applied in small terminals [3]. The outage probability in dual SIR-based SC and SSC with correlated and uncorrelated branches over Rayleigh fading channels has been evaluated and published in [9]. Moreover, analytical studies investigating outage probability and other performance measures of SC with the algorithm of determining based on SIR in correlated Nakagami-m and Weibull fading channels are reported in literature [10] and [11], respectively.

In opposition to analysis presented in [11], in this paper dual SSC diversity system in the presence of interference is obtained, instead of SC which processes both branches. The statistics of system's output SIR of correlated Weibull desired signal and interference envelopes is presented. The new form of formulae for the probability density function (pdf), cumulative distribution function (cdf) and average output SIR of the SSC are analytically defined. Numerical results are based on proposed analytical analysis and confirmed by Monte Carlo simulations.

II. CHANNEL MODEL

We model fading and interference by Weibull distribution which is flexibile and often used for describing urban environments in cases for which Rayleigh distribution is inadequate [12-13]. The desired signal as well as interfering one received by ith antenna in Weibull fading environment can be written as [5]

$$y(t) = x(t)\exp\left(j(2\pi f_c t + \Phi(t) + \phi(t))\right) \quad (1)$$

with f_c being the carrier fraquency, $\Phi(t)$ the desired information signal, $\phi(t)$ the random phase uniformly distributed in $[0, 2\pi)$ and $x(t)$ a Weibull distributed random amplitude process [14]

[1]Aleksandra M. Cvetkovic is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: aleksandra.cvetkovic@elfak.ni.ac.rs

[2]Jelena A. Anastasov is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: anastasovjelena@gmail.com

[3]Goran T. Đorđević is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: goran@elfak.ni.ac.rs

[4]Mihajlo Č. Stefanović is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: mihajlo.stefanovic@ elfak.ni.ac.rs

$$p_x(x)=\frac{\beta}{\Omega}x^{\beta-1}\exp\left(-\frac{x^\beta}{\Omega}\right) \quad (2)$$

where $\Omega=\mathrm{E}[x^\beta]$ (E[.]) denoting expectation, parameter β being Weibull fading parameter (β >0). When parameter β increases, fading severity decreases. By setting β=1 Weibull distribution becomes exponential and by setting β=2 Weibull becomes Rayleigh [14].

Let us consider the system where the fading amplitude of the desired $R(t)$ and interfering $r(t)$ signal have Weibull distribution. Then the PDF for instantaneous SIR, denoted by $z=R^2/r^2$, has the form

$$p_z(z)=\frac{\beta\Omega_d}{2\Omega_i}\frac{z^{\beta/2-1}}{\left(1+\frac{\Omega_d}{\Omega_i}z^{\beta/2}\right)^2} \quad (3)$$

where $\Omega_d=\mathrm{E}[R^\beta]$, $\Omega_i=\mathrm{E}[r^\beta]$. The cdf of instantaneous SIR is:

$$F_z(z)=\frac{z^{\beta/2}}{\frac{\Omega_d}{\Omega_i}+z^{\beta/2}} \quad (4)$$

Since $\mathrm{E}[R^2]=\Omega_d^{2/\beta}\Gamma(1+2/\beta)$, $\mathrm{E}[r^2]=\Omega_i^{2/\beta}\Gamma(1+2/\beta)$ [14], and $S=\mathrm{E}[R^2]/\mathrm{E}[r^2]=(\Omega_d/\Omega_i)^{2/\beta}$ ($\Omega_d/\Omega_i=S^{\beta/2}$) are the average SIRs equal at two input branches of the SSC, then the dual SIR-based SSC output pdf (3) and cdf (4) can be respectively obtained as:

$$p_z(z)=\frac{\beta}{2S^{\beta/2}}\frac{z^{\beta/2-1}}{\left(1+\left(\frac{z}{S}\right)^{\beta/2}\right)^2} \quad (5)$$

and

$$F_z(z)=\frac{1}{1+\left(\frac{S}{z}\right)^{\beta/2}} \quad (6)$$

In diversity system where the anntennas are unsufficiently appart, the envelopes of desired signals and interference are correlated. In this paper dual diversity system, commonly used in practise, is considered. Also, similarly to some already published papers [3], [9], the balanced SIRs at two input branches are assumed.

The joint probability density function of instantaneous input SIRs $z_1=R_1^2/r_1^2$ and $z_2=R_2^2/r_2^2$, for unbalanced case, $\Omega_{d_1}/\Omega_{i_1}=\Omega_{d_2}/\Omega_{i_2}=S^{\beta/2}$ with $\Omega_{d_1}=\mathrm{E}[R_1^\beta]$, $\Omega_{i_1}=\mathrm{E}[r_1^\beta]$, $\Omega_{d_2}=\mathrm{E}[R_2^\beta]$, $\Omega_{i_2}=\mathrm{E}[r_2^\beta]$ can be finally expressed as [11]

$$p_{z_1z_2}(z_1,z_2)=\sum_{i_1,i_2=0}^{\infty}\frac{\rho^{i_1+i_2}(1-\rho)^2\beta^2}{4(i_1!i_2!)^2S^{\beta(1+i_1)}}\Gamma^2(2+i_1+i_2)z_1^{\beta(1+i_1)/2-1} z_2^{\beta(1+i_1)/2-1}\left(\left(1+\left(\frac{z_1}{S}\right)^{\beta/2}\right)\left(1+\left(\frac{z_2}{S}\right)^{\beta/2}\right)\right)^{-(2+i_1+i_2)}, \quad (7)$$

where $\Gamma(.)$ is the Gamma function [15, eq. (8.310/1)] and ρ correlation coefficient.

III. SWITCH AND STAY COMBINING

In this paper dual SSC diversity system with correlated envelopes of desired and interfering signals are observed with the algorithm of determining based on SIR. We assume that the first branch is selected. If the instantaneous SIR falls below a given threshold, the switch selects the second branch. So, the system processes the selected branch until the instantaneous SIR falls below a given threshold when switch selects the second one. Cumulative distribution function (cdf) of output instantaneous SIR can be obtained as:

$$F_{ssc}(z)=\Pr(z_T\le z_1\le z)+\Pr(z_2<z_T\wedge z_1\le z) \quad (8)$$

Differentiating previous equation, the output SIR's probability density function (pdf) is expressed as [16]:

$$p_{ssc}(z)=\begin{cases} r_{ssc}(z), & z\le z_T \\ r_{ssc}(z)+p_z(z), & z>z_T \end{cases} \quad (9)$$

where z_T is previously determinated threshold, $p_z(z)$ defined by equation (5) and $r_{ssc}(z)$

$$\begin{aligned} r_{ssc}(z)&=\int_0^{t_T}p_{z_1z_2}(z,z_2)\mathrm{d}z_2= \\ &\sum_{i_1,i_2=0}^{\infty}\frac{\rho^{i_1+i_2}(1-\rho)^2\beta}{2(i_1!i_2!)^2S^{\beta(1+i_1)}}\Gamma^2(2+i_1+i_2)z^{\beta(1+i_1)/2-1} \\ &\times z_T^{\beta(1+i_1)/2}\left(1+\left(\frac{z}{S}\right)^{\beta/2}\right)^{-(2+i_1+i_2)} \\ &\times{}_2F_1\left(1+i_1,2+i_1+i_2,2+i_1,-\left(\frac{z_T}{S}\right)^{\beta/2}\right), \end{aligned} \quad (10)$$

with ${}_2F_1(a,\ b;\ c;\ x)$ being the Gaussian hypergeometric function [15, eq. (9.100)].

The cumulative distribution function, after some manipulations of equation (8) can be written as [3]:

$$F_{ssc}(z)=\begin{cases} F_{z_1z_2}(z,z_T), & z\le z_T \\ F_z(z)-F_z(z_T)+F_{z_1z_2}(z,z_T), & z>z_T. \end{cases} \quad (11)$$

with $F_z(z)$ defined by equation (6) while cumulative distribution function is $F_{z_1z_2}(z,z_T)$

$$\begin{aligned} F_{z_1z_2}(z,z_T)&=\sum_{i_1,i_2=0}^{\infty}\frac{\rho^{i_1+i_2}(1-\rho)^2\Gamma^2(2+i_1+i_2)}{(i_1!i_2!)^2S^{\beta(1+i_1)}(1+i_1)^2}z^{\beta(1+i_1)/2}z_T^{\beta(1+i_1)/2} \\ &\times{}_2F_1\left(1+i_1,2+i_1+i_2,2+i_1,-\left(\frac{z}{S}\right)^{\beta/2}\right) \\ &\times{}_2F_1\left(1+i_1,2+i_1+i_2,2+i_1,-\left(\frac{z_T}{S}\right)^{\beta/2}\right). \end{aligned} \quad (12)$$

IV. AVERAGE OUTPUT SIR

Average output SIR is useful parameter for system performances evaluation in enviromentes of wireless

communication in the presence of cochannel interference. Average output SIR, $\overline{S}_{out}$, at the SC output may be evaluated as [10], [3]

$$\overline{S}_{out} = \int_0^\infty z p_{ssc}(z)\mathrm{d}z = \int_0^\infty z r_{ssc}(z)\mathrm{d}z + \int_{z_T}^\infty z p_z(z)\mathrm{d}z = I_1 + I_2 . \quad (13)$$

Substituting (5) and (10) in (13), the infinite-range integrals can be obtained in the form of infinite series:

$$I_1 = \sum_{i_1,i_2=0}^{\infty} \frac{\rho^{i_1+i_2}(1-\rho)^2\,\Gamma(2+i_1+i_2)\Gamma(1+i_1+2/\beta)\Gamma(1+i_2-2/\beta)}{(i_1!i_2!)^2\,S^{\beta(1+i_1)/2-1}(1+i_1)} \times z_T^{\beta(1+i_1)/2}\,{}_2F_1\left(1+i_1,2+i_1+i_2,2+i_1,-\left(\frac{z_T}{S}\right)^{\beta/2}\right), \quad (14)$$

$$I_2 = \frac{\beta}{\beta-2} S^{\beta/2} z_T^{-\beta/2+1}\,{}_2F_1\left(2,1-\frac{2}{\beta},2-\frac{2}{\beta},-\left(\frac{S}{z_T}\right)^{\beta/2}\right). \quad (15)$$

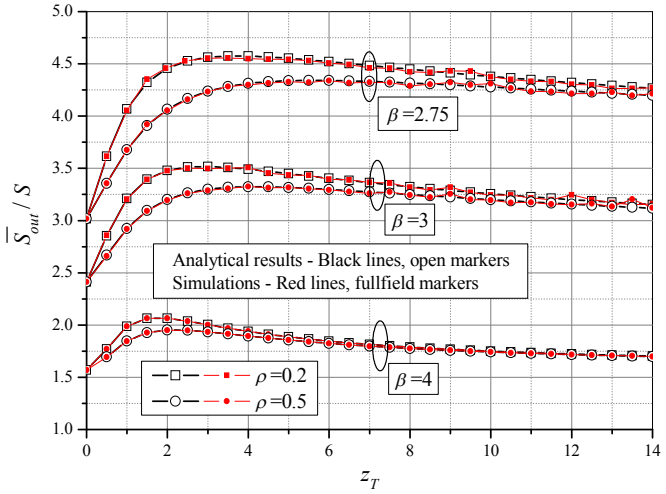


Fig. 2. Normalized output SIR versus threshold z_T

Fig.1 shows the first branch normalized output SIR $\overline{S}_{out}/S$, versus threshold for several values of Weibull paremeter (β) and correlation coefficient (ρ). It is obvious here that diversity gain decreases with an increase of the correlation coefficient. Also, when fading parameter β increases (fading severity decreases), small increase or decrease in correlation coefficient does not have significant effect on $\overline{S}_{out}/S$. Simulations are performed using the C^{++} programming language. The values of $\overline{S}_{out}/S$ are calculated based on over 10^8 correlated Weibull fading samples. From Fig. 1 is evident good match between the theory and simulations.

Differentiating equation (13), the optimal value for threshold when output SIR has maximum, can be obtained. In this case, the optimal threshold can be evaluated numerically using *root-finding* method which is allowed in many software packages.

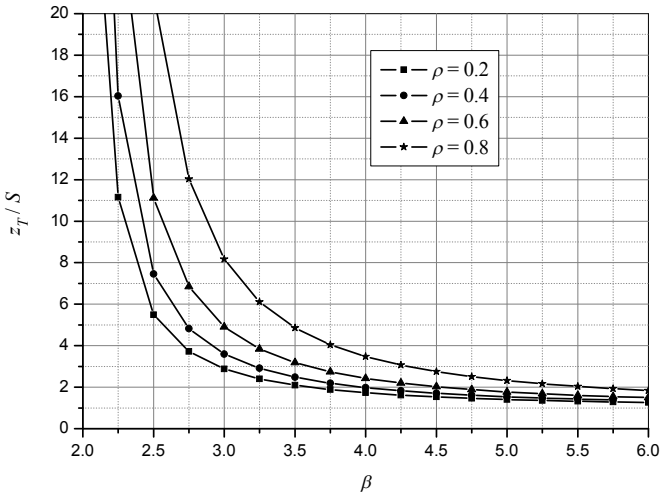


Fig. 2. Normalized optimal threshold versus Weibull feding parameter β

The effects of Weibull fading parameter and correlation coefficient on optimal evaluated threshold are described in Fig. 2. Actually, in Fig. 2 the normalized optimal threshold is plotted versus fading parameter β for several values of correlation coefficient ρ. The optimal threshold decreases when fading parameter increases, as it is expected. At the other side, when correlation coefficient increases, optimal threshold also increases.

V. OUTAGE PROBABILITY

Outage probability is a measure of quality in transmission over fading channels in wireless communication systems. This performance measure is commonly used in wireless communication systems especially in environments where co-channel interference is present.

Outage probability is defined as the probability which the instantaneous SIR at the output of SSC combiner falls below a certain given threshold λ (protection ratio) and using equation (11) can be expressed as:

$$P_{out} = \int_0^\lambda p_{ssc}(z)\mathrm{d}z = F_{ssc}(\lambda) . \quad (16)$$

In Fig. 3, the outage probability is plotted versus normalized threshold z_T/S for several values of fading parameter β. The normalized optimal threshold (z_T=λ), when outage probability has minimum, can be obtained from Fig. 3. In this case, for z_T=λ, SSC combining method becomes SC combining method. A very good match between analytical results and simulations is evident.

The outage probability versus normalized parameter S/λ for different Weibull fading parameters β and for optimal switching threshold value minimizing the outage probability is shown in Figure 4. When Weibull fading parameter increases (fading severity decreases), the outage probability also decreases. As it is expected, an increase in correlation coefficient does have effect on P_{out} (outage probability also increases).

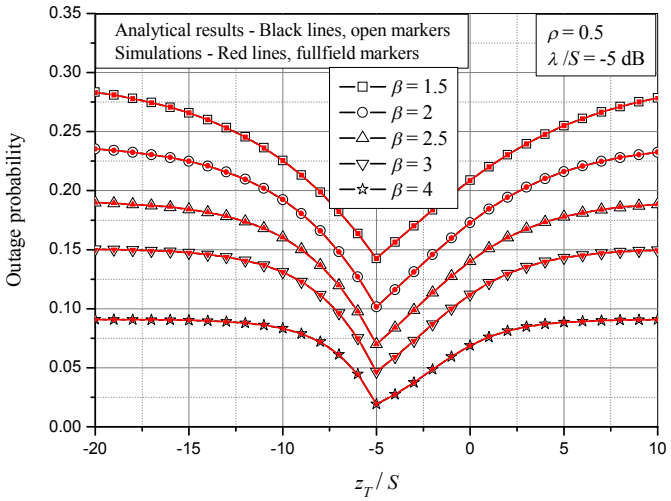


Fig. 3. Outage probability versus normalized threshold z_T/S

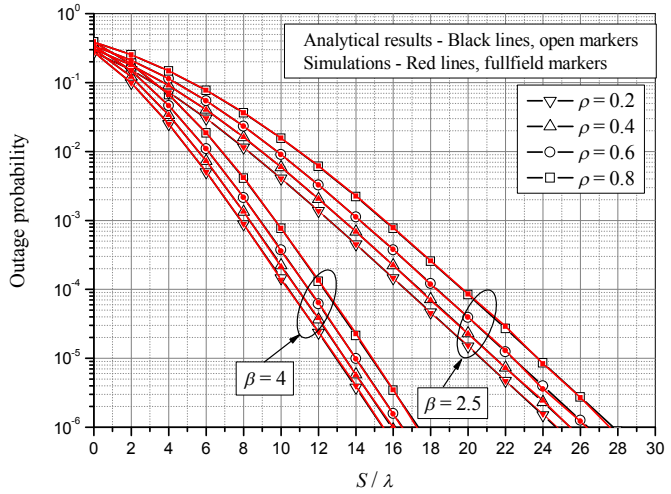


Fig. 4 Outage probability versus normalized input SIR

VI. Conclusion

In this paper analytical results and required simulations of output SIR statistics of SSC diversity system in the presence of correlated Weibull fading were presented. Useful formulae for the probability density function (pdf), cumulative distribution function (cdf) of SSC output SIR and average output SIR in the form of infinite series were derived. Using these new formulae and by observing dual SSC diversity system, the outage probability and other performance measures were graphically presented. The effects of the fading severity, optimal threshold and level of correlation to the performance measures were observed. Also, a very good match between analytical results and simulations in this paper was confirmed.

Acknowledgement

This paper was supported in part by the Ministry of Science of Serbia within the project "Development and realization of new generation software, hardware and services based on software radio for specific purpose applications" (TR-11030).

References

[1] G. L. Stuber, *Principles of Mobile Communication*, Boston, MA: Kluwer, 2001.

[2] W. C. Jakes, *Microwave Mobile Communications*, 2nd edition, Piscataway, NJ: IEEE Press, 1994.

[3] M. K. Simon and M. S. Alouini, *Digital Communication Over Fading Channels*. New York: Wiley, 2000.

[4] H. Yang and M. S. Alouini, "Outage probability of dual-branch diversity systems in presence of co-channel interference," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 310–319, March 2003.

[5] A. A. Abu-Dayya and N. C. Beaulieu, "Outage probabilities of diversity cellular systems with cochannel interference in Nakagami fading," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 4, pp. 343–355, November 1992.

[6] Y. D. Yao and A. U. H. Sheikh, "Outage probability analysis for microcellmobile radio systems with cochannel interferers in Rician/Rayleigh fading environment," *IEEE Electronic Letters*, vol. 26, no. 13, pp. 864–866, Jun 1990.

[7] Q. T. Zhang, "Outage probability of cellular mobile radio in the presence of multiple Nakagami interferers with arbitrary fading parameters," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, pp. 661–667, August 1995.

[8] A. A. Abu-Dayya and N. C. Beaulieu, "Outage probabilities of cellular mobile radio systems with multiple Nakagami interferers," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 757–768, November 1991.

[9] K. Sivanesan and N. C. Beaulieu, "Precise outage analysis of selection diversity and switched diversity in bandlimited microcellular systems with cochannel interference," *IEEE Transactions on Communications*, vol. 55, no. 10, pp. 2002–2010, October 2007.

[10] G. K. Karagiannidis, "Performance Analysis of SIR-based Dual Selection Diversity Over Correlated Nakagami-m Fading Channels," *IEEE Transactions on Vehicular Technology.*, vol. 52, pp. 1207-1216, September 2003.

[11] M. Č. Stefanović, D. M. Milović, A. M. Mitić, M. M. Jakovljević, "Performance analysis of system with selection combining over correlated Weibull fading channels in the presence of cochannel interference," *AEÜ - International Journal of Electronics and Communications,* vol. 62, no. 9, 1 October 2008, pp. 695-700

[12] F. Babich and G. Lombardi, "Statistical Analysis and Characterization of the indoor Propagation Channel," *IEEE Transactions on Communications*, vol. 48, pp. 455–464, March 2000.

[13] G. Tzeremes and C. G. Christodoulou, "Use of Weibull Distribution for Describing outdoor Multipath Fading," *Antennas and Propagation Society International Symposium*, vol. 1, pp. 232-235, June, 2002.

[14] N. C. Sagias, G. K. Karagiannidis, "Gaussian Class Multivariate Weibull Distributions: Theory and Applications in Fading Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3608-3619, October 2005.

[15] I. S. Gradsteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th edition, New York: Academic, 2000.

[16] A. A. Abu-Dayya and N. C. Beaulieu, "Analysis of switched diversity systems on generalized-fading channels," *IEEE Transactions on Communications*, vol. 42, no. 11, pp. 2959–2966, November 1994.

Antenna Diversity Multi User Detection Algorithm for Synchronous CDMA System

Ilia Georgiev Iliev[1] Marin Nedelchev[2]

Abstract – **A spatial diversity reception assisted multiuser code-division multiple-access detector based on consecutively search algorithm is proposed. Algorithm performances and the precision are theoretical studied. Functional dependence between number of the iterations and error rate performance are derived. Computer simulations show high computational speed and acceptable BER precision of the algorithm. For 10 users and spreading factor 31 a computational complexity is reduced 85 times in comparison with the optimum multiuser detector using full search.**

Keywords – **MUD, CDMA, Antenna Diversity, DS.**

I. Introduction

Main problem of the CDMA systems is the Multiple Access Interference – MAI, caused by the distortion of the orthogonality of the signals by every one user. The presence of MAI reduces the signal-to-interference ratio and therefore decreases the noise performance of the mobile communication system. As it is well known, from the loss of information point of view, the correlation receiver is optimal when MAI is missing. There exist multiple of methods for reduction of MAI, but one of most effective is the parallel multiuser detection [1]. The optimal receiver for parallel multiuser detection (MUD) is based on the maximum likelihood criterion (ML) [1].

Finding of the optimal decision is connected with checking of all possible combinations of the transmitted symbols. Consequently the huge number of calculations is a drawback of the ML MUD. They increase exponentially with the active users. This seriously makes difficult its application in the conventional mobile communication systems, despite of the high computational power of the current DSPs.

There are multiple suggestions, in the literature, of methods and algorithms for suboptimal receiving that decrease the computational complexity for detection. In most of the cases they are a trade-off between the computational complexity and the quality factors of the receiver [2,3]. There are suggestions for MUD involving parametric optimization. The maximum a posteriori probability (MAP) [1] is used as a cost function for optimization. In the case of MUD the cost function is discrete, intermittent, non-differential, non-unimodal. Therefore it is appropriate to use the methods for random search, genetic algorithm, evolutionary strategy etc [4,5,6]. Their application reduces the number of calculations in comparison to the optimal receiver with ML for MUD.

The paper investigates the algorithm possibilities for MUD that uses discrete consecutive search. The starting point for optimization is found by correlative reception. The algorithm accuracy and its computational speed are bettered, if antenna diversity with maximum ration combining is used.

[1,2] Ilia Georgiev Iliev and Marin Veselinov Nedelchev –are with Dept. of Radiocommunication and Videotechnologies in Faculty of Telecommunication in TU –Sofia, N8, Kliment Ohridski bul., 1700 Sofia, Bulgaria. E-mail: igiliev@tu-sofia.bg

II. Optimal MUD in synchronous CDMA system

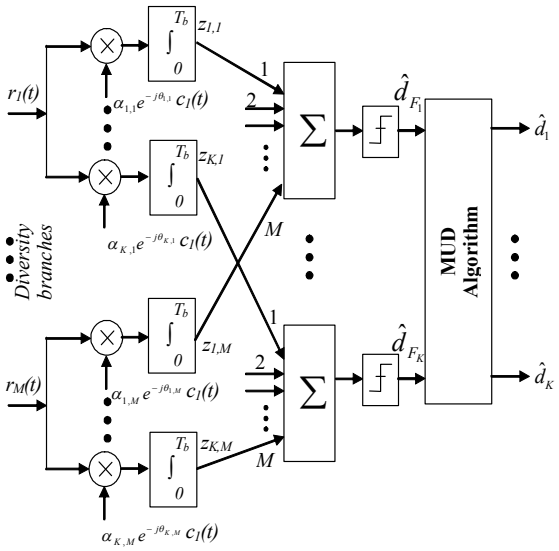


Fig.1

The block schematic of the system is shown on Fig.1. The processing is in baseband and consists of M number of receivers for special diversity. It is supposed that the diversity makes the signals from different receivers are statistically independent. K users transmit synchronously signals with direct spread spectrum (DSS) and 2PSK modulation. Every branch is a correlation receiver with output the received vector $\boldsymbol{z}$, with elements:

$$\boldsymbol{z_m} = [z_1, z_2, ...z_K]^T = \boldsymbol{R A_m E d} + \boldsymbol{n_m} \qquad (1),$$

where m is the receiver m=1..M, $\boldsymbol{d} = [d_1, d_2, ..., d_K]^T$ is a vector column that consists the value of the transmitted symbol with duration T_b coming from the k user. The symbols are presented in binary code $d_k \in \{-1, +1\}$. $\boldsymbol{R}$ is the e cross correlation KxK matrix, which coefficients are the values of the normalized cross correlation functions of the code sequences:

$$R_{ij} = \frac{1}{N}\int_0^{T_b} c_i(t)c_j(t)dt \qquad (2).$$

$n_m(t)$ is the realization of complex additive white Gaussian noise (AWGN) in the input of the m-th receiver with independent real and imaginary component and each of them has $\sigma^2 = No/2$ [W/Hz]. $\boldsymbol{n_m}$ is a AWGN after the m-th corellator as a vector column $\boldsymbol{n_m} = [n_{1,m}, n_{2,m}, ..., n_{K,m}]^T$ with covariance matrix equal to: $\boldsymbol{R}_n = 0.5N_o\boldsymbol{R}$. k-th element is:

$$n_{k,m} = \int_o^{T_b} n_m(t)c_k(t)dt \,.$$

$\boldsymbol{A_m} = diag[\alpha_{1,m}e^{j\theta_{1,m}}, \alpha_{2,m}e^{j\theta_{2,m}}, ..., \alpha_{K,m}e^{j\theta_{K,m}}]$ is a diagonal matrix and the elements in the main diagonal are the complex transmission coefficients of the channel for each user in m-th channel. The channel is supposed to be with slow Raleigh fading. The amplitudes have Raleigh distribution; the phase difference distribution in the channel is uniform in the interval [0,2π]. It is assumed that the channels for each user are statistically independent. $\boldsymbol{E} = diag\left[\sqrt{E_1}, \sqrt{E_2}, ..., \sqrt{E_K}\right]$ is a diagonal matrix. $\sqrt{E_k}$ is the symbol energy for the k-th user, $\boldsymbol{c}$ – is a matrix with rows the spreading sequence for each user and $c_k^{(n)} \in \{-1, +1\}$. The length of the sequences is equal to N - $N = T_b / T_c$, T_c is the element duration.

The optimal MUD is done, when the MAP criterion is applied [1]. The maximum of the correlation of the received signal with every one of the transmitted signals id searched for. The logarithmic function of likelihood for the m-th receiver is expressed in matrix form as:

$$\Psi_m(\boldsymbol{d}) = 2\Re(\boldsymbol{d}^T\boldsymbol{E}\boldsymbol{A_m}^*\boldsymbol{z}) - \boldsymbol{d}^T\boldsymbol{E}\boldsymbol{A_m}\boldsymbol{R}\boldsymbol{A_m}^*\boldsymbol{E}\boldsymbol{d} \quad (3).$$

The symbol $()^*$ denotes complex conjugate value, and $()^T$ – transposed matrix. The maximum ratio combining diversity (MCR) requires the decision for the transmitted symbols to be taken on the base of the quantity statistics:

$$\Psi(\boldsymbol{d}) = \sum_{m=1}^{M}\Psi_m(\boldsymbol{d}) \qquad (4).$$

The decision for the transmitted symbols is:

$$\hat{\boldsymbol{d}} = \arg\left\{\max_{\boldsymbol{d}}\left[\Psi(\boldsymbol{d})\right]\right\} \qquad (5).$$

III. POSSIBILITIES OF THE ALGORITHM FOR SUBOPTIMAL RECEIVING WITH CONSECUTIVE SEARCH

A. Algorithm for consecutive search..

Finding of optimal decision in MUD is considered as an optimization problem using multiparameter discrete, non-unimodal cost function-(4). Because of the channels are statistically independent $\Psi_i(\boldsymbol{d}) \neq \Psi_j(\boldsymbol{d})$ за $i \neq j$, $\Psi(\boldsymbol{d})$ is the multi criteria cost function. This paper proposes the use of algorithm for consecutive search, detailed and analyzed in [9]. The optimization parameters are elements of a vector with the symbols coming from the users $\boldsymbol{d}$. The output data from the correlation receivers, when hard decision is made, is:

$$\hat{\boldsymbol{d}}_{Fm} = [\hat{d}_{F_1,m}, \hat{d}_{F_2,m}, ..., \hat{d}_{F_K,m}] = sign\left\{\Re(\boldsymbol{A_m}^*\boldsymbol{z})\right\} \quad (6).$$

The start point for optimization is determined after MCR:

$$\hat{\boldsymbol{d}}_F = sign\left\{\sum_{m=1}^{M}\Re(\boldsymbol{A_m}^*\boldsymbol{z})\right\} \qquad (7).$$

If in the package $\boldsymbol{d}$ has one error bit, then it may be corrected according to the decision criterion (5). It is necessary to do only (K+1) computations of the cost function (4) with vectors from the set M_d that have Hemming distance with the vector $\hat{\boldsymbol{d}}_F$ unit:

$$M_d = \left\{d : H_d\left(\hat{\boldsymbol{d}}_F, \boldsymbol{d}\right) = 1\right\} \qquad (8).$$

One iteration of the algorithm is related to a search in the vicinity of a point in the K –dimensional space with Hemming distance unit. This is realized as for every one step of the optimization process consecutively changes one bit of the vector $\boldsymbol{d}$ (it changes the value of the bit) and for every change computes the cost function. Then the algorithm chooses the maximum value of (6). The next vector for i-th iteration is:

$$\boldsymbol{d}^{(i+1)} = \arg\left\{\max_{\boldsymbol{d}\in\boldsymbol{M_d}}\left[\Psi(\boldsymbol{d}^{(i)})\right]\right\} \qquad (9).$$

Searching for optimal decision is confined after satisfying at least of one of both criterions. The first confines the number of the iterations L and the second is:

$$\max_{\boldsymbol{d}\in\boldsymbol{M_d}}\left[\Psi(\boldsymbol{d}^{(i+1)})\right] \leq \max_{\boldsymbol{d}\in\boldsymbol{M_d}}\left[\Psi(\boldsymbol{d}^{(i)})\right] \qquad (10).$$

The search for optimal decision is stopped unconditionally by the strong criterion (10). This reduces the number of computations of the cost function (4), but at the expense the accuracy of the algorithm.

Б. Theoretical Algorithm Performance

The symbols in the output of the receiver are parallel and the received vector $\hat{\boldsymbol{d}}$ is a package consists of K number symbols. The user channels are independent. The error probability for the k-th bit when diversity reception and maximal combining ratio for Raleigh fading channel is:

$$Pe_{k_{MAI}} = \left(\frac{1-\mu}{2}\right)^M \sum_{m=0}^{M-1}\binom{M-1+m}{m}\left(\frac{1+\mu}{2}\right)^m \qquad (11),$$

$$\mu = \sqrt{\frac{SINR}{1+SINR}} \text{ и } SINR = E_{b_1} / \left(N_o + 2\sigma_R{}^2 \sum_{k=2}^{K} E_{b_k} \right).$$

This formula is for single user detection and M branches for diversity. SNIR is the signal-to-noise plus interference ration. It is assumed that the MAI has Gaussian distribution with dispersion $\sigma_R{}^2$. This is the dispersion of the values of the cross correlation functions. If it is used an algorithm for power equalization and management the SNIR is transformed into:

$$SINR = 1/\left[N_o + (K-1)/N\right].$$

Let the spreading sequences are random bipolar with uniform value distribution. The number of error symbols in a packet q has binomial distribution and the probability is determined by [9]:

$$P_K(q) = C_K^q (Pe_{k\,MAI})^q (1 - Pe_{k\,MAI})^{K-q} \qquad (12).$$

Let assume that the cost function (4) decreases with increasing the Hemming distance to the optimal solution in (5). If errors in the package are more than one, this procedure can be repeated consecutively as many times as there are errors in the packet. Each iteration in the consecutive search reduces the number of errors in the package with one. This reduction of the error, the lower limit of the error probability for one bit, for one user could be determined by the formula:

$$P_e = \sum_{q=L+1}^{K} \frac{q-L}{K} C_K^q (Pe_{k_{MAI}})^q (1 - Pe_{k\,MAI})^{K-q} \qquad (13),$$

where L is the number of the iterations for consecutive search.

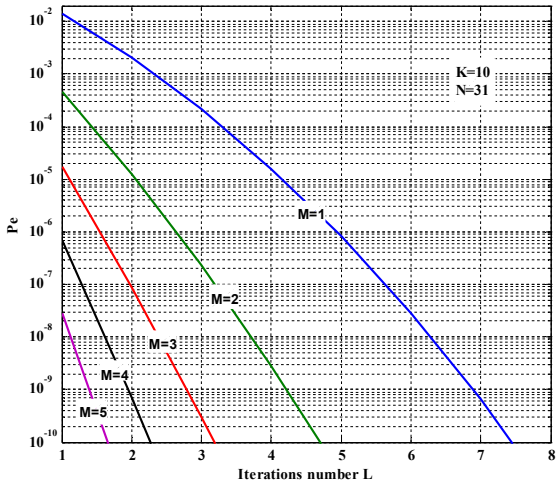


Fig.2

Fig.2 shows the dependence of the error probability against the number of iterations for different number of branches for diversity receiving M. The number of active user is K=10. Fig.3 shows the error probability for one bit in function of the active users when a parameter is M and a fixed number of iterations L=4. If M=5, it is seen that the algorithm accuracy is satisfactory from the practical point of view- Pe<10^{-6}. The results from both figures are derived when AWGN impact is neglected. It is seen from Fig.2 that the increase of M, the necessary iteration in the algorithm for achieving specified error probability reduces. The dependences from (13), Fig.2 and Fig.3 give possibilities for a specified error probability and specified number of active users to define the minimum number of iterations of the algorithm L. Then, adaptive changes the number of calculations depends on the load of the mobile networks. It is seen that the algorithm of the consecutive search reduces significantly the number of computations of the cost function- because of the strict criterions for detection stopping.

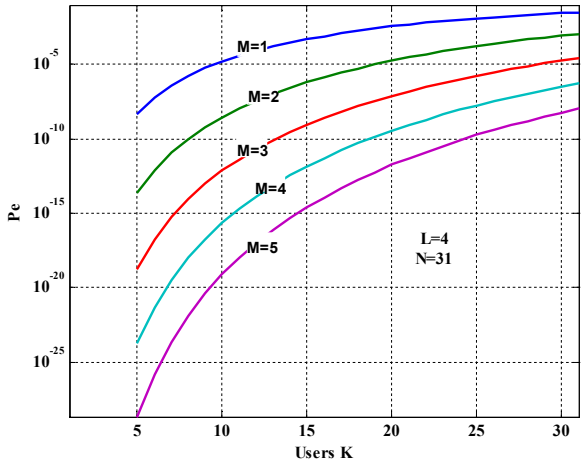


Fig.3

IV. SIMULATION RESULTS

The algorithm is simulated in MATLAB. The processing gain of the signature sequences is N=31, and the sequences are randomly generated. Perfect power control and CIR estimation is assumed. The channel is modelled as AWGN with slow Raleigh fading and $E[\alpha^2{}_{k,m}] = 1 \; \forall m,k$.

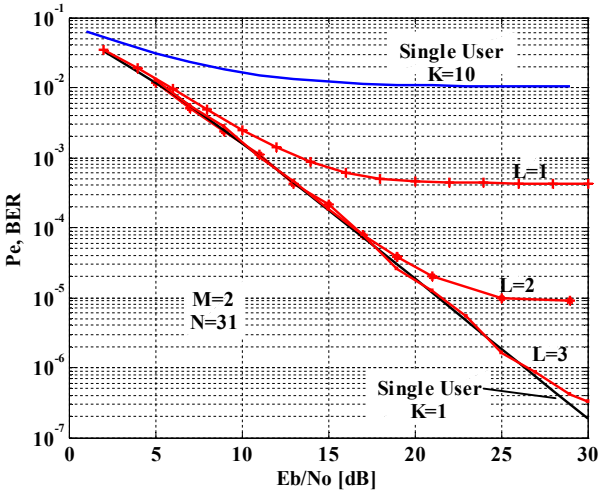


Fig.4

Fig.4 shows the measures BER against the average value of Eb/No for K=10 and different number of L=1,2,3 - M=2. The results show that the increase of the number of iterations it is approaches the theoretical curve for full compensation of MAI or single receiving for K=1. BER is limited from below and the value coincide with the one calculated with formula (13). The results show that the diversity receiving increases the speed of the algorithm, because the starting point is close to the minimum. The accuracy is bettered because the probability of "sticking" of the algorithm in a local minimum is low.

A. Computational complexity of the algorithm

In [8] is made a comparison of the computational complexity of some conventional algorithms for MUD – common correlation receiver, decorrelation receiver for MUD, optimal receiver for MUD [1] and MUD with genetic algorithm [8]. The computational complexity is determined on the base of the number of multiplications and sums for calculation of the likelihood logarithmic function (3) for detection of K bits. For the decorrelation receiver the computational complexity is proportional to $Q=K^3$. For the optimal algorithm the number of computations Q with the function (4) are $Q=2^K$. The computational complexity of the proposed algorithm is defined as the average number of computations of the cost function $\Psi(\boldsymbol{d})$.

For K=10 the Table 1 compare the computational number of fitness function (4) for proposed algorithm against specified conventional multiuser detectors.

TABLE 1

Decorrelator MUD $Q\sim K^3$	Optimum MUD $Q=2^K$	GA MUD [8] Q=PY	Proposed MUD (14) Q
1000	1024	300	12.3

The commputional complexity of proposed algorithm is measured by simulation when the number of iterations are L=3. The results show that the proposed algorithm has tens or hundreds of times less computational complexity compared to the known detectors, when meeting the practical needs of decision accuracy.

V. CONCLUSION

The results show that the diversity reception helps in certain extent to improve the error performance in MUD. Thanks to its increased accuracy of detection due to reduction of the probability for "sticking" of the algorithm in a local minimum. Because of the independence of the channels, it is reduced the number of the necessary iterations of the algorithm, which is important in its practical application. As a result of the derived formulas in the paper, it is possible to determine the number of necessary iterations for a specified BER and number of active users. This allows flexible management and economical use of the computational resources. The simulation results show that the number of the computations for the proposed algorithm is ten times less than the known optimal and suboptimal detectors.

VI. REFERENCES

[1] S. Verdú, Multiuser Detection. New York: Cambridge Univ. Press,1998.

[2] R. Lupas and S. Verd´u, "Linear multiuser detectors for synchronous code-division multiple-access channels," IEEE Trans. Inform. Theory, vol. 35, pp. 123-136, Jan. 1989.

[3] Z. Xie, R. Short, and C. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," IEEE Journal on Selected Areas in Communications, vol. 8, pp. 683-690, May 1990

[4] C. Ergun and K. Hacioglu, "Multiuser detection using a genetic algorithm in CDMA communications systems," IEEE Trans. Commun., vol. 48, pp. 1374–1383, Aug. 2000.

[5] M. J. Juntti, T. Schl¨osser, and J. O. Lilleberg, "Genetic algorithms for multiuser detection in synchronous CDMA," in IEEE International Symposium on Information Theory – ISIT'97, (Ulm, Germany), p. 492, 1997.

[6] K. Yen and L. Hanzo, "Hybrid genetic algorithm based multi-user detection schemes for synchronous CDMA systems," in Proceedings of the IEEE Vehicular Technology Conference (VTC), (Tokyo, Japan), May 15-18,2000.

[7] [7] Peng Hui Tan, Lars K. Rasmussen, Multiuser Detection in CDMA—A Comparison of Relaxations, Exact, and Heuristic Search Methods, IEEE TR. WIRELESS COMM., VOL. 3, N 5, SEPT. 2004

[8] K. Yen and L. Hanzo, Antenna-Diversity-Assisted Genetic-Algorithm-Based Multiuser Detection Schemes for Synchronous CDMA Systems, IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 51, NO. 3, MARCH 2003

[9] Ilia G.Iliev, Marin V. Nedelchev, Performance Analysis of a suboptimal multiuser detection algorithm, ICEST 2007, Macedonia, pp.531 – 534.

An Analysis of the Possibilities for Using Lattices for the Interpolation of the Exact Signal Vector Position of QAM Constellations

Roumen Kountchev[1] Dimitar Bojchev[2] Dobri Dobrev[3]

Abstract – **In this paper it is made an analysis of using rectangular and hexagonal lattices to define the exact vector position of height order QAM signal super constellation, passed trough to the cable communication channel with nonlinearity.**

Keywords – **Signal vector position estimation, Height order QAM modulation, and Super constellation deformation.**

I. INTRODUCTION

In wide band multi channel hybrid optical and coaxial cable communication systems there are some modules which causes nonlinearity [1], [2]. In the optical part the most significant is external MZ modulator [3], [4]. In the coaxial one it is non the ideal transfer function of the cable amplifiers [5]. Both of them are the reason for nonlinear distortions – generation intermodulation contributions, which are fallow in the spectrum of transmitted signal. In another way, the nonlinearity causes deformation of QAM super constellation. In this paper we will consider the static effect of transfer nonlinearity. A relatively small deformation of a high order QAM constellation can reduce the Euclidian, noise protection distance between some signal vector position and this is the reason for degradation of the symbol error ratio (SER), for the corresponding signal points – fig.1. Bearing in mind the above the signal becomes more vulnerable to the channel noise process, which can be modelate by normal distribution [3], [4], [5], [6]. For this reason estimating exact error vector (EV) magnitude, e.g. its modulation components must be made by a representative ensemble of realizations of the corresponding vector position to eliminate noise influence. Literal estimation EV for each symbol will be taking relatively a lot of time and system resource.

A special interest is to obtain some of the exact vector positions by the interpolation of another statistical estimated. This will be reducing calculation time and system resource.

[1] Roumen Kountchev is a professor with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: rkountch@tu-sofia.bg.

[2] Dimityr Bojchev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: dbojchev@tu-sofia.bg.

[3] Dobri Dobrev is a professor with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: dobrev@tu-sofia.bg.

II. INTERPOLATION BY USING RECTANGULAR LATTICE

The simplest case is two dimensional rectangular lattice. The

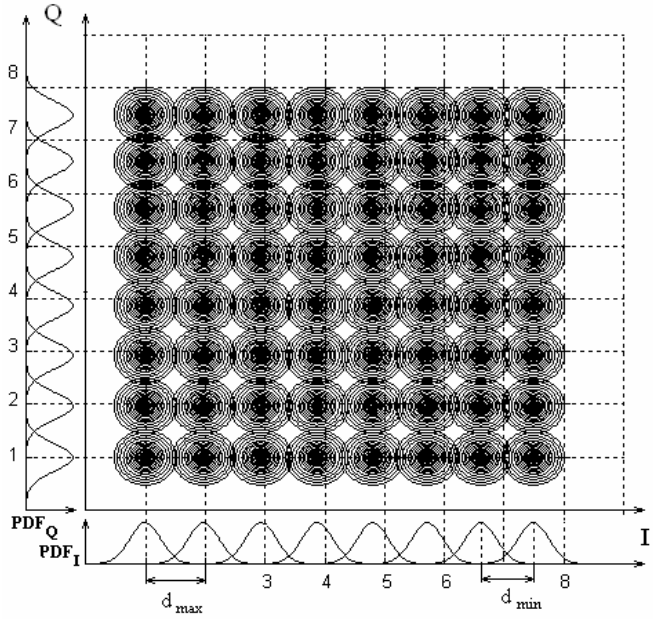


Fig.1. 256 QAM super constellation passed through to the Gaussian channel with nonlinearity – first quadrant

position of the interpolated point is practically is the geometric middle between neighbor points, for a rectangular regular lattece. The polar coordinates can be calculated by:

$$\varphi_k = arctg \frac{I_{k-1} + I_{k+1}}{Q_{k-1} + Q_{k+1}} \tag{1}$$

$$M_k = \sqrt{\frac{(I_{k-1} + I_{k+1})^2}{4} + \frac{(Q_{k-1} + Q_{k+1})^2}{4}} \tag{2}$$

Relatively to the origin it can be written:

$$\overrightarrow{V_X} = \frac{1}{4}\sum_{k=1}^{4} w_k.(I_k\ Q_k) \tag{3}$$

Where w_k is the weigh of corresponding vector I_k and Q_k are the vector inphase and the quadrature modulation components.

In the general case the value of all weighs is one. Without nonlinearity this lattice gives exact results. In the practical

case there are two errors. First is the statistic estimation error. It does not depend of the vector position or of nonlinearity. Second is the error of interpolation. It depends not only of both of them, but also of estimation error of primary neighbor points – (4).

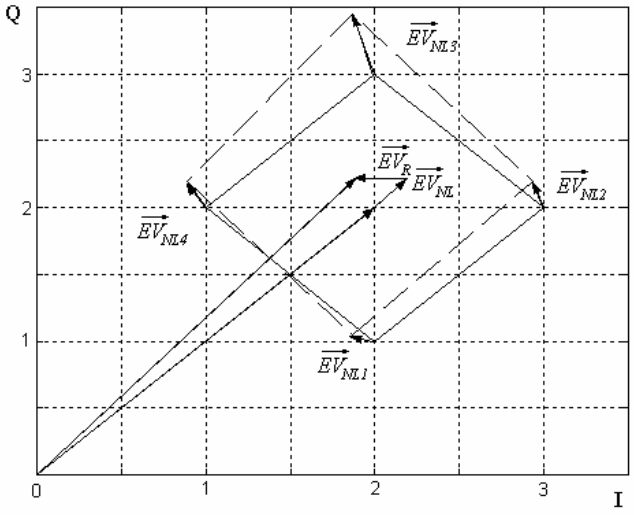


Fig. 2 The Error Vector of interpolation of one signal vector position in nonlinear sheet

$$\Delta(i,j) = \Delta_L(i,j,\Delta_S) + \Delta_S \tag{4}$$

For this reason interpolation error computing is a very important problem. In fig. 2 is shown the interpolation of one signal vector point.

To analyze the interpolation error first we will assume that the statistical error is zero, e.j. the primary signal points are exact. Second, the analisys will be made relatively to the modulation components. Thus the absolute interpolation error of each modulation component will depends only of nonlinearity and vector position.

$$\Delta_I = I(i) - \frac{I(i-1) + I(i+1)}{2} \tag{5}$$

But if the nonlinearity is represented by third order power series expansion [7] and $I=i$ it can be written as follows:

$$\Delta_I = i + a_2.i^2 + a_3.i^3 - \frac{i-1+a_2(i-1)^2 + a_3(i-1)^3}{2} + \frac{i+1+a_2(i+1)^2 + a_3(i+1)^3}{2} \tag{6}$$

After the required mathematical operations, the expression (6) becomes:

$$\Delta_I = -a_2 - 3a_3.i \tag{7}$$

As seen in equation (7) the dependence is linear. It is more convenient to accept the module of (7). The relatively error can be calculated:

$$\delta_I = \frac{a_2 + 3a_3.i}{i + a_2.i^2 + a_3.i^3}.100[\%] \tag{8}$$

In the worst practical case the numerical values of polynomial coefficients, for example can be: $a_2 = -4{,}472.10^{-3}; a_3 = 1{,}186.10^{-6}$. It is shown in fig.1. In this case the maximal relatively error is δ_I =0,44%. The absolute module error vector can be calculated:

$$\Delta_{M_{i,q}} = \sqrt{(a_2 + 3a_3.i)^2 + (a_2 + 3a_3.q)^2} \tag{9}$$

The relatively error will be:

$$\delta_{M_{i,q}} = \sqrt{\frac{(a_2 + 3a_3.i)^2 + (a_2 + 3a_3.q)^2}{\left(i + a_2.i^2 + a_3.i^3\right)^2 + \left(q + a_2.q^2 + a_3.q^3\right)^2}} \tag{10}$$

For the considered case δ_M =0,32%.

In fact δ_M is the value of interpolation error vector (IEV). In dB it can be expressed:

$$\overrightarrow{IEV} = 10.\lg\left[\frac{\left(i + a_2.i^2 + a_3.i^3\right)^2 + \left(q + a_2.q^2 + a_3.q^3\right)^2}{(a_2 + 3a_3.i)^2 + (a_2 + 3a_3.q)^2}\right] \tag{11}$$

For the current example IEV=49,9dB. But this is not a real value of error. To make this analysis we first neglect the error of statistical processing. Thus we assumed that the primary vector positions are exact, but it is not precisely. In this way the real error has to be computed. It can be done by considering the case which primary points that the definition is with maximal error, e.j. it must to be considered the worst case.

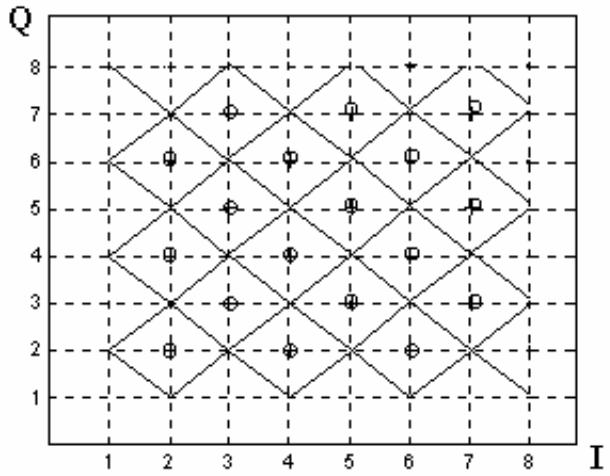


Fig.3 Signal point interpolation with rectangular lattice – first quadrant of 256 QAM super constelation

Using formula (6) it can be expressed the absolute inphase (quadrature) component error, given of non-literal estimation of primary points:

$$\Delta_{Is} = \frac{1}{4}\sum_{k=1}^{4}\Delta_{I\,k} \qquad (12)$$

For a current analysis we will use the regular rectangular lattice and the statistical estimated error will be simulated by nonlinear processing over the primary signal points. Supposing that three of the points are exact and one is determined with error – fig.4. Thus the relatively module error

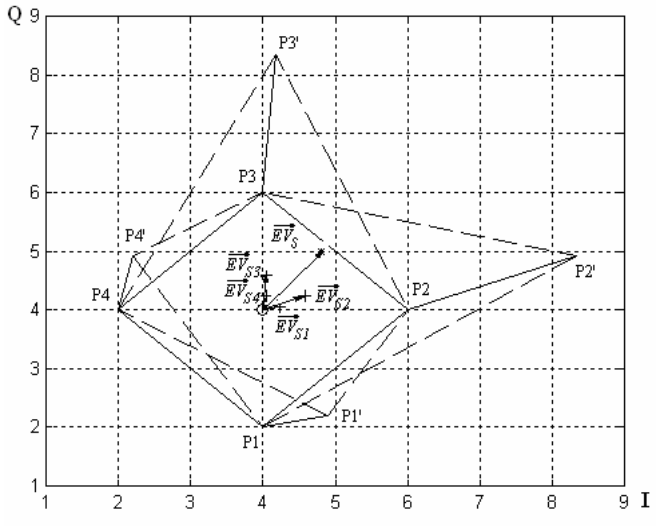


Fig. 4. Reflection an error estimation of primary into interpolated point

will be:

$$\delta_{Ms_{i,q}} = \frac{1}{2}\cdot\sqrt{\frac{\Delta_I^2 + \Delta_Q^2}{\left(i + a_2.i^2 + a_3.i^3\right)^2 + \left(q + a_2.q^2 + a_3.q^3\right)^2}} \qquad (13)$$

Practically with formula (13) it can be calculated the module of relatively error caused by absolute error estimation of one signal point. It can be seen that the lattice decreases the statistical estimation error, reflecting in to the interpolated point. The module of full statistical error vector is a sum of four vector errors. Thus the formula (13) becomes:

$$\delta_{Ms_{i,q}} = \frac{1}{2}\cdot\sqrt{\frac{\left(\sum_{k=1}^{4}\Delta_{I_k}\right)^2 + \left(\sum_{k=1}^{4}\Delta_{Q_k}\right)^2}{\left(i + a_2.i^2 + a_3.i^3\right)^2 + \left(q + a_2.q^2 + a_3.q^3\right)^2}} \qquad (14)$$

Finally the module of full relatively error can be obtained by the sum between (14) and (10):

$$\delta_{M_{i,q}} = \sqrt{\frac{\left(a_2 + 3.a_3.i\right)^2 + \left(a_2 + 3.a_3.q\right)^2 + 0{,}25.\left[\left(\sum_{k=1}^{4}\Delta_{I_k}\right)^2 + \left(\sum_{k=1}^{4}\Delta_{Q_k}\right)^2\right]}{\left(i + a_2.i^2 + a_3.i^3\right)^2 + \left(q + a_2.q^2 + a_3.q^3\right)^2}} \qquad (15)$$

III. INTERPOLATION BY USING HEXAGONAL LATTECES

To apply hexagonal lattice for rectangular QAM, first we must make a shifting of odd or even vector positions of I or Q modulation component of QAM super constellation in half of Euclidian distance (noise protection distance). It must to be noted that it are necessary six neighbor points to calculate interpolated position – fig.5.

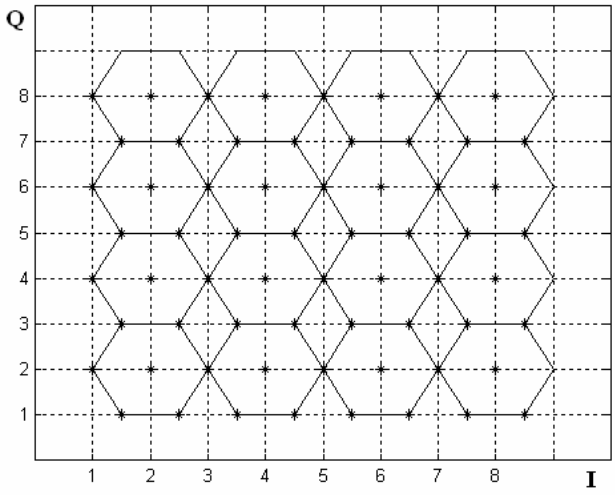


Fig.5. Interpolation signal vector position, using hexagonal lattice.

Using expression (3), for hexagonal lattice can be written:

$$\overrightarrow{V_X} = \frac{1}{6}\sum_{k=1}^{6} w_k.\left(I_k\, Q_k\right) \qquad (16)$$

Without nonlinearity, this lattice is also exact but in nonlinearity space the interpolation error is smallest in compare with rectangular lattice, because interpolated point depends of six neighbor positions instead of four.

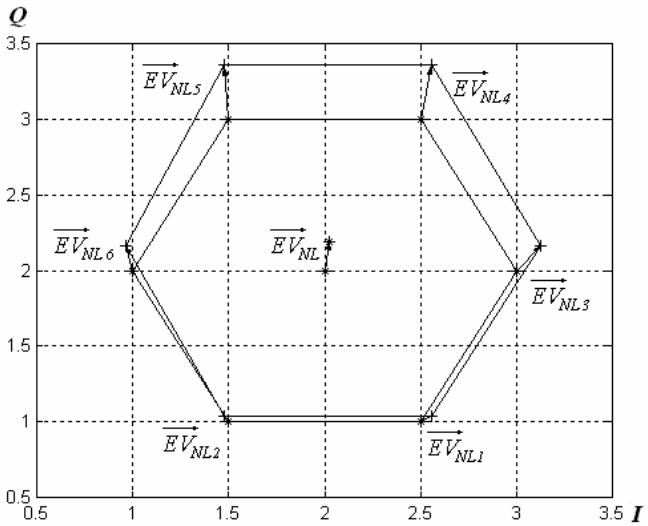


Fig.5. Interpolation signal vector position, using hexagonal lattice in nonlinearity space

For this reason in current analysis it can be neglect. Using formula (14) can be written:

$$\delta_{Ms_{i,q}} = \sqrt{\frac{0{,}17.\left[\left(\sum_{k=1}^{4}\Delta_{I_k}\right)^2 + \left(\sum_{k=1}^{4}\Delta_{Q_k}\right)^2\right]}{\left(i + a_2.i^2 + a_3.i^3\right)^2 + \left(q + a_2.q^2 + a_3.q^3\right)^2}} \quad (17)$$

It can be seen that the hexagonal lattice give more exact result, but it is necessaries six neighbor points. Also it mast be done primary shifting of some of vector positions in a half of Euclidian noise protection distance, which is additional processing. For this reasons the hexagonal lattice is not efficient for considering goal. Another important problem is interpolation a points, located in the end of constellation. The required primary points can be recovered not exactly, e.g. it will be done with significant error. The problem is very complicate and has to be discussed.

IV. MULTI LEVEL INTERPOLATION

Of Fig. 2 and Fig. 5 can be seen that the resultant error vector *EV* of interpolated point is smallest than maximal same of primary points. This fact can be used for two purposes. Firs it can be done a reducing the statistical error, e.g. decreasing. Second, by a lot of iterations the space can be made relatively linear, because the lattice seeks to equalize the point distance. Both of these purposes can be achieved by multi level interpolation. An example is depicted in Fig.6.

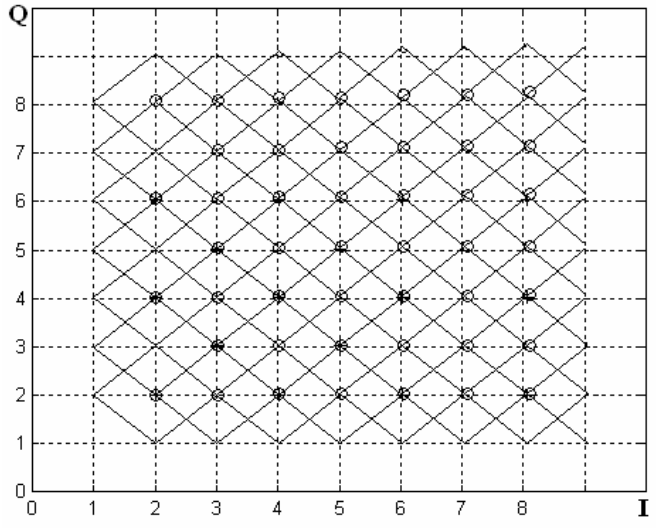


Fig.6. An example of two level interpolation, using rectangular lattice in first and second level

In first level it must be selected a primary and then will be obtained interpolated positions. In second level the interpolated points becomes primary. In this way it can be calculated another positions, which were a primary in first iteration, so called second interpolation level. Thus they can be obtained more exact in comparison with statistical estimation only, because the interpolated vector position depends of four neighbor points. Also it can be use a combination between rectangular and hexagonal lattice. For example first interpolation level can be done with a rectangular lattice. After shifting of any determinate vector positions how it was considered earlier, the second interpolation level can be done by using a hexagonal lattice. The effect of this processing will depends of each concrete case, of iterations number and the choice of the lattices kind.

V. CONCLUSION

In this paper an analysis has been done of the possibilities for using rectangular and hexagonal lattices for the interpolation of the vector position of square QAM super constellation after the transition of a signal trough to the Gaussian channel with nonlinearity. We have worked out relations for estimating the full error after interpolation for rectangular (15) and hexagonal lattice (17). A practical case has been considered where the nonlinearity is represented by a third order power series expansion. It has also been shown that the value of the error depends not only on nonlinearity, but also on the current signal vector position. Finally has given an idea for decreasing a statistical estimation error by using multi level interpolation.

REFERENCES

[1] Xiong F. *"Digital Modulation Techniques*" Artech House Inc. 2000.

[2] Zhang J., A.B. Sharma *"Design of Efficient Optical Fiber Communication Systems for CATV/HDTV distribution"* IEEE Trans. Consumer Electronics, vol. 43, no. 3, pp.453-461, August 1997.

[3] Jordanova L., V. Topchiev *"Nonlinear Distortions of the Signals in Broad Band Cable Communication Systems, Using Conventional Mach-Zehnder Modulator"* pp.35-40 "E+E" 1-2/ 2008.

[4] Jordanova L., V. Topchiev **"***Research of the Characteristics of an Optical Channel of HFC CATV Systems"* pp.149-154, Electronics'2006 September, Sozopol, Bulgaria

[5] Gibson Jerry D. *"Communications Handbook"* IEEE and CRC Press 1997.

[6] Ferdinandov Er., Boriana Pachedjieva, Kalin Dimitrov "*Optical communication systems"* Sofia 2007.

[7] Sloane N.J. *"Tables of Sphere Packing and Spherical Codes"* IEEE Trans. Information Theory, vol. IT-27, no. 3, pp.327-337, 1981.

[8] Wenzel D., Joachim Speidel *"Modulation of Conventional TV Signals in All-Digital CATV Network Head Edns"* IEEE Trans. Consumer Electronics, vol. 46, no. 1, pp.116-126, February 2000.

[9] Nuno M. B., Francisko A. B. "*On the BER Performance of Hierarchical M-QAM Constellations With Diversity and imperfect channel estimation"* IEEE Trans. Communications, vol. 55, no. 10, pp.1852-1856, October 2007.

[10] Han J. Hyung-Myung Kim *"Joint Optimization of VQ Codebooks and QAM Signal Constellations for AWGN Channels"* IEEE Trans. Communications, vol. 49, no. 5, pp.816-825, May 2001.

[11] Lei J. Wen Gao *"Backward-Compatible Solution for next Generation DVB-C System"* IEEE Communications Society ICC 2008 proceedings pp.1962-1966, May 2001.

The second-order statistical measures of SC Macrodiversity System over independent Weibull Fading Channels

Dragana S. Krstic[1], Stefan R. Panic [2], Jelena A. Anastasov [3], Goran Lj. Stamenovic [4] and Dusan M. Stefanovic [5]

Abstract – **The average level crossing rate and average fade duration of SC (Selection Combining) macrodiversity system's output signal in the presence of independent Weibull fading are analyzed. The SC macrodiversity system consists of two microdiversity systems and selection is based on their output signal's average power values. Both microdiversity systems are composed of SC diversity systems with uncorrelated branches in the presence of Weibull fading. Average power of the output signal is modeled by Gamma distribution. Analytical results for LCR (Level crossing rate) and AFD (Average Fading Duration) of the output signal envelope are graphically presented.**

Keywords – **Average level crossing rate, Average fade duration, Selection Combining, Weibull fading.**

I. INTRODUCTION

The effect of fading on the performance of wireless communications systems has received a great deal of research interest due to the possible use of space diversity technique [1], [2]. The rapid and random fluctuations of the signal envelope and phase in a radio channel are caused with two propagation phenomena: multipath scattering and shadowing. The multipath components include a LOS component and many scattered components, while the effects of shadowing can be modeled as LOS shadow fading or multiplicative shadow fading. The amplitude of the LOS is often assumed to be a lognormal random variable or a gamma random variable. On the other hand, there are many researches based on analyzing fast fading. Fast fading can be modeled by several distributions including Weibull distribution [2], [3].

[1]Dragana S. Krstic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, e-mail: dragana.krstic@elfak.ni.ac.rs

[2]Stefan R. Panic is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: stefanpnc@yahoo.com

[3]Jelena A. Anastasov is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: anastasovjelena@gmail.com

[4]Goran Lj. Stamenovic is with Tigar, Pirot, 18300 Pirot, Serbia, e-mail: goran.stamenovic@tigar.com

[5]Dusan M. Stefanovic is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: dusan.stefanovic@itcentar.co.yu

For all diversity techniques: equal gain combining (EGC), maximal ratio combining (MRC), selection combining (SC), and generalized selection combining (GSC), the receiver processes the obtained diversity signals to maximize the system's power efficiency. Among these diversity techniques, SC is the least complicated [4].

All of mentioned techniques can be applied in the presence of fast fading. When there is shadowing (slow fading), the use of macrodiversity system is suggested. Macrodiversity system simultaneously combines signals from the outputs of microdiversity systems and thus reduces the effect of shadowing.

Other than the outage probability which is essential to determine several systems' performance measures, there are some the second-order statistical measures also used in describing system's performance. The average level crossing rate (LCR) and the average fade duration (AFD) are the system second-order statistical measures, which reflect the correlation properties of its input/output signals. The average LCR can be used to estimate the rate of occurrence at which the envelope of the received signal crosses a specified level.

The level crossing rate and average fading duration are used in modeling and designing wireless communication systems. Actually, these second-order statistical measures are related to criterion used to assess error probability of packets of distinct lenght and to determinate parameters of equivalent channel, modeled by a Markov chain with defined number of states [5].

In this paper the second-order statistical measures (AFD and LCR) at the output of macrodivesity system over the independent Weibull fading channels are presented. Two microdiversity systems with SC receivers are obtained. The selection is based on picking maximal signal's average power value at the output of each SC receiver. Average powers of the output signals are modeled by Gamma distribution. Some numerical results of the average level crossing rate and the average fade duration are presented.

I. MODEL OF THE SYSTEM

Let r be the received sampled envelope and $\dot{r}$ its derivative with respect to time, with joined probability density function (pdf) $p_{\dot{r},r}(\dot{r},r)$. The average LCR of the envelope ratio at threshold R represents the average number of times, per time unit, the stationary fading process $R(t)$ crosses threshold R in the positive direction, and is mathematically defined by formula

$$N_r(r) = \int_0^\infty \dot{r} p_{\dot{r},r}(\dot{r}, r) d\dot{r} \quad (1)$$

The AFD is defined as the average time that the envelope ratio remains below the threshold R, and is determinated as ratio of output signal's envelope cumulative distribution function (cdf) and the average LCR:

$$T_r(r) = \frac{F_r(r)}{N_r(r)} \quad (2)$$

Weibull fading channel model has been often used to analyze performance in mobile communication environment. Weibull distribution is an adequate distribution in describing multipath waves propagating in the nonhomogenous environments in which Rayleigh distribution is not an adequate choice.

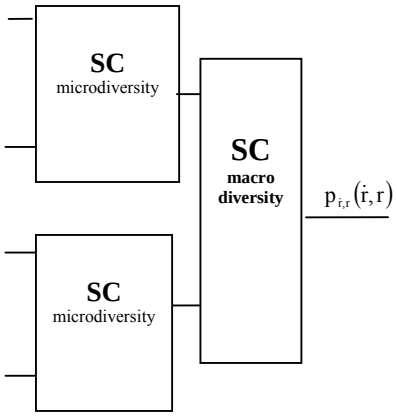


Fig. 1. Model of system

In Fig. 1 macrodiversity system contains of two microdiversity systems is presented. The independent Weibull fading at the input of both microdiversity systems is appeared. SC receivers with two input branches are obtained. The probability density function and cumulative distribution function of input signals' envelope [3]:

$$p_{r_{i1}}(r_{i1}/\Omega_1) = \frac{\alpha_1}{\Omega_1} r_{i1}^{\alpha_{i1}} \exp\left(-\frac{r_{i1}^{\alpha_1}}{\Omega_1}\right), r_{i1} \geq 0, i = 1,2 \quad (3)$$

$$F_{r_{i1}}(r_{i1}/\Omega_1) = 1 - \exp\left(-\frac{r_{i1}^{\alpha_1}}{\Omega_1}\right) \quad (4)$$

$$p_{r_{i2}}(r_{i2}/\Omega_2) = \frac{\alpha_2}{\Omega_2} r_{i2}^{\alpha_{i2}} \exp\left(-\frac{r_{i2}^{\alpha_2}}{\Omega_2}\right), r_{i2} \geq 0, i = 1,2 \quad (5)$$

$$F_{r_{i2}}(r_{i2}/\Omega_2) = 1 - \exp\left(-\frac{r_{i2}^{\alpha_2}}{\Omega_2}\right) \quad (6)$$

parameters α_1 and α_2 are Weibull fading parameters ($\alpha_1>0$, $\alpha_2>0$) and they express fading intensity measures. When values of those parameters increase the fading intensity decreases. For the special case when $\alpha_1=2$ and $\alpha_2=2$ proposed Weibull distributions become Rayleigh distribution, so our analysis has high level of generality. The mean powers from the input branches of each microdiversity system are defined as $\Omega_1=E(r_{i1}^{\alpha_1})$ and $\Omega_2=E(r_{i2}^{\alpha_2})$.

II. ANALYSIS

SC combiners at the output of both microdiversity systems pick the maximal signal's mean power which can be mathematically presented as:

$$p_{r_1}(r_1/\Omega_1) = p_{r_{11}}(r_1/\Omega_1) F_{r_{21}}(r_1/\Omega_1) + p_{r_{21}}(r_1/\Omega_1) F_{r_{11}}(r_1/\Omega_1) \quad (7)$$

$$p_{r_2}(r_2/\Omega_2) = p_{r_{12}}(r_2/\Omega_2) F_{r_{22}}(r_2/\Omega_2) + p_{r_{22}}(r_2/\Omega_2) F_{r_{12}}(r_2/\Omega_2) \quad (8)$$

The probability density function of derivates $\dot{r}$ of the received signals r at the output of both microdiversity systems, with respect to time, are Gaussian pdfs [5]:

$$p(\dot{r}_j) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_{r_j}} \exp\left(\frac{\dot{r}_j^2}{2\hat{\sigma}_{r_j}^2}\right), j = 1,2 \quad (9)$$

where $\dot{r}_i$ presents derivatives of received microdiversity envelopes and can be expressed as [5]:

$$\dot{r}_j = \begin{cases} \dot{r}_{1j}, \dot{r}_{1j} > \dot{r}_{2j} \\ \dot{r}_{2j}, \dot{r}_{2j} > \dot{r}_{1j} \end{cases}, j = 1,2 \quad (10)$$

For isotropic scattering, $\dot{r}$ is a Gaussian distributed random variable with zero mean and variance can be expressed as [6]:

$$\dot{\sigma}_{r_j}^2 = \left(\frac{2\pi f_d}{\alpha_j}\right) \Omega_j r_j, j = 1,2 \quad (11)$$

where f_d is a Doppler shift frequency.

Whereas pdfs of signal envelopes r_i at the output of microdiversity systems and pdfs of its derivatives $\dot{r}_i$ are statistically independent, joint pdf of r_i and $\dot{r}_i$ is given by:

$$p_{r_j\dot{r}_j}(r_j, \dot{r}_j/\Omega_j) = p(r_j/\Omega_j) p(\dot{r}_j) =$$

$$= 2\frac{\alpha_j}{\Omega_j} r_j^{\alpha_j} \exp\left(-\frac{r_j^{\alpha_j}}{\Omega_j}\right)\left(1 - \exp\left(-\frac{r_j^{\alpha_j}}{\Omega_j}\right)\right) \cdot$$

$$\cdot \frac{1}{\sqrt{2\pi}\hat{\sigma}_{r_j}} \exp\left(\frac{\dot{r}_j^2}{2\hat{\sigma}_{r_j}^2}\right) \quad (12)$$

At the output of suggested macrodiversity system is also SC combiner. It picks the maximum microdiversity's output average signal power, which mathematically can be defined by formulae:

$$p_{r,\dot{r}}(r,\dot{r}) = \int_0^\infty d\Omega_1 \int_0^{\Omega_1} d\Omega_2\, p_{r_1\dot{r}_1}(r,\dot{r}/\Omega_1) P_{\Omega_1\Omega_2}(\Omega_1\Omega_2) + \int_0^\infty d\Omega_2 \int_0^{\Omega_2} d\Omega_1\, p_{r_2\dot{r}_2}(r,\dot{r}/\Omega_2) P_{\Omega_1\Omega_2}(\Omega_1\Omega_2) \quad (13)$$

$$F_r(r) = \int_0^\infty d\Omega_1 \int_0^{\Omega_1} d\Omega_2 F_{r_1}(r/\Omega_1) P_{\Omega_1\Omega_2}(\Omega_1\Omega_2) + \int_0^\infty d\Omega_2 \int_0^{\Omega_2} d\Omega_1 F_{r_2}(r/\Omega_2) P_{\Omega_1\Omega_2}(\Omega_1\Omega_2) \quad (14)$$

There is shadowing at the input of microdiversity systems as well as at the input of macrodiversity system. The slow fading is statistically independent due to sufficient input antenna spacing. In this case we achieve the highest decrease of the fading influence on system's performances. The slow fading is modeled by joint Gaussian pdf [7]:

$$p_{\Omega_1,\Omega_2}(\Omega_1,\Omega_2) = p_{\Omega_1}(\Omega_1) p_{\Omega_{21}}(\Omega_2) = \frac{1}{\Gamma(c_1)} \frac{\Omega_1^{c_1-1}}{\Omega_{01}^{c_1}} \exp\left(\frac{-\Omega_1}{\Omega_{01}}\right) \frac{1}{\Gamma(c_2)} \frac{\Omega_2^{c_2-1}}{\Omega_{02}^{c_2}} \exp\left(\frac{-\Omega_2}{\Omega_{02}}\right) \quad (15)$$

where Ω_{0i} present mean values at the inputs of macrocombiner, while c_1 and c_2 present orders of Gamma distributions, and they determine measures of the shadowing present in the channel.

III. NUMERICAL RESULTS

After substituting (12), (13) and (15) in (1), we obtain following expressions for normalized level crossing rate, LCR:

$$N_r(r)/f_d = \int_0^\infty H_1 \gamma\left(c_2, \frac{\Omega_1}{\Omega_{02}}\right) \Omega_1^{c_1-1} \exp\left(-\frac{\Omega_1}{\Omega_{01}}\right) \dot{\sigma}_{r_1}^2 d\Omega_1 + \int_0^\infty H_2 \gamma\left(c_1, \frac{\Omega_2}{\Omega_{01}}\right) \Omega_2^{c_2-1} \exp\left(-\frac{\Omega_2}{\Omega_{02}}\right) \dot{\sigma}_{r_2}^2 d\Omega_2 \quad (16)$$

with $\gamma(a,x)$ incomplete Gamma function [8] and

$$H_1 = \frac{1}{\Gamma(c_1)\Omega_{01}^{c_1}\Gamma(c_2)\Omega_{02}^{c_2}} 2\frac{\alpha_1}{\Omega_1} r^{\alpha_1} \exp\left(-\frac{r^{\alpha_1}}{\Omega_1}\right)\left(1-\exp\left(-\frac{r^{\alpha_1}}{\Omega_1}\right)\right) \quad (17)$$

$$H_2 = \frac{1}{\Gamma(c_1)\Omega_{01}^{c_1}\Gamma(c_2)\Omega_{02}^{c_2}} 2\frac{\alpha_2}{\Omega_2} r^{\alpha_2} \exp\left(-\frac{r^{\alpha_2}}{\Omega_2}\right)\left(1-\exp\left(-\frac{r^{\alpha_2}}{\Omega_2}\right)\right) \quad (18)$$

Normalized LCR for various values of system's parameters is presented in Fig.2 LCR is normalized by maximal Dopler shift frequency f_d. We have considered the case of balanced macrocobiner's inputs, $\Omega_{01}=\Omega_{02}$. Numerical results for LCR (and AFD) are presented in the function of normalized signal level. Signal level is normalized with the sqare root of mean power of Gamma distributed signal , $\rho = z/\sqrt{\Omega_{01}}$.

In practice it is usauall that treshold level at the receiver, r is set to the value, that is smaller than the root mean value Ω_{01}, in order to obtain reasonable indication of outage. Because of that, from practical point of view, more intrasting are results for the values of normalized signal level that follows $\rho<0$ dB [9]. We can observe from Fig. 2, that for the normalzed signal levels, which are $\rho<0$ dB, larger values of Weibull paremeters α_1 and α_2 lead to smaller LCR due to smaller fading severity caused by the increment of α_1 and α_2. Also we can observe that LCR has smaller values in observed range in the presence of shadowing with smaller values of parameters c_1 i c_2.

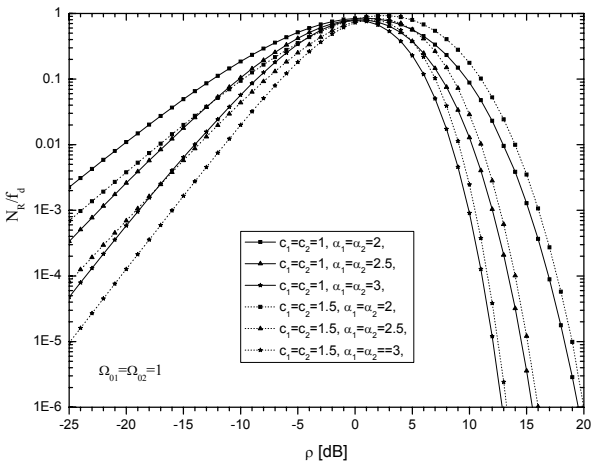


Fig. 2. Normalized LCR in the function of normalized signal level for various values of system's parameters

After substituting (14), (15) in (2), and following simillar mathematical procedure, we obtain following expressions for normalized average fade duration, AFD:

$$T_R(z) = \frac{\int_0^\infty \left(1-\exp\left(\frac{r^{\alpha_1}}{\Omega_1}\right)\right) \frac{1}{\Gamma(c_1)\Gamma(c_2)} \frac{\Omega_1^{c_1-1}}{\Omega_{01}^{c_1}} \exp\left(-\frac{\Omega_1}{\Omega_{01}}\right) \gamma(\Omega_1/\Omega_2, c_2) d\Omega_1}{N_z(z)} + \frac{\int_0^\infty \left(1-\exp\left(\frac{r^{\alpha_2}}{\Omega_2}\right)\right) \frac{1}{\Gamma(c_1)\Gamma(c_2)} \frac{\Omega_2^{c_2-1}}{\Omega_{02}^{c_2}} \exp\left(-\frac{\Omega_1}{\Omega_{01}}\right) \gamma(\Omega_2/\Omega_1, c_1) d\Omega_2}{N_z(z)} \quad (19)$$

Normalized AFD for various values of systems' parameters is presented in Fig.3.

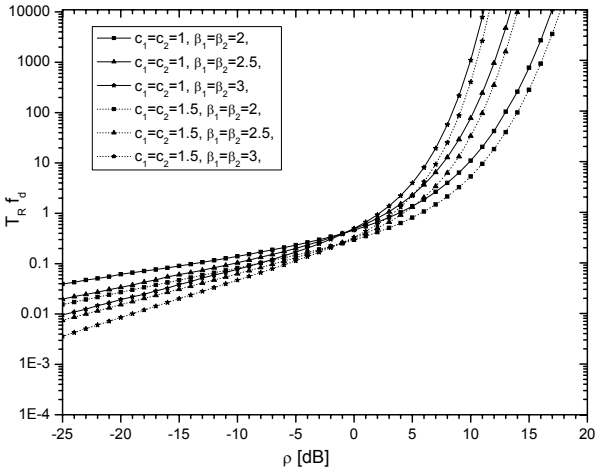


Fig. 3. Normalized AFD in the function of normalized signal level for various values of system's parameters

We can observe from Fig. 3, that for the normalzed signal levels, which are $\rho<0$ dB, larger values of Weibull paremeters α_1 and α_2 lead to higher values of AFD Also we can observe that AFD has higher values in observed range in the presence of shadowing with smaller values of parameters c_1 and c_2.

IV. CONCLUSION

Numerical results for LCR and AFD of SC macrodiversity system in the presence of independent Weibull fading at the microdiversity inputs are presented in this paper. Analysis has been performed for the case when selection on macro level is obtained considering mean power of the signlas at the input branches of macrocombiner. The shadowing was modeled by statistically independent Gamma distributions. Second order statistics has been analyzed for various values of systems' parameters.

REFERENCES

[1] G. Stuber, *Principles of Mobile Communication,* Boston: Kluwer Academic Publishers; 2000.

[2] W. C. Y. Lee, *Mobile communications engineering*, Mc-Graw-Hill, New York, 0-7803-7005-8/01, 1992.

[3] D. Drajić, *Uvod u statističku teoriju telekomunikacija*, Akademska misao, *Beograd*, 2003. (in serbian)

[4] E. A. Neasmith, N. C. Beaulieu, "New results in selection diversity", IEEE Trans. Commun., vol. 46, pp:695-704, 1998.

[5] X. Dong, and N. Beaulieu, ″Average Level Crossing Rate and Average Fade Duration of Selection Diversity", IEEE Communications Letters, Vol. 5. No.10, October 2001.

[6] A. Mitic, M. Jakovljevic, "Second-Order Statistics in Weibull-Lognormal Fading Channels", TELSIKS 2007, Conference Proceedings, Nis, Serbia, September 26-28, 2007.

[7] P. M. Shankar, "Analysis of microdiversity and dual channel macrodiversity in shadowed fading channels using a compound fading model," Int. J. Electron. Commun. *(AEU)*, 62, pp. 445-449, 2008.

[8] I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series, and products*, Academic Press, New York, 1980.

[9] M. Stefanović, A. Mitić, D. Pavlović, "Poređenje statističkih karakteristika signala u kanalu sa Vejbulovim fedingom za različite tehnike kombinovanja" , Infoteh, Jahorina, Conference Proceedings, 2007. (in serbian)

Influence of Cochannel Interference on SC Diversity System over Rician Fading

Aleksandra Panajotović[1], Mihajlo Stefanović[2], Dragan Drača[3], Daniela Milović[4], Nikola Sekulović[5], Dušan Stefanović[6]

Abstract – **Dual selection combining (SC) diversity system operating over correlated Rician fading channels is considered in this paper. It is assumed interference-limited environments where cochannel interference has Rayleigh statistic. Previous obtained probaility density functions (PDF) of SC output signal-to-interference ratio (SIR) have been used to analyse system peformance for different modulation schemes. Numerical results are presented to show effects of both fading severity and branch correlation.**

Keywords – **SC diversity system, Rician fading, Cochannel interference, Average symbol error probability.**

I. INTRODUCTION

In a mobile radio system, the received signal may suffer from both fading and shadowing. Fading is due to multipath propagation and shadowing is due to topographical variations of the transmission path. Microdiversity reception, such as spase diversity, is well known method used to combat multipath fading, while macrodiversity reception can be used to mitigate the effects of shadowing [1-3]. Microdiversity system with selection combining (SC) technique is considered in this paper, as technique having the least complexity. Traditionally, in SC the combiner chooses the branch with the highest signal-to-noise ratio, wich corresponds to the strongest signal if equal noise power is assumed among the branches [4]. Therefore, SC does not require all or some of the channel state informations from all the received signals.

In cellular land mobile radio, the received signal suffers cochannel interference (CCI), which also arises in mobile satellite communication channels. In some systems, where CCI is more significant than the front-end Gaussian noise [5], SC selects the branch with the highest signal–to-interference ratio (SIR) [6, 7].

[1]Aleksandra Panajotović is with Faculty of Electrical Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail:aleksandra.panajotovic@elfak.ni.ac.rs.

[2]Mihajlo Stefanović is with Faculty of Electrical Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[3]Dragan Drača is with Faculty of Electrical Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[4]Daniela Milović is with Faculty of Electrical Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[5]Nikola Sekulović is with Faculty of Electrical Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[6]Dušan Stefanović is with High Technical School of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

Rayleigh, Nakagami-m, Rice and Weibull statistical models are the most frequently used in communication system to describe the environment. There is assumption that, all signals, desired and undesired, have the same statistical characteristics. For example, in [8], both signals have Nakagami-m statistics, while in [9] they are subject to Weibull fading. In both papers, desired and undesired signals are correlated. Such assumption is quiet reasonable for medium to large cell systems. For instance, in a microcellular environment, an undesired signal from a distant cochannel cell may well be modeled by Rayleigh statistics, but Rayleigh fading may not be a good assumption for desired signal since a line-of-sight (LoS) path may exist within a microcell [10-12]. Therefore, different fading statistics are needed to characterize the desired and undesired signals in a microcellular radio system. Also in small-sized hand-hold terminal equipped with multiple antennas (space diversity), it is important to investigate the effects of branch correlation.

The performance of dual branch SC diversity receiver operating over correlated Rician fading channels in the presence of Rayleigh CCI had been studied in this paper. Actually, the average symbol error probability (ASEP), as important performance metric, is obtained for following modulation techniques binary phase shift keying (BPSK), binary frequency shift keying (BFSK) and M-ary quadrature amplitude modulation (M-QAM).

II. AVERAGE SYMBOL ERROR PROBABILITY

The Rician distribution is often used to model propagation path consisting of one strong direct LoS signal and many randomly reflected and usually weaker signals. Such fading environments are typically encountered in some microcellular systems [10, 13]. For the case when diversity antennas are not placed sufficiently apart, correlation arises between diversity branches. Then, desired signal envelopes experience correlative Rician fading with joint PDF [14]

$$p_{r_1 r_2}(r_1, r_2) = \frac{r_1 r_2 (1+K)^2}{\beta^2 (1-r^2)} \exp\left(-\frac{(r_1^2 + r_2^2)(1+K) + 4K\beta(1-r)}{2\beta(1-r^2)} \right) \sum_{k=0}^{+\infty} \varepsilon_k \cdot \cdot I_k\left(\frac{r_1 r_2 r (1+K)}{\beta(1-r^2)} \right) I_k\left(\frac{r_1}{(1+r)} \sqrt{\frac{2K(1+K)}{\beta}} \right) I_k\left(\frac{r_2}{(1+r)} \sqrt{\frac{2K(1+K)}{\beta}} \right) \quad (1)$$

where r is branch correlation coefficient, β is average power of r_1 and r_2 defined as $\beta = \overline{r_1^2}/2 = \overline{r_2^2}/2$, K is Rice factor defined as the ratio of the signal power in the dominant component over the scattered power, $\varepsilon_k = 1\,(k=0)$, i.e.

$\varepsilon_k = 2$ ($k \neq 0$) and $I_k(\cdot)$ is modified Bessel function of the first kind and *k*-th order.

The envelope of CCI on diversity branches is Rayleigh distributed because of its multipath propagation over large distance [6]. Its correlative bivariate PDF, due to insufficient antenna spacing, is expressed by

$$p_{A_1,A_2}(A_1,A_2) = \frac{A_1 A_2}{\sigma_A^4(1-r^2)} \exp\left(-\frac{A_1^2 + A_2^2}{2\sigma_A^2(1-r^2)}\right) I_0\left(\frac{A_1 A_2 r}{\sigma_A^2(1-r^2)}\right) \quad (2)$$

where average power of CCI is $\sigma_A^2 = \overline{A_1^2}/2 = \overline{A_2^2}/2$.

In interference-limited fading environments selection combiner chooses and outputs the branch with largest SIR, i.e. $\mu_{sc} = \max\{r_1/A_1, r_2/A_2\}$. Analytical expression of PDF of SC output SIR in the form of infinite series was obtain in [15]. Equation (3), shown at the bottom of the page, presents this PDF and it is substantional to study wireless performance criteria such as channel capacity, ASEP, average output SIR, etc.

The ASEP, $\overline{P}_{se}$, can be evaluated directly by averaging the conditional symbol error probability, $P_{se}(\mu)$, over PDF of μ_{SC} [8, 16]:

$$\overline{P}_{se} = \int_0^\infty P_{se}(\mu) p_{\mu_{sc}}(\mu) d\mu \quad (4)$$

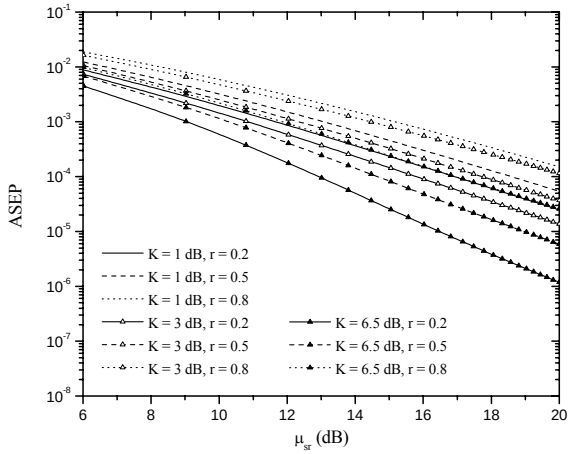
$P_{se}(\mu)$ is defined, for some modulation schemes, as:

$$P_{se}(\mu) = A\,erfc\left(\sqrt{B\mu^2}\right) \quad (5)$$

where $erfc(\cdot)$ is the complementary error function and *A*, *B* are constants the values of which depend on the specific modulation scheme under consideration.

III. NUMERICAL RESULTS

Using the previous mathematical analysis, various performance evaluation results have been obtained by means of numerical techniques and will be presented in this section. Such results include ASEP performance for different modulation schemes presented for different channel conditions. The proposed infinite series representations of (3) can be efficiently used to study important performance criteria, such as ASEP. The main problem in these infinite series expressions may be their convergence. However, obtained numerical results show that $p_{\mu_{SC}}(\mu)$ converges rapidly and number of sum terms need to achieve significant accuracy of $p_{\mu_{SC}}(\mu)$ depend on both Rice factor and branch correlation coefficient.



. Fig. 1. Average symbol error probability for BPSK system versus input average signal-to-interference ratio.

$$p_{\mu_{SC}}(\mu) = \exp\left(-\frac{2K}{1+r}\right) \sum_{k,\,p,\,n,\,l,\,m=0}^{+\infty} \frac{2\varepsilon_k K^{p+l+k}(K+1)^{2k+2n+p+l+2}(1-r)^{p+l+k+2}\Gamma(n+p+m+k+2)\Gamma(n+l+m+k+2) r^{2m+2n+k} \mu^{4n+4k+2p+2l+3}}{\beta^{2k+2n+p+l+2}(1+r)^{p+l+k-2} n!\,p!\,m!\,l!\,\Gamma(m+1)\Gamma(l+k+1)\Gamma(n+k+1)\Gamma(p+k+1)}$$

$$\cdot \left(\frac{\sigma_A^{2n+2l+2k-2m}\, {}_2F_1\left[n+l+k+m+2, n+l+k+1, n+l+k+2, -\frac{\sigma_A^2(K+1)}{\beta}\mu^2\right]}{(n+l+k+1)\left(\frac{1}{\sigma_A^2} + \frac{\mu^2(K+1)}{\beta}\right)^{n+p+k+m+2}} + \right. \quad (3)$$

$$\left. + \frac{\sigma_A^{2n+2p+2k-2m}\, {}_2F_1\left[n+p+k+m+2, n+p+k+1, n+p+k+2, -\frac{\sigma_A^2(K+1)}{\beta}\mu^2\right]}{(n+p+k+1)\left(\frac{1}{\sigma_A^2} + \frac{\mu^2(K+1)}{\beta}\right)^{n+l+k+m+2}} \right)$$

The average symbol error probability defined by (3), (4) and (5) is evaluated via numerical integration using the well-known mathematical software package *Mathematica 5.2*. The numerical results present ASEP plotted against input average SIR ($\mu_{sr} = \beta / \sigma_A^2$) for some modulation schemes. Figs. 1, 2 and 3 depict the effect of both branch correlation and fading severity on the error performance of considered SC diversity system. From all these figures one can conclude that increase of branch correlation coefficient and/or fading severity deteriorate system performance. Moreover, proposed results show that BPSK system exhibits the best performance. Although M-ary modulation techniques is attractive for use in bandlimited channels, because these techniques achieve better bandwidth efficiency. However, M-ary signaling results in poorer error performance. It is evident if we compare Fig. 3 with Figs. 1 and 2. Also, comparison of results from these figures shows that BFSK system has the greatest immunity against changing of *K* and *r*.

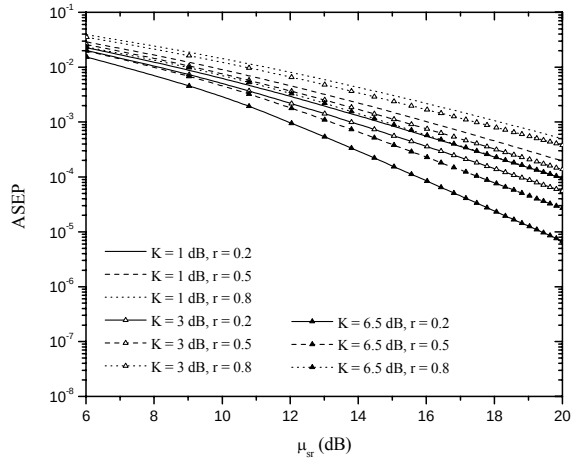


Fig. 2. Average symbol error probability for BFSK system versus input average signal-to-interference ratio.

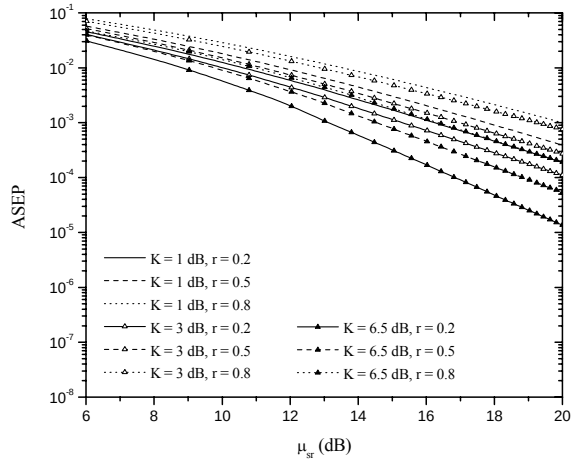


Fig. 3. Average symbol error probability for 4-QAM system versus input average signal-to-interference ratio.

IV. CONCLUSIONS

In summary, we presented a performance analysis of dual SC diversity system for various digital modulation schemes and over a fading channels encountered in real-life scenarios. Obtained results describe ASEP dependence on branch correlation coefficient and fading severity. They show that system's performance improves when Rice factor increases (fading severity decreases) and/or correlation coefficient decreases. The publication of these error system performance curves and comparisons will alow the diversity system designer to make the best choice in planing of wireless system.

REFERENCES

[1] W. C. Jakes, *Microwave Mobile Communications*, New York, Wiley, 1974.
[2] W. Lee, *Mobile Communications Engineering*, New York, McGraw-Hill, 1982.
[3] A. A. Abu-Dayya, N. C. Beaulieu, "Micro- and Macrodiversity NCFSK (DPSK) on Shadowed Nakagami-Fading Channels", IEEE Trans. Commun., vol. 42, no. 9, pp. 2693-2702, 1994.
[4] M. K. Simon, M. –S. Alouini, *Digital Communication Over Fading Channels*, New York, Wiley, 2000.
[5] D. C. Cox, "Cochannel Interference Considerations in Frequency Reuse Small-Coverage-Area Radio Systems", IEEE Tran. Commun., vol. COM-30, no. 1, pp. 135-142, 1982.
[6] S. Okui, "Effects of CIR Selection Diversity with Two Correlated Branches in the m-Fading Channel", IEEE Trans. Commun., vol. 48, no. 10, pp. 1631-1633, 2000.
[7] V. A. Aalo, J. Zhang, "On the Effect of Cochannel Interference on Average Error Rates in Nakagami-Fading Channels", IEEE Commun. Lett., vol. 3, no. 5, pp. 136-138, 1999.
[8] G. K. Karagiannidis, "Performance Analysis of SIR-Based Dual Selection Diversity Over Correlated Nakagami-m Fading Channels", IEEE Trans. Commun., vol. 52, no. 5, pp. 1207-1216, 2003.
[9] M. Stefanović, D. Milović, A. Mitić, M. Jakovljević, "Performance Analysis with Selection Combining Over Correlated Weibull Fading Channel in the Presence of Cochannel Interference", International Journal AEÜ, vol. 62, no. 9, pp. 695-700, 2008.
[10] R. J. Bultitude, G. K. Bedal, "Propagation Characteristics on Microcellular Urban Mobile Radio Channels at 910 MHz", IEEE J. Select. Areas Commun., vol. 7, no. 1, pp. 31-39, 1989.
[11] F. Adachi, K. Ottno, "Block Error Probability for Noncoherent FSK with Diversity Reception in Mobile Radio", Electron. Lett., vol. 24, no.24, pp. 1523-1525, 1988.
[12] Y. D. Yao, A. U. H. Sheikh, "Investigations into Cochannel Interference in Microcellular Mobile Radio Systems", IEEE Trans. Veh. Technol., vol. 41, no. 2, pp. 114-123, 1992.
[13] R. Steele, "The Cellular Environment of Lightweight Handheld Portables", IEEE Commun. Mag., vol. 27, no. 7, pp. 20-29, 1989.
[14] M. K. Simon, "Comments on "Infinite-Series Representations Associated with the Bivariate Rician Distribution and Their Applications"", IEEE Trans. Commun., vol. 54, no. 8, pp. 1511-1512, 2006.
[15] A. Panajotović, M. Stefanović, D. Drača, "Performance analysis with selection combining over correlated Rician fading channel in the presence of cochannel interference," International Journal AEU, (accepted for publication), doi: 10.1016/jaeue.2008.08.001, 2008.
[16] N. C. Sagias. D. A. Zogas, G. K. Karagiannidis, "Selection diversity receivers over nonidentical Weibull fading channels", IEEE Trans. Vech. Technol., vol. 54, no. 6, pp. 2146-2151, 2005.

Channel Capacity of a System in Shadowed Fading Channels with Micro- and Macrodiversity Reception

Nikola Sekulović[1], Dragan Drača [2], Aleksandra Panajotović [3], Zorica Nikolić[4], Časlav Stefanović[5]

Abstract - **This paper studies wireless communication system following microdiversity to mitigate the effects of short-term fading and macrodiversity processing to reduce shadowing effects. *L*-branch maximal-ratio combining (MRC) is implemented at the micro level (single base station) and selection combining (SC) with two base stations (dual diversity) is implemented at the macro level. This model assumes a Rician density function for the envelope of the received signal and a gamma distribution to model the average power to account for the shadowing. Closed-form expression for the probability density function (PDF) of the signal-to-noise ratio (SNR) after the diversity combining at the micro and macro level is obtained. The derived PDF is applied to study channel capacity of proposed system.**

Keywords - **Shadowing, fading, microdiversity, macrodiversity, channel capacity**

I INTRODUCTION

In wireless communication systems, the received signal can be affected by both short-term fading and long-term fading (shadowing). Short-term fading is the result of multipath propagation while shadowing is the result of large obstacles between transmitter and receiver.

The reliability of communication over the wireless channels can be improved using diversity techniques, such as space diversity [1], [2]. Diversity techniques at single base station (microdiversity) reduce the effects of short-term fading. Impairments due to shadowing can be mitigated using macrodiversity techniques which employ the processing of signals from multiple base stations. The use of composite micro- and macrodiversity has recently received considerable interest due to fact that it simultaneously combats both short-term fading and shadowing. A composite multipath/shadowed fading environment modeled either as Rayleigh-lognormal, Rician-lognormal or Nakagami-lognormal is considered in [3-5].

[1]Nikola Sekulović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: sekulani@bankerinter.net.

[2]Dragan Drača is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[3]Aleksandra Panajotović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[4]Zorica Nikolić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[5]Časlav Stefanović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

The use of lognormal distribution to model the average power which is random variable due to shadowing doesn't lead to a closed form solution for the probability density function (PDF) of the signal-to-noise ratio (SNR) at the receiver. A compound fading model uses a gamma distribution to account for shadowing instead of the lognormal distribution [6], [7]. This model incorporates short-term fading and shadowing and provides an analytical solution for the PDF of the SNR facilitating the analysis of wireless systems.

In this paper, system following micro- and macrodiversity reception in correlated gamma shadowed Rician fading channels is considered. Closed-form expression for the PDF of the SNR is obtained and used to study channel capacity of proposed system.

II SYSTEM MODEL AND CHANNEL CAPACITY

Macrodiversity composed of two geographically distributed microdiversity systems (base stations) per cell operating over gamma shadowed Rician fading channels is analyzed in this paper. *L*-branch maximal-ratio combining (MRC) is implemented at the micro level (single base station) and selection combining (SC) with two base stations (dual diversity) is implemented at the macro level. Signals at antennas in single base station are independent. Because of fact that shadowing has a larger correlation distance, the two base stations are treated to have nonzero correlation.

The instantaneous SNR at the MRC output of the *i*th base station is given by $X_i = \left(E_b / N_0\right)\sum_{j=1}^{L} A_{ij}^{\,2} \quad (i=1,2)$, where E_b is the transmitted signal energy per information bit, N_0 is the single-sided power spectral density of the additive white Gaussian noise and A_{ij} are statistically independent envelopes of the faded signals received on the *j*th branch. The PDF of the X_i is given by [4]

$$f_{X_i}\left(X_i|Y_i\right) = \frac{K+1}{Y_i}\exp\left(-\frac{(K+1)X_i}{Y_i} - KL\right)\left(\frac{(K+1)X_i}{KLY_i}\right)^{\frac{L-1}{2}} \times I_{L-1}\left(2\sqrt{\frac{KL(K+1)X_i}{Y_i}}\right), \quad i=1,2 \tag{1}$$

where Y_i is the average power at the base station, K is Rice factor defined as the ratio of the signal power in dominant component over the scattered power and $I_n(\cdot)$ is modified Bessel function of the first kind and *n*th order. The conditional

nature of the PDF in Eq. (1) illustrates the existence of shadowing.

The PDF of the SNR at the output of a dual-port selection based macrodiversity system is defined as

$$f_X(X)=\int_0^\infty dY_1\int_0^{Y_1} f_{X_1}(X|Y_1)f_{Y_1Y_2}(Y_1,Y_2)\ dY_2+ \\ +\int_0^\infty dY_2\int_0^{Y_2} f_{X_2}(X|Y_2)f_{Y_1Y_2}(Y_1,Y_2)\ dY_1= \\ =2\int_0^\infty f_{X_1}(X|Y_1)dY_1\int_0^{Y_1} f_{Y_1Y_2}(Y_1,Y_2)\ dY_2 \quad (2)$$

The compound PDF fading model uses a gamma distribution for shadowing. In that case, the joint PDF of Y_1 and Y_2 can be expressed as [8], [9]

$$f_{Y_1Y_2}(Y_1,Y_2)=\frac{\rho^{-\frac{c-1}{2}}}{\Gamma(c)(1-\rho)Y_0^{c+1}}(Y_1Y_2)^{\frac{c-1}{2}}\exp\left(-\frac{Y_1+Y_2}{Y_0(1-\rho)}\right) \\ \times I_{c-1}\left(\frac{\sqrt{4\rho Y_1Y_2}}{Y_0(1-\rho)}\right) \quad (3)$$

where ρ is the correlation between Y_1 and Y_2, c is the order of gamma distribution, Y_0 is related to the average power of Y_1 and Y_2, and $\Gamma(\cdot)$ is gamma function. Substituting Eqs. (1) and (3) in Eq. (2) and using [10, Eqs (3.351/1) and (3.471/9)], for integer values of c the following expression is obtained

$$f_X(X)=\frac{4\exp(-KL)}{\Gamma(c)} \\ \times\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(K+1)^{\frac{L+c+i+j}{2}}(KL)^i Y_0^{-\frac{c+L+i+j}{2}}X^{\frac{i+j+c+L-2}{2}} \\ \times\frac{\rho^j}{(1-\rho)^{\frac{L-c+i+j}{2}}}\frac{(c+j-1)!}{i!j!\Gamma(i+L)\Gamma(j+c)} \quad (4) \\ \times\left\{K_{c-L-i+j}\left(2\sqrt{\psi}\right)-\sum_{l=0}^{c+j-1}\frac{1}{2^{\frac{c-L-i+j+l}{2}}}\frac{\psi^{\frac{l}{2}}}{l!}K_{c-L-i+j+l}\left(2\sqrt{2\psi}\right)\right\}$$

where $K_n(\cdot)$ is modified Bessel function of the second kind of order n [10], [11] and $\psi=\frac{(K+1)X}{Y_0(1-\rho)}$.

Channel capacity is an important performance measure in design of future digital communications systems since it provides an upper bound of maximum transmission rate. Considering a signals' transmission of bandwidth BW over the additive white Gaussian noise (AWGN) channel, the average channel capacity can be obtained averaging the Shannon capacity over the PDF of X, i.e.,

$$\bar{C}=BW\int_0^\infty \log_2(1+X)f_X(X)dX \quad (5)$$

III NUMERICAL RESULTS

In Fig. 1, the normalized average channel capacity is plotted as a function of Y_0 using Eq. (5). As it was expected, C improves with an decrease of correlation coefficient and increase of number of diversity branches at the micro level. The normalized channel capacity of system with no macrodiversity is also plotted. It can be deduced clear conclusion that the combination of micro- and macrodiversity provides significantly performance improvement. In Fig. 2, the normalized average channel capacity is plotted in relation to the Y_0 for several values of Rice factor and order of gamma distribution. We observe here that an increase of K and c leads to an improvement of the system performance.

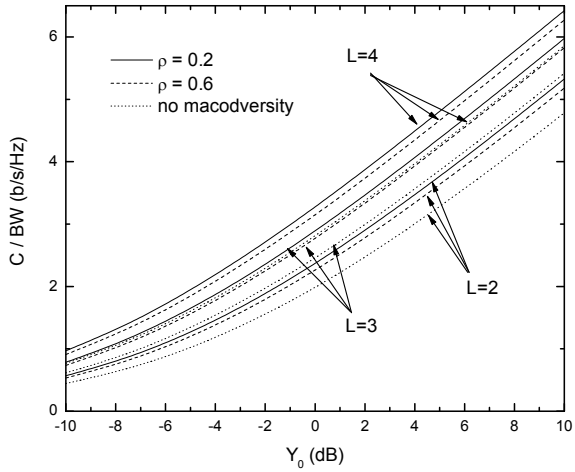


Fig. 1 Normalized average channel capacity versus Y_0 *(dB)* with *K=2.4 dB* and *c=2*

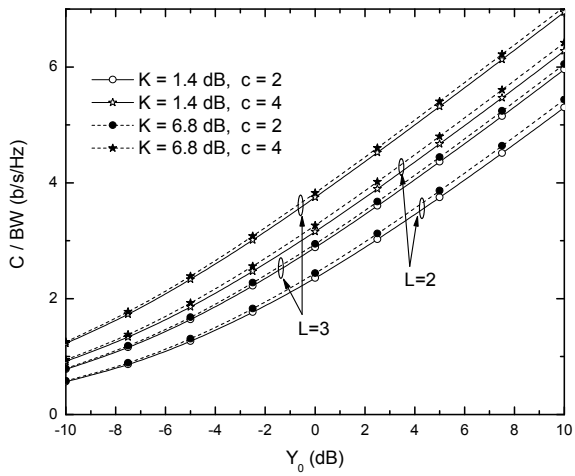


Fig. 2 Normalized average channel capacity versus Y_0 *(dB)* with ρ *=0.2*

IV CONCLUSION

Using a compound PDF model, system with micro- and macrodiversity reception in gamma shadowed Rician fading channels has been analyzed. Closed-form expression for the PDF of the SNR after diversity combining at the micro and macro level is obtained and used to study the average channel capacity in Shannon's sense of proposed system. Numerical results have been graphically presented, showing the effect of number of branches, correlation coefficient, Rice factor and order of gamma distribution on the system performance. It was also shown that composite micro- and macrodiversity provides significantly performance improvement which was the foreground task of this paper.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Science of Serbia within the project "Development and realization of new generation software, hardware and services based on software radio for specific purpose applications" (TR-11030).

REFERENCES

[1] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 1st ed. New York: Wiley, 2000.

[2] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.

[3] F. Hansen and F.I. Mano, "Mobile Fading-Rayleigh and Lognormal Superimposed", IEEE Trans. Vehic. Tech., vol. 26, pp. 332–335, 1977.

[4] J. Zhang and V. Aalo, "Effect of Macrodiversity on Average-Error Probabilities in a Rician Fading Channel with Correlated Lognormal Shadowing", IEEE Trans. Commun. vol. 49, pp. 14-18, 2001.

[5] E. K.Al-Hussaini, A.M. Al-Bassiouni, H. Mourad and H. Al-Shennawy, "Composite Macroscopic and Microscopic Diversity of Sectorized Macrocellular and Microcellular Mobile Radio Systems Employing RAKE Receiver over Nakagami Fading plus Lognormal Shadowing Channel", Wireless Personal Communications, vol. 21, pp. 309–328, 2002.

[6] P. M. Shankar, "Error rates in generalized Shadowed Fading Channels", Wireless Personal Communications, vol. 28, pp. 233-238, 2004

[7] P. M. Shankar, "Analysis of microdiversity and dual channel macrodiversity in shadowed fading channels using a compound fading model", Int. J. Electron. Comm., vol. 62, pp. 445-449, 2007.

[8] S. Yue, TBMJ Ouarda, B. Bobee, "A review of bivariate gamma distributions for hydrological application", J. Hydrol. vol. 246, pp. 1-18, 2001.

[9] E. Xekalaki, J. Panaretos and S. Psarakis, "A Predictive Model Evaluation and Selection Approach - The Correlated Gamma Ratio Distribution" ; STOCHASTIC MUSINGS: PERSPECTIVES FROM THE PIONEERS OF THE LATE 20TH CENTURY, J. Panaretos, ed., Laurence Erlbaum, Publisher, USA, pp. 188-202, 2003. Available at SSRN: http://ssrn.com/abstract=947067]

[10] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Academic, New York, 5th edn., 1994.

[11] M. Abramovitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover publications, New York, 1972.

The Performances of Generalized Selection Combiner in the Presence of Generalized-K Fading Channels

Mihajlo Č. Stefanović[1], Petar Nikolić[2], Dragana S. Krstić[3], Goran Lj. Stamenović [4] and Srdjan Milosavljević

Abstract **In this paper the performances of generalized selection combining (GSC) will be observed. Combiner has three inputs. The two strongest signals are isolated and added. For this combiner the probability density function for the output signal and the error probability will be determined. Generalized-K (K_G) fading is present at the input.**

Keywords – **Generalized selection combining, Generalized K fading, Probability density function, Bit error probability.**

I. INTRODUCTION

The shadow effect is a factor that degrades the system performances in mobile telecommunication systems at the most. It derogates the power of transmitted signal. When a received signal experiences shadow effect or fading during transmission, its envelope and phase fluctuate over time.

In wireless communication systems various techniques for reducing fading effect and influence of shadow effect are used. Such techniques are diversity reception, dynamic channel allocation and power control. Upgrading transmission reliability and increasing channel capacity without increasing transmission power and bandwidth is the main goal of diversity techniques.

Diversity reception, based on using multiple antennas at the receiver, (space diversity, with two or more branches), is very efficient methods used for improving system's quality of service, so it provides efficient solution for reduction of signal level fluctuations in fading channels. Multiple received copies of signal could be combined on various ways, and most popular of them are maximal ratio combining (MRC), equal gain combining (EGC), selection combining (SC) and switch and stay combining (SSC).

Maximal-Ratio Combining (MRC) is one of the most widely used diversity combining schemes whose SNR is the sum of the SNR's of each individual diversity branches. MRC is the optimal combining scheme, but its price and complexity are higher. Also, MRC requires cognition of all channel parameters and admit in the same phase all input signals, because it is the most complicated for realization ([1]-[3]).

With SC receiver, the processing is performed at only one of the diversity branches, which is selectively chosen, and no channel information is required. That is why SC is much simpler for practical realization. In general, selection combining, assuming that noise power is equally distributed over branches, selects the branch with the highest signal-to-noise ratio (SNR) that is the branch with the strongest signal ([4]-[6]).

The fading influence to the system performances is considered in [7]. The most often Rayleigh, Rice, Nakagami, Weibull and log-normal fading are considered.

Relatively new models in communications over fading channels are *K* and Generalized-*K* distributions [8], which in the past have been widely used in radar applications [9], [10]. The Generalized-*K* distribution has two shaping parameters, and as a consequence includes the *K* distribution as a special case. Moreover, it is sufficiently generic as it is able to incorporate most of the fading and shadowing effects observed in wireless communication channels and hence seems to be appropriate for the generic modeling of fading channels [11].

In [8], the generalized-K distribution was presented as analytically simpler than the Nakagami-lognormal or Suzuki distributions and general enough to approximate them, as well as several others including Rayleigh and Nakagami-*m*. Moreover, in the same work, the Amount-of-Fading (AF) and the Average Bit-Error-Rate (ABER) for the special case of the binary phase swift keying (BPSK) modulation, were derived. However, detailed performance analysis for the SNR statistics of a receiver operating over Generalized-*K* fading channel has not yet been published in the open technical literature and this is the topic of the current work.

A detailed performance analysis of generalized selection combining GSC (2, *L*) diversity receivers operating over generalized-gamma fading channels is presented in [12]. For this class of receiver a novel closed-form expression for the moment's output signal-to-noise ratio is derived. Furthermore, infinite series representation for the moments-generating and the cumulative distribution function are obtained. The proposed mathematical analysis is accompanied by various performance evaluation results. These theoretical results are complemented by equivalent computer simulated results, which validate the accuracy of the proposed analysis.

In the paper[13] a simple closed form expression for the average SNR of this generalized selection combining for independent identically distributed diversity channels is found, which is upper bounded by the average SNR of maximal ratio combining and lower bounded by the average SNR of conventional selection combining. In this paper the

[1]Mihajlo Stefanovic is with Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: mihajlo.stefanovic@elfak.ni.ac.rs

[2]Petar Nikolić is with Tigar Tyres, Nikole Pašića, 18300 Pirot, Serbia, e-mail:p.nikolic@tigartyres.com

[3]Dragana Krstić is with Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail:dragana.krstic@elfak.ni.ac.rs

[4]Goran. Stamenovic is with Tigar, Pirot, 18300 Pirot, Serbia, e-mail: goran.stamenovic@tigar.com

[5]Srdjan Stanković is with Faculty of Electronic Engineering,

presence of log-normal fading at the input of generalized selection combiner will be observed.

We will derive the expressions for the probability density function for the output signal, characteristic function for the output signal, the moments for the output signals and the error probability. We will present the error probability results graphically for different signal and fading parameters values

II. SYSTEM PERFORMANCES

The model of the complex GSC combiner with three inputs, considered in this paper, is shown in Fig. 1.

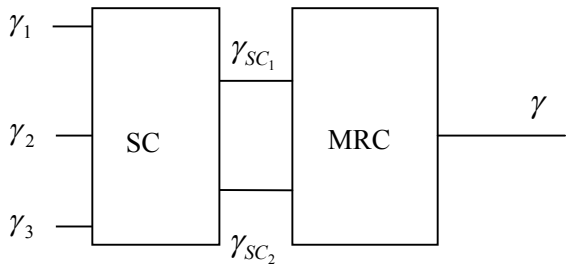


Fig. 1. Model of the GSC combiner with three inputs

The signals at the combiner input are γ_1, γ_2 and γ_3. The combiner chooses first the two strongest input signals from their three input signals. The SC combiner output signals are γ_1 and γ_2. These signals are added in the MRC combiner. The MRC combiner output signal is γ. That is the complex (GSC) combiner output signal and it is:

$$\gamma = \gamma_{SC_1} + \gamma_{SC_2} \tag{1}$$

For this complex GSC combiner we derive the probability density for the output signal, characteristic function, moments for the output signal and the error probability.

The probability density of the input signals, γ_1, γ_2 and γ_3, are:

$$p_{\gamma_1}(\gamma_1) = \frac{2\left(\frac{k_1 m_1}{\bar{\gamma}_1}\right)^{(k_1+m_1)/2} \gamma_1^{(k_1+m_1-2)/2}}{\Gamma(m_1)\Gamma(k_1)} K_{k_1-m_1}\left[2\left(\frac{k_1 m_1}{\bar{\gamma}_1}\gamma_1\right)^{\frac{1}{2}}\right], \quad \gamma_1 \geq 0 \tag{2}$$

$$p_{\gamma_2}(\gamma_2) = \frac{2\left(\frac{k_2 m_2}{\bar{\gamma}_2}\right)^{(k_2+m_2)/2} \gamma_2^{(k_2+m_2-2)/2}}{\Gamma(m_2)\Gamma(k_2)} K_{k_2-m_2}\left[2\left(\frac{k_2 m_2}{\bar{\gamma}_2}\gamma_2\right)^{\frac{1}{2}}\right], \quad \gamma_2 \geq 0 \tag{3}$$

$$p_{\gamma_3}(\gamma_3) = \frac{2\left(\frac{k_3 m_3}{\bar{\gamma}_3}\right)^{(k_3+m_3)/2} \gamma_3^{(k_3+m_3-2)/2}}{\Gamma(m_3)\Gamma(k_3)} K_{k_3-m_3}\left[2\left(\frac{k_3 m_3}{\bar{\gamma}_3}\gamma_3\right)^{\frac{1}{2}}\right], \quad \gamma_3 \geq 0 \tag{4}$$

where $m_i \geq 1/2$ is the Nakagami-m shaping parameter, k_i is shadowing shaping parameter and $\bar{\gamma}_i$ is average SNR for input channels [14].

The joint probability density of the SC combiner output signals γ_{SC_1} and γ_{SC_2} is:

$$\begin{aligned} p_{\gamma_{SC_1}\gamma_{SC_2}}(\gamma_{SC_1}\gamma_{SC_2}) = \\ = p_{\gamma_1}(\gamma_{SC_1})\cdot p_{\gamma_2}(\gamma_{SC_2})\cdot F_{\gamma_3}(\gamma_{SC_2}) + \\ + p_{\gamma_2}(\gamma_{SC_1})\cdot p_{\gamma_1}(\gamma_{SC_2})\cdot F_{\gamma_3}(\gamma_{SC_2}) + \\ + p_{\gamma_1}(\gamma_{SC_1})\cdot p_{\gamma_3}(\gamma_{SC_2})\cdot F_{\gamma_2}(\gamma_{SC_2}) + \\ + p_{\gamma_3}(\gamma_{SC_1})\cdot p_{\gamma_1}(\gamma_{SC_2})\cdot F_{\gamma_2}(\gamma_{SC_2}) + \\ + p_{\gamma_2}(\gamma_{SC_1})\cdot p_{\gamma_3}(\gamma_{SC_2})\cdot F_{\gamma_1}(\gamma_{SC_2}) + \\ + p_{\gamma_3}(\gamma_{SC_1})\cdot p_{\gamma_2}(\gamma_{SC_2})\cdot F_{\gamma_1}(\gamma_{SC_2}) \end{aligned} \tag{5}$$

where: $\gamma_{SC_1} \geq \gamma_{SC_2}$.

In this manner all possible combinations of input signals values, γ_1, γ_2 and γ_3, are taken into consideration.

For: $\gamma_i \geq \gamma_j \geq \gamma_k$ SC combiner output signals are:

$$\gamma_{SC_1} = \gamma_i; \quad \gamma_{SC_2} = \gamma_j$$

where $i=1,2,3;$ $j=1,2,3;$ $k=1,2,3;$ $i \neq j \neq k$

We obtain the cumulative probability density of the input signals, γ_1, γ_2 and γ_3 as [14]:

$$\begin{aligned} F_{\gamma_i}(\gamma_i) = \pi \csc[\pi(k_i - m_i)]\cdot \\ \cdot\left[\frac{(k_i m_i \gamma_i/\bar{\gamma}_i)^{m_i}\,{}_1F_2(m_i;1+m_i-k_i,1+m_i;m_i k_i \gamma_i/\bar{\gamma}_i)}{\Gamma(k_i)\Gamma(1+m_i-k_i)\Gamma(1+m_i)} - \right. \\ \left. - \frac{(k_i m_i \gamma_i/\bar{\gamma}_i)^{k_i}\,{}_1F_2(k_i;1-m_i+k_i,1+k_i;m_i k_i \gamma_i/\bar{\gamma}_i)}{\Gamma(m_i)\Gamma(1-m_i+k_i)\Gamma(1+k_i)}\right] \end{aligned} \tag{6}$$

$i=1,2,3.$

If we replace the expressions (2)-(4) and (6) into (5) we have the joint probability density of the SC combiner output signals γ_{SC1} and γ_{SC2}, (for $\gamma_{SC_1} \geq \gamma_{SC_2}$), in too long form for this place.

The probability density of the GSC combiner output signal γ is:

$$p_\gamma(\gamma) = \frac{1}{\pi}\int_0^\gamma p_{\gamma_{SC_1}\gamma_{SC_2}}(\gamma - \gamma_{SC_2}, \gamma_{SC_2})\, d\gamma_{SC_2} \tag{7}$$

Finally, the error probability is [7]:

$$P_b(e)=\frac{1}{\pi}\int_0^{\infty}\int_0^{\pi/2}\exp\left(-\frac{g\gamma}{\sin^2\phi}\right)p_\gamma(\gamma)d\phi d\gamma \qquad (8)$$

III. NUMERICAL RESULTS

After some substitutions in the (8) the error probability becomes:

$$P_b(e)=\frac{1}{\pi}\int_0^{\infty}\int_0^{y_1}\int_0^{\pi/2}\exp\left(-\frac{gy_1}{\sin^2\phi}\right)\exp\left(-\frac{gy_2}{\sin^2\phi}\right)p_{y_1y_2}(y_1y_2)d\phi dy_1dy_2 \qquad (9)$$

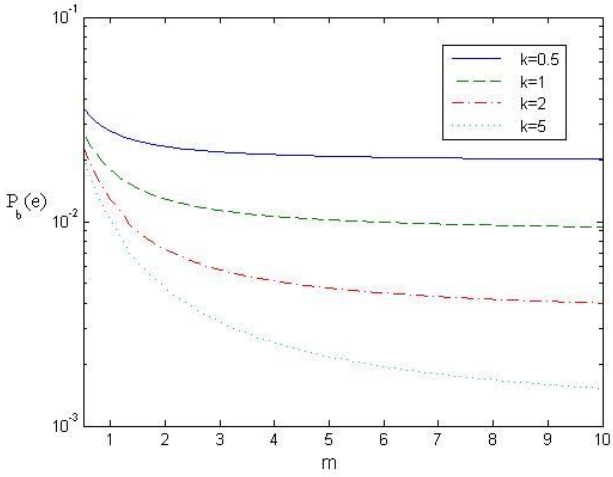


Fig. 2. *Pb(e)* versus parameter *m* for *k =0.5; 1; 2;5,* $\bar{\gamma}=5$ for GSC combiner.

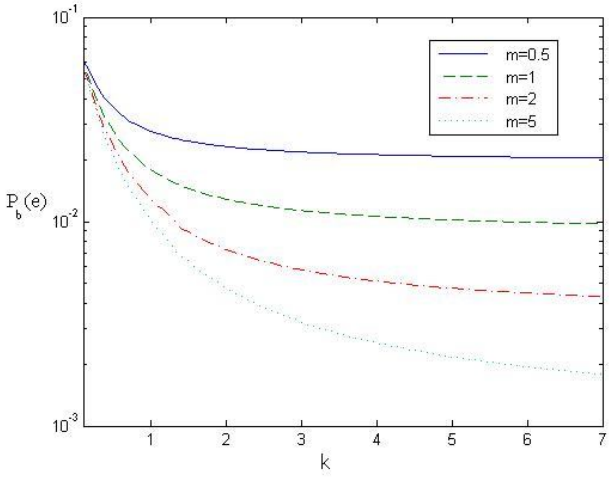


Fig. 3. *Pb(e)* versus parameter *k* for *m=0.5; 1; 2;5,* $\bar{\gamma}=5$ for GSC combiner

In this section we considered BPSK modulation scheme. For this case parameter *g* in above equation is equal to 1.

The main values of the input signals amplitudes are:

$$p_\gamma(\gamma)=\frac{1}{\pi}\int_0^{\gamma}p_{\gamma_{SC_1}\gamma_{SC_2}}(\gamma-\gamma_{SC_2},\gamma_{SC_2})d\gamma_{SC_2} \qquad (10)$$

where $\gamma_{SC_1}\geq\gamma_{SC_2}$.

Figures 2, 3 and 4 show *Pb(e)* versus parameters *m*, *k* and $\bar{\gamma}$, respectively, for different values of other parameters.

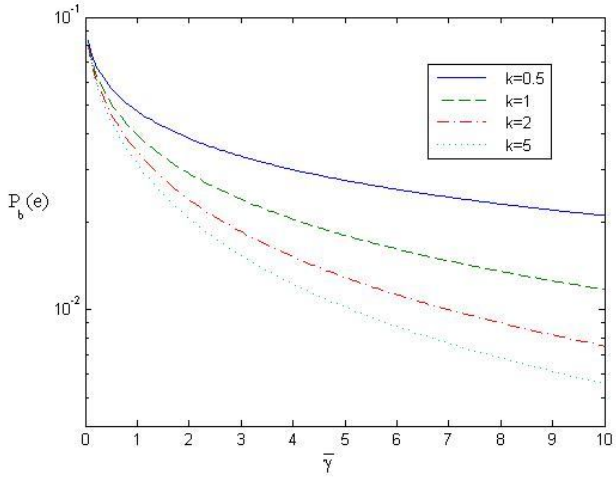


Fig. 4. *Pb(e)* versus parameter $\bar{\gamma}$ *for m=1, k =0.5; 1; 2;5,* for GSC combiner

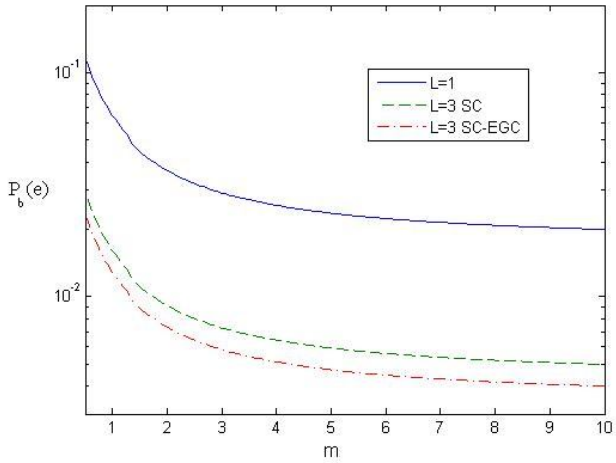


Fig. 5. *Pb(e)* versus parameter *m* for *k =2,* $\bar{\gamma}=5$

In Figures 5, 6 and 7 we can see *Pb(e)* versus parameters *m*, *k* and $\bar{\gamma}$, respectively, for different values of other parameters, for *L=3* when we have only SC combiner, then when we have complex GSC combiner and the case for one channel transmission, *L=1*.

From all these figures we can note that excellent agreement exists between theoretical and computer calculated results.

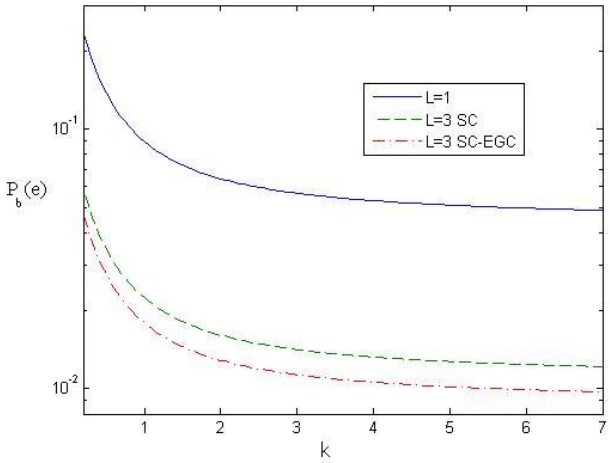


Fig. 6. *Pb(e)* versus parameter parameter *k* for *m =1,* $\overline{\gamma} = 5$.

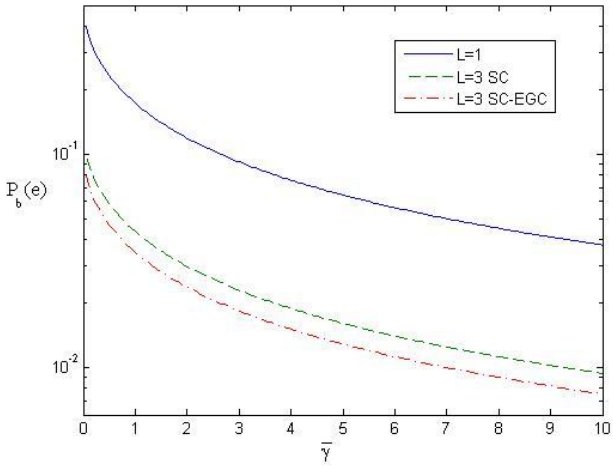


Fig. 7. *Pb(e)* versus parameter $\overline{\gamma}$ *for m=1, k =2.*

IV. CONCLUSION

In this paper diversity system with complex GSC combiner with three inputs is considered. The joint statistical characteristics of SC combiner output signal are derived, and then the probability density for the complex combiner output signal in the presence of generalized-*K* fading. The probability density can be used for the system error probability and for the outage probability calculating. Generalized selection combining (GSC) is considered as an alternative diversity scheme for bridging the performance emptiness between two classical diversity schemes, maximal ratio combining (MRC) and selection combining (SC). As compared to optimal MRC combining, GSC has reduced system complexity, while as compared to SC, which is one of the simplest diversity schemes, it has improved performance.

In the past the performance of GSC diversity receivers has been analyzed for various fading models, including Rayleigh, Nakagami-*m* and Weibull. Some attention is given to the practical important class of GSC (2, *L*), where among *L* available resolvable path the two strongest are adaptively combined. Since this class of receivers is an effective compromise achieving very good performance with reduced implementation complexity, it is considered also here in the presence of generalized-*K* fading and when number of branches is three: *L*=3. We can note the excellent agreement between theoretical and computer calculated results.

REFERENCES

[1] A. Annamalai, C. Tellambura and V. K. Bhargava, "Error performance of M-ary QAM with MRC diversity reception in a Nakagami fading channel", IEEE Int. Symp. Wireless Communications Dig., pp.44, May 1998.

[2] D. Krstić, M. Stefanović, "The statistical characteristics of the MRC diversity system output signal", Electronics and Electrical Engineering, No.1(73), pp. 45-48, January 2007.

[3] K. Noga, "The performance of binary transmition in slow Nakagami fading channels with MRC diversity", IEEE Trans, Commun., vol. 46, pp. 863-865, July 1998.

[4] K. Sivanesan and N. C. Beaulieu, "Exact BER analysis of bandlimited BPSK with EGC and SC diversity in cochannel interference and Nakagami fading", IEEE Commun. Lett ., vol. 8, pp. 623-625, Oct. 2004.

[5] D. Krstić, G. Stamenović, P. Nikolić, M. Stefanović, "Statistical Characteristics of Output Signal from Dual Diversity SC Combiner for Demodulation of BPSK Signals", INTERNATIONAL SCIENTIFIC CONFERENCE UNITECH'08, 21–23. November 2008, GABROVO

[6] M. Stefanović, D. Krstić, J. Anastasov, S. Panić, A. Matović, "Analysis of SIR-based Triple SC System over Correlated α-μ Fading Channels", to be presented at The Fifth Advanced International Conference on Telecommunications, AICT 2009, Venice/Mestre, Italy, May 24-28, 2009.

[7] M. K. Simon, M. S. Alouni, *Digital Communication over Fading Channels,* Second Edition, Wiley Interscience, New Jersey, pp. 586, 2005.

[8] P. M. Shankar, "Error rates in generalized shadowed fading channels," Wireless personal communication, vol. 00, pp. 1-6, Jan 2004.

[9] D. R. Iskander and A. M. Zoubir, "Estimation of the parameters of the *K*-distribution using higher order and functional moments", IEEE Trans. Acrosp. Electron. System, vol. 35, no. 4, pp. 1453-1457, Nov. 1999.

[10] S. Chitroub, A. Houacine and B. Sansal, "Statistical characterisationand modelling of SAR images", Elsevier, Signal Processing, vol. 82, pp. 69-92, 2002.

[11] P. S. Bithas, N. C. Sagias, P. T. Mathiopoulos, G. K. Karagiannidis, A. A Rontogiannis, "Digital Communications over Generalized-K Fading Cannes", in Proc. International Workshop on Satellite and Space Communications, pp. 684–687, Sept. 2005.

[12] P. S. Bithas, N. C. Sagias,P. T. Mathiopoulos, "GSC diversity receivers over generalized-gamma fading channels", IEEE Commun. Letters, vol. 11, No. 11, pp. 1-3, November 2007.

[13] N. Kong, L. B. Milstein, "Average SNR of a generalized diversity selection combining scheme", IEEE Commun. Letters, vol. 3, No. 3, pp. 57-59, March 1999.

[14] P. S. Bithas, N. C. Sagias, P. T. Mathiopoulos, G. K. Karagiannidis, A. A Rontogiannis, "On the Performance Analysis of Digital Communications over Generalized-*K* Fading Channels", IEEE Commun. Letters, vol. 10, No. 5, pp. 353-355, May 2006.

Session

RADIO COMMUNICATION SYSTEMS – 2

Central Control Unit for the Outdoor Unit of IMTEL Digital Radio Relay Systems Series B

Nemanja Mitrović[1], Dragan Obradović[2] and Miroslav Perić[3]

Abstract – **Today, very complex system are not rare thing, so that fact have been made their central control units very complex itself. This paper have assign, to tell about one realization of central control unit which is a part outdoor unit of IMTEL digital radio relay systems. It's a little bit hardware, followed by complex software.**

Keywords –**Control Unit, Digital Radio Relay System.**

I. INTRODUCTION

Important assign in developing devices is to made final (release) product as much universal and easy for upgrade. The central control units that will be described in this paper have been imagined as same unit for couple similar devices.

Block diagram of Digital radio relay systems (DRRS) [1] are shown on Fig. 1. The DRRS have near and far side. Each side of the DRRS is consisting of one indoor unit (IDU) and one (1+0 version) or two (version with 1+1 protection mode) outdoor units (ODU). IDU and ODU are interconnected with single coaxial cable which transitions power supply for ODU and data signals in both ways.

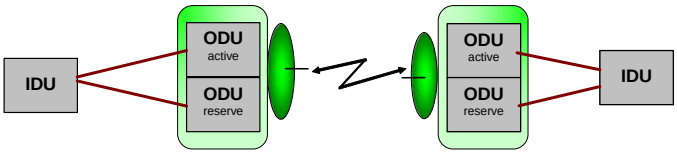


Fig. 1. Digital radio relay system block diagram

ODU block diagram is shown in Fig. 2. ODU consists of following physically modules: baseboard module, microwave module (MW transmitter with direct IQ modulator and MW receiver) microwave synthesizers and diplexer. The baseboard module is central part of ODU and it is designed to be used for more frequency bands (e.g. from 7 to 38 GHz). The baseboard module is integrated from following sub modules:

- Quad diplexer, which extracts from coaxial cable power supply and base band signal from IDU and inserts IF signal and control signals to IDU;
- DC/DC converters;
- Base band signal processing;
- Controlled IF synthesizers;
- IF automatic gain control amplifiers;
- Central control unit (CCU).

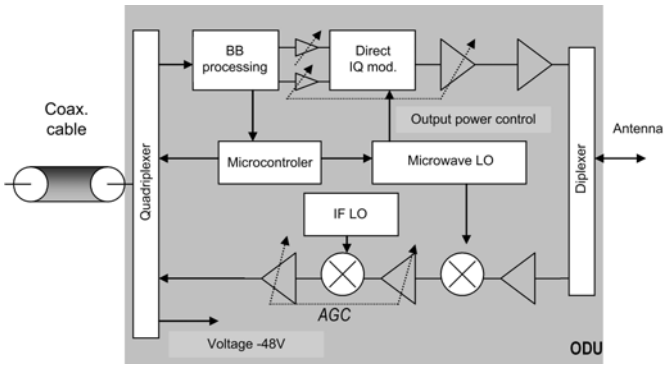


Fig. 2. Outdoor unit block diagram

The central control unit is very important for properly function of ODU and it makes possible to use same baseboard module for various frequency bands by using different configuration parameters. The central control unit has following basic functions:

- control transmitter and receiver frequency,
- control intermediate frequency,
- monitor alarms and other parameters,
- adjust transmitter power level,
- executing test modes,
- communication with IDU and other ODUs,
- maintain startup configuration parameters.

In the following sections will be described hardware and software realization of the central control unit.

II. HARDWARE

The complete ODU hardware was designed with attention to minimalism negative influence between modules. This is especially important for baseboard module which has many different types of components which can generate undesired signals (DC/DC converters, control unit, digital logic, etc.). Also, all components in ODU, should work in extended temperature range form -30C to +60C.

The central control unit is based on SILABS C8051F121 microcontroller [2], [3], and its realization on baseboard PCB module is shown on Fig. 3.. This microcontroller was chosen because it has integrated many peripherals like digital IOs, analog-digital and digital-analog converters, on-chip oscillator, watch-dog timer, dual UARTs, flash and RAM memory.

Since, this is DRRS that operate on very high frequencies, all controlled line are behaving as duct so there is coming to crosstalk between lines. There are couple solutions for this problem. First, filtering those lines, second make those lines

[1]Nemanja Mitrović is with IMTEL Komunikacije A.D., Bulevar Mihajla Pupina 165-B, 11070 Belgrade, Serbia, E-mail: nemanja@insimtel.com.

[2]Dragan Obradović is with IMTEL Komunikacije A.D., Bulevar Mihajla Pupina 165-B, 11070 Belgrade, Serbia, E-mail: obrad@insimtel.com.

[3]Miroslav Perić is with IMTEL Komunikacije A.D., Bulevar Mihajla Pupina 165-B, 11070 Belgrade, Serbia, E-mail: micha@insimtel.com.

short, third include first two and move peripheral on separate board. Most results could be reached, combining those solutions. To avoid more problems with this listen in between lines interface should have at least number of line as it possible.

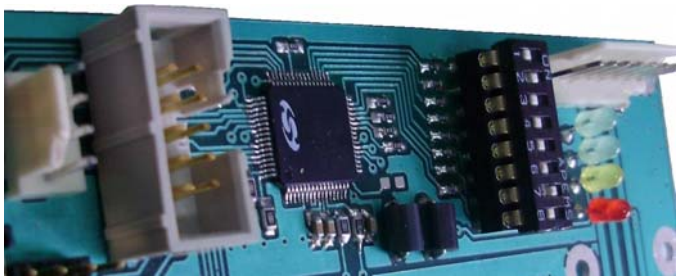


Fig. 3. Central Control Unit

In this case, for controlling main synthesizer [1], [4], [5], [6], [7], [8], which determine transmitter an receiver frequency, are used all solution. Main synthesizer is on separate board, with filtering lines on both sides. Length of lines is something that you can't reduce to zero, but you can make them smaller. Main synthesizer consists of Direct Digital Synthesis (DDS) for fine setting frequency and Phase Lock Loop (PLL) for rough setting frequency.

Both parts demand four controlling lines. As it shown on Fig. 4., those eight controlling lines, in this case are reduced to two lines. Microcontroller drives with two lines I^2C expander [10], which drives DDS and PLL with their own SPI interface [7], [8]. This is double serialization, but here, also, exists parallelization between I^2C expander [10] and I^2C EEPROM [9], which is filled with information about synthesizer. This method reduce undesired notice from digital components

Local synthesizer only consists of PLL [8], since it has only two frequencies that are depending on type DRRS. This synthesizer is on same board as CCU so its interface has short, filtered lines. Interface is SPI [8] that used three controlled lines.

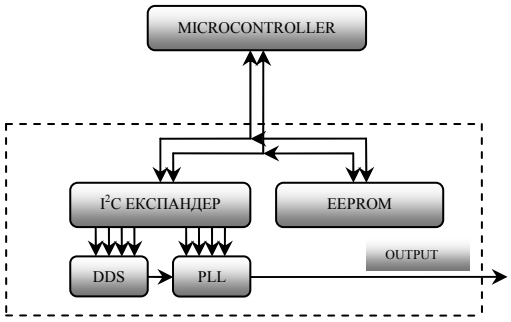


Fig. 4. Block diagram main synthesizer interface

IDU-ODU communication, on ODU side, is obtained by on UART0 port. UART1 port is used for terminal mode.

CPLD performs main task for baseband processor. Interface between CPLD and microcontroller is made with 4-bit data bus with control and address lines because distance is very short. Microcontroller can access to registers which are implemented in CPLD for settings and reading parameters.

CCU, also, work as acquisition center for information about bit rate transfer between two DRRSs. If there is some error in bit rate transfer CCU light on some diode and create alarms that can be read on IDU.

III. SOFTWARE

At the start of this paper was written that CCU hardware and software is imagined as same unit for couple similar devices. This CCU is same for DRRS that worked on 7GHz, 13GHz, 18GHz, 23GHz, 26GHz and 38GHz. At first time, developer should configure wanted device, and device will be ready for use. Next time when device is turned on it will be started as it was configured last time that it works. This is obtained by saving configuration of all devices that CCU controlling in flash memory of microcontroller SILABS C8051F121. Fig. 5. is showing CCU software block diagram.

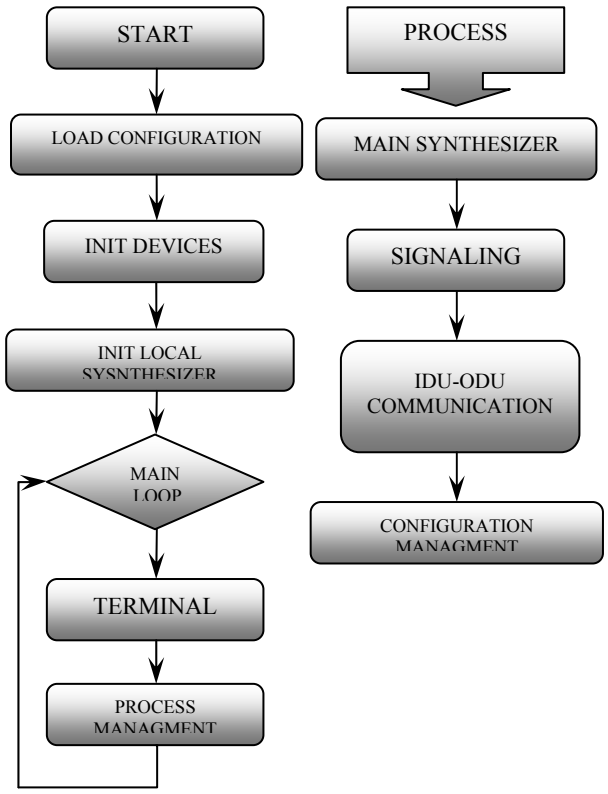


Fig. 5. CCU software block diagram

On the left side of the Fig. 5. is shown general software block diagram and on the right side of the Fig. 3. is shown main process. After loading last saved configuration, main program initialize controlled devices and set frequency of local synthesizer that depends on DRRS type. As it can be seen in main loop swaps controlled process, then save any change in configuration, and if it was demanded, writes something on the terminal.

A. Main and local synthesizer

Main synthesis is imagined as multi channel synthesis. Hardware setting channels is obtained by 8-bit switcher that is connected to the microcontroller, which controlled synthesis.

It is easy to see that using this kind control number of channels is limited by bits of switcher, but in here this is only secondary way setting channels. Better way setting channels is via terminal, but the bet way setting channel is via radio relay connection that is obtained by IDU-ODU communication. Only limit in number of channels is maximal and minimal frequency that microwave parts (diplexers, MW filters) of ODU can support.

CCU software has a function that calculates right frequency considering DRRS type, frequency band, sub band and desired channel. Setting channel demands order in setting DDS and PLL. Setting DDS must always go first.

Using two instead ten lines for controlling main synthesizer put us to double serialization and parallelization. From software opinion, double serialization has two levels. First level is software I^2C interface and second level is SPI that used as substructure first level - I^2C interface. Second level of software is controlling [7] and PLL [8] independent. On first level there is parallelization between I^2C expander [10] and I^2C EEPROM [9]. This is easier part, thanks to characteristic of I^2C interface. On I^2C interface can be connected several peripherals that are assign to different addresses. Double serialization and paralellization is shown on Fig. 6..

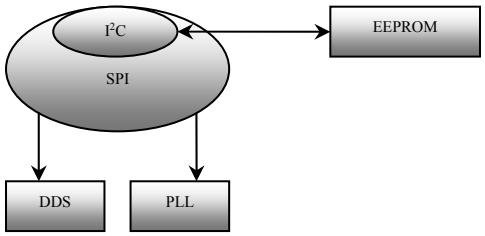


Fig. 6. Double serialization and parallelization

In the EEPROM are stored information about type, model and calibration data for synthesizer which are used from CCU software for correct calculation.

Local synthesis is setting at the startup and can be changed only via terminal mode, when DRRS type and frequency band is chosen. It has only PLL and is controlled using SPI interface.

B. Signaling

Another function of CCU is collecting information about all subsystems. Microcontroller is connected via its pins to ODU radio relay communication hardware. That way microcontroller collects errors, alarms and processes them. Processing alarms is reducing to preparing message, software filling signaling structure for IDU-ODU communication. Also, microcontroller drives led diodes that indicate some alarms. One more function of signaling module is reading temperature of microcontroller, and monitoring AGC voltages in IQ and IF section. Temperature and AGC voltages are analog signal that are converted to digital using analog to digital converters (AD converters) that are integrated in SILABS C8051F121 microcontroller.

C. IDU-ODU communication

Main purpose of IDU-ODU communication is sending signaling and configuration parameters to IDU or to PC computer which are equipped with software for DRRS remote monitoring and control RRUNet [5]. From this software is also possible to setup some ODU parameters like transmitter and receiver frequency, DRRS capacity, test modes, etc.

Communications between IDU and ODU is realized with asynchrony communication with baud rate 9600 bit/s by messages with are received and transmitted form UART0 port. Each messages had destination and source addresses, type of messages and data field. With this protocol is possible to communication between PC and ODU where IDU only forward messages. Also, one ODU can communicate with other ODU on same or opposite side of DRRS. This ODU-ODU communication can be used for adaptive transmits power control (ATPC) and for realization does hot-standby protection system where only one ODU transmitter in 1+1 configuration is active in same time.

D. Terminal

Terminal mode is obtained via UART1 [7], [8] of microcontroller. UART1 routine is processing in microcontroller interrupt routine. That way processing terminal mod is lower priority and does not disturbs other more important CCU jobs.

```
******
BB IF blok, Firmware V0.20, (c)2007-08 IMTEL Komunikacije a.d.

>UCITAVANJE... U REDU.
> UCITANA DEFAULT KONFIGURACIJA

> UCITANA DEFAULT KONFIGURACIJA IZ FLASH-A

>CHANGING CHANNEL...
> CHANNEL 1
>info

 BB IF blok (c)2007-08 IMTEL Komunikacije V0.20

Tip ODU/Podopseg                        - RRU23B/NIZI
S/N ODU jedinice                        - 0/1
S/N lokalnog oscilatora                 - 0
Najnizi radni kanal lokalnog oscilatora - 1
Najvisi radni kanal lokalnog oscilatora - 1
Radni kanal lokalnog oscilatora         - 2
Referentna ucestanost                   -      0
Radna ucestanost Tx/Rx                  -      0 /      0
MedjuUcestanost                         - 1008 MHz
Test rezim                              - NORMALNI

>sig


Alarm NON SYNC - ON
Alarm BER      - ON
Alarm Tx       - ON
Alarm Rx       - ON
Temperatura mikrokontrolera je:          0
AGC napon u IF grani (sirov) je:         0
AGC napon u IF grani (obradjen) je:      0
AGC napon u IQ grani je:                 0
```

Fig. 7. Start terminal look with two commands

Most information about peripherals and CCU itself can be readied from terminal as witch device is configured, does it work as higher or lower half. From terminal can be readied operate channel and corresponding frequency, intermediate

frequency, device serial number and is it set some test regime. Also, terminal mod can shown controller temperature, AGC voltages in both IQ and IF section and all others alarms. More, terminal can be used for setting commands to peripherals, as is changing channels. After any change to system, terminal output buffer is filling with data, so when main program come to terminal serve some warning will be shown on terminal. Start terminal look with two commands, command for viewing ODU information (INFO) and command for viewing ODU states (SIG), is shown on Fig. 7.. Primary function of terminal mode is for use in initial ODU setup and testing without using IDU.

IV. CONCLUSION

Central Control Unit is built as small complex unit. The hardware and software was designed to assign universal unit, what CCU is, since it is same for DRRS in different frequency bands.

Negative influences (noise, crosstalk, etc.) of central control unit to other ODU parts are minimized with different techniques such as relatively short control lines or inserting simple RC filter into lines.

New features like adaptive transmit power control, new test modes, can be added to ODU only with upgrade of CCU software or with minor hardware modifications.

At last, this central control unit is tested and running in many IMTEL digital radio relay systems.

REFERENCES

[1] Perić Miroslav, Perić Dragana, Obradović Dragan, Orlić Vladimir, "Realisation of Signal Processing at Intermediate Frequency and in Base Band in Digital Radio Relay Systems Imtel Komunikacije Series B", 52. conference ETRAN, Palić, Serbia, June 2008.

[2] C8051F120/1/2/3/4/5/6/7 – 100MIPS, 128 kB Flash, 12-bit ADC, 64-Pin Mixed-Signal MCU, Silison Labratories Products Inc., 4635 Boston Lane, Austin, TX 78735, 2004.

[3] *C8051F120/1/2/3/4/5/6/7 – Mixed Signal ISP Flash MCU Family,* Silison Labratories Products Inc., 4635 Boston Lane, Austin, TX 78735, 2004.

[4] Nemanja Mitrović, Dragan Obradović, Sinisa Tasić, Zeljko Gajić, „Frequency synthesizer control (DDS and PLL) based on SiLabs C8051F121 microcontroller“, TELFOR, Belgrade, 25-27. November 2008.

[5] J. VANKKA and K. A. I. HALONEN, “Direct Digital Synthesizers: Theory, Design and Applications”, Engineering and Computer Science Serial, Springer International, Hardcover, 2001.

[6] R. E. BEST, “Phase-Lock Loops – Design, Simulation and Application”, 5th ed., Ed. New York: McGraw-Hill, 2004.

[7] 400MSPS 14-bit, 1.8V CMOS Direct Digital Synthesizer – AD9951, Analog Devices Inc., 2008.

[8] *PLL Frequency Synthesizer ADF4108,* Analog Devices Inc., Norwood, MA 02062-9106, U.S.A., 2006-2007

[9] *Two-Wire Automotive Temperature Serial EEPROM,* Atmel Corporation, 2325 Orchard Parkway, San Jose, CA 95131, U.S.A. 2008.

[10] PCA9502 – 8-bit I/O expander with I^2C-bus/SPI interface, Philips Inc., Rev. 03, 2006.

[11] Miloje D. Zečević, Dragan D. Obradović, Branko Radan: "RRUNet Lite – New Software for Radio Relay Network Management and Supervision", XIV Telecommunication forum, Belgrade , 21.-23. November 2006.

Applications of 2.4 GHz Microwaves to WLAN Development at the University of Beira Interior Campus

José A. R. Pacheco de Carvalho[1], Hugo Veiga[2], Paulo A. J. Gomes[3], António D. Reis[4], and Cláudia F. F. P. R. Pacheco[5]

Abstract – **Wireless communications are developed and WLAN's created at the University Campus, including point-to-point and point-to-multipoint links to interconnect the Poles and buildings. Microwaves in the 2.4 GHz band and the IEEE 802.11b and 802.11g standards are used. The SNR quality of the links has been tested. It is shown how it can be experimentally improved. Several results are presented and discussed.**

Keywords – **Microwaves, antennas, SNR, WLAN, point-to-point links, point-to-multipoint links.**

I. INTRODUCTION

The Beira Interior University (UBI), located at Covilhã city, comprises four main Poles (I, II, III and IV), as well as medical pedagogical spaces at the regional Hospitals. The University network is a MAN. All the Poles, Faculties, Departments, Centres, Services, including all pedagogical spaces, are interconnected through the Informatics Centre (at Pole I) in a star topology for ~6000 users. Access is provided to the national academic network RCCN-RCTS/Internet at 20 Mbps, using IP over ATM. The Poles have been interconnected through dedicated circuits at low speeds up to 2 Mbps and IP routers. The e-UBI Project [1] has been developed, within the national initiative e-U Virtual Campus, to give users access to information in any place of the University Campus. WLAN's (Wireless LAN's) were created using the IEEE 802.11b and IEEE 802.11g standards, where microwaves in the 2.4 GHz band permit speeds up to 11 Mbps and 54 Mbps, respectively [2]. The LAN has been improved [3] to permit the required quality of the wireless network, enabling electronic devices to communicate along the whole University. The main networking equipments are Smart Switch Routers 8600, Matrix E7 and Matrix E1 level 2/3/4 switches with routing capabilities, from Enterasys Networks [4]. The main access networking equipments are level 2 Vertical Horizon switches and ~220 level 2/3/4 RoamAbout R2 Access Points (AP's) [4]. Both out-door and in-door WLAN's were developed for outside and inside of buildings, respectively. Two types of out-door WLAN's were necessary: point-to-point configurations, using two out-door directional antennas (and AP's in LAN-to-LAN endpoint bridge mode), and point-to-multipoint arrangements using an out-door omni directional antenna (for a central AP in LAN-to-LAN multipoint bridge mode) and out-door directional antennas (for AP's in endpoint bridge mode). Out-door point-to-point links were intended to interconnect the four main Poles, over medium distances (up to 5.5 km for the Pole II-Pole IV link, with a repeater), and buildings within the same Pole at 11 Mbps and 54 Mbps, for shorter distances. Out-door directional antennas of various types and manufacturers (14 dBi Yagi [4], 18 dBi planar [5], 21 dBi parabolic grid [5], 23 dBi parabolic grid with 2300-2600 MHz band pass [6]) were used (Table II). Fig.1 shows the general scheme of the links. Out-door point-to-multipoint links were established at Pole II to permit communications of the Staff Residence, the Old Residence, the Sports Building and the Yellow Residence with the Rectory Building, which was the Pole II central point for communications. The relative positions of these buildings are visible in Fig. 2. A 7 dBi omni directional out-door antenna [4] was mounted at the Staff Residence to communicate at 11 Mbps with out-door directional antennas of the 14 dBi Yagi type [4], located in each of the other buildings. The distances involved were 52-108 m. The in-door WLAN's, developed to complement the LAN's, covered the interior of the buildings. AP's were set to workgroup bridge mode. In-door omni directional antennas were useful for improving coverage. Starting from topography of the buildings, site surveys were made to optimize the locations of antennas, using a portable computer having an 802.11 a/b/g wireless card and Network Stumbler software [7] to measure signal to noise ratios SNR.

II. EXPERIMENTAL DEVELOPMENT AND RESULTS

A. Out-door Links

A communications channel has an important dependence of capacity per unit bandwidth, this is the channel capacity (C, in bps) to channel bandwidth (W, in Hz) ratio, on signal to noise ratio (S/N, power of the signal over power of noise). Shannon's formula gives an upper limit for this dependence, $C/W=\log_2(1+S/N)$, which represents channel efficiency [8].

[1]José Pacheco de Carvalho is with the Informatics Centre and the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: pacheco@ubi.pt.

[2]Hugo Veiga, and [3]Paulo Gomes are with the Informatics Centre and the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mails: hveiga@ubi.pt, pgomes@ubi.pt.

[4]António Reis is with the Remote Detection Unit at the University of Beira Interior, and with the Department of Electronics and Telecommunications/Institute of Telecommunications, at the University of Aveiro, 3810 Aveiro, Portugal,E-mail: adreis@ubi.pt.

[5]Cláudia Pacheco is with the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: a17597@ubi.pt.

TABLE I
1ST FRESNEL ZONE MAXIMUM DIAMETERS FOR THE LINKS BETWEEN POLES

Link	Distance (m)	1st Fresnel zone maximum diameter (m)
Pole I-Pole II	350	6.6
Pole II-Pole III	1500	13.7
Pole II-Repeater	2500	17.7
Repeater-Pole IV	3000	19.4

TABLE II
CHARACTERISTICS OF THE DIRECTIONAL AND OMNI DIRECTIONAL ANTENNAS

Parameters	Yagi, 14dBi (Enterasys)	Planar, 18 dBi, PW9618 (Phasak)	Parabolic grid, 21 dBi, PW9321 (Phasak)	Parabolic grid, 23 dBi, band pass, GD24BP-23 (Pacific Wireless)	Omni directional, 7 dBi, Roam About (Enterasys)
Frequency Range	2400-2485 MHz	2400-~2500 MHz	2400-~2500 MHz	2400-2485 MHz	2400-2485 MHz
Gain	14dBi	18 dBi	21 dBi	23 dBi	7 dBi
3 dB beam width, V or H	30.8º, 31.4ºH	17 º V, 22 º H	13 º	8 º V, 10 º H	---
Front to back ratio	> 20 dB	>= 25 dB	28 dB	>= 25 dB	---
VSWR	1.5	<1.5	<=1.4	1.5	<2
Max. Input Power	~5 W	50 W	100 W	100 W	---
Impedance	50 Ω	50 Ω	50 Ω	50 Ω	50 Ω
Dimensions	45.7 cm	37 x 25.2 x 4.2 cm	68 cm diameter	72.4 x 91.4 cm	45.7 cm

Therefore it was essential for the quality of wireless communications, to guarantee high *S/N* ratios. Point-to-point wireless links between two LAN's were set through a pair of equipments of the type switch/router, access point (AP) and out-door directional antenna. They were dimensioned so as to obtain clear, unobstructed line-of-sight paths between emitters and receivers. The Fresnel zones are ellipsoids whose foci are located at these points, having maximum diameters d_n at half the way. For a distance d along the link, d_n for the nth Fresnel zone is [9] $d_n = (n\lambda d)^{1/2}$, with $d >> d_n$, where $\lambda = 0.125$ m for $f =$ 2.4 GHz. It was taken into account the recommended clearance factor of at least 0.6 of the first Fresnel zone radius [10]. Given the topography of Covilhã city, which is very mountainous and has several valleys, strategically located buildings were chosen for installing the out-door antennas, considering the calculated values of Table I [11]. It was decisive to identify and minimize causes of electromagnetic interferences. The presence of various 2.4 GHz beams, existing in the urban and surrounding areas of Covilhã city, was detected through a portable computer having a 802.11 a/b/g wireless card, a directional out-door antenna and Network Stumbler software [7]. For the 13 ETSI frequency channels in the 2400-2485 MHz band, spacing of at least 3 channels were chosen to minimize interferences.

TABLE III
SNR RATIO FOR THE LINKS BETWEEN POLES AND POLE II BUILDINGS VERSUS TECHNOLOGY AND ANTENNA TYPE

Link	Distance (m)	Technology	Antenna type	(SNR) dB
Pole I-Pole II	350	IEEE 802.11b, 11 Mbps	Yagi, 14 dBi	11-26
Pole I-Pole II	350	IEEE 802.11g, 54 Mbps	Yagi, 14 dBi	0-18
Pole I-Pole II	350	IEEE 802.11b, 11 Mbps	Parabolic grid, 23 dBi, band pass	38-42
Pole I-Pole II	350	IEEE 802.11g, 54 Mbps	Parabolic grid, 23 dBi, band pass	37-38
Pole II-Pole III	1500	IEEE 802.11b, 2 Mbps	Yagi, 14 dBi	19-26
Pole II-Repeater	2500	IEEE 802.11b, 11 Mbps	Parabolic grid, 21 dBi + planar 18 dBi	22-28
Repeater-Pole IV	3000	IEEE 802.11b, 11 Mbps	Planar 18 dBi + parabolic grid, 21 dBi	20-35
Pole II-Repeater (with 2 x 1W amp.)	2500	IEEE 802.11b, 11 Mbps	Parabolic grid, 21 dBi + planar 18 dBi	45-51
Repeater-Pole IV (with 2 x 1W amp.)	3000	IEEE 802.11b, 11 Mbps	Planar 18 dBi + parabolic grid, 21 dBi	43-55
Pole II-Repeater (with 2 x 1W amp.)	2500	IEEE 802.11g, 24 Mbps	Parabolic grid, 21 dBi + planar 18 dBi	38-44
Rectory Building - Staff Residence	97	IEEE 802.11b, 11 Mbps	Yagi, 14 dBi + Omni directional, 7dBi	33
Staff Residence-Yellow Residence	52	IEEE 802.11b, 11 Mbps	Omni directional, 7dBi + Yagi, 14 dBi	30
Staff Residence- Old Residence	75	IEEE 802.11b, 11 Mbps	Omni directional, 7dBi + Yagi, 14 dBi	39
Staff Residence-Sports Building	108	IEEE 802.11b, 11 Mbps	Omni directional, 7dBi + Yagi, 14 dBi	30

Polarization matching was implemented and verified in detail. Other precautions were also taken in setting up the necessary equipments, including the construction of a repeater station at a relatively isolated location of the Cova da Beira valley, as the Poles II and IV were not in line-of sight [12]. The backbone of the LAN is Gigabit Ethernet over optical fibre (full duplex 1000-Base-SX and 1000-Base-LX). The WLAN uses Enterasys RoamAbout R2 AP's having RoamAbout 802.11b and 802.11 a/b/g 15 dBm radio cards. The AP's are connected through 100-Base-TX to level 2/3/4 switches capable of IPv4 routing, Enterasys Matrix E1 [4]. For reasons of performance, WEP (Wireless Equivalent Privacy) was not activated in the AP's. Initial measurements of signal to noise ratio SNR, have shown that long distances (determined by

locations of buildings) and electromagnetic interferences in urban areas (where the 2.4 GHz band is increasingly used) are particularly critical. Table II shows the characteristics of the directional antenna types chosen to overcome this problem and improve SNR [11]. The main SNR results, obtained by using the Enterasys RoamAbout AP Manager software [4], are presented in Table III. The best quality was for the 350 m Pole I-Pole II link at 54 Mbps and parabolic grid, 23 dBi, band pass antennas. Thus, a second link of this type was implemented to support the considerable traffic existing between both Poles. For long distance 2.5 km (Pole II-Repeater) and 3.0 km (Repeater-Pole IV) links, 14 dBi Yagi antennas and 11 Mbps did not give the required quality. The combination of 21 dBi parabolic grid and planar 18 dBi antennas at 11 Mbps, for these links, was essential and proved as very efficient. As shown later, its quality was further improved. Table II contains the characteristics of the 7 dBi omni directional antenna used for the Pole II point-to-multipoint wireless links. Again, the AP's were set without WEP activation. The same software as for the point-to-point links was used to measure SNR for every link. Table III shows good results at 11 Mbps, given the relatively short distances and antenna types. All the links mentioned have been additionally monitored through the MRTG (Multi Router Traffic Grapher) traffic measurement software tool [13]. Further tests over the links have been made [14]. They have improved quality of WLAN and LAN communications at the University.

B. Further developments for the Pole II-Pole IV link

In order to improve quality of the Pole II-Pole IV communications, each link Pole II-Repeater and Repeater-Pole IV was equipped with a pair of 1W (30 dBm) out-door, professional, bi-directional Wi-Fi 802.11b/g compatible amplifiers, available from Hyperlink Technologies [15]. They had a receive band pass filter, a low-noise receiver amplifier and a transmit power amplifier. Active power control was available for monitoring output power and providing maximum output power regardless of cable length. The amplifiers operated in bi-directional, half-duplex, time division duplex mode, having been designed for burst half-duplex operation. The aim was to achieve better SNR ratios for the links. The receive band pass characteristic eliminated the unwanted frequency band. Thus, the amplification acted on the band of interest. Figs. 3 and 4 show details of the new arrangement for the Pole II-Pole IV link. Table III shows the new results obtained at 11 Mbps and 24 Mbps, using 1W amplifiers. Details are given in Figs. 5 and 7. It is clear that the results are much better than those obtained without amplifiers. Figs. 5 and 6 permit comparison at 11 Mbps with and without amplifiers, respectively. Fig. 7 shows that 802.11g-24 Mbps is possible at high SNR values. It was also the case, not shown, of the Repeater-Pole IV link. The slight difference is that at 24 Mbps this link is not so stable, compared with the Pole II-Repeater link given the larger distance (3 km versus 2.5 km) and electromagnetic interference. Further tests over the links have been made [14].

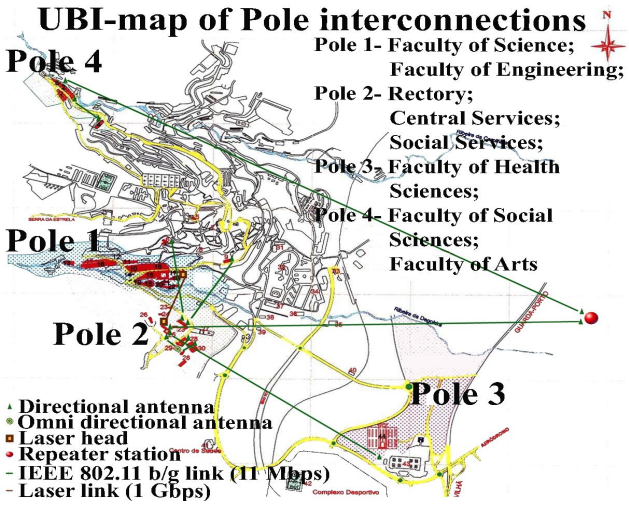


Fig. 1- Scheme of the wireless links between Poles and buildings of the University Campus.

Fig.2- Point-to-multipoint wireless links for Pole II buildings.

Health safety conditions in relation to Wi-Fi have been discussed [16]. The reference safety limit for long term exposure to 2.4 GHz electromagnetic fields is 61 V/m [17], which corresponds, in free-space, to a power density of 9.9 W/m2. This value, for free-space and far-field conditions, would be reached at 1 m and 0.71 m distances from the Pole II (and Pole IV) and the Repeater antennas, respectively. Still, these values are less than the corresponding Fraunhofer distances [9]. As power density varies with inverse square distance, for 10 m and 7.1 m (in the Fraunhofer regions) 0.1 W/m2 arise, which are 1% of the reference safety limit. The newly implemented link has permitted the flow of the increasingly growing traffic between Pole IV and Pole I.

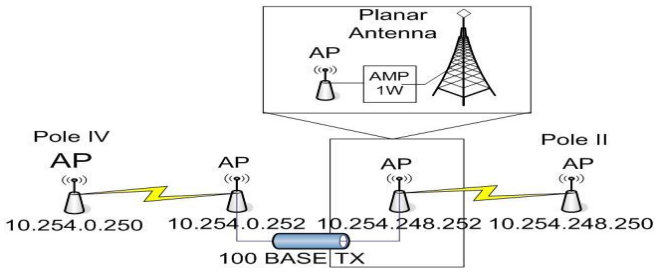


Fig.3- Scheme of the Pole II-Pole IV link, with 1W out-door amplifiers.

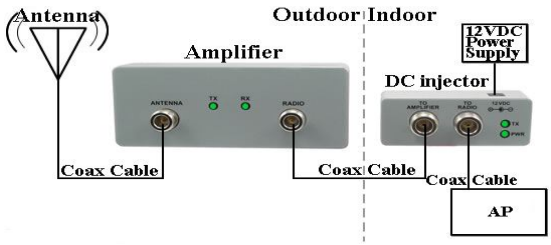


Fig.4- Out-door amplifier connection diagram [15].

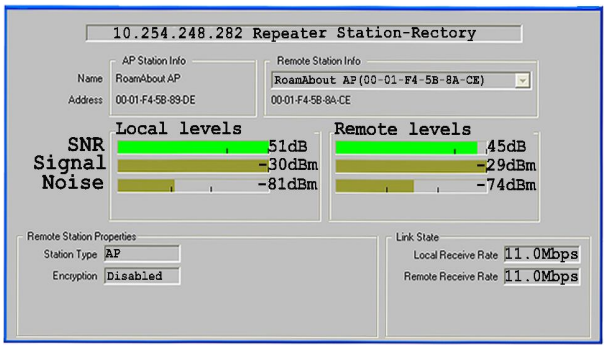


Fig.5- Pole II-Repeater link with 21 dBi parabolic grid and 18 dBi planar antennas, 2x1W amplifiers, 802.11b – 11 Mbps.

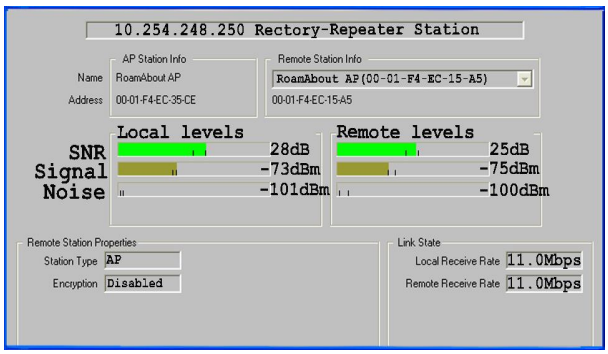


Fig.6- Pole II-Repeater link with 21 dBi parabolic grid and 18 dBi planar antennas, no amplifiers, 802.11b – 11 Mbps.

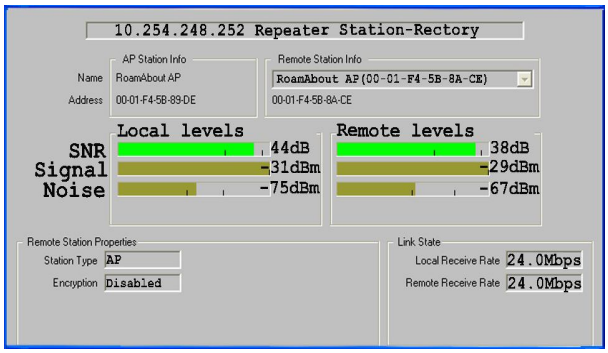


Fig.7- Pole II-Repeater link with 21 dBi parabolic grid and 18 dBi planar antennas, 2x1W amplifiers, 802.11g – 24 Mbps.

III. Conclusion

New wireless links, both point-to-point (including long distances) and point-to-multipoint have been developed and successfully tested at the University Campus. The new developed wireless infrastructure has been important for complementing the wired network. It has been important for electronic devices to communicate along the whole University Campus. Further work is in progress to further improve the University communications.

Acknowledgement

Supports from University of Beira Interior, POSC, UMIC and FCT/POCI2010 are acknowledged.

References

[1] Centro de Informática da Universidade da Beira Interior, "Projecto de Candidatura Institucional à Iniciativa Campus Virtuais: e-UBI (Universidade da Beira Interior)", (Proj. Rep. 30/4/2003), Universidade da Beira Interior, Covilhã, Portugal, Abril de 2003.

[2] http://standards.ieee.org.

[3] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, P. M. G. Fonseca, H. Veiga, N. Sampaio, A. D. Reis, "A Rede Geral de Informática da Universidade da Beira Interior - Desenvolvimento e Perspectivas", in Proc. Engenharia'2005, , pp. 267-272, Universidade da Beira Interior, Covilhã, Portugal, Novembro de 2005.

[4] http://www.enterasys.com.

[5] http://www.phasak.com.

[6] http://www.pacwireless.com.

[7] http://www.stumbler.net.

[8] C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press, 1949 (reprinted 1998).

[9] T. S. Rappaport, *Wireless Communications Principles and Practice* (2nd ed.). Upper Sadle River, NJ: Prentice-Hall, Inc., 2002, ch. 4.

[10] R. L. Freeman, *Telecommunications Transmission Handbook* (4th ed.). NY: John Wiley & Sons, Inc., 1998, ch. 5.

[11] J. A. R. Pacheco de Carvalho, A. D. Reis, P. A. J. Gomes, H. Veiga, "Experimental Development of Wireless LAN's at the Beira Interior University Campus", in Proc. SEON2006-IVth Symposium on Enabling Optical Networks and Sensors, pp. 63-64, FEUP-UP, Porto, Portugal, June 2006.

[12] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, H. Veiga, A. D. Reis, N. Sampaio, P. M. G. Fonseca, "Desenvolvimento Experimental de Ligações Sem-Fios Ponto-a-Ponto Entre os Pólos da Universidade da Beira Interior", in Proc. Engenharia'2005, pp. 289-294, Universidade da Beira Interior, Covilhã, Portugal, Novembro de 2005.

[13] http://mrtg.hdl.com.

[14] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, H. Veiga, A. D. Reis, "Development of a University Networking Project", in Encyclopedia of Networked and Virtual Organizations, Goran D. Putnik, Maria Manuela Cunha, Eds. Hershey, PA (Pennsylvania): IGI Global, pp. 409-422, 2008.

[15] http://www.hyperlinktech.com.

[16] T. M. Crespo, C. Salema, "Design and Implementation of Wi-Fi Networks: Remarks on an Experience", in Proc. 5th Conference on Telecommunications, a1095, cr1095 Tomar, Portugal, April 2005.

[17] Portaria nº 1429/2004, Diário da República-I Série-B, nº 275, pp. 6834-6838, 23 Novembro de 2004.

Performance Measurements of IEEE 802.11 a, g Laboratory Point-to-Point Links

José A. R. Pacheco de Carvalho[1], Paulo A. J. Gomes[2], Hugo Veiga[3],

Cláudia F. F. P. Ribeiro Pacheco[4], Nuno Marques[5], António D. Reis[6]

Abstract – **The importance of wireless communications has been growing. Performance is a very relevant issue, leading to more reliable and efficient communications. Laboratory measurements are made about several performance aspects of Wi-Fi (IEEE 802.11 a, g) point-to-point links. A contribution is given to performance evaluation of this technology, using available access points from Enterasys Networks (RBTR2). Detailed results are presented and discussed, namely at OSI levels 4 and 7, from TCP, UDP and FTP experiments: TCP throughput, jitter, percentage datagram loss and FTP transfer rate.**

Keywords – **WLAN, Wi-Fi, Point-to-Point Links, Wireless Network Laboratory Performance Measurements.**

I. INTRODUCTION

Wireless communications are increasingly important for their versatility, mobility and favorable prices. It is the case of microwave based technologies, e.g. Wi-Fi. The importance and utilization of Wi-Fi have been growing for complementing traditional wired networks. Wi-Fi has been used both in ad hoc mode and infrastructure mode. In this case an access point, AP, is used to permit communications of Wi-Fi devices with a wired based LAN through a switch/router. In this way a WLAN, based on the AP, is formed. Wi-Fi has reached the personal home, forming a WPAN, allowing personal devices to communicate. Point-to-point and point-to-multipoint configurations are used both indoors and outdoors, requiring specific directional and omnidirectional antennas. Wi-Fi uses microwaves in the 2.4 and 5 GHz frequency bands and IEEE 802.11a, 802.11b and 802.11g standards [1]. As the 2.4 GHz band becomes increasingly used and interferences increase, the 5 GHz band has received considerable interest, although absorption increases and ranges are shorter.

Nominal transfer rates up to 11 (802.11b) and 54 Mbps (802.11 a, g) are permitted. CSMA/CA is the medium access control. Wireless communications, wave propagation [2,3] and WLAN practical implementations [4] have been studied. Detailed information is available about the 802.11 architecture, including performance analysis of the effective transfer rate. An optimum factor of 0.42 was presented for 11 Mbps point-to-point links [5]. Wi-Fi (802.11b) performance measurements are available for crowded indoor environments [6]. Performance has been a very important issue, resulting in more reliable and efficient communications. New telematic applications present special sensitivities to performances, when compared to traditional applications. Application characterization and requirements have been discussed for cases such as voice, Hi Fi audio, video on demand, moving images, HDTV images, virtual reality, interactive data, static images, intensive data, supercomputation, electronic mail, and file transfer [7]. E.g. requirements have been quoted as: for video on demand/moving images, 1-10 ms jitter and 1-10 Mbps throughput; for Hi Fi stereo audio, jitter less than 1 ms and 0.1-1 Mbps throughputs.

Several performance measurements have been made for 2.4 GHz Wi-Fi [8], as well as WiMAX and high speed FSO [9,10]. In the present work further Wi-Fi (IEEE 802.11 a,g) results arise, through OSI levels 4 and 7. There is interest in comparing two technologies working in the 5 GHz and 2.4 GHz bands, as at 2.4 GHz interferences have grown up. Performance is evaluated in laboratory measurements of point-to-point links using available equipments.

The rest of the paper is structured as follows: Chapter II presents the experimental details i.e. the measurement setup and procedure. Results and discussion are presented in Chapter III. Conclusions are drawn in Chapter IV.

II. EXPERIMENTAL DETAILS

The measurements used Enterasys Networks RBTR2 level 2/3/4 access points [11], equipped with IEEE 802.11 a/b/g radio cards similar to the Agere-Systems model 0118 type, and firmware version 6.08.03, and 100-Base-TX/10-Base-T Allied Telesis AT-8000S/16 level 2 switches [12]. The configuration of the access points was for minimum transmitted power i.e. micro cell, point-to-point, LAN to LAN mode, using the antenna which was built in the card. Interference free communication channels were used in the communications. WEP encryption was not activated. No

[1]José Pacheco de Carvalho is with the Informatics Centre and the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: pacheco@ubi.pt.

[2]Paulo Gomes, [3]Hugo Veiga, and [5]Nuno Marques are with the Informatics Centre and the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mails: pgomes@ubi.pt, hveiga@ubi.pt, nmarques@ubi.pt.

[4]Cláudia Pacheco is with the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: a17597@ubi.pt.

[6]António Reis is with the Remote Detection Unit at the University of Beira Interior, and with the Department of Electronics and Telecommunications/Institute of Telecommunications, at the University of Aveiro, 3810 Aveiro, Portugal, E-mail: adreis@ubi.pt.

power levels above the minimum were required, as the access points were very close (30 cm).

The laboratory setup is shown in Fig. 2. TCP and UDP experiments at OSI level 4, were as mentioned in [10], permitting network performance results to be recorded. Both TCP and UDP are transport protocols. TCP is connection-oriented. UDP is connectionless, as it sends data without ever establishing a connection. For a TCP connection, TCP throughput was obtained. For a UDP communication with a given bandwidth parameter, UDP throughput, jitter and percentage loss of datagrams were obtained. TCP packets and UDP datagrams of 1470 bytes size were used. A window size of 8 kbytes and a buffer size of the same value were used for TCP and UDP, respectively. One PC, with IP 192.168.0.2 was the Iperf server and the other, with IP 192.168.0.6, was the Iperf client. Jitter, which represents the smooth mean of differences between consecutive transit times, was continuously computed by the server, as specified by RTP in RFC 1889 [13]. This scheme was also used for FTP measurements, where FTP server and client applications were installed in the PCs with IPs 192.168.0.2 and 192.168.0.6, respectively.

The PCs were portable computers running Windows XP. They were configured to maximize the resources allocated to the present work. Also, batch command files were written to enable the TCP, UDP and FTP tests. The results were obtained in batch mode and written as data files to the client PC disk.

III. Results and Discussion

The access points were configured, for each standard IEEE 802.11 a, g, with typical fixed transfer rates. For every fixed transfer rate, data were obtained for comparison of the laboratory performance of the links, measured namely at OSI levels 4 and 7 using the setup of Fig. 2. For each standard and every nominal fixed transfer rate, an average TCP throughput was determined. This value was used as the bandwidth parameter for every corresponding UDP test, giving average jitter and average percentage datagram loss. The main results are shown in Figs. 3-6.

In Fig. 3, polynomial fits were made to the TCP throughput data, where R^2 is the coefficient of determination. It is seen that the best TCP throughputs are for 802.11a. The average values are 12.92 and 12.60 Mbps for 802.11a and 802.11g, respectively. In Figs. 4 and 5, the data points representing jitter and percentage datagram loss were joined by smoothed lines. In Fig. 4, the jitter data are on average lower for IEEE 802.11g (1.8 ms) than for 802.11a (2.1 ms). Fig. 5 shows that, generally, the percentage datagram loss data (1.2 % on average) agree reasonably well for both standards.

At OSI level 7 we measured FTP transfer rates versus nominal transfer rates configured in the access points for the IEEE 802.11a, g standards. Every measurement was the average for a single FTP transfer, using a binary file size of 100 Mbytes. The results thus obtained are represented in Fig. 6. Polynomial fits to data were made for the implementation of each standard. It was found that the best FTP performance was for 802.11a.

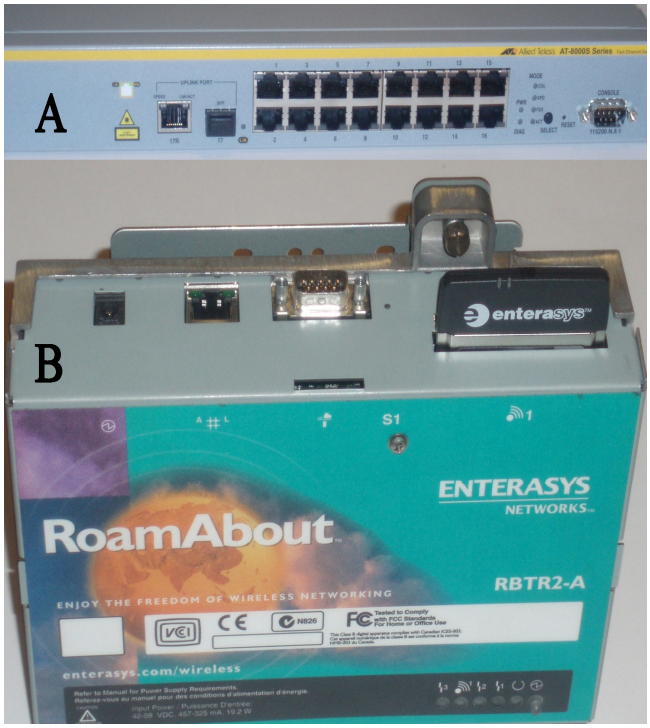


Fig. 1- Switch (A) [12] and Access Point (B) [11].

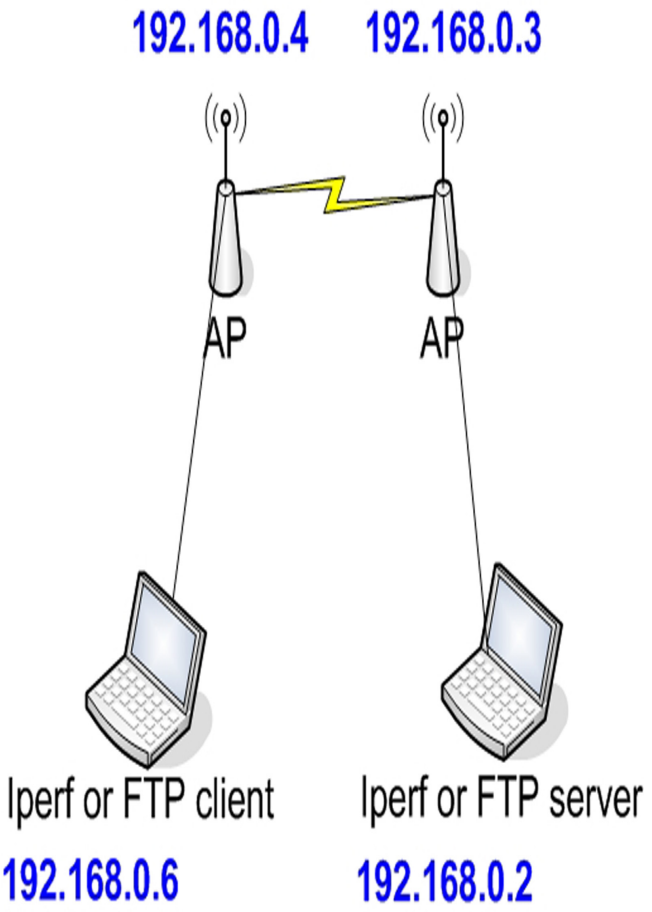


Fig.2- Laboratory setup scheme.

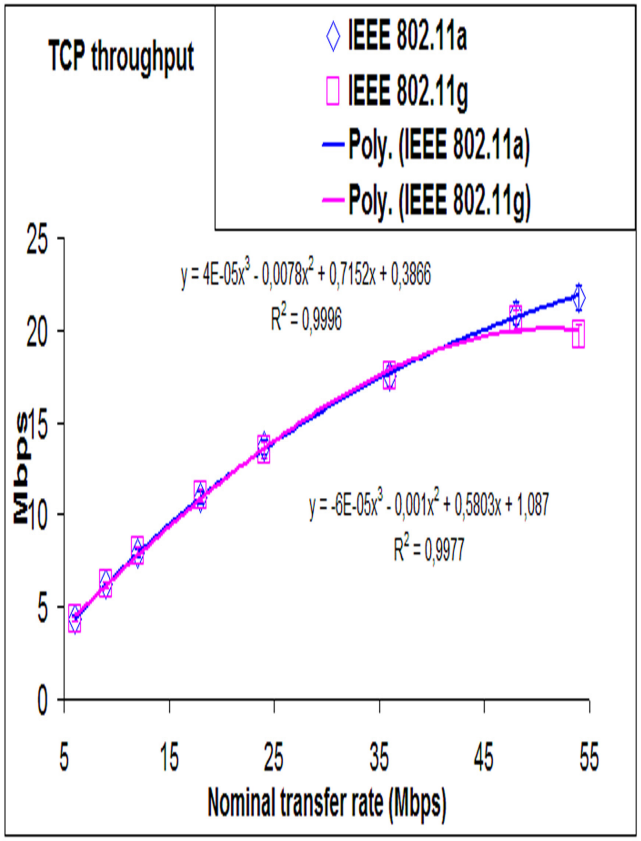


Fig.3- TCP throughput versus technology and nominal transfer rate.

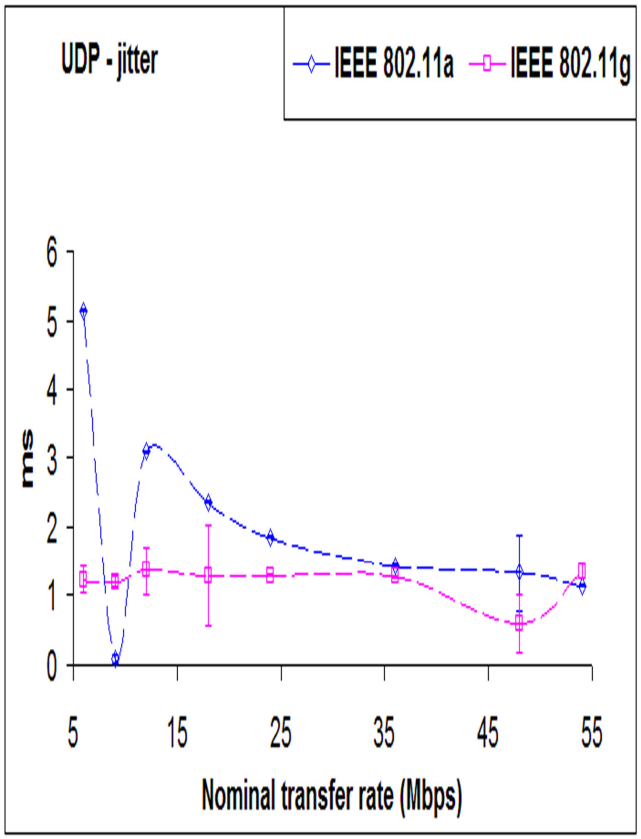


Fig.4- UDP – jitter results versus technology and nominal transfer rate.

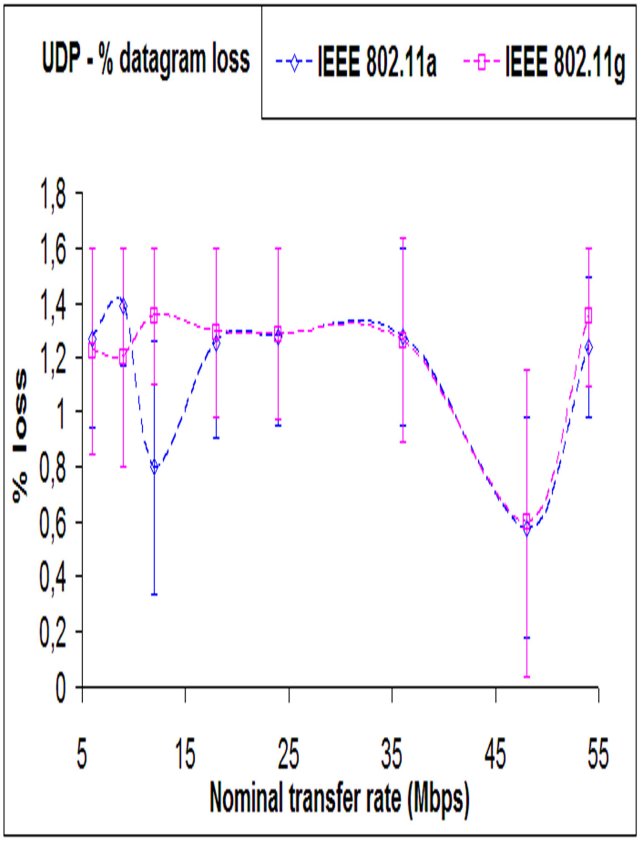


Fig.5- UDP – percentage datagram loss results versus technology and nominal transfer rate

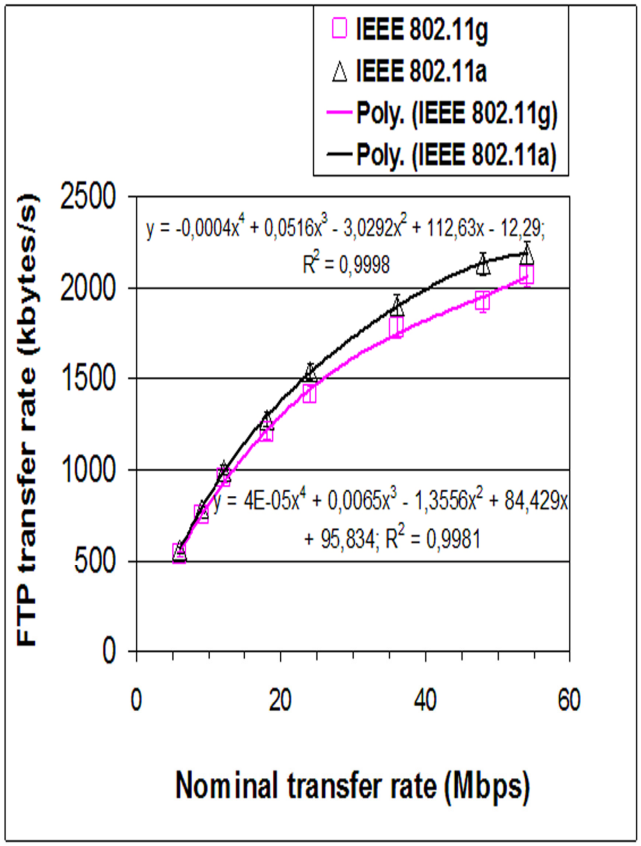


Fig.6- FTP transfer rates versus technology and nominal transfer rate.

IV. CONCLUSION

In the present work a simple laboratory setup was planned that permitted systematic performance measurements of available access point equipments (RBTR2 from Enterasys) for Wi-Fi (IEEE 802.11a, g) in point-to-point links. Through OSI level 4, TCP throughput, jitter and percentage datagram loss were measured and compared for each standard. The best TCP throughput was found for 802.11a. The average jitter was lower for 802.11g. For the percentage datagram loss, a reasonably good agreement was found for both standards. At OSI level 7, the best FTP performance was for 802.11a. Additional performance measurements either started or are planned using several equipments, not only in laboratory but also in outdoor environments involving, mainly, medium range links.

ACKNOWLEDGEMENT

Supports from Universidade da Beira Interior and FCT (Fundação para a Ciência e a Tecnologia)/POCI2010 (Programa Operacional Ciência e Inovação) are acknowledged. We acknowledge Enterasys Networks for their collaboration.

REFERENCES

[1] Web site http://standards.ieee.org
[2] J. W. Mark, W. Zhuang, *Wireless Communications and Networking*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2003.
[3] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed., Prentice-Hall, Inc., Upper Saddle River, NJ, 2002.
[4] W. R. Bruce III, R. Gilster, *Wireless LANs End to End*, Hungry Minds, Inc., NY, 2002.
[5] M. Schwartz, *Mobile Wireless Communications*, Cambridge University Press, 2005.
[6] N. Sarkar, K. Sowerby, "High Performance Measurements in the Crowded Office Environment: a Case Study", In Proc. ICCT'06-International Conference on Communication Technology, pp. 1-4, Guilin, China, 27-30 November 2006.
[7] E. Monteiro, F. Boavida, *Engineering of Informatics Networks*, 4th ed., FCA-Editor of Informatics Ld., Lisbon, 2000.
[8] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, H. Veiga, A. D. Reis, "Development of a University Networking Project", in Encyclopedia of Networked and Virtual Organizations, Goran D. Putnik, Maria Manuela Cunha, Eds. Hershey, PA (Pennsylvania): IGI Global, pp. 409-422, 2008.
[9] J. A. R. Pacheco de Carvalho, H. Veiga, R. Costa, P. A. J. Gomes, A. D. Reis, "A Contribution to Experimental Performance Evaluation of Point-to-Point WiMAX Links", In Proc. ISSPIT 2008-8th IEEE International Symposium on Signal Processing and Information Technology, pp. 150-153, Sarajevo, Bosnia and Herzegovina, December 16-19, 2008.
[10] J. A. R. Pacheco de Carvalho, H. Veiga, P. A. J. Gomes, C. F. F. P. Ribeiro Pacheco, A. D. Reis, "Experimental Performance Study of a Very High Speed Free Space Optics Link at the University of Beira Interior Campus: a Case Study", In Proc. ISSPIT 2008-8th IEEE International Symposium on Signal Processing and Information Technology, pp. 154-157, Sarajevo, Bosnia and Herzegovina, December 16-19, 2008.
[11] Web site http://www.enterasys.com
[12] Web site http://www.alliedtelesis.com
[13] Network Working Group, "RFC 1889-RTP: A Transport Protocol for Real Time Applications", http://www.rfc-archive.org

Application of above 1Gbit/s millimeter wave radio links for realization of IP networks in urban areas

Perić Dragana[1], Perić Miroslav[2], Šunjevarić Milan[3]

Abstract. **In this paper network planning example is given for IP network realized with high capacity radio links at millimeter wave frequency band 80 GHz. Using real equipment characteristics taken from links currently available in the market, and network planning procedure, interference and unavailability parameters are calculated and discussed.**

Keywords: **Millimeter wave link, Network planning, Interference, Unavailability**

I. INTRODUCTION

Fixed radio links operating in the 71-76 GHz and 81-86 GHz bands are good solution for the future market demands for increasingly high bandwidth access, in particular for Internet-based applications. They could be deployed much quicker and in certain cases are more cost efficient than the wired networks, and as such these bands provide sufficient bandwidth for terrestrial fixed links to compete or complement the fiber optic based access networks [1].

In the proposed scenario of using the 71-76 GHz and/or 81-86 GHz band for Fixed Services, availability objectives in the order of 99.99% with the average European rain rates may be satisfied by very high capacity (up to 10 Gbit/s) links with some 1-2 km hop lengths (line-of-sight conditions); There is also slight attenuation variation between the two bands (71-76 GHz and 81-86GHz), which make possible their paired use. These systems would allow a rapid and effective deployment of broadband capacity in areas where fiber optic cables are not available or are not cost effective. The main features of operating fixed radio systems in this part of the spectrum may be summed up as follows:

- Availability of wide bandwidths, allowing for the low cost of traffic
- Possibility of multiple channel frequency re-use, thanks to highly directional antenna beams
- Feasibility of deploying radio links is much easier in comparison to alternative wire-bound solutions
- Ability to ensure high security because of low possibility of interference/capture of signals

The use of the spectrum between 70 to 100 GHz is the only viable solution for fixed links to achieve the above objectives.

In following, network planning example is given in network that is realized with 80 GHz radio links with capacity 1 Gbit/s. Links connect IP routers settled in nodes of this network. Attention is paid to the main network planning issues: fulfilling availability objectives and interference influence.

II. NETWORK PLANNING EXAMPLE

A. Millimeter wave links

As an example network we decided to use characteristics of gigabit Bridgewave millimeter wave digital radio AR80 Adapt Rate at 80 GHz frequency band (Table 1) [2] which completely European standards for millimeter wave equipment channel allocation [1], characteristics [3] and antennas [4]. These devices have adaptive rate and can operate at 1Gbit/s or at 100Mbit/s during the rainfalls when received signal level becomes too low.

TABLE 1. EQUIPMENT CHARACTERISTICS

Manufacturer	Bridgewave, USA
Device	AR80
Frequency band	71-74 paired with 81-84GHz
Bit rate	1Gbit/s (adaptive to 100Mbit/s)
Modulation type	DBPSK
FEC	Reed Solomon (204,188)
Hop gain (BER=1E-12)	for 0.3m antenna 166dB (177dB) for 0.6m antenna 180dB (191dB)
Output power	17dBm
Receiver threshold (BER=1E-12)	-63dBm (-74dBm)
Central frequency	72.5 / 82.6GHz
RF bandwidth	1400MHz (140MHz)
Antenna gain	43dBi (0.3m), 50dBi (0.6m)
Adaptive threshold	hysteresis -59/-57dBm

As we can see in Table 1, devices could operate at single Rx/Tx frequency pair and in normal operation they occupy 1400MHz of RF spectra. As a consequence of usage of very powerful FEC we have very small difference between receiver thresholds with different throughputs. Therefore, entire device operation have three states: 1Gbit/s, 100Mbit/s and unavailable without BER degradation.

[1]Perić Dragana is with the Imtel Komunikacije a.d., Bulevar Mihajla Pupina 165B, 11070 Belgrade, Serbia, e-mail: dragana@insimtel.com.

[2]Perić Miroslav is with the Imtel Komunikacije a.d., Bulevar Mihajla Pupina 165B, 11070 Belgrade, Serbia, e-mail: micha@insimtel.com.

[3]Šujevarić Milan is with the Imtel Komunikacije a.d., Bulevar Mihajla Pupina 165B, 11070 Belgrade, Serbia, e-mail: micosun@ptt.rs.

According to [4] antennas have considerable gain, and very good side lobes suppression (Fig 1.) which are more than 35dB when angle of arrival is higher than 5°.

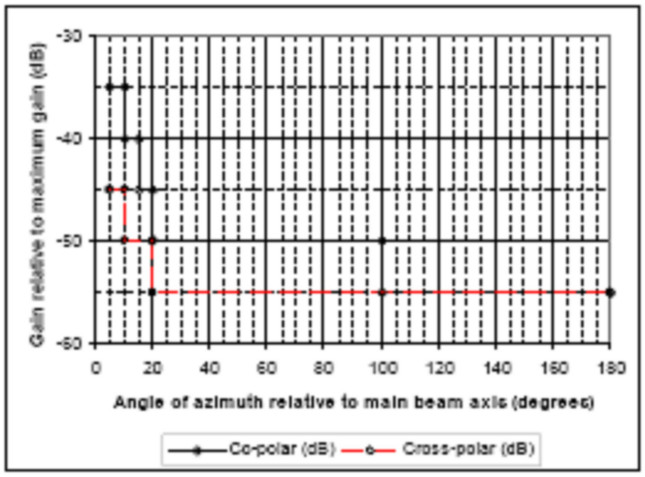


Fig 1. Antenna radiation pattern according to EN302217

B. Network area, topology and interference

Classical approach taken from lower bands [5] suggest that on the same network node all transmitters should operate at the same sub-band to prevent interference in near zone and third order intermodulation. Since there is only single frequency pair available, the crucial in network planning is that on the same node all transmitters should operate at the same frequency. When taking into account line of sight condition, this gives us serious restrictions in selecting possible hops.

For illustration planning of small network is presented. Network consists of 11 nodes and 14 links. Nodes are chosen to be located on the highest buildings in Novi Beograd and Zemun, and therefore network is from "real world" (Fig. 2.). On this area of the city live about 400000 people. Each node should be connected to the at least other two nodes, so there should be at least one the backup route.

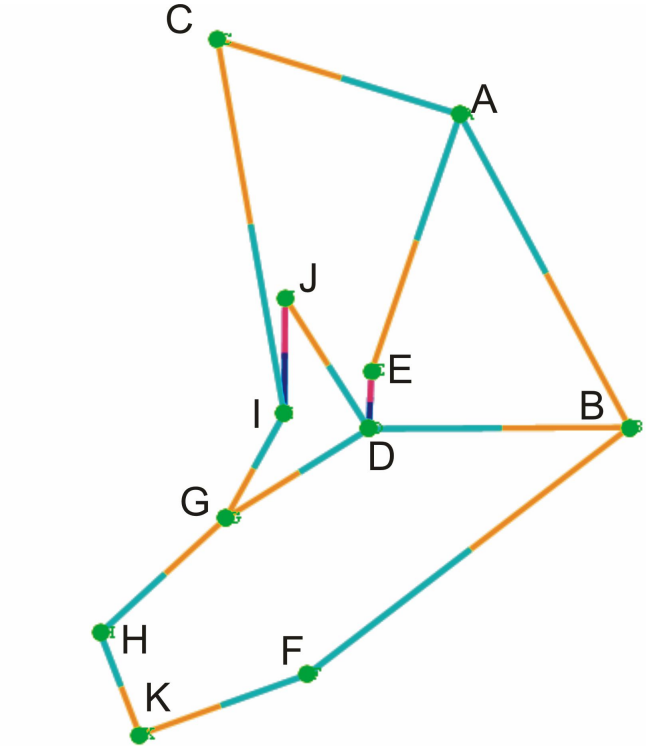


Fig.3. Network topology

One possible solution is given in Fig. 3. where nodes A,

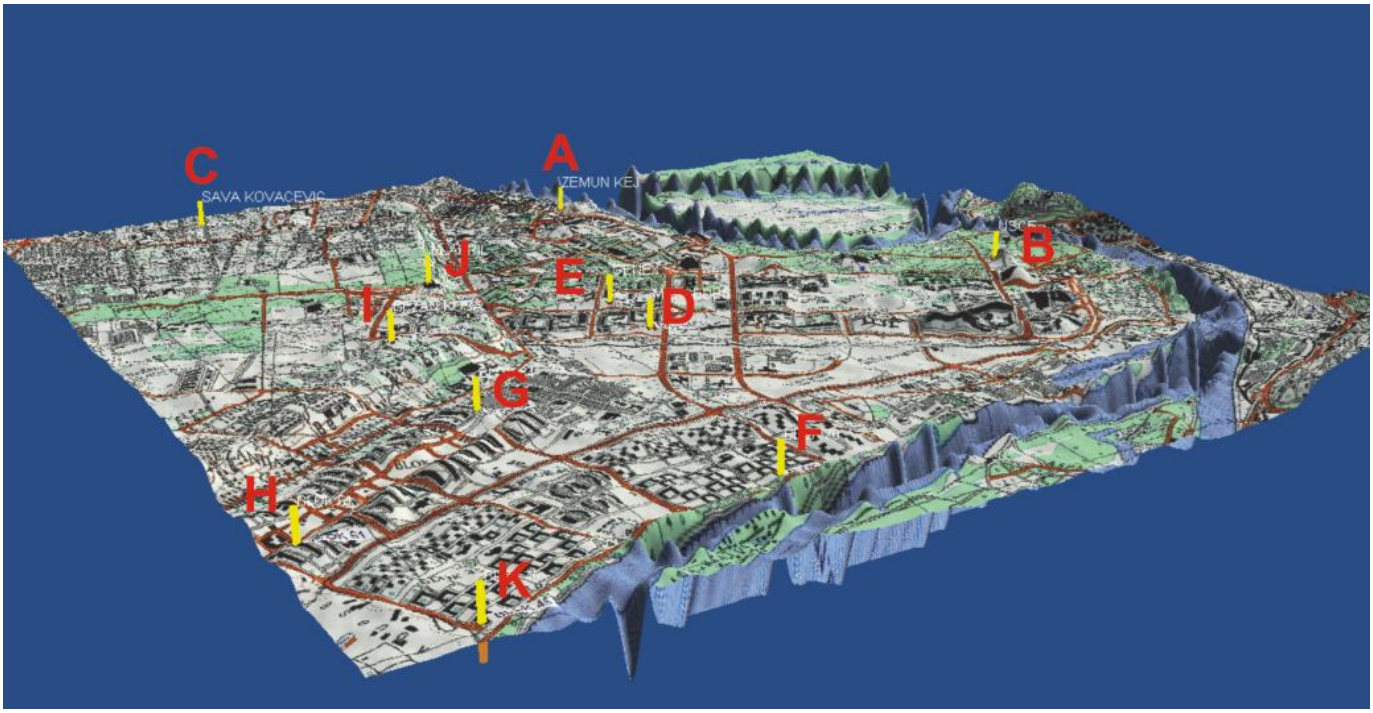


Fig. 2. Node locations on Novi Beograd and Zemun

D, F, H and I have transmitters at 72.5GHz while the rest have transmitters at 82.5GHz. The short hop E-D, on which both nodes are located on very high buildings, has very important role in solving these problem. However short hop is potentially very strong interference source and it should operate at different polarization from the rest of the network.

Classical interference calculation that takes into account only antenna radiation patterns in azimuth plane and 3D terrain model without buildings [6] in this 80GHz urban case gives very pessimistic calculation. However it could be used for the first approximation, since if networks operate in this model they will certainly operate in practice. Such calculation shows that I-J hop is also very strong interferer, and should operate at different polarization.

TABLE 2. INTERFERENCE CALCULATION

General					Interference			Strongest interferer		
Rx	Tx	d (km)	n_{r0} (dBm)	a_{FM} (dB)	I (dBm)	a_{FMI} (dB)	Δa (dB)	Tx	Rx	I_i(dBm)
F	B	3.8	-38.8	24.2	-80.9	20	4.2	K	F	-82.9
B	F				-82.3	20.8	3.4	D	B	-85.5
I	C	3.3	-37.6	25.4	-75.3	17	8.4	J	I	-77.2
C	I				-83.6	22.7	2.7	A	C	-85.1
A	B	3.1	-37	26	-80.2	21.4	4.6	E	A	-84.8
B	A				-81.4	22.1	3.9	D	B	-85.5
D	B	2.6	-35.5	27.5	-74	18	9.5	E	D	-76.2
B	D				-81.6	23.7	3.8	A	B	-87
C	A	2.5	-35.1	27.9	-81.9	24.3	3.6	A	E	-85.1
A	C				-80.9	23.7	4.2	E	A	-84.8
E	A	2.4	-34.8	28.2	-75.1	19.6	8.6	D	E	-76.2
A	E				-79	22.7	5.5	G	I	-82.4
F	K	1.7	-31.8	31.2	-81	27.1	4.1	K	H	-81.8
K	F				-76.2	23.6	7.6	H	K	-77.2
H	G	1.6	-31.3	31.7	-77.3	25	6.7	G	D	-81.9
G	H				-76.8	24.6	7.1	I	G	-79.2
G	D	1.6	-31.3	31.7	-74.6	22.7	9	I	G	-78
D	G				-73	21.3	10.4	E	D	-76.2
J	D	1.4	-30.1	32.9	-76.9	25.8	7.1	D	G	-80.1
D	J				-74.8	24.1	8.8	E	D	-76.2
I	G	1.1	-28	35	-74.6	26	9	G	D	-78
G	I				-75.5	26.8	8.2	I	C	-80.8
H	K	1	-27.2	35.8	-76.7	28.6	7.2	K	F	-77.2
K	H				-78.9	30.3	5.5	F	K	-81.8
I	J	1	-27.2	35.8	-81.6	32	3.8	G	I	-83
J	I				-84.6	33.5	2.3	D	J	-85.1
E	D	0.5	-21.2	41.8	-88	40.6	1.2	A	E	-89.8
D	E				-81.5	38	3.8	J	D	-85.1

In Table 2., results of simplified interference calculation are given. We can notice that the strongest interference comes from hops that share the same node.

As expected, hops J-I and C-I are critical due to small angle between routes, and receiver threshold degradation is very high, without influence of different polarizations. This would have very strong influence on C-I hop performance due to its significant length. Similar situation happens on node D and node G.

Interference at hops that share the same node could be reduced by careful installation of antenna system, with usage of additional protection at antenna tower of building like wall or metal shield. As we notice from table 2. around 10dB of additional attenuation is sufficient.

Next hop overreach case like H-G-D-B in this topology is present, but less significant. When effects of urban area are taken into account (Fig 2), next hop overreach hardly could happen. Considering total amount of interference and the strongest interferer, we see that it frequently differs for more than 2dB, that means that more than one significant interference source effect is very present in this network. In this case area coverage by buildings and trees could be of great help since they provide additional attenuation in this cases.

We must notice that only azimuth antenna radiation pattern is used, and not 3D model. Such additional attenuation due to different antenna elevations unfortunately could not be much significant. Only for very short hops this could give significant attenuation of more than 10dB (e.g. on 0.5km hop antenna height difference of about 30m gives about 3.5° which is for sure more than 10dB of additional attenuation).

Therefore real link performance should be calculated for both cases with and without interference, and in many cases link behavior would be much closer to without interference case. Therefore light licensing method is welcome. Some countries have already adopted it.

C. *Network performance*

We can notice that receiver level in absence of fading n_{r0} is much higher than usually at lower bands, and fading margins without interference is considerably high from 24 to more than 40dB.

Classical method for hop availabilities calculation is described in [8]. We assumed rain rate 42mm/h to be exceeded in 0.01% of time, which is according to [7] the worst case for Serbian territory. At that rain rate specific attenuations are 16.4 dB/km and 15.1 dB/km for horizontal and vertical polarization at 72.5GHz and 17.1dB/km and 15.9dB/km at 82.5GHz. For both frequencies atmosphere attenuation is slightly less than 0.5dB/km. Under these assumptions we calculated hop availabilities in both directions for individual links.

TABLE 2.PERCENTAGE OF UNAVAILABLE TIME PER HOPS AND DIRECTIONS

Node		Without interference		With interference	
1	2	2→1	1→2	2→1	1→2
F	B	5.96E-02	5.30E-02	9.12E-02	7.46E-02
I	C	4.06E-02	3.59E-02	1.00E-01	4.67E-02
A	B	3.38E-02	2.99E-02	5.33E-02	4.39E-02
D	B	2.03E-02	1.79E-02	5.59E-02	2.58E-02
C	A	1.58E-02	1.80E-02	2.23E-02	2.70E-02
E	A	1.40E-02	1.60E-02	3.45E-02	2.74E-02
F	K	5.26E-03	4.53E-03	7.80E-03	9.86E-03
H	G	4.27E-03	3.67E-03	8.35E-03	7.57E-03
G	D	3.67E-03	4.27E-03	9.43E-03	1.28E-02
J	D	2.25E-03	2.64E-03	4.67E-03	6.57E-03
I	G	1.03E-03	8.61E-04	2.72E-03	2.10E-03
H	K	6.91E-04	5.71E-04	1.50E-03	1.04E-03
I	J	8.97E-04	7.73E-04	1.32E-03	9.78E-04
E	D	1.34E-05	1.83E-05	1.67E-05	3.45E-05

Without interference condition of less than 0.05% unavailability time (ITU-T G.826, access level) is satisfied for almost all hops shorter than about 3km. In these cases operation at lower bit rate (100Mbit/s), in order to reduce receiver threshold, would happen very rarely (less than one hour annually).

For longer hops and in presence of interference such cases would happen much frequent, and therefore usage of adaptive modulation and bit rate is highly recommended.

We also should note that since the network is used for IP traffic there is possibility of usage of path diversity [5] and routers could act as protection devices. However all commonly used routing protocols like RIP, OSPF and EIGRP does not have fast enough reaction time [9], and there is space for more improvements [10].

III. CONCLUSION

Described planning and analysis of hypothetical network at urban area shows that there is great potential of usage 80GHz band for very cost effective method for realization of Gbit/s IP network. The most important thinks that network planner should take care is interference at hops that share the same node. Despite of very high antenna side lobes attenuation, high interferer level is still very present, and additional counter-measurements like usage of buildings concrete blocks as obstacle between interfering link antennas.

Performances of hops shorter than about 3km satisfies ITU-T access level requirements without additional protection method like adaptive bit rate and/or modulation. For longer hops such methods are highly recommended.

Addition performances could be gain at network level by path diversity, for which IP routers, running link state routing protocols, could acts as protection devices.

ACKNOWLEDGEMENT

This work is partially financed by Serbian Ministry for Science and Technology. Special thanks to Mr Steve Odell and Mr Jim Mc Gowan from BridgeWave corporation for data about millimeter wave devices.

REFERENCES

[1] ECC Recommendation (05)07:"Radio frequency channel arrangements for fixed service systems operating in the bands 71-76 GHz and 81-86 GHz, 2009.

[2] Bridgewave AR80 Datasheet, www.bridgewave.com

[3] EN 302 217-4-2 V1.3.1 (2007-10), "Fixed Radio Systems; Characteristics and requirements for point-to-point equipment and antennas; Part 4-2: Harmonized EN covering essential requirements of Article 3.2 of R&TTE Directive for antennas", ETSI Standard, 2007.

[4] ETSI TS 102 524 V1.1.1 (2006-07) "Fixed Radio Systems; Point-to-Point equipment; Radio equipment and antennas for use in Point-to-Point Millimeter wave applications in the Fixed Services (mmwFS) frequency bands 71 GHz to 76 GHz and 81 GHz to 86 GHz", 2006.

[5] Ivanek F., editor, *Terrestrial Microwave Communications,* Artech House, 1989.

[6] Petar Jovanović, Miroslav Perić, Dragan Obradović, Dragana Perić, Zoran Mićić, Branko Radan: *CARRD.NET – Software for Planning Radio-Relay Networks,* EUROCON 2005*,* Belgrade, 2005.

[7] ITU-R rec. P.837-2: "Characteristics of precipitation for propagation modeling", 1999.

[8] ITU-R rec. P.530-11, "Propagation data and prediction methods required for design of terrestrial line of sight systems, 2005.

[9] Perić D., Perić M., Petrović G., "*Performance of Different Methods of Protection in 60 GHz Radio Networks with IP Traffic*", International Symposium on Performance Evaluation of Computer and Telecommunication Systems SPECTS 2007*,* San Diego, CA, USA, July 16-18

[10] Dragana Perić, Miroslav Perić, Grozdan Petrović: Redundant Topology in Computer Network Realized with Millimeter Wave Radio Links, 14th IST Mobile and Wireless Communications Summit, Drezden, 19-23. jun 2005.

Concept of signal processing in ultra-high capacity (1Gbit/s) millimeter wave IP digital radio

Perić Miroslav[1], Perić Dragana[2], Obradović Dragan[3] and Orlić Vladimir[4]

Abstract – **In this paper we describe concept of base band and intermediate signal processing for ultra high (around 1Gbit/s) milimeter wave links. The dominant propagation effect in this frequencies is rain attenuation. Thanks to IP packet transmission adaptive capacity could be easily implemented, which gives tradeoff between capacity and receiver threshold. In this paper we focused on adaptive DBPSK and Pi/4-DQPSK modulation. We proposed the realization that avoids costly DSP components for extremely high bitrates, and focus on analogue realization, which is nowadays much cheaper for this application. System block diagrams and results of experimental design idea verification are given.**

I. INTRODUCTION

Ultra high speed millimeter wave radio links are most appropriate for urban solution, business and ISP IP networks, because of their quick installation and reconfiguration, smaller interference and much higher capacity than in lower frequency bands. The price of such wireless link is usually much more cost effective than leasing or deploying fiber optics link.

In millimeter wave link there are several bands allocated for point to point link purpose:

1. Band 58-63GHz, which is in many country license free, with only limitation of max 55dBm EIRP, minimal antenna gain of 30dBi, and maximum output power of 10dBm [1]. The main drawback of this band is peak of oxygen attenuation (about 15dB/km) which restricts hop length to up to 2km. [4][5]
2. Band 64-66GHz, known as high density network [2], where channel bandwidth is limited to multiples of 30MHz up to 240MHz.
3. Band 71-76 paired with 81-86GHz, which is the most suitable band for point to point link application, since channel spacing is multiple of 250MHz up to 2GHz [3]. In this band oxygen attenuation is not much severe (less than 1dB/km) and hop lengths up to 10km could be achieved.

The most significant propagation effect in these bands is rain attenuation, which dominantly determines individual link availability [5]. Thanks to packet organization of IP traffic usage of adaptive modulation, capacity and/or FEC could be very efficient method for increasing of link availability at expense of link capacity in small amounts of time when rain event occurs. We must note that only drastically reduction of receiver threshold (about 10dB, or more) would have significant effect in these frequencies due to rain intensity distribution and rain attenuation [4][5].

Having these regulations and propagation effects in mind the first and the third band are of great interest since simple modulation schemes could be used. For practical realization of modem it is very important that in these cases components primary designed for fiber-optics and/or SATA communication in PC computer [6], together with lower cost FPGA [7] and MMIC [8] could be used. Since sample rates are very high (even up to 2Gs/s), costly digital signal processing DSP (analogue/digital-ADC and digital/analogue-DAC converters and FPGA multiplier resources) should be avoided due to extremely high price. Instead analogue components, together with differential logic (typically PECL or LVDS) logic devices should be used.

Therefore, we focused on DBPSK and Pi/4 DQPSK with noncoherent demodulation due to modem simplicity, and relatively good BER performance [9]. Low spectrum efficiency of these modulations is not of importance for this application.

II SYSTEM BLOCK DIAGRAM

A. Entire system

One of the most important assumptions is that system should implement adaptive capacity, modulation and/or coding schemes, that yields to different capacity at radio level. Therefore Ethernet interface should perform bridging function. From the bridge point of view entire radio acts as contra directional transmission system that gives to the bridge gaped clock to which it responds with parallel date. Considering novel bridge (like Maxim-IC DS33X162) [6] chips we assumed 8 bit parallel bit stream. After this follows the framer together with Reed Solomon coder which are easy to implement even in lower end FPGA [7].

The second stage is base band signal processor at transmitter side, which is the most critical part for maintain low cost of the entire system. We intend to use external parallel to serial converter with multiplication factor from 8 to 16 (like Maxim MAX9205 and MAX9207) [6], which would give final Gbit/s level data stream from parallel symbol level data obtained from low cost FPGA, like Xilinx Spartan III [7]. Some details about parallel signal processing could be found in [10].

[1]Perić Miroslav, is with IMTEL Komunikacije AD, Bulevar Mihaila Pupina 165b, 11070 Novi Beograd, Serbia, e-mail: micha@insimtel.com
2Perić Dragana, is with IMTEL Komunikacije AD, Bulevar Mihaila Pupina 165b, 11070 Novi Beograd, Serbia, e-mail: dragana@insimtel.com
[3]Obradović Dragan, is with IMTEL Komunikacije AD, Bulevar Mihaila Pupina 165b, 11070 Novi Beograd, Serbia, e-mail: obrad@insimtel.com
[4]Orlić Vladimir, is with IMTEL Komunikacije AD, Bulevar Mihaila Pupina 165b, 11070 Novi Beograd, Serbia, e-mail: cheggy@insimtel.com

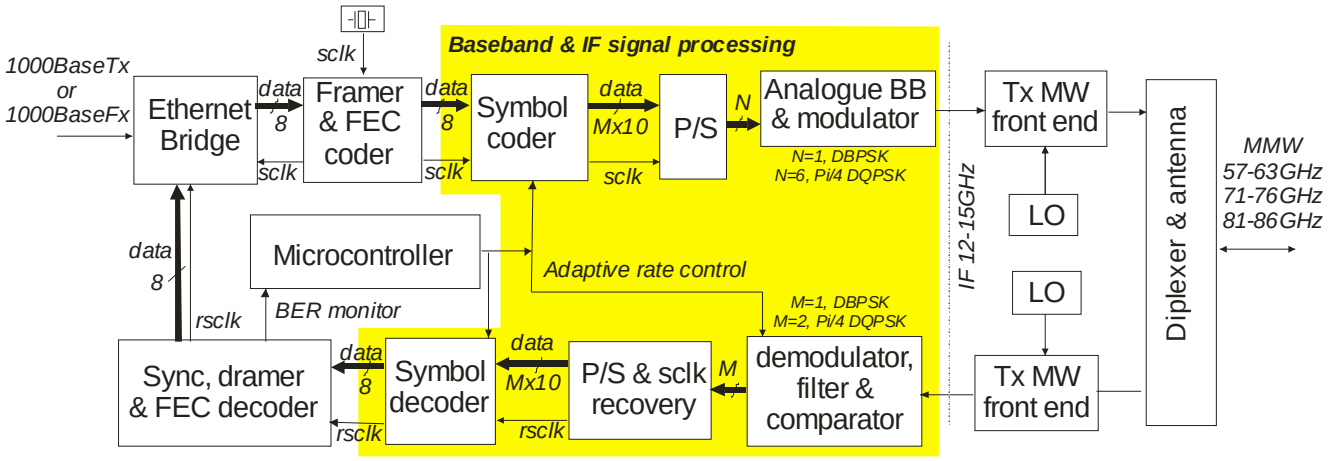


Fig 1. System block diagram

High bandwidth analogue operational amplifiers are used for interfacing between LVDS logic and microwave mixers, and may be got from several manufacturers [6][11].

In order to maintain moderate ratio between modulation bandwidth and central frequency, modulation is performed at 12-15GHz intermediate frequency. At these frequencies we can find high quality MMIC mixer, IQ modulator and demodulator devices [8], or they may be custom designed.

Than follows the mm wave part of the system that up converts the intermediate frequency signal to desired mm-wave frequency from 60 to 90 GHz. In the receiver direction after the antenna follows classical receiver front-end consisting of LNA, down converting mixer, filter and IF frequency AGC block, which are not subject of this paper.

The receiver IF is in the same band as the transmitter. Due to extreme high bandwidth of the signal we assume non-coherent demodulation by differential receivers. After the demodulation follows the serial/parallel converters with clock recovery circuits that converts incoming data into parallel data stream. Than we perform symbol decoding and give receiver 8-bit parallel data stream.

The final stage of the receiver is frame synchronizer with FEC decoder. This block also performs pseudo BER monitoring and alarming. The very important part of the system is microcontroller subsystem that monitors system performance and decides about modulation/capacity changes. This system uses several bits of the frame for internal single hop communication, while for TNM purposes SNMP could be implemented.

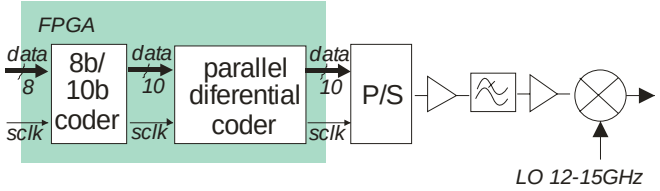


Fig. 2. DBPSK modulator with serializers

B. DBPSK modem with adaptive rate

The easiest to implement modulation scheme is DBPSK. On the transmitter side (Fig. 2.) the incoming data stream is coded by 8B/10B code in order to remove DC and low frequency components. Than follow parallel realization differential coder and serializer with x10 multiplication factor with differential output. Signal from digital differential output is than converted to analogue level and filtered by high bandwidth operational amplifier, buffered and fed into microwave mixers IF port, that acts as a modulator.

Modulator operation at lower bit rate is done at FPGA level since it requires only bit repetition. Additional filtering is unnecessary since, the system works in this condition very small amount of time during the rain falls, which is less than 0.05% of time on well planned hop.

The receiver (Fig. 3.) is construct as differential receiver. After amplification signal is unequally splited into LO side (high power, attenuation about 1dB) and RF side (low power, attenuation about 10dB). Than it is fed into two delay lines each tailored to Tb of specific bit rate: Tsh – symbol duration for high bit rate (about 1Gbit/s, therefore Tsh=1nS) and Tsl – symbol duration for lower bit rate (about 100Mbit/s, and therefore Tsl=10ns). Delay line of Tsh could be realized by meander strip-line technology, while Tsl could be realized by coaxial cable.

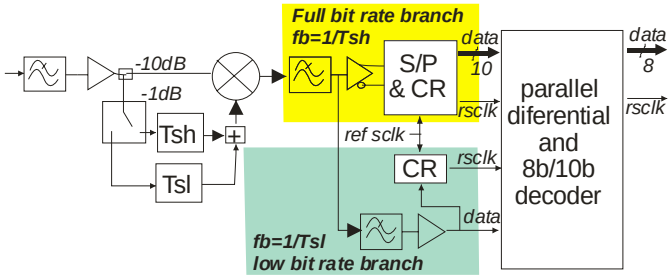


Fig 3. DBPSK demodulator for adaptive bit rate with deserializers and clock recovery

After the mixer follows the buffer and a filter section. The first filters section corresponds to Niquist filtering at the highest bit rate, and it is followed by comparator circuit and deserializer with clock recovery circuitry. For clock recovery circuitry it is necessary to obtain reference clock from local crystal oscillator. Then follows parallel symbol decoder and 8b/10b decoder.

For lower bit rate operation there are additional filter, comparator and clock recovery circuitry, that avoids deserializer and are fed directly to FPGA.

C. Pi/4 DQPSK modem

To achieve better spectrum efficiency we decide to use Pi/4 DQPSK modulation since it could be also differentially demodulated. Similarly to DBPSK adaptive bit rate is also performed by different delay lines.

Conventional Pi/4-DQPSK modulator is described in [10] (Fig. 4.). The main part of modulator is IQ modulator which modulation base band inputs are:

$$u_k = (u_{k-1}I_k - v_{k-1}Q_k)/\sqrt{2},$$
$$v_k = (u_{k-1}Q_k + v_{k-1}Q_k)/\sqrt{2} \quad (1)$$
$$I_k, Q_k \in \{-1,1\}, u_k, v_k \in \{-\sqrt{2}, -1, 0, 1, \sqrt{2}\}$$

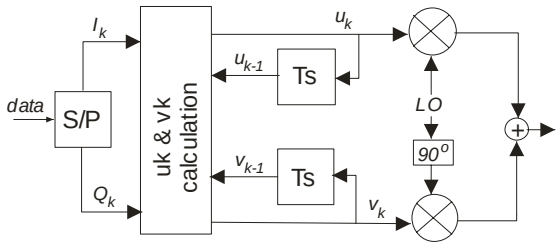


Fig. 4. conventional Pi/4-DQPSK modulator

In conventional realization may be easily obtained by simple DSP and ADC. On Gbit/s such realization would be extremely expensive and therefore we decide to implement it by analogue arithmetic cells made of operational amplifiers. Each level of u_k and v_k could be represented by three bits which in differential logic are represented by +1 and -1, or simple "+" and "-". In that way:

$$u_k = A1 \cdot U_{A1} + A2 \cdot U_{A2} + A3 \cdot U_{A3},$$
$$v_k = A1 \cdot V_{A1} + A2 \cdot V_{A2} + A3 \cdot V_{A3}, \quad (3)$$
$$A1 = 1/\sqrt{2}, A2 = 1/2, A3 = (\sqrt{2}-1)/2$$

The values of U and V for each constellation symbol are given in Table 1.

TABLE 1. SERIALISER DIFFERENTIAL OUTPUT POLARITY FOR SYMBOL GENERATION

cons sym.	u_k	v_k	U_{A1}	U_{A2}	U_{A3}	V_{A1}	V_{A2}	V_{A3}
0	1	1	+	+	-	+	+	-
1	0	$\sqrt{2}$	+	-	-	+	+	+
2	-1	1	-	-	+	+	+	-
3	$-\sqrt{2}$	0	-	-	-	+	-	-
4	-1	-1	-	-	+	-	-	+
5	0	$-\sqrt{2}$	+	-	-	-	-	-
6	1	-1	+	+	-	-	-	+
7	$\sqrt{2}$	0	+	+	+	+	-	-

*Zero can be coded as {+,-,-} or {-,+,+}

The complete Pi/4-DQPSK modulator consists of FPGA signal processor that besides 8B/10B coding and splitting signal into I and Q branch also calculates three bit representation of u_k and v_k symbols. Calculating of uk and vk according to equations (1) could be done by simple, and speed efficient look-up table with 8 inputs and 6 outputs. Since serializers are 10 bit wide, there is necessity for simultaneous computing 10 successive u_k and v_k, which is still does not take much of FPGA resources.

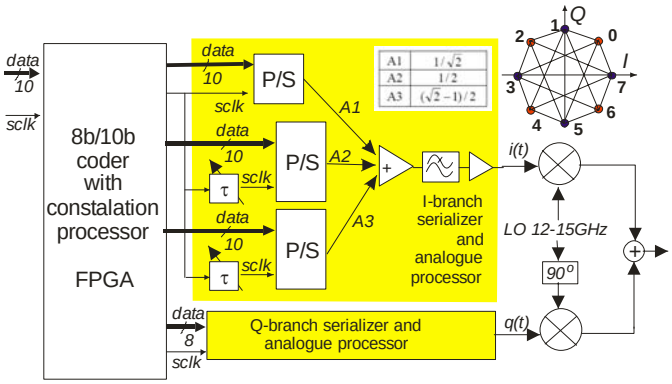


Fig. 5. Ultra high bit-rate Pi/4-DQPSK modem

In this design very critical is phase adjustment of clock multipliers in serializer circuitry. Therefore additional delay tuning circuits should be implemented.

Unlike the complexity of the modulator, the demodulator (Fig. 6.) is much simpler and does not differs much from classical concept given in [9]. Similarly to DPSK case, I and Q outputs of IQ demodulator are filtered, compared with zero and fed into deserializer circuitry with clock recovery. The phase adjustment of two independent deserializer is now much simpler, since it is done at symbol level. It could be completely implemented in FPGA by using delay locked loop –DLL circuitry [7].

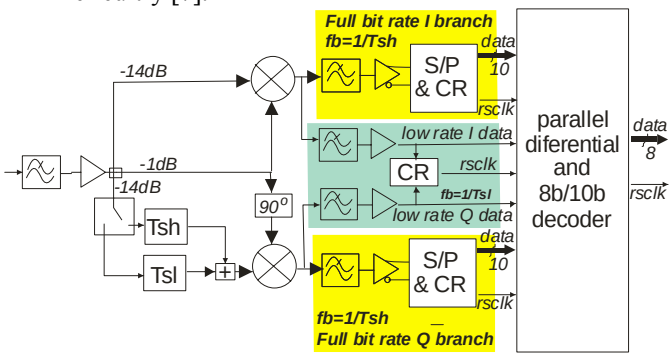


Fig. 6. Differential Pi/4 demodulator

Also, similarly to DBPSK case, operation at lower bit rates at the transmitter side is implemented as symbol repetition, while at the receive side it is done as different delay line at IF frequency and parallel, low bit rate branch at base band level.

If we carefully look the hardware structure we can notice that DBPSK case is actually included Pi/4-DQPSK case. In that way, realization of Pi/4-DQPSK give us not only adaptive rate possibility, but also an adaptive modulation.

III EXPERIMENTAL DESIGN IDEA VERIFICATION

In order to verify our concept we have made an experiment for DBPSK case (Fig 7.). For signal generation we use Anritzu ME520 pattern generator for which clock is provided by Rhode Schwartz SM300 signal generator. The bit rate is varying from 80 to 200Mbit/s. For the DBPSK modulator we used Hittite HMC412 double balanced mixer [], which LO at 12GHz is provided by signal generator HP8673B. The differential DBPSK receiver is constructed by custom design microwave wideband amplifier, that gives total signal power of about +10dBm and than is spited by 3dB

splitter and fit to receiver HMC412 mixer. As a delay line to LO port of the mixer we used various length coaxial cables. The received signal is monitored on 1GHz bandwidth oscilloscope Agilent MSO7104 (Fig 8.).

As expected, on demodulator we got the transition pulse on each data transition in transmitted data stream. The pulse width is correlated by coaxial cable delay line length. Therefore, if the delay line should have delay line should be smaller, *Ts* to prevent inter-symbol interference. To got the highest demodulated signal level the delay should be close to *Ts* as much as possible. According to this, the delay line is not much critical and may result in some small additional implementation loss. The smallest measurable pulse width, with acceptable quality was about 2ns which corresponds to 500MHz bandwidth. This proves proper choice if mixer devices.

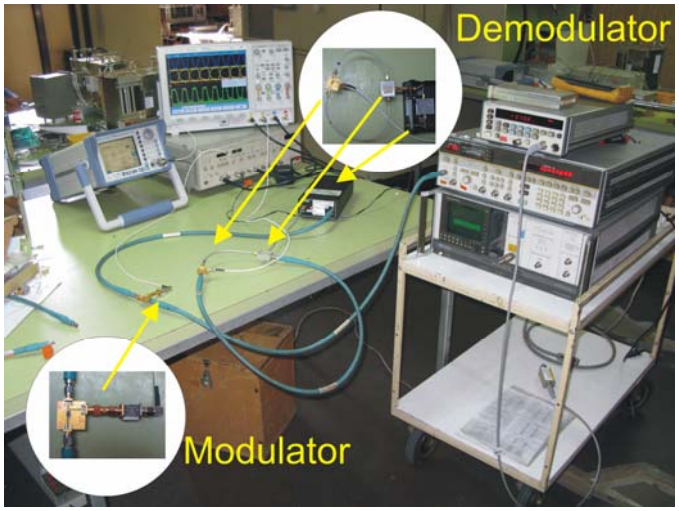


Fig 7. Laboratory test of DBPSK IF subsystem

On given method find the most critical is LO level fed to demodulator mixer. This problem could not be solved by simple adding additional amplifier due to additional delay insertion. Therefore proposed unequal coupler is the proper solution.

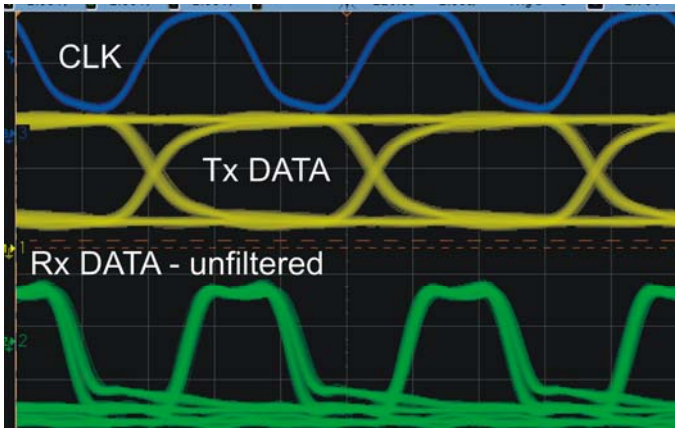


Fig 8. Measured signals

Finally we try external filtering of demodulated signal by transferring measured data from oscilloscope to PC. We simulate in Matlab analogue filter response and obtained acceptable quality eye-pattern, with not significant zero-crossing jitter (Fig. 9.).

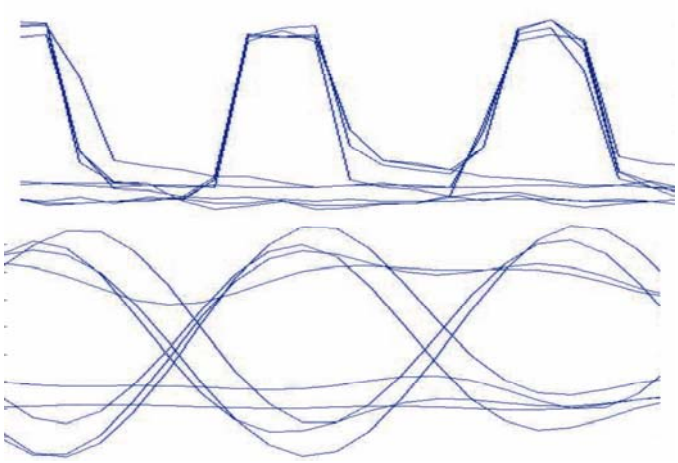


Fig 9. Filtered demodulated signals

IV Conclusion

Experimental design idea verification proved that described concept of IF and baseband signal processing for ultra-high capacity millimeter wave radio is realizable with relatively low cost of entire system. The base band signal processing is not crucial for DBPSK is simple, and not critical. In realization of Pi/4-DQPSK case we expect the most significant problems at transmitter part, especially with timing adjustment of serializers.

V Acknowledgement

We want to thank our colleagues Dr Branka Jokanović, Siniša Jovanović, Nebojša Pupovac and Nenad Pejović for providing us microwave components for the experiment. This work is partialy financed by Serbian Ministry for Science and Technological Development of Serbia.

References

[1] ECC Recommendation (09)01: "Use of the 57 - 64 GHz frequency band for point-to-point fixed wireless systems", 2009
[2] ECC Recommendation (05)02: "Use of the 64-66 GHz frequency band for fixed service", e2009.
[3] ECC Recommendation (05)07:"Radio frequency channel arrangements for fixed service systems operating in the bands 71-76 GHz and 81-86 GHz, 2009.
[4] ITU-R rec. P.837-2: "Characteristics of presipication for propagation modeling", 1999.
[5] ITU-R rec. P.530-11, "Propagation data and prediction methods required for design of terestrial line of sight systems, 2005.
[6] www.maxim-ic.com
[7] www.xilinx.com
[8] www.hittite.com
[9] Xiong F., *Digital Modulation Techniques,* Artech House, 2000.
[10] M.Perić, D. Obradović, D. Perić, V. Orlić: " One easy-to-implement method for BER performance testing of uncoded ultrahigh capacity (Gbit/s) radio link", ICEST Conference, Niš, 2009.
[11] www.analog.com

Session

RADIO COMMUNICATION SYSTEMS – 3

Estimating losses from transient and intersymbol distortions in hybrid fiber-coaxial television network

Krasen Angelov[1], Kiril Koitchev[2], Natalia Varbanova[3]

Abstract – **Modern hybrid cable television networks use high-speed optical trunk lines as backbone network. When estimating the optimum number of amplifying sections within the optical trunk line, the power potential of the system is used. This potential is the difference between the maximum signal level and the signal losses inside the transmission line and termination devices. These are the losses caused by the signal fading, the losses resulting from transient and intersymbol distortions plus the losses caused by additive noise. The primary goal is to offer a practical approach for estimating the signal losses, resulting from transient and inter-symbol distortions, depending on the physical parameters of the transmission line and the optical equipment in use.**

Keywords – **HFC CATV, WDM, optical amplifier, photodiode receiver, signal level, nonlinear distortion**

I. INTRODUCTION

As a transmission medium, fiber optic cables are widely used in backbone networks for signal transmission in hybrid fiber-coaxial television networks. Wavelength division multiplexing systems are also widely used when building optical trunk–lines [1,2,3]. Structural diagram of an optical system for transmission with wavelength multiplexing is shown in fig. 1.

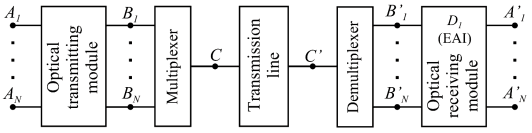


Fig. 1. Structural diagram of an optical system for data transmission with wavelength-division multiplexing

The transmission line can be presented as amplifying sections connected in series each of which has a structure that is shown in fig. 2 [2,3,8]. p_S is the signal level in the transmission line, $p_S - A_{NL}$ and $p_S - A_{ISD}$ are respectively the differences between the system signal level and the group signal protectability level from the total power of nonlinear transient distortions and the losses due to intersymbol distortions.

[1]Krasen Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@dir.bg

[2]Kiril Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

[3]Natalia Varbanova is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: nataliavarbanova@abv.bg

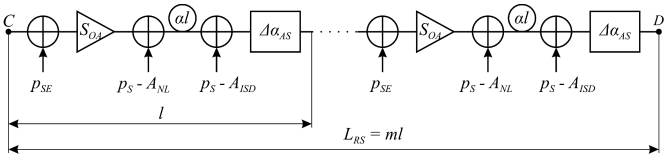


Fig. 2. Structural diagram of amplifying sections

The need of periodic amplification of signal is causally related to amplitude losses and the dispersion expansion of code pulses. Nonlinear distortions in optical fibers are caused by phenomena arising from inadmissible high levels of transmitted signals [5,9]. Effects related with light scattering are characterized by threshold level which is highly sensitive to the length of the amplifying section. Nonlinear effects related with self-phase modulation and four wave mixing characterize systems with wavelength division multiplexing and exert a great overall influence [5,6].

Optimum number of amplifying sections in optical trunk-lines of HFC networks can be implemented by applying an iterative approach. Through series of iterations a solution is found which for a set number of amplifying sections m_j for the current iteration will yield percentage error that satisfies the inequality

$$\delta_{Cj} = \left|\Delta A_E(m_j) / A_{N_E}(m_j)\right| \le 0{,}1\,, \qquad (1)$$

where $\Delta A_E(m_j)$ is the difference between the maximum possible and total of real signal losses given at the elecrtonical amplifier input (EAI) (fig. 1); $A_{N_E}(m_j)$ – is the minimum admissible signal level at the amplifier input.

Maximum admissible total losses of signal in optical network can be expressed as

$$\Delta A_{\max} = A_{\max} - A_N\,, \qquad (2)$$

where $A_{\max}$ – is the maximum possible protectability of signal from thermal noise at the amplifier input (EAI) with no optical amplifier and ideal devices, A_N – is the norm of signal protectability level which can be estimated if the error ratio of the regenerator is determined.

Total real losses of signal within the optical system can be expressed as

$$\Delta A_\Sigma = \Delta A_S + \Delta A_{TD} + \Delta A_{ISD} + \Delta A_{AN}\,, \qquad (3)$$

where ΔA_S – are the signal losses due to signal fading in the transmission line; ΔA_{TD} – losses due to all kinds of transient distortions (linear in the termination devices and nonlinear in the optical amplifiers); ΔA_{ISD} – is for the losses due to intersymbol distortions (in the termination devices and the transmission line); ΔA_{AD} – are losses due to all kinds of additive noise (in the termination devices and transmission line), with the exception of the thermal noise of the electronic amplifier which is taken into consideration when estimating $A_{\max}$. Therefore, in order to achieve a realizable optical

transmission system with wavelength division multiplexing there should be fulfilled the condition of

$$\Delta A_{\max} \geq \Delta A_{\Sigma}, \quad (4)$$

By substituting Eqs.(2) and (3) in Eq.(4) we get

$$A_{\max} - \Delta A_S - \Delta A_{TD} - \Delta A_{ISD} - \Delta A_{AN} \geq A_N. \quad (5)$$

Then the difference between maximum possible and the total real signal losses considering the number of amplifying sections m will be

$$\Delta A(m) = \Delta A_{\max}(m) - \Delta A_{\Sigma}(m). \quad (6)$$

Conversion of Eq. (6) by using Eq. (5) for the input of electronic amplifier will appear as

$$\begin{aligned}\Delta A_E(m) = {} & A_{\max_E}(0) - \Delta A_{S_E}(m) - \Delta A_{TD_E}(m) \\ & - \Delta A_{ISD_E}(m) - \Delta A_{AN_E}(m) - A_{N_E}(m) \geq 0,\end{aligned} \quad (7)$$

where the maximum signal level in the transmission line, taking into consideration the losses in the termination devices (the receiving and transmitting optical modules, multiplexers and demultiplexers) is estimated as

$$A_{\max_E}(0) = A_{\max_E} - \Delta A_{S_E}(0) - \Delta A_{TD_E}(0) - \Delta A_{ISD_E}(0). \quad (8)$$

In Eq. (8) $A_{\max_E}$ is the maximum level of signal ($S_{APD} = 0$).

For the purpose of analysis in Eqs. (7) and (8) of the influence of individual types of losses, they are conditionally divided into two separate groups: those depending on the number of amplifying sections m in the transmission line and those which do not depend on the number of amplifying sections (i.e. the losses at the termination devices when m=0).

Complete analysis of Eqs. (7) and (8) to determine all components which are set in the expressions. The aim of this paper is to determine the last two components of Eq. (7) – the losses caused by transient $\Delta A_{TD}(m)$ and intersymbol distortions $\Delta A_{ISD}(m)$ in the transmission line.

It is advisable to make estimations for the signal losses for the electronic amplifier input (fig. 1), i.e. index "E" refers to EAI.

II. Estimation of Losses Caused by Transient and Intersymbol Distortions in the Transmission Line

A. Estimation of signal protectability level from the transient distortions in function of the number of channels in a single optical amplifier

If the average power of the signal in N channels at the optical amplifier output is constant, then the value in the k channel of signal protectability level from transient distortions which are composite third order (CTB) distortions arising in the optical amplifier [1] will be

$$\begin{aligned}A_{TD}(f_k) = {} & 10\lg\left(NC_1P_S g_1(f_k)\right)/\left(N^3C_3P_S^3 g_3(f_k)\right) = \\ & = A_{NL} + 10\lg\left(g_1(f_k)/g_3(f_k)\right),\end{aligned} \quad (9)$$

where $A_{NL} = 10\lg((NC_1P_S)/(N^3C_3P_S^3))$ is the group signal protectability level from the total power of nonlinear transient distortions; NC_1P_S and $N^3C_3P_S^3$ are the real values of signal power at the output of the optical amplifier; C_1 and C_3 are decomposition coefficients of the expression for the output power of the optical amplifier in series concerning the exponents of the input power; $g_1(f_k)$ and $g_3(f_k)$ are the normalized spectrums from first and third order for channel k with central frequency f_k. These transient distortions in optical amplifiers are referred to as four-wave mixing (FWM) [2,5]. For normalized spectrums of composite third order distortions in optical amplifier it is possible to write

$$g_3(f) = g_2(f) * g_1(f), \quad (10)$$

where * means a convolution of normalized spectrums, $g_2(f)$ – normalized spectrum of nonlinear second order distortions (CSO), $g_1(f)$ – normalized spectrum of the output signal. It can be expressed as the sum of positive and negative spectrums which are necessary for considering the convolution of all contributing frequencies that occur as lower and upper sidebands of the spurious amplitude modulation caused by the nonlinearity of the optical amplifier

$$g_1(f) = \sum_1^N g(f_k) + \sum_N^1 g(-f_k) = 0{,}5/N. \quad (11)$$

The normalized second order spectrum is composed of following three components

$$g_2(f_x) = g_{21}(f_{x1}) + g_{22}(f_{x2}) + g_{23}(f_{x3}). \quad (12)$$

All components in Eq. (12) can be expressed by the variable x which is changed by 1 from –(N–1) to (N–1). This also is valid for the frequencies f_{x1}, f_{x2}, f_{x3} within the discrete frequency interval $\Delta = |f_k - f_{k\text{-}1}|$ for each k.

The first component of the second order normalized spectrum will be

$$g_{21}(f_{x1}) = \left(n - |x|\right)/\left(4N^2\right), \quad (13)$$

where the frequency corresponding to x is determined as $f_{x1} = -(f_1 + f_N) - x\Delta$ in the frequency range from $-2f_N$ to $-2f_1$.

The second component then will be

$$g_{22}(f_{x2}) = \left(n - |x|\right)/\left(2N^2\right), \quad (14)$$

and the frequency corresponding to x is determined as $f_{x2} = x\Delta$ in the frequency range from $-(f_N - f_1)$ to $(f_N - f_1)$.

The third component will be

$$g_{23}(f_{x3}) = \left(n - |x|\right)/\left(4N^2\right), \quad (15)$$

and frequency corresponding to x: $f_{x3} = (f_1 + f_N) + x\Delta$ in the frequency range from $2f_1$ to $2f_N$.

The output spectrum of third order transient distortions is convolution of the obtained normalized second order spectrums with the output spectrum [2,7,8].

The normalized third order spectrum with number k obtained by the convolution of $g_{22}(f_{x2})$ with $f_{x2} > 0$ and $g_1(f)$ with $f > 0$, can be expressed as

$$g'_{32}(k) = g'_{32}(f_{32}) = \left(S_{N-k}^{N-2}\right)/\left(2N^2\right), \quad (16)$$

where each number of the channel k has a corresponding channel frequency f_k; S_{N-k}^{N-2} is the sum of the members of the arithmetic progression from $(N-k)$ to $(N-2)$.

It is evident that the operating range of frequencies must be selected k from *2* to N in accordance with the minimum third order frequency spectrum.

The sum of the terms of the arithmetic progression from a to b will be

$$S_a^b = 0{,}5(a+b)(b-a+1), \quad (17)$$

therefore for Eq. (16) we obtain

$$g'_{32}(k) = (S_{N-k}^{N-2})/(2.N^2) = 0{,}25(2N-2-k)(k-1)/N^2 . \quad (18)$$

Then the normalized third order spectrum in a channel numbered by k, resulting from the convolution of $g_{22}(f_{x2})$ with $f_{x2} < 0$ and $g_1(f)$ with $f > 0$, will be:

$$g''_{32}(k) = g''_{32}(f_{x2}) = (S_{k-1}^{N-2})/(2N^2) = \\ = 0{,}25(N-3+k)(N-k)/N^2, \quad (19)$$

where S_{k-1}^{N-2} is the sum of the terms of the arithmetic progression from $(k-1)$ to $(N-2)$.

In this case it is evident that for the work frequency range the value of k should be within the limits from 1 to $(N-1)$.

Accordingly, from Eqs. (18) and (19) the target normalized third order spectrum will be

$$g_{32}(k) = g'_{32}(k) + g''_{32}(k) = \left(S_{N-k}^{N-2} + S_{k-1}^{N-2}\right)/\left(2N^2\right) = \\ = 0{,}5\left(N^2 - 5N - 2k^2 + 2k(N+1) + 2\right)/N^2. \quad (20)$$

The operating frequency range corresponds to a value of k from 1 to N.

The consecutive convolution of the other two components $g_{23}(f_{x3})$ and $g_{21}(f_{x1})$ with $g_1(f)$ is expressed by

$$g_{31}(k) = g_{33}(k) = 0{,}25 g_{32}(k) . \quad (21)$$

From Eqs. (20) and (21) it follows that the full normalized third order spectrum will be

$$g_3(k) = g_{31}(k) + g_{32}(k) + g_{33}(k) = \\ = 0{,}75\frac{N^2 - 5N - 2k^2 + 2k(N+1) + 2}{N^2}. \quad (22)$$

Thus, by substituting Eq. (22) in Eq. (9) for $A_{TD}(k)$ we get:

$$A_{TD}(k) = A_{NL} - 10\lg\left(0{,}75\frac{N^2 - 5N - 2k^2 + 2k(N+1) + 2}{N^2}\right). \quad (23)$$

The dependency shown in fig. 3 is drawn from the analytically obtained expression Eq. (23). It is shown for three different values of the number of the channels N with A_{NL} = 26,7*dB* (depending on the parameters of the optical amplifier).

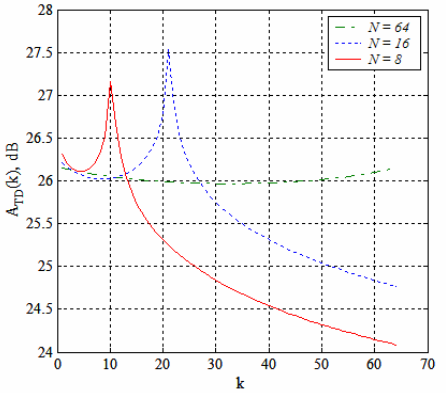


Fig. 3. Signal protectability from transient distortions as a function of the number of channels in a single optical amplifier

As is evident from fig. 3 the minimum of $A_{TD}(k)$ is obtained for the channel with number $k = (N+1)/2$ for odd values of N. Theoretically if N is even, the minimum of $A_{TD}(k)$ will be for two channels with numbers $k = N/2$ and $k = (N/2) + 1$. It is assumed that estimations according an average number of the channel are conditionally written as k_{av}.

Then the value of the signal protectability from transient distortions will be

$$A_{TDav} = A_{NL} + 1{,}25 - \\ - 10\lg\left(\frac{N^2 - 5N - 2k_{av}^2 + 2k_{av}(N+1) + 2}{N^2}\right), \quad (24)$$

where the value *1,25dB* is the increment of signal protectability from transient distortions due to the expansion of the normalized third order spectrum in the non-operative frequency bands situated to the left or right of the operating range. In this way part of the composite distortions and noises are filtered through the transparency window of the optical fiber. Therefore, the maximum value $A_{TD}(k)$ is reached at $k = 1$ and $k = N$.

B. Estimation of signal losses due to transient distortions in function of the number of amplifying sections

The value of the losses due to the non-linear transient distortions in m number of optical amplifiers can be obtained by means of the following equivalent formula

$$\Delta A_{TD_E}(m) = -20\lg(1 - 10^{-0{,}05 A_{TD\,OA}(m)}) . \quad (25)$$

In order to estimate $A_{TD\,OA}(m)$, it could be assumed that the level of signal at the output of each amplifying section (fig. 4) drops by $\Delta\alpha$, and the level of all third order non-linear products (which generate transient distortions in narrowband devices such as optical amplifiers) decreases by $3\Delta\alpha$.

Accordingly, if there is a defined signal level diagram (where p_S is the signal level in the transmission line), the signal protectability is increased by in every following amplifying section. In this way if the signal protectability from nonlinear transient distortions at the output of the first optical amplifier is A_{NL}, at the output of the second optical amplifier the signal protectability will be A_{NL} +2$\Delta\alpha$, at the output of the third optical amplifier: A_{NL} +4$\Delta\alpha$, and at the output of the m-th optical amplifier: A_{NL} +2(m –1)$\Delta\alpha$.

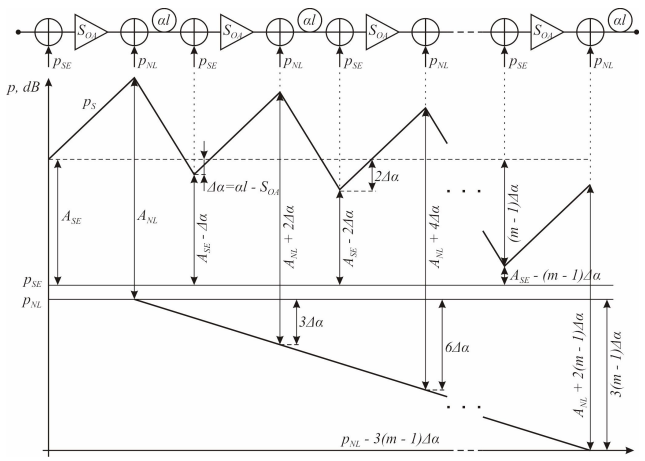


Fig. 4. Signal protectability diagram: A_{SE} – from spontaneous emitting, A_{NL} – from transient noise

The value of equivalent signal protectability from transient distortions when the number of optical amplifiers is m will be:

$$A_{NL}(m) = -10\lg(10^{-0,1A_{NL}} + 10^{-0,1(A_{NL}+2\Delta\alpha)} + .. + 10^{-0,1(A_{NL}+2(m-1)\Delta\alpha)}) =$$
$$= A_{NL} - 10\lg(1 + 10^{-0,12\Delta\alpha} + 10^{-0,14\Delta\alpha} + ...10^{-0,12(m-1)\Delta\alpha}) = \quad (26)$$
$$= A_{NL} - 10\lg S_G,$$

where $S_G = \frac{1-10^{(-0,2.m\Delta\alpha)}}{1-10^{(-0,2.\Delta\alpha)}}$ is the sum of terms of the geometric progression.

Taking into account Eq. (26), $A_{TD\,OA}(m)$ estimated for the middle channel will be

$$A_{TD\,OA}(m) = A_{NL}(m) + 1,25 - 10\lg\left(\frac{N^2 - 5N - 2k_{av}^2 + 2k_{av}(N+1) + 2}{N^2}\right), dB \quad (27)$$

Therefore the value $A_{TD_E}(m)$ for the middle channel will be obtained by means of substitution of Eq. (27) in Eq. (25).

Based on the analytically obtained expressions Eqs. (27) and (25) in fig. 5 is shown the dependence of losses caused by transient distortions as a function of the number of optical amplifiers. It is shown for the middle channel for three values of N when $A_{NL} = 26,7dB$ (depending on the parameters of the optical amplifier) and $\Delta\alpha = 0,3$ (depending on the parameters of the optical amplifier and optical fiber).

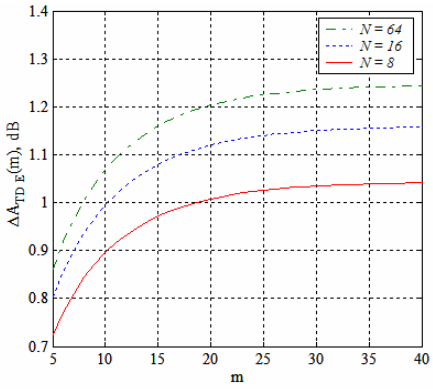


Fig. 5. Losses due to nonlinear transient distortions as a function of the number of optical amplifiers

C. Estimation of the signal level due to intersymbol distortions in function of the number of amplifying sections

Intersymbol distortions decrease the range of the eye-diagram $(1 - K_{ISD_U})$ times. The value of intersymbol distortion ratio multiplied by the voltage at the input of the electronic amplifier is:

$$K_{ISD_U} = \left(K_{ISD_{mE}}\right)^{0,5}, \quad (28)$$

where $K_{ISD_{mE}} = K_{ISD_{mO}} = \left(10^{-0,14A_{ISD}(m)}\right)$ is the transient distortions ratio according electrical (for the input of the electronic amplifier) and optical power (for the transmission line).

Thus as is the case with signal losses due to transient distortions, the signal losses due to intersymbol distortions in the transmission line will be

$$\Delta A_{ISD}(m) = -20\lg(1 - 10^{-0,05A_{ISD}(m)}), \quad (29)$$

where $\Delta A_{ISD}(m) = 41/(1 + 10^{-7} l^2 m^2)$ s signal protectability from inter-symbol distortions caused by dispersion distortions inside the optical fiber; $l = (\Delta\alpha + S_{OA})/\alpha$ (l – is the length of amplifying section, α – is the kilometric attenuation ratio of optical fiber, S_{OA} – amplification of optical amplifier in the transparency window, $\Delta\alpha$ – is the difference between attenuation in an amplifying section with length of *1km* and the amplification of the optical amplifier

III. Conclusion

The approach with dividing losses into dependent on and independent from the number of amplifying sections leads to substantial simplification of the assigned task. This solution is based on the physical characteristics of the transmission line (which depend on the type of optic fiber) and the physical characteristics of transmitting and receiving optical module, multiplexers and demultiplexers.

In this paper we propose an approach for estimating only losses from transient and intersymbol distortions in hybrid television network. Estimating the difference between the maximum possible and total of real signal losses given at the EAI and the maximum signal level in the transmission line, taking into consideration the losses in the termination devices is also associated with the need of complete analysis of losses due to all kinds of additive noise and the signal losses due to signal fading in the transmission line. This complete analysis would allow estimation of optimum number of amplification sections and appears as a goal of an ongoing research.

References

[1] Наний О., Основы технологий спектральнога мултиплцексирования каналов передачи (WDM), LIGHTWAVE russian edition №2, стр.47-52, 2004.

[2] Фердинандов Е., Оптични комуникационни системи, Техника, София, 2007.

[3] Hybrid/Fiber Coax (HFC) and Dense Wavelength Division Multiplexing (DWDM) Networks, IEC Web ProForum Tutorials, 2000.

[4] K. Angelov, K. Koitchev, S. Sadinov, An Investigation of Noise Influences in Optical Transmitters and Receivers in Cable TV Networks, ICEST 2006, Proc. of Papers, pp.102-105, Sofia, Bulgaria, 2006.

[5] K. Angelov, K. Koitchev, S. Sadinov. Influence of Optical Fiber Nonlinear Effects in HFC Television Networks with WDM Multiplexing. ICEST 2007, Proc. of Papers, pp.287-290, Ohrid, Macedonia, 2007.

[6] M. R. Philips, K. L. Sweeney, Distortion by Stimulated Brillouin Scattering Effect in Analog Video Lightwave Systems, OSA TOPS, System Technologies, vol.12, pp.182-185, 1997.

[7] O. Panagiev, Analytical approach for determination of the composite nonlinear distortions and dynamic range of HFC networks' signals. COMPUTER SCIENCE 2008, part III, pp.863-868, Kavala, Greece, 2008.

[8] R. Freeman, Fiber-Optic Systems for Telecommunications, John Wiley & Sons, New York, 2002.

[9] V. Topchiev, L. Jordanova, Analysis of the reasons for limiting the dynamic range of the signals in CATV systems, ICEST 2006, Proc. of Papers, pp.98-101, Sofia, Bulgaria, 2006.

EDFA Application in WDM CATV Systems

Lidia Jordanova[1] and Valentin Topchiev[2]

Abstract – **Research of the influence of the erbium-doped fiber length and optical input power on the gain spectrum and noise figure of EDFA, working with milti-wavelength signals, is conducted and the results are presented in this paper. The gain spectrum in full C-band when EDFA is pumped with laser at 980 nm and 1480 nm, is analyzed. The EDF length optimum value, when pumping at 980 nm, is defined so that maximum optical output power and gain flatness are attained. The influence of optical input power on the gain spectrum and noise figure is investigated and its maximum variation is determined.**

Keywords – **WDM CATV system, EDFA, EDF lengh, Gain flatness, noise figure.**

I. Introduction

Erbium-doped fiber amplifier (EDFA) is mostly applicable for signal amplification in C-band (1525-1565 nm) due to high signal gain, wide waveband, low noise figure (3-5 dB) and low price.

EDFAs are made by doping the silica fiber with erbium ions. These ions can absorb light energy injected from laser source and reemit it in the range of input signal wavelength, due to stimulated light emission. The highest quantum efficiency is attained when pumping at 980 nm or 1480 nm.

The EDFA is designed to operate in saturation when single-wavelength signal amplifies. This way, when the input signals level varies in wide range, the output power and noise figure remain stable. It is well-known that the average inversion population is very low (< 0.69) in saturation regime operation, which is precondition for maximum transformation from pump to signal power.

One of the main issues when EDFA amplifies multi-wavelength signal is the gain flatness in the used wavelength range. Researches show that maximum gain flatness is achieved when the average inversion population is in the range from 0.75 to 0.8 [1]. This means that EDFA, which operates in saturation, is not suitable to work in WDM systems. To ensure high gain flatness, it is necessary to retain high average value of inversion population. There are two ways to obtain that – by using higher pump power or shorter erbium-doped fiber. In the second case, this leads to gain decrease.

The goal of the researches in this paper is to optimize EDFA parameters in order to obtain high gain flatness and low noise figure when it operates in C-band. Since this amplifier is used in WDM CATV system, we are interested in the influence of the total input power on the amplifier characteristics.

[1]Lidia Jordanova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: jordanova@tu-sofia.bg.

[2]Valentin Topchiev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: vtsc@mail.bg.

II. Mathematical description of the processes in an EDFA

EDFAs are designed to operate in three main pumping schemes – forward, backward and bidirectional. The forward pumping provides lowest noise figure (NF), while the backward – the highest saturated output power. As the parameter noise figure is one of the most important for EDFA, the researches are made using forward pumping only.

Fig. 1 shows simplified EDFA block scheme which consists of: pump laser, connected to EDF by wavelength division multiplexor (WDM), optical isolators for separation of the amplifier input and output, an optical waveband filter that decreases the pump and ASE noise power at the amplifier output.

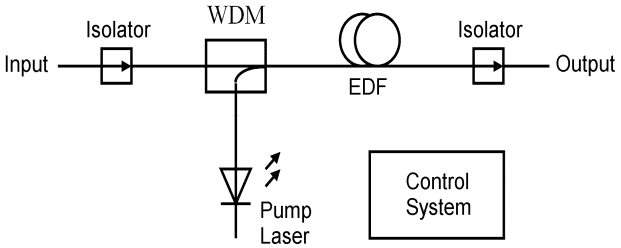


Fig. 1. EDFA Scheme

The processes in EDFA can be described by two types of mathematical equations – the rate equations, which define the transitions between energy states, and the propagation equations, which characterize signal (P_s), pump (P_p) and ASE (P_{ASE}) power evolution along the active fiber [2].

On the base of the rate equations, a formula that calculates the number N_2 of the erbium ions in excited state can be written as:

$$N_2 = \left(\sum_i \frac{\tau \sigma^a_{\nu_i}}{Ah\nu_i} \Gamma_s P_{s_i} + \sum_j \frac{\tau \sigma^a_{\nu_j}}{Ah\nu_j} \Gamma_{\nu_j} P_{ASE_j} + \frac{\tau \sigma^a_p}{Ah\nu_p} \Gamma_p P_p \right) N \text{ x}$$

$$\left[\sum_i \frac{\tau\left(\sigma^a_{\nu_i} + \sigma^e_{\nu_i}\right)}{Ah\nu_i} \Gamma_s P_s(\nu_i) + \sum_j \frac{\tau\left(\sigma^a_{\nu_j} + \sigma^e_{\nu_j}\right)}{Ah\nu_j} \Gamma_{\nu_j} P_{ASE}(\nu_j) + \right. \quad (1)$$

$$\left. + \frac{\tau\left(\sigma^a_p + \sigma^e_p\right)}{Ah\nu_p} \Gamma_p P_p + 1 \right]^{-1},$$

where τ is the life time of electrons in exited state, σ^e and σ^a – emission and absorption cross section, A – effective area of erbium fiber, $h\nu$ – photon energy, Γ – overlap factor, N – erbium ions concentration and ν is the signal light frequency.

The indexes used in formula are refer to the signal (*s*), pump power (*p*), the number of the signal (*i*) and the ASE noise power (*j*) spectrum terms.

Since the EDF length usually does not exceed 20 m, the signal attenuation is neglectfully low and the propagation equations can be described as follows:

- about signal and pump powers:

$$\frac{dP(\lambda)}{dz} = \Gamma(\lambda)P(\lambda)\left[N_2\sigma^e(\lambda) - N_1\sigma^a(\lambda)\right], \quad (2)$$

- about ASE noise power:

$$\frac{dP^{\pm}_{ASE}(\lambda)}{dz} = \pm\Gamma(\lambda)P^{\pm}_{ASE}(\lambda)\left[N_2\sigma^e(\lambda) - N_1\sigma^a(\lambda)\right] \pm \sigma^e(\lambda)N_2\Gamma(\lambda)P_0(\lambda). \quad (3)$$

The parameters used in the formulas given above are related to different wavelength λ. The number of erbium ions in ground state N_1 and the part of ASE noise power that propagates along with the signal, can be calculated by:

$$N_1 = N - N_2, \qquad P_0 = 2h\nu\Delta\nu. \quad (4)$$

When we calculate P_{ASE}, we take into consideration just the part of power that propagates to the signal direction P^{+}_{ASE}, which can be defined with the following expression:

$$P_{ASE} = n_{SP}h\nu(G-1)\Delta\nu, \quad (5)$$

where $n_{SP} = N_2/(N_2 - N_1)$ inversion population factor, $\Delta\nu$ – waveband of optical filter и G – amplifier gain.

III. Fiber parameters and pump power selection

The simulation researches are conducted by using erbium-doped fiber, which has the following parameters: fiber type – Al-Ge-Er-SiO_2; erbium ions concentration N = 0,7.10 $^{-19}$ cm $^{-3}$; life time of electrons in exited state τ = 10 ms; overlap factor Γ(1535–1565) = 0.40, Γ(1480) = 0.43 and Γ(980) = 0.64.

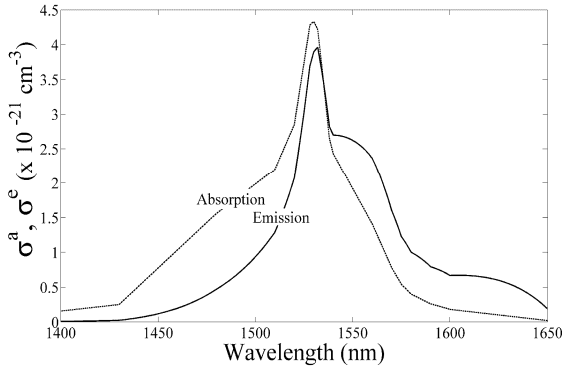


Fig. 2. Спектрални характеристики на σ^a и σ^e

On the Fig. 2 absorption (σ^a) and emission (σ^e) cross sections as a function of wavelength are presented. To create absorption cross section curve, experimental data is used [3], while the emission cross section is calculated by using McCumber equation [4]

$$\sigma^e = \sigma^a e^{\left(31.232 - \frac{48.007}{\lambda[nm]}\right)}. \quad (6)$$

To define the pump laser power, the dependence of EDFA parameters (G, P_{ASE} and NF) on the pump power attained in [5], is used. On the base of made analysis, we obtain that the gain is higher than 17 dB, when the value of input signal is approximately 1 mW, and the needed pump power is at about 97 mW. For that reason, in researches conducted below, P_p is assumed to be equal to 100 mW.

IV. Simulation and results

A. Gain Spectrum

In order to investigate the amplifier gain that is designed on the base of the described above erbium-doped fiber, multi-wavelength signal is launched in the amplifier input. This signal consists of 41 single-wavelength signals with channel spacing of 1 nm, placed on the wavelength range from 1525 to 1565 nm and each of them has – 16 dBm power (or total input power is equal to 1.025 mW). The numerical simulations are made with pump laser source having 100 mW output power in two cases: when λ_p = 980 nm and λ_p = 1480 nm.

The EDF length varies at about its optimum value that is attained in [5] in case the amplifier operates with 1550 nm single-wavelength signal. The main goal of this investigation is to define the optimum value of EDF length so that the gain is higher than 17 dB and the gain flatness is maximal in largest possible wavelength range.

The amplifier gain can be calculated by the following expression:

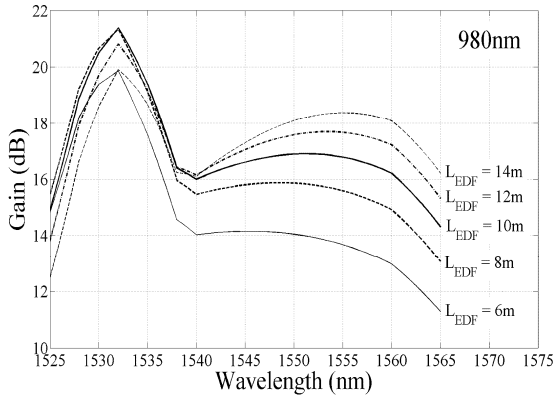
$$G = P_s(L_{EDF})/P_s(0), \quad (6)$$

where $P_s(0)$ and $P_s(L_{EDF})$ are input and output signal optical power calculated by using formula (2).

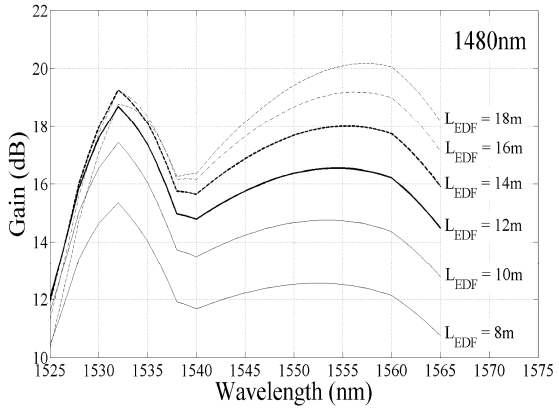
The results given on Figs 3a) and 3b) show that EDFA appliance in the range of 1525 to 1540 nm is not appropriate, because the gain flatness in not sufficient (about 8 dB). In the range from 1540 to 1560 nm, the gain spectrum is relatively flat, which is the reason for applying this amplifier in WDM systems. Furthermore, the shorter is L_{EDF}, the higher is the gain flatness. At the same time the gain value decreases too much.

From Fig. 3a) we can easily define that when L_{EDF} is approximately equal to 10.5m, a gain at about 17 dB is obtained so that the gain flatness is lower than 1 dB in the range from 1540 to 1560 nm. If the pump laser operates at 1480 nm the required gain value of 17 dB can be achieved at $L_{EDF} \approx$ 14 m. In this case the gain flatness is too low (2.5 dB). Therefore such amplifier is not appropriate for multi-

wavelength signal amplification and further researches are conducted just in case λ_p = 980 nm.



a)



b)

Fig. 3. Gain Spectrum, when multi-wavelength signal is amplified and 100 mW with 980 nm a) and 1480 nm b) is pumped

B. Defining of L_{EDF} and G optimal values

In the previous simulation more than 70 % of the pump power is used for amplification of the input signals at about 1530 nm. Therefore, the optimum values of the L_{EDF} and G in the range of 1540 to 1560 nm cannot be determined. For that reason it is necessary a second simulation to be performed and the input signal spectrum is limited to the range that is interesting for us.

The results shown on Fig. 4 are attained by launching EDFA input with 21 single-wavelength signals with channel spacing of 1 nm, placed on the wavelength range from 1540 to 1560 nm and each of them has – 13 dBm power (or total input power is equal to 0 dBm).

It can be clearly seen that the maximum gain flatness is obtained when L_{EDF} is equal to 10 m. Then the gain is 17.2 ± 0.4 dB and the quantum efficiency of the optimized amplifier is approximately 59 %, which is a typical value for EDFA applied for single-wavelength amplification.

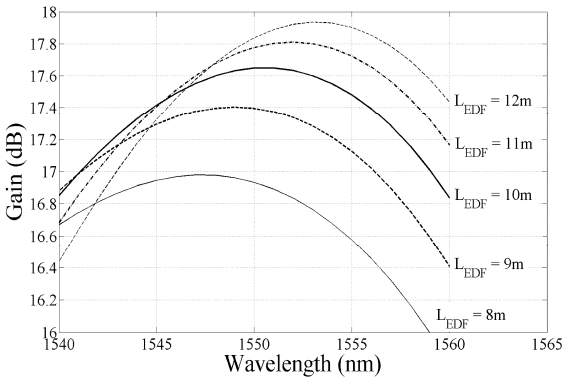


Fig. 4. Gain Spectrum when pump 100 mW with 980 nm and L_{EDF} is from 8 m to 12 m

B. Influence of the optical input power

The results till now show that when a chosen erbium-doped fiber is used, then EDFA in the range from 1540 to 1560 nm can be designed. When the pump laser operates at 980 nm and its output power is equal to 100 mW and the total input power is 0 dBm, the gain flatness remains lower than 0.4 dB.

When the amplifier operates in real WDM CATV system, the optical input power usually differs from the optimized value. To clarify the EDFA behavior, when the total input power changes, a research with multi-wavelength signal containing 21 single-wavelength signals within the optimized range, is conducted. The power of each single-wavelength changes by step of 3 dB in the range of – 19 до – 7 dBm, therefore the total input power of the amplifier is respectively: – 6, – 3 , 0, 3 and 6 dBm.

The obtained results of the simulation are presented on Fig. 5. It can be seen that any deviation of the optical input power from the optimum ($P_{in\,1}$ = – 13 dBm, respectively. P_{in} = 0 dBm), leads to decrease of the gain flatness. This can be explained by the deviation of the average level of inversion population from its optimum value (N_2 = 0.75), that can be calculated by formula (1). Nevertheless, even if the total level of input signals changes twice, i.e. P_{in} = ± 3 dBm, the gain changes in the examined range not more than 1 dB.

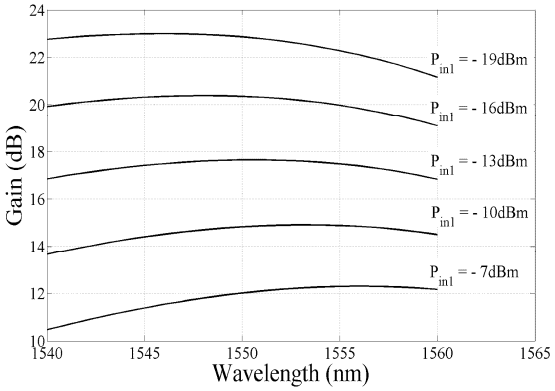


Fig. 5. Gain Spectrum when pump 100 mW with 980 nm and $P_{in\,1}$ is from – 19 dBm to – 7 dBm

The analysis of the obtained results show that the total output power remains almost constant. This guarantees a stable output level (± 0.5 dB), which allows easier calculations of the optical channel budget.

C. Noise Figure

On Fig. 5 are given results of the simulation which show how changes in the input power and its wavelength affect the noise figure (NF). The value of this parameter can be determined by the following formula:

$$NF(\lambda) = \frac{2P_{ASE}(\lambda)}{G(\lambda)h\nu\Delta\nu}. \quad (7)$$

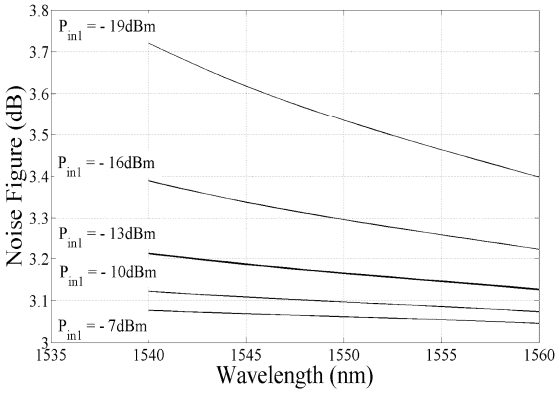


Fig. 6. Noise Figure Spectrum when pump 100 mW with 980 nm and P_{in1} is from – 19 dBm to – 7 dBm,

It is obvious that the investigated parameter slightly exceeds the quantum level of 3 dB and its value remains lower than 3.8 dB even in the most unfavorable case. The higher level of noise in the short wavelengths of the examined range is due to the higher values of gain cross section and the tendency of the erbium ions to emit photons spontaneously in this spectrum range, respectively.

V. CONCLUSION

The performed simulations show that the researched EDFA scheme can ensure flat gain of the optical signals at wavelengths from 1540 nm to 1560 nm. When an EDFA is launched with 21 single-wavelength signals with channel spacing of 1 nm, placed on the wavelength range from 1540 to 1560 nm and each of them has – 13 dBm power (or total input power is equal to 0 dBm), then the obtained amplifier gain is 17.2 ± 0.4 dB. In the same time the gain flatness remains approximately ± 0.5 dB, even if the total input power changes twice. This ensures stable output level (± 0.5 dB) and noise figure lower than 3.8 dB.

REFERENCES

[1] P. Becker, N. Olsson and J. Simpson, "Erbium-Doped Fiber Amplifier – Fundamental and Technologies", Academic Press. 1997.

[2] N. Suong and Ph. Hop, "Simulation of the Gain Characteristics of EDFA", ISEEE 2005, Conference Proceeding, pp. 37-41, HCM City, Vietnam, 2005.

[3] E. Desurvire, C. Giles, J. Zyskind and all of N.J., "Erbium-Doped Fiber Amplifier", US Patent 5027079, 1991.

[4] N. Mohan, "Erbium-Doped Fiber Amplifier: Modeling and Simulation Using VHDL-AMS", Course Project Report ECE-770, Ontario.

[5] L. Jordanova and V. Topchiev, "Optimizing the Parameters of Amplifiers used in the Optical Channel of CATV Systems", Telecom 2008, Conference Proceedings, pp. 206-211, Varna, Bulgaria, 2008.

[6] "Optical Amplifier: EDFA", EECS 290Q, CCH Lecture 3, September 25, 2001.

[7] St. Pinter, J. Jiang and X. Fernando, "A Dynamic Multi-Wavelength Simulink Model for EDFA", Conference Proceedings, CCECE 2004 - CCGEI 2004, Niagara Falls, May 2004.

[8] Li Qian, "Experiment on Erbium-Doped Fiber Amplifiers", Advanced Labs for Special Topics in Photonics (ECE 1640H), University of Toronto, April, 1998.

Software Receiver for Return Path Signals in Cable Television Networks

Lidia T. Jordanova[1] and Dobri M. Dobrev[2]

Abstract - **In this paper, a new software receiver for return path signals using flexible digital signal processing is presented. The receiver is based on direct digitization of the entire 5-65 MHz upstream spectrum. The digitized signal is passed to a digital front-end that performs baseband conversion, filtering and decimation. After that, the signal is send to the digital signal processor (DSP) where demodulation, derandomization and FEC decoding is performed. The architecture of the receiver shown allows the implementation of DOCSIS and DVB standards on a single hardware platform. In addition, the architecture is optimized for an implementation with an application-specific integrated circuits (ASICs).**

Keywords – **CATV system, Return Path Receiver, DOCSIS, DVB-C, digital front-end, DSP.**

I. INTRODUCTION

Modern CATV system are characterized by the application of digital and optical technologies. Cable distribution networks are bi-directional that makes it possible for additional services (such as Internet access, VoD, VoIP etc.) to be provided to the subscribers. Two-way transmission of high-speed interactive services is performed by Cable Modem Terminal System (CMTS) that is located in the headend or the hub. Cable Modem (CM) or Set-Top-Box (STB) is used in order to receive the data packets addressed to the subscriber and to transmit the data to the CMTS.

Two major systems were developed for the delivery of high-speed interactive services across cable networks: DOCSIS and DVB (DVB-C and DVB-RCC). These two systems are incompatible, although there is an extension to DOCSIS called Euro-DOCSIS, which adapts DOCSIS to the European cable environment, but changes are only at the physical layer. Therefore, there is a need to develop devices that can handle both standards. The application of such programmable and flexible devices in the headend and STBs allows implementation of multiple standards on the same hardware platform.

When building the software headend architecture a solutions similar to the software radio principle have been applied [1-2]. A flexible digital signal processing in downstream direction at the headend is rather straightforward as the hedend acts as transmitter of a certain number of data signals. Therefore the signal processing of the received upstream signal at the headend is most challenging because of the point-to-multipoint architecture of existing CATV networks.

The investigations in this work are focused on the digital front-end of the return path receiver used in the modular software headend architecture. The purpose was to choose the most suitable scheme solution and to specify the requirements to the single functional units.

II. FUNCTIONAL REQUIREMENTS

The required signal processing elements for upstream transmission according to the standards DOCSIS and DVB-RCC are [3-5]:

1) combination of time division multiple access (TDMA/A-TDMA/S-CDMA) and frequency division multiple access (FDMA);

2) upstream frequency range: 5 - 42 MHz (for DOCSIS) and 5 - 65 MHz (for DVB);

3) supported channel widths: 200 kHz, 400 kHz, 800 kHz, 1.6 MHz, 3.2 MHz, 6.4 MHz (for DOCSIS) and 1 MHz, 2 MHz and 4 MHz (for DVB);

4) modulation schemes: QPSK or 16QAM (for DOCSIS) and differential QPSK and 16QAM (for DVB);

5) differential-coded and Gray-coded symbol mapping for DOCSIS;

6) spectral shaping: square-root raised cosine with roll-off factor $\alpha = 0.25$ (for DOCSIS) and $\alpha = 0.3$ (for DVB);

7) randomization: polynomial $1 + x^{14} + x^{15}$ (for DOCSIS) and $1 + x^5 + x^6$ (for DVB) with programmable seed value for DOCSIS;

8) forward error correction (FEC): Reed-Solomon (RS) decoder with variable input length and variable error protection - 18 … 255 bytes data input, 0 … 10 bytes error protection (for DOCSIS) and RS(59,53) decoder (for DVB);

9) upstream burst: programmable preamble with variable length (for DOCSIS), unique word (for DVB).

This multitude of requirements can be performed when most of the headen's functionality is realized in software.

One of the main causes to worsen communications over the reverse path channel of a CATV system is the noise funneling effect. Noise and interference couple into the network due to poor shielding, loose contacts, mismatches, etc. As this happens in all branches of the network, these influences accumulate while they propagate through the tree-and-branch network and sum up in the headend. In result, the carrier-to-noise ratio (CNR) and the carrier-to-intermodulation product ratio (CIR) at the receiver input of the CMTS are reduced to an unacceptable value and communications over the reverse path channel get worse or simply break off. Hence, when designing the reverse path it is of great importance to provide such values of parameters CNR and CIR that ensure given bit

[1] Lidia T. Jordanova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: jordanova@tu-sofia.bg.

[2] Dobri M. Dobrev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: dobrev@tu-sofia.bg.

error ratio (BER). The service delivery over CATV networks requires different BER values varying in the range of 10^{-4} to 10^{-7}.

Because of the frequency dependence of the transfer function and the noise, the return channels have different CNRs. This can be taken into account by applying different constellation sizes on the channels. For low SNR, a robust scheme like QPSK is appropriate while for high SNR higher order QAM schemes are more suited. A necessary condition is that neither the transfer function nor the noise power vary over time. This is fairly true for the broadband and the narrowband noise, but not for the impulse noise. As these influences are not localized in the time or frequency domain, the modulation scheme alone can offer no remedy, but channel coding is the proper antidote. With forward error correction (FEC), the interference due to the impulse noise can be combated.

III. SOFTWARE HEADEND ARCHITECTURES

A new concept that combines the digital signal processing requirements for both standards, the so-called software headend, was introduced in [6-7]. This concept is derived from the software radio principle that was introduced in mobile communications.

There are three main types of software headend that each use different techniques to separate the upstream channels – modular, parallel and FFT-based.

The modular software headend architecture manifolds the signal channel structure M times to support M upstream channel. All digital signal processing units for one upstream channel are grouped into module. This architecture offers most flexibility regarding the number of upstream channels because modules can simply be added or removed. The main disadvantage of this solution is the large amount of identical functional units if many upstream channels have to be supported.

Fig 1 shows the digital signal processing elements required to demodulate and decode one specific channel. Here the input signal is the completely digitized upstream spectrum that includes the return signals of all active software terminals (cable modems and STBs). The sampling rate of the incoming complete signal is selected to be about 150 MHz. In the digital down converter (DDC) a channel selection and down-conversion of the selected RF signal to baseband is performed. Channel selection is done by means of numerically controlled oscillator (NCO), whose frequency and phase are adjusted from signal formed in the carrier recovery functional unit. The resulting baseband in-phase (I) and quadrature (Q) streams are filtered by digital low-pass filters (DLPF), to form the required baseband digital signals. These filters have to provide all possible roll-off factors and bandwidths. In addition to filtering, the DLPFs also reduce the sampling rate to values, required by the Nyquist condition. The decimation factor D depends on the bandwidth of the selected signal.

The signal detector tests the baseband digital signal for a specific pattern than occurs when the preamble is received. For DOCSIS, the preamble is chosen in such a way that the transmitted symbol sequence is composed of P symbols alternating in the first and third quadrant of the complex plane. For DVB, the fixed pattern and the fixed length of the unique word can be used only. When the preamble passes through the detector, a peak with linear increase and decrease can be observed. The position of the pick indicates the beginning of an upstream burst and is used for timing recovery. The timing recovery functional unit passes data to the resampling block that regains the exact sampling point of time. After resampling, the blocks for symbol decision, demaping and differential decoding follow. Finally, the functional units for derandomization and a FEC decoding supply the output bitstream of the selected channel.

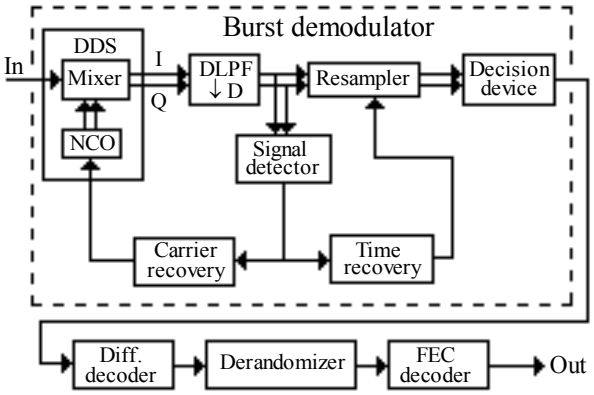


Fig. 1. Single module of the modular software headend

The functional units that are used in the parallel software headend perform digital signal processing of all upstream channels. Each unit has got M parallel inputs that are switched to the selected functionality (each upstream channel requires its own set of parameters). As it is shown in Fig. 2, the complete digitized upstream spectrum passes first through M-channel filterbank with decimation. This block has to be flexible, because the matched filters support the different bandwidth and the central frequencies have to be adjustable to currently used carrier frequencies. In the burst demodulator the following operations are performed: downconversion of the selected RF signal to baseband, signal detection and resampling.

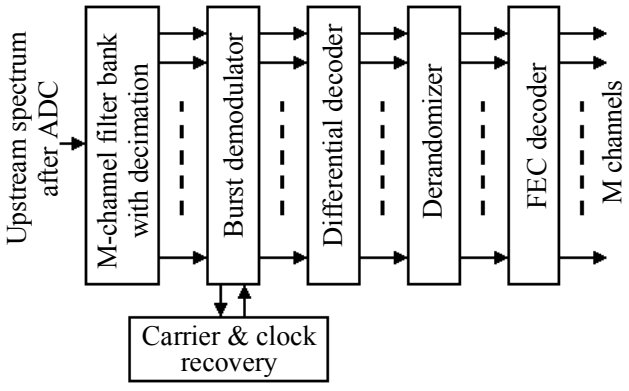


Fig. 2. Parallel software headend

It is evident, that in contrast to the modular software headend the filtering and decimation in this architecture are pulled out of the burst demodulator. This results in a modified

burst demodulator with less complexity that is completely clocked with the decimated sampling frequency. The main advantage of the parallel software headend is that only one implementation of each functional unit is required, resulting in less hardware amount but significant faster circuits.

The FFT-based software heandend makes use of one or more fast Fourier transforms to separate the upstream channel.

IV. Block diagram of the return path receiver

It is well known that the heterodyne receiver can easily adapt itself to many different standards requirements achieving a very good sensitivity and selectivity. However the need of the large number of external components, i.e. the image rejection filter, and the complexity of the structure causes problems if a high level of integration is necessary and flexibility features have to be implemented.

These disadvantages of the heterodyne receiver can be avoided through using architectures that allow many of the conventional receiver functions, such as channel selection, demodulation and decoding to be implemented (performed) by digital signal processing. Therefore it is necessary to convert the complete upstream signal to digital by the analog-to-digital converter (ADC). As the bandwidth of the received signal is 65 MHz, this would require an ADC with a sample rate of a least 130 MHz. It is well known, that for conversion speeds of 1 MHz to 100 MHz, the pipelined ADC finds its best position. The resolution N of an ideal ADC can be calculated from the following equation:

$$CNR_p[\text{dB}] = 6.02N + 1.76 \text{ dB}, \quad (1)$$

where SNR_p is the peak value of this parameter.

A general block diagram of the return path receiver is shown in Fig. 3. The complete upstream spectrum in the range of 5 to 65 MHz is selected by a bandpass filter (BPF) and after amplification is digitized through a wideband ADC. In this receiver, the ADC is operated at a 153.6 MHz sampling rate with resolution of 10 bit/sample. A variable gain amplifier (VGA) is used to adjust the received signals to the dynamic range of the ADC. After A/D conversion, the complete signal is sent to the fully digital front-end.

The first function of the digital front-end is to convert the spectrum of a desired signal to baseband. Channel selection is done by means of numerically controlled oscillator (NCO), whose frequency is controlled by the DSP. The baseband signal is then passed to a cascaded integrator comb (CIC) filter, where digital filtering and decimation is performed. For the different bandwidths the sampling rate can be decreased by different decimation factors D_k ($k = 1, \ldots, 8$). The final stage of the digital front-end is the matched filter, which performs square-root raised-cosine (SRRC) Nyquist filtering. This filter has to provide all possible roll-off factors and bandwidths required for the standards DOCSIS and DVB-RCC. The required bandwidth, roll-off factor and decimation factor are adjusted programmable by means of digital signal processor (DSP).

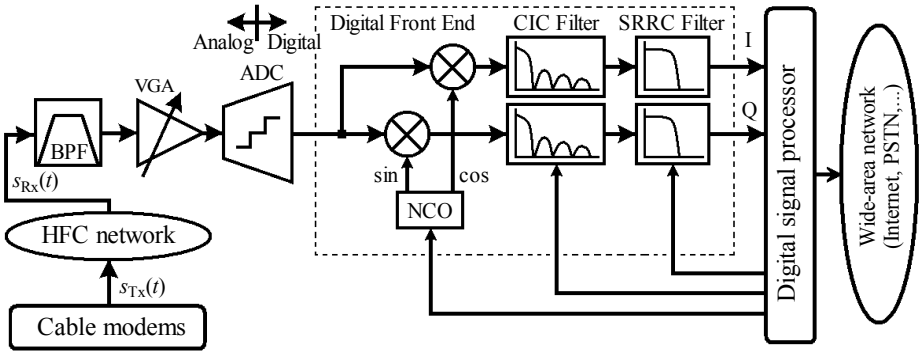


Fig. 3. Block diagram of the return path receiver

The DSP unit includes digital demodulator, derandomizer and FEC decoder. The basic function of the digital demodulator is to perform carrier and timing recovery, channel equalization, ingress noise cancellation, and make symbol decision and demapping. More information about this functional unit and applied techniques for signal processing is given in [7-8].

V. Detailed description of the digital frond-end

In Fig. 4 a simplified block diagram of the NCO is shown. The NCO contains sine and cosine generators which can be viewed simply as ROM-based Look Up Tables (LUT) that perform the following functions:

$$\begin{aligned} \sin[n] &= \sin[2\pi n/N] \\ \cos[n] &= \cos[2\pi n/N], \end{aligned} \quad (2)$$

where: n is the address input to the LUT, N is the number of samples in the LUT. Incrementing n from 0 to $N-1$ causes the LUT to output one complete cycle of amplitude values for the sine and cosine functions. The time required to increment n from 0 to $N-1$ is the period of the sine and cosine waveforms produced. The LUT address is incremented once each cycle of the clock by an amount equal to the phase word input. The phase angle is accumulated and stored in the phase

accumulator register. The output of the phase accumulator is used to address the sine and cosine LUT's.

The frequency of the sine and cosine waveforms f_{NCO} is defined as follows [9]:

$$f_{NCO} = f_{clk} * \text{phase}\left[(k-1)\ldots 0\right]/2^k , \quad (3)$$

where: f_{clk} is the frequency of the input clock, phase[(k -1)…0] is k-bit tune data. The frequency tuning resolution is given by:

$$\Delta f_{NCO} = \pm 0.5 f_{clk}/2^k . \quad (4)$$

Two digital multipliers are used to heterodyne, or mix, the input data samples with the NCO quadrature waveforms. The downconversion process has translated the signal of interest (a purely real signal located at 5-65 MHz) down to baseband (a complex signal located at 0 Hz).

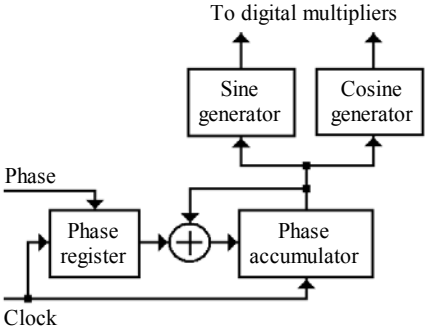


Fig. 4. Simplified block diagram of the quadrature NCO

Here presented digital frond-end contains two filter types. The decimating CIC filter is composed of three sections. The first section consists of digital integrators and operates at the A/D sampling rate. After integrator section follows a rate change, or decimation, section. The decimated samples feed into the comb filter section, which operates at the sample rate divided by D_k. As the CIC filter does not require any multipliers and is therefore useful at high sampling rates, its filtering characteristics are severely limited. That is way its essential function is to decrease the sampling rate and to keep passband aliasing within specified limits.

The system function for the CIC filter, referenced to the high speed sampling rate, is found by combining the transfer function of the integrator and comb sections, i.e. [9]

$$H(z) = \frac{\left(1-z^{-D}\right)^P}{\left(1-z^{-1}\right)^P} = \left[\sum_{m=0}^{D-1} z^{-m}\right]^P . \quad (5)$$

where: P is the number of stages, D is the decimation factor. The frequency response for the CIC filter is given by

$$H(f) = \left(\frac{\sin(\pi f)}{\sin(\pi f/D)}\right)^{2P} . \quad (6)$$

where f is the sampling frequency f_s relative to the decimated rate, f_s/D.

The square-root raised cosine (SRRC) filter is used primarily for isolating the signal of interest by attenuating out-of-band signal energy. Its frequency-domain description is piecewise function, given by

$$H(f) = \begin{cases} T^{1/2}, & f \le (1+\beta)/2T \\ \left(\frac{T}{2}\right)^{1/2}\left\{1+\cos\left[\frac{\pi T}{\beta}\left(f-\frac{1-\beta}{2T}\right)\right]\right\}^{1/2}, & \frac{1-\beta}{2T} < f < \frac{1+\beta}{2T} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where $0 \le \beta \le 1$ is the roll-off factor and $T = 1/f_s$. The bandwidth of a SRRC filter can be calculated as

$$BW = \frac{1}{2T}(1+\beta) . \quad (8)$$

VI. Conclusion

The return path receiver presented in this paper provides very good sensitivity and selectivity and satisfies the high quality requirements that exist in professional equipment for CATV headends. The receiver is based on direct digitization of the upstream channel spectrum and makes use of highly sophisticated algorithms for timing and carrier recovery, channel equalization, and ingress noise cancellation. Using this concept, a 4-input/16-output receiver can be integrated in a single chip, which also includes 4 downstream modulators and critical medium access control (MAC) functions. A software-controlled switch integrated in the chip allows sending to each one of the 16 digital front-ends and demodulators the desired input port signal, and then each digital front-end selects the desired upstream channel. In addition to the increased CMTS density, such receiver architecture also offers cable operators full flexibility in network planning and handling the evolution of their cable plants without human intervention at hubs and cable head-ends.

References

[1] J. Mitola, "Software Radio Architecture", John Wiley & Sons, 2000.
[2] V. Giannini, J. Craninckx and A. Baschirotto, "Baseband Analog Circuits for Software Defined Radio," Springer, 2008.
[3] ETSI standard ES 201 488 v1.1.1, "DOCSIS Radio Frequency Interface Specification," Nov. 2000.
[4] DOCSIS Radio Frequency Interface Specification, version 1.0: SP-RFII05, Nov. 1999, version 1.1: SP-RFIv1.1-I06, Dec. 2000.
[5] ITU-T Rec. J.83, "Digital multi-programme systems for television, sound and data services for cable distribution," Apr. 1997.
[6] A. Braun, J. Speidel, H. Krimmel, "The Software Headend Architecture - A New Approach for Multi-Standard CATV Headends," IEEE Conference on Consumer Electronics, Los Angeles, Digest of Technical Papers, pp. 156-157, Jun. 2002.
[7] A. Braun, J. Speidel, H. Krimmel, "A Programmable, Flexible Headend for Interactive CATV Networks," IEEE, ICCE, Los Angeles, Juni, pp. 255-259, 2002.
[8] F. Buda, E. Lemois, H. Sari "An advanced CMTS receiver architecture for DOCSIS 1.X/2.0 cable access networks", Wed, 11 Jun 2003.
[9] Д. Добрев, Л. Йорданова "Радиокомуникационна техника", СИЕЛА, С., 2006.

Presentation of the Results of Measuring Characteristics of Power Line Installations in the Signals Transmission

Jasmina Mandić-Lukić[1], Bojan Milinković[2], Nenad Simić[3], Nedžad Hadžiefendić[4]

Abstract – **The subject of this paper is the presentation of research results of contemporary electrical cables characteristics in the function of signals transmission in the frequency range from 1 to 30 MHz, which is required by standard ETSI TR 102 494 for communications within the residential and business buildings. Results of analysis of the standard power line network topologies in the typical residential and business facilities are presented, also. Specific phenomena that are manifested in the signals propagation through electrical network are pointed out.**

Keywords – **Research results, Power line network, Residential and business facilities.**

I. INTRODUCTION

Power line networks are designed and implemented according to the criteria of optimization of basic function enforcement – transfer of electricity to final consumers – a variety of devices in the business and residential facilities. This fact causes that the installations, in the frequency bands for transmission of the information and multimedia signals, have characteristics that are, more or less different from the characteristics of networks specifically built for these signals. The most influential differences are constructive and technological characteristics of the cables, installation topologies in the business and residential facilities, variations of installation parameters caused by the changes in the consumer states, as well as the presence of specific additive interferences and noises.

II. ELECTRICAL INSTALLATION CABLES

The main elements of the electrical installation cables are copper wires which with insulating layers around them form the structures, hereinafter called core. One or more cores in the same sheath form cable. The standard dimensions of the cables are expressed as part of wire cross sections, given in mm^2. Cables with three or five cores with cross sections of 1.5, 2.5, 4 or 6 mm^2 are applied in most residential and business facilities. In cable forming process cores are led straight and parallel, as opposed to telecommunications cables, where the twisting is done in pairs or quadruplets.

Insulating layers around the core are made of dielectric materials, primarily of extruded polyethylene and polyvinyl – chloride. In general, cables with polyethylene isolation are very superior for the transmission of high frequency signals (up to 30 MHz), because the dielectric losses in this material are approximately two order of magnitude smaller than in the polyvinyl – chloride. However, as the lengths of cables in the system are small, order about 10 meters, cables with polyvinyl – chloride insulation are fully satisfying.

For assessing the functionality of the installation cables as the telecommunications transmission medium it is necessary to determine their characteristic parameters in the specified range of frequencies. The basic parameters are the transfer function, characteristic impedance and propagation velocity of the signal. In further analysis, as well as in experimental verifications, cables are treated as homogeneous two – conductor transmission lines, regardless of the actual number of cores in them. Calculations were made for the cable type 3x2.5mm^2. Cross – section of this cable type is given in Fig. 1.

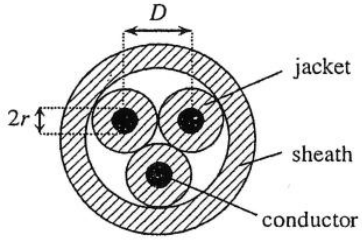


Fig. 1. Cable cross – section [2]

Primary distributed parameters are determined by standard formulas 1 – 6 [2]:

$$R = \sqrt{\frac{2\mu_r\mu_0 f}{\pi\sigma r^2}}\left[\frac{\left(D/2r\right)}{\sqrt{\left(D/2r\right)^2-1}}\right] \quad (1)$$

$$L_{ex} = \frac{\mu_r\mu_0}{\pi}\cosh^{-1}\left(\frac{D}{2r}\right) \quad (2)$$

$$L_{in} = \frac{R}{2\pi f} \quad (3)$$

$$L = L_{in} + L_{ex} \quad (4)$$

[1] Jasmina Mandić-Lukić is with Energoprojekt – Entel, Bulevar Mihaila Pupina 12, 11000 Belgrade, Serbia,
E-mail: jmlukic@ep-entel.com.
[2] Bojan Milinković is with Energoprojekt – Entel, Bulevar Mihaila Pupina 12, 11000 Belgrade, Serbia,
E-mail: bmilinkovic@ep-entel.com.
[3]Nenad Simić is with Faculty of Electrical Engineering, Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia.
[4]Nedžad Hadžiefendić is with Faculty of Electrical Engineering, Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia.

$$c = \frac{\pi\varepsilon_r \pi\varepsilon_0}{\cosh^{-1}\left(D/2r\right)} \quad (5)$$

$$G = 2\pi f C \tan\delta \quad (6)$$

On the primary parameters basis the terms for the attenuation constant, propagation velocity and characteristic impedance are delivered and given by formulas 7, 8 and 9 [2]:

$$\alpha = \frac{R}{2Z_0} + \frac{GZ_0}{2} \quad (7)$$

$$c = \frac{c_0}{\sqrt{\varepsilon_r}} \quad (8)$$

$$c_0 = \sqrt{\frac{L}{C}} \quad (9)$$

Eq. (7) can be expressed in the form of:

$$\alpha\left(dB/m\right) = \alpha_1^{-5}\sqrt{f} + \alpha_2^{-9} f \quad (10)$$

For standard cable PP-Y 3x2.5 relevant information are:

$2r = 1.8\text{mm}$, $D = 43.6\text{mm}$, $\varepsilon_r = 3.5$, $\tan\delta = 0.08$.

For such a cable, equation for path attenuation may be presented:

$$\alpha\left(dB/m\right) = 2{,}3 * 10^{-5}\sqrt{f} + 3{,}3 * 10^{-9} f \quad (11)$$

In Eqs. (10) and (11) frequency is introduced in Hz. In the same way it is found that value of characteristic impedance is 87Ω and propagation velocity is $0.53c_0$.

With additional calculating it is established that Eq. (11) can be used, with satisfactory accuracy, for cables with cross – sections of 1.5mm^2 and 4mm^2. Values of parameter (D/2r), which has a dominant influence on the results, are changed within range of ±10 % and this makes the variation of coefficients values in Eq. (11) below 5%.

Results shown in [2] can be used in favor of previous attitude. For cables NYM 3x1.5 mm^2 and NYM 3x2.5 mm^2 (standard VDE – 0250) had been carried out calculations and measurements of path attenuations in the frequency band 2.5 – 30 MHz . In the lower frequency band, even up to 10 MHz, values of path attenuations in both cables are identical. With further increase of frequency, a slightly greater increase of attenuation is noticed in cable 3x1.5 mm^2, and on 25 MHz attenuation difference is 0.04 dB/m.

Experimental verification of the calculation according to the Eq. (11) had been done by measurements on the standard installation cables PPY 3x1.5 mm^2 , PPY 5x2.5 mm^2 and PPY 3x4 mm^2. Measurements were carried out in homogeneous segments of cables, 25 meters length.

A fully satisfactory concordance of the calculation and measurement results is determined, as well as data from the cited literature. Values of path attenuation are within the limits of 0.04 dB/m at the frequency of 1 MHz to 0.24 dB/m at the frequency of 30 MHz . The average value of the characteristic impedance is 90Ω with variations of ±3Ω. Propagation velocity is very close to 150000 km/s, which means that the wavelength is approximately two times smaller than in free space.

The diagrams of calculating value of path attenuation dependence of frequency, the results of measurements in the mentioned installation cables, as well as for representative results from the literature, are shown in Fig. 2.

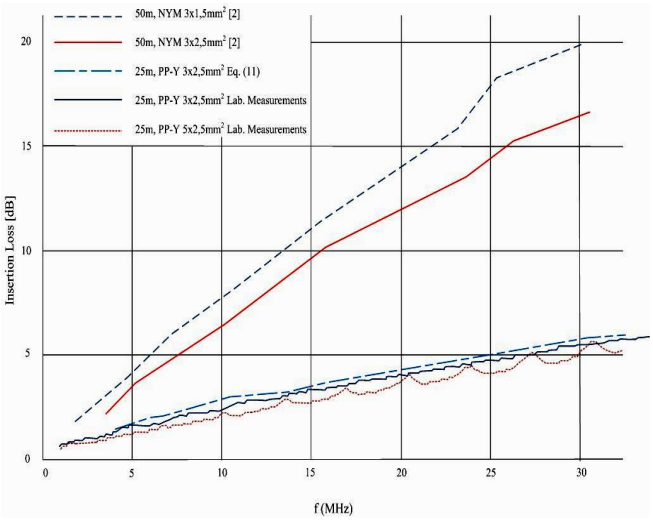


Fig.2 Insertion loss in function of frequency and type of power line cables

On the basis of presented material it follows the conclusion that the installation cables can be used for quality communications at distances of several tens and up to a hundred meters. Specific limits are determined by spanned attenuations of devices, installation topology, as well as communication modes.

III. TOPOLOGIES OF POWER LINE INSTALLATIONS

The standard power line installations in the facilities have basically the tree structure with two levels of branching. At the entrance to the facility (as a rule) is the main distribution cabinet where the first branching is done and from this point cables are led to residential or business units. Distribution boards exist in the each of these units where the second, local branching is done and from these boards cables are led to the individual consumers. Further discussion is limited to communications within the units, residential or business, i.e. to the local communication networks.

Local installations always have more or less extended structures with multiple branching points where the branches lengths are usually related with wavelengths of the signals. In

the Fig. 3 is shown, as an example, single line diagram in a typical medium – sized residential unit.

The central point of this electrical installation is a distribution board where a large number, 15 – 20 or even more, of separate electrical circuits are met. These circuits are realized of the same number of three – core or five – core cords through which various consumers are powered. Often, these cables branch further to individual consumers, lighting bodies or power sockets.

In principle, it is considered that local network should provide the possibility of communication between any two plug points, which means that signal injected in one point should have, in all other plug points, characteristics which are sufficient for the quality of communication.

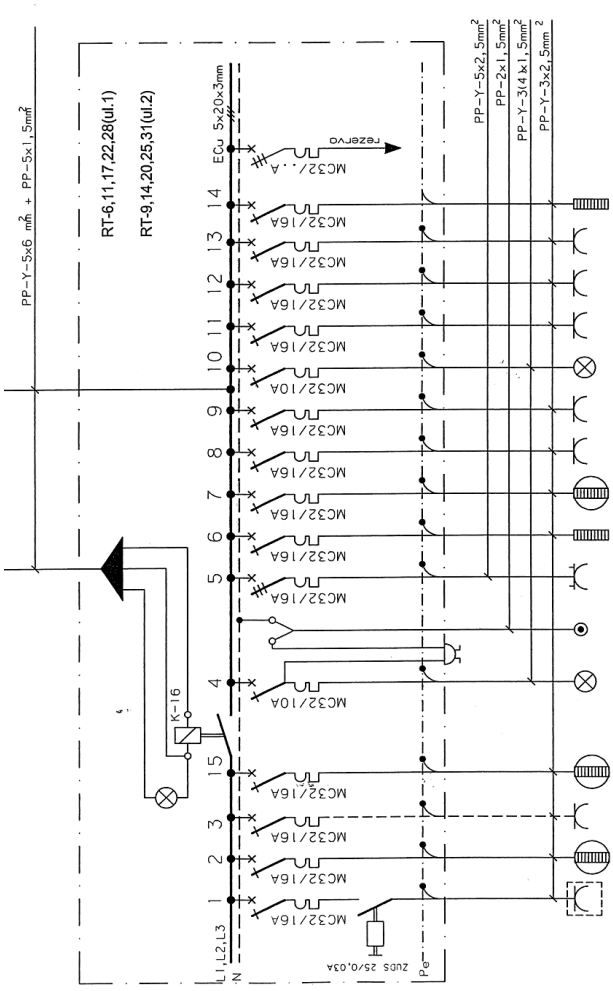


Fig. 3. Single line diagram of residential unit

In general, during the transmission between two plug points, a signal passes through several different branching points, and each of them is the discontinuity that degrades transfer conditions. Particularly adverse effect is manifested in the cases when the cables, which are branching during the way, are relatively short and the ends of them are open circuit or short-circuit. At the frequencies where the length of branches is close to multiples of signal wavelength quarters, input impedance of a branch becomes very small and represents a short-circuit for a transferred signal.

In the band 10 – 30 MHz and in the conductors with relative permittivity 3 – 4, wavelength is within the limits of, approximately, 5 and 15 meters, which means that each branch of length 1 – 4m is a potential "trap" for the signal. The results of measurement on the laboratory model are shown in [3], which had branches of length 1.6 – 2m and at the ends of them were open circuit state. In Fig. 4 is shown a transfer function against frequency diagram of that model. Large drop of transfer function module (i.e. increase of attenuation) is located around the frequency of 24 MHz, in which the wavelength quarter is very close to length of branch 1.6m.

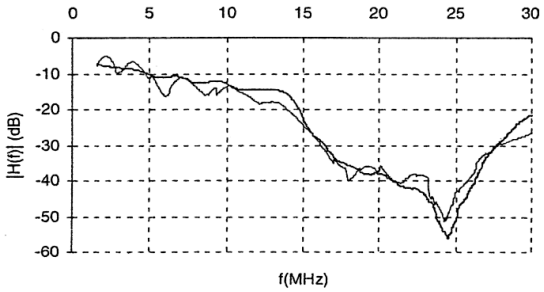


Fig. 4. Amplitude response of the sample channel; soft line represents simulation, and bold one corresponds to measured data [3]

Particularly important place in this consideration belongs to distribution board, which is a local communication network node. On the distribution board connectors a large number (up to and over 20) of conductors that belong to a separate electrical circuits are met. From the aspect of communication network that is a parallel connection of the same number of lines with very different nature and values of input impedances. Calculating of resultant impedance value in this point is quite complex and the value itself is exposed to variations due to changes in the installation states. The measurements showed that thermal component has an average value of approximately 10Ω, while the reactance changes in the limits of +/- a few tens of Ohms.

Low values of impedances in these points, as well as variations in the function of frequency, are the main causes for additional attenuations on the connections between the installation points that are at different electrical circuits. It has been determined that these attenuations are, on average, 15 – 20 dB, with a dominant frequency dependence.

The results of simulations, calculations and measurements in the laboratory model, which approximately corresponds to the configuration of smaller residential unit, are presented in [4]. Analysis and measurements were performed in the frequency band from 5 to 30 MHz. Graphical display of results is given in the Fig. 5. Large variations of transfer function module are evident in the range of about 30 dB. Variations are quasi-periodic with a period of about 5 MHz, and the average value has a tendency to descend with increasing of frequency in the range of approximately 5 dB.

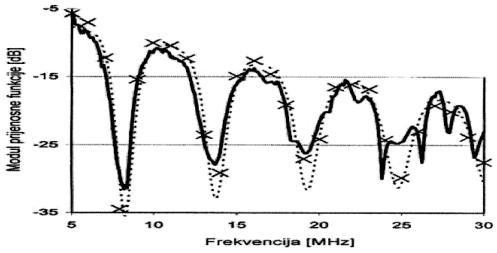


Fig. 5. Graphical display of results for laboratory model; dotted line represents simulation, bold line is for analytical measurements and crosses stand for laboratory measurements [4]

It is concluded that the same approach can be applied to the building as a whole. Single line diagram of a typical building part is shown in the Fig. 6. In this case the node of network is the main power distribution board, while the terminal points are the distribution units in the flats. It is also noted that the same approach can be applied for the standard administrative and business buildings, where the nodes are usually in the power distribution boards on every floor of the building.

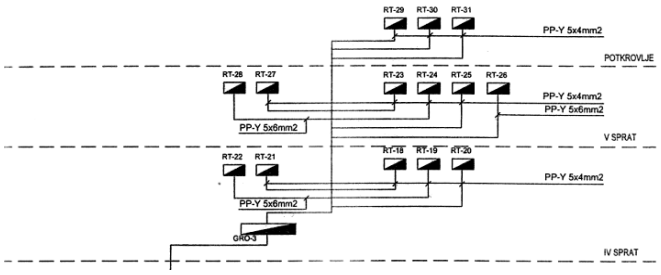


Fig. 6. Single line diagram of a building floor

IV. CONCLUSION

In the paper the results of the first research phase in a program [5], [6] are presented. The objectives of research in this phase were the identification of power line installations properties as a medium for transfer information and multimedia signals in frequency bands from 1 to 30 MHz. It was found that the standard installation cables, as homogeneous lines, can ensure quality transmission of signals to the lengths of several tens, and over a hundred meters.

Crucial influence on the potential of the whole electrical installation, viewed as a medium for transfer, have the local topology, and in particular the nodes – distribution boards and distribution cabinets. Accordingly, the main target of further work should be finding the ways of conditioning these segments of installations, and that will lead to the most efficient usage of the installation potential.

REFERENCES

[1] T. Esmailian, F.R. Kschischang, P.G. Gulak, "In-building power lines as high-speed communication channels; channel characterization and a test channel ensemble", 2003.

[2] I.C. Papaleonidopoulos, C.G. Karagiannopoulos, N.J. Theodorou, "Evaluation of the two-conductor HF transmission-line model for symmetrical indoor triple-pole cables", 2006.

[3] I.C. Papaleonidopoulos, C.G. Karagiannopoulos, N.J. Theodorou, "Travelling-wave multipath simulation of two-conductor HF signalling over indoor power-line networks and RMS-delay-spread dependence", 2006.

[4] Dr. Sc. Dubravko Sabolić, "Model propagacije signala u razdjelnim elektroenergetskim mrežama u frekvencijskom području od 5 do 30 MHz", Zagreb, Energija 4, 2005, HEP.

[5] "Unapredjenje sistema i razvoj tehničkih uredjaja za prenos informatičkih i multimedijalnih signala posredstvom elektroenergetskih instalacija i niskonaponskih mreža", Ministarstvo nauke Repuplike Srbije, istraživački program 2008 – 2010.

[6] J.M. Lukić, D.Pokorni, N. Simić "Characteristics of Low-Voltage Installations as Transport Medium in Office Buildings", ICEST 2008.

[7] Standards:
- ETSI TR 102 494,
- VDE – 0250.

Investigation of the impact of CSO/CTB/CNR parameters in designing and operating CATV networks

Stanimir Sadinov[1], Krasen Angelov[2], Kiril Koitchev[3], Ivelina Balabanova[4]

Abstract **–This article reviews the impact of CSO (*Composite Second Order*) and CTB (*Composite Triple Beat*) upon the channel spectrums. Both coaxial cables and laser optical fiber lines are used as transmission lines which makes it necessary to account for and investigate the effect of pure Gaussian noise within the systems as well as the errors due to it by measuring the parameter of carrier-to–noise ratio (CNR). All CSO/CTB and CNR data are related to the noise level and are directly connected with the active and passive elements used in the system. Their correct disposition and the tunings made in them contribute largely to improved service quality and use of network.**

Keywords **– CATV, CSO/CTB/CNR spectrums, signal level, nonlinear distortion**

I. INTRODUCTION

Availability of a large number of channels in present day CATV networks or of some other types of signals with marked carrier frequencies in channel multiplexing brings about to considerable nonlinear signal interaction. For most cable TV frequency plans products which have resulted from beat, accumulate in given frequencies and to a large extent affect network quality. Usually discrete CSO (*Composite Second Order*) and CTB (*Composite Triple Beat*) composite inter-modulation products are the ones which exert the greatest impact. The CSO/CTB spectrum is a sum total of the products of the nonlinear interaction of all possible combinations of second and third order input signal frequencies and the level of nonlinearity of amplifiers. It contains hundreds and thousands of inter-modulation frequencies distributed along the entire frequency band of the cable TV network. There are many cut-and-try formulae used to calculate only one value: the worst value of CSO/CTB from the test plan, however, all these formulae are of various coefficients and therefore are difficult to summarize.

For the purpose of analyzing a network it is necessary to have not only the worst value of CSO/CTB in the frequency band, but also the concrete distribution of that spectrum over the channels.

[1]Stanimir Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H.Dimitar St., 5300Gabrovo, Bulgaria, E-mail: murry@tugab.bg

[2]Krasen Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@mail.bg

[3]Kiril Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

[4]Ivelina Balabanova is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: ivstoeva@yahoo.com

Very often a concrete inclination (slope) of amplitude frequency response (AFR) components is set for the purpose of improving network parameters. In this case the calculation of CSO/CTB spectrum will get more complicated. In some cases this inclination (slope) can cause worsening of inter-modulation noise depending on the type amplifier used. For this reason the use of cut-and try formulas in calculating the value of CSO/CTB is very problematic.

In this connection there arises the actual task to calculate the full spectrum of CSO/CTB for randomly selected operational frequency plan with random non-uniformity of the input and output signals for a certain network elements.

Another interesting task is the search for optimum plan that will ensure the lowest value for CSO/CTB in the channels.

II. CALCULATION OF CSO/CTB SPECTRUM OF NON-LINEAR ELEMENTS

It is assumed that the input signal of the non-linear element (the amplifier) contains *N* (*Continues Wave – CW*) *RF* (*Radio Frequency*) carriers:

$$x = \sum a_i \cos(\omega_i + \varphi_i). \quad (1)$$

For the given element it is possible to determine the non-linear transfer function *A* which depends on element's characteristic and does not depend on the type and parameters of the input signal. By means of this function it is possible to express the output signal *y* as some function of the input signal *x*. This can be rendered as $y = Ax$.

It is very difficult to draw such transfer function even for the simplest type of amplifiers given the considerable number of spurious parameters. Nevertheless, the function can be synthesized for the given non-linear element on the grounds of the results from the tested spectrum of CSO/CTB in any plan. Since such a transfer function does not depend on the type of input signal (and the non-uniformity of input AFR as well) it is possible to calculate the full output spectrum including the CSO/CTB spectrum for random frequency plan of channels which can be different, due to various needs, from the frequency plan of channels during their first test and synthesis of the non-linear transfer function. An algorithm for calculation of coefficients of such non-linear transfer function is considered in [1, 3].

Fig.1 and fig.2 present CSO/CTB spectrums which are calculated for amplifier No2 (Model NWA-M1). Short columns are the control values for CSO/CTB tested in lab settings. Channels are indicated by vertical grey stripes.

To calculate CSO/CTB spectrums and noises in a laser optical cable TV network it is possible to use the physical model of the laser with accounting of its physical parameters in accordance with the results from the tests of CSO/CTB spectrums. It is also necessary to take into account the impact

of the spurious modulation within the laser (chirp) as well as the availability of dispersion inside the optical fiber. This allows to calculate the full spectrum of the inter-modulation CSO/CTB products and the noises in the cable TV network for all frequency ranges. For lasers with direct modulation it is also possible to use the above method for calculation of CSO/CTB spectrums whereas lasers with external modulation require a little different approach and accounting of additional factors.

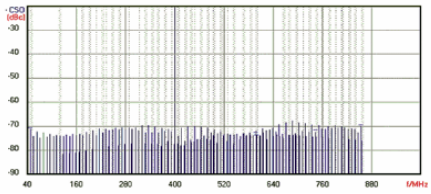


Fig.1.Calculated CSO spectrum for amplifier №2

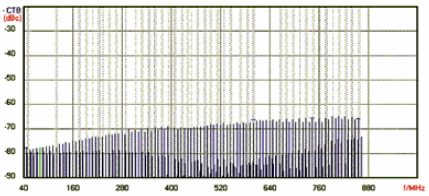


Fig.2.Calculated CTB spectrum for amplifier №2

III. IMPACT OF CSO/CTB/CNR SPECTRUMS ON BER IN DIGITAL QAM CHANNELS

The use of MPEG-2 (*Motion Picture Expert Group*) standard for compression and transmission of TV image allows the simultaneous transmission of up to ten TV programs in the *8MHz* frequency band. The idea of MPEG-2 standard is based on transmission which in turn is based on changing parts of TV image calculated as interim image frames and using Reed-Solomon (RS) error correction codes which allow correction of definite number of erroneously received symbols and quadratic amplitude modulation QAM. Initially digital data (presented as ones and zeros) is integrated into packages of definite length and synchronizing symbols and then is filled in with a series of correction symbol which are calculated for the concrete package. QAM modulation is done in accordance with the definite data symbol (the combination from ones and zeros) and two orthogonal I and Q signals with concrete combinations for amplitude and phase for each of them. In the receiving end of the line takes place demodulation QAM of the signal and correction of erroneous symbols by means of FEC scheme (*Forward Error Correction* – progressive error correction). Such presentation of output data allows for substantial reduction of required frequency band for transmission. Correction codes such as Reed-Solomon allow to correct a certain number of symbols in the package. For example in a packet with Reed-Solomon code (*204*, *188*) it is possible to correct up to *8* faulty reception symbols. A very important characteristic of quality of transmission of digital signals is the *BER* (*Bit Error Rate*) parameter – the ratio between error bits and the total number of transmitted bits.

In terms of history at first *BER* deterioration was connected with white Gaussian noise. Reed-Solomon code allows effective correction of errors caused by Gaussian noise. After that a number of papers showed the large share of pulse noise as well [4, 5] combined with white Gaussian noise for BER [1, 2] deterioration in laser optical lines with mixed AM/QAM channels.

A number of articles [4, 5, 7] present measured time characteristics of the sum total of noises and the inter-modulation products within the optical line and in the frame of QAM channel of intermediate frequency. Despite that QAM signal is decoded not at intermediate frequency in the *8MHz* band, but in the basic frequency band *0 – 4,0 MHz*, in which I and Q components are separated back by means of mixing the signal in the middle of QAM frequency band at intermediate frequency as that of the local oscillator (heterodyne) signal. In this way the basic frequency band of the channel for I and Q components proves to be twice as large as compared with the band for QAM high frequency channel., therefore ,the frequency spectrum and its relevant time characteristic will be quite different. Time and spectrum measurements in the basic frequency band *0 – 4 MHz* for 64-QAM signal for mixed AM/QAM channel plan are dwelt upon and presented in [7].

Fig.3 presents an experimental formulation for measurements in a hybrid AM/64-QAM transmission system. QAM channels are generated by 64-QAM modulator according to the Digital Video Broadcasting standard. The packet contains Reed-Solomon (*204*, *188*) code. Bit transmission rate is e *41,25Mbps*, symbol rate is *6,875Msym/s*, and the channel band – *7,8MHz*. The matrix generator generates unmodulated CW carriers according to the European frequency plan from *48,25MHz* to *855,25MHz* with different number of channels (*60* to *90*). The levels of 64-QAM channels are in accordance with the standard; by 10dB lower than CW carriers. Combination of CW and QAM signals creates and an input signal for coaxial cable TV line or a modulated signal for DFB laser when measurements takes place in an optical line.

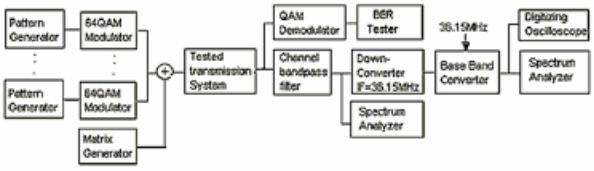


Fig.3. Experimental formulation for measurements in a hybrid AM/64-QAM transmission system

Output high frequency signal from the measuring transmission system (from the output of the optical receiver is for optical transmission system and from the output of the last amplifier is for coaxial transmission system) is diverted to QAM demodulator and BER-meter and also to the other branch of the diagram above which is connected with frequency and time measurements in the QAM channel. The channel is split by means of a band filter and additional attenuator for the purpose of minimizing the impact of adjacent channels. White noise impact is measured by means

of spectral analyzer also taking into account *CNR* and CSO/CTB spectrum in the channel.

The step-down converter transfers the high frequency channel in the medium frequency band whose centre is at 36,15MHz. Then follows signal mixing of medium frequency 36,15MHz inside the heterodyne converter in the basic frequency band whereby the medium frequency signal is converted into basic frequency band *0 – 4 MHz*. In this way the individual I or Q component of the QAM signal is modeled. Time characteristics of distortions are measured by means of digital oscilloscope with a triggered scanning starting at a point when the signal reaches certain amplitude. Spectral analyzers are used to monitor the frequency spectrum of distortions in the basic frequency band. Signals in the measured channel as well as those in the adjacent channels are switched off when measuring noise and interferences. All CSO/CTB and CNR data are related to the level of CW signal.

Generally, an optical transmission system consists of a DFB immediately modulated laser, optical single mode fiber of *6km* in length, and an optical receiver. A coaxial transmission line contains three cascade amplifiers with attenuators as equivalent to the coaxial line. First is to be investigated the impact of Gaussian noise in the coaxial system. Errors that are due to mere Gaussian noise with *CNR* greater than *32dB* (in this case all CW channels are switched off) are corrected very well by the FEC circuit as $BER = 10^{-5}$ before the FEC circuit and $BER = 10^{-10}$ after it.

Like the optical lines discussed in [6, 7], the situation also changes here when there are discrete CSO/CTB products (switching on of channels). With 65 CW and 10 QAM channels what is obtained is that in a 64-QAM channel of *434MHz*, *CSO/CTB = 52dB* и *CNR = 48dB* there will be $BER = 10^{-6}$ before and $BER = 10^{-7}$ after FEC. In his particular case when CW channels are on then the efficiency of FEC drops down to *0*, i.e. there is no correction of faulty reception symbols. In this case efficiency boundary of forward error correction is observed not only for the modulated laser, but in the coaxial system with amplifiers. Consequently, it gives ground for the assumption that the efficiency boundary of FEC for multichannel AM/QAM system for cable TV displays similar nature both in coaxial and optical transmission system.

Fig.4 and Fig.5 show spectrums of inter-modulation products in a basic frequency band in QAM channels of coaxial and optical line. After conversion inside the basic frequency band, CSO/CTB products form harmonic order with first harmonic $f_1 = 0,25MHz$.

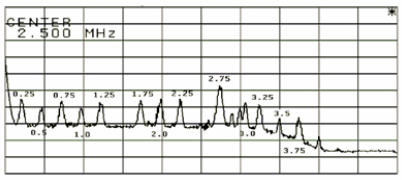


Fig.4. Coaxial transmission system–harmonics in basic frequency band of 434MHz channel

The highest level harmonic has a frequency of $f_m = 2,75MHz$. Harmonic sequences in fig.4 and fig.5 will be formed inside the basic frequency band with time sequence having a period $T = 4\mu s$. The values in fig. 4 and fig.5 are given in MHz; the horizontal divisions are of *0,5MHz* whereas those in vertical direction are *10dB*.

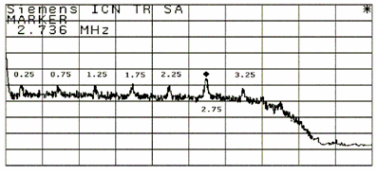


Fig.5. Optical transmission system –harmonics in the basic frequency band in a 858MHz channel

Fig.6 and fig.7 present results from time measurements in the basic frequency band for the a QAM channel. It is evident from the dependences that in the optical and the coaxial line the spectrum of CSO/CTB inter-modulation products in QAM channels form inside the basic frequency band of QAM channel some periodic pulse sequence with period $T = 1/f_1$ where f_1 is the first harmonic in the basic band of the digital channel. This is cyclic pulse sequence caused by the spectrum of CSO/CTB and appears to be the reason for the decrease of efficiency of the error correction circuit.

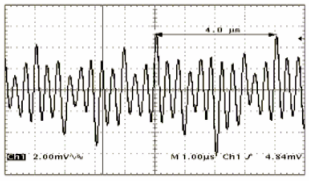


Fig.6. Measurement of interferences in time progression inside the basic frequency band of 434MHz channel (coaxial transmission system)

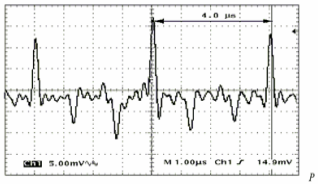


Fig.7. Measurement of interferences in time progression inside the basic frequency band of 858MHz channel (optical transmission system)

The first harmonic inside the basic frequency band of 64-QAM channel for European and American frequency plan of TV channels is almost always equal to *0,25MHz*. In reality phase ratios in the input sequence are changed with time and therefore the amplitude of pulses should also be changed. Fig.8 shows a burst of pulses inside the basic frequency band for *858MHz* channel.

The space between individual divisions on fig.8 is *10µs*. The duration of the measured burst is 60 - 90µs, which is two times more than the data packet.

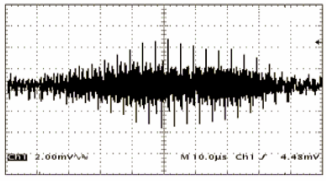


Fig.8. Burst of pulses inside the basic frequency band of 858MHz channel for coaxial transmission system

IV. INVESTIGATION OF INTER-MODULATION NOISE IN COAXIAL AND OPTICAL LINEAR ELEMENTS OF CABLE TV NETWORK

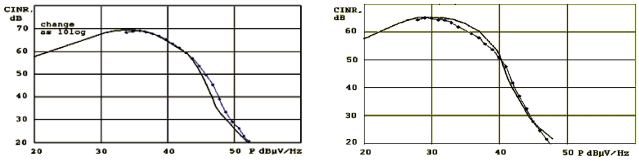
Inter-modulation products of these channels will have noise-like spectrum due to , as is the case with the CSO/CTB spectrum of discrete inter-modulation products, non-linearities of second and third order. Since the frequency spectrum of QAM or QPSK channels is sufficiently uniform in the channel, it is possible to regard the input signal of the laser or the reverse channel amplifier as white noise. The non-linearity of laser or the amplifier will bring to the formation of its own broad band noise caused by the second or third order inter-modulation interactions for the input signal frequencies. There is no multi-channel system for measurement for the reverse channel frequency range. Accordingly, there arises the question how can non-linear, distortions and the non-linear elements' noise be determined.

In [2, 7] there is an interesting method proposed for measurement of inter-modulation noise and estimation of the quality of the optical laser system for the reverse channel; that is the Notch-filter method which has become the standard one for the reverse channel. At the input of the measured optical line or amplifier a broadband noise burst is fed by noise generator which has a notch cut out by means of a narrowband rejection filter with a high level of bounding (about *80dB*). From this output the signal is transferred to a spectrum analyzer.

On fig.9 are shown curves measured (thick line) and calculated (point line) in accordance with CINR for amplifiers for the reverse channel as a function to the density of the output channel.

Fig.9 a) is about amplifier №1 with frequency band *5* to *30 MHz* and frequency of the notch-filter: *18MHz*.

Fig.9 b) is for amplifier №2 with frequency band *5* to *65 MHz* and frequency of the notch-filter: *35MHz*.



a) **b)**

Fig.9. *CINR = f(P_{out})* a)for amplifier №1, b) for amplifier №2

It is evident that the increase of the frequency band of the reverse channel amplifier causes a shift of the largest value of CINR towards the side of the lower signal levels. The reverse channel amplifier cannot work in areas of quick loss of CINR. The initial point of the quick loss of CINR can be regarded as the upper limit of the admissible density of the output (input) signal in the reverse channel.

V. CONCLUSION

It can be said in conclusion that the peculiarity of inter-modulation interferences is in the fact that their levels depend largely not only on the size and complexity of cable TV network, but also on the number of channels, their frequency plan and the level of signals. By forming pulse sequence of video frequencies the CSO/CTB spectrum brings about to characteristic screen interferences whilst in the case with MPEG-2 channels can substantially deteriorate the effectiveness of the error correction system.

In the reverse channel there are noise-like inter-modulation interferences which may sharply rise up in number provided there is an increase in the signal level above a certain limit in the reverse channel. Accounting of the influence of the inter-modulation interferences is done by means of the Notch-filter method.

There are certain challenges in calculation of reverse channel noises and interferences when designing return connection channel and they are connected with the necessity of accounting the impact of accumulated noises which have penetrated from any point in the network as well as with recorded noises within the entire network.

ACKNOWLEDGEMENT

This paper has been sponsored under the auspices of the "Planning of Multimedia Telecommunications Networks with Traffic and QoS Management" project – a part of the "SCIENTIFIC INVESTIGATIONS" Fund, contract ВУ-ТН-105/2005.

REFERENCES

[1] Angelov K., K. Koitchev, S. Sadinov, N. Varbanova. Technical aspects of upstream channel design in broadband cable television network. Journal Materials, Methods & Technologies – International Scientific Publications, Vol.2, Part 2, 2008, ISSN 1313-2539, pp.110-131.

[2] Panagiev O., Intermodulation composite distortions theoretically research in HFC networks. ICEST, Proc. Of Papers, vol.1, Ohrid, 24-27 June 2007, pp.325-328.

[3] Germanov V., "Calculating the CSO/CTB Spectrum of CKT Amplifiers and optical Receivers", IEEE Trans.on Broadcasting,vol.44 No.3 pp.363-366, Sept. 1998.

[4] Maeda K., H. Nakata, K Fujito, "Analysis of BER of 16QAM Signal in AM/16QAM Hybrid Optical Transmission System", Electron. Lett.vol.29, No.7, pp.640-642, April 1993.

[5] Tsushima H., K. Kitamura, "Preclipping Experiment for subcarrier-multiplexed AM/64QAM optical transmission", Electron. Lett. vol.31, No.21, pp.1863-1865, Oktober 1995.

[6] Ovadia S., C. Lin,"Performance Characteristics and Applications of Hybrid AM-VSB/MQAM Video Lightwave Transmission Systems", J. Lightwave Technology, vol 16, No. 7, pp.1171-1185, July 1998.

[7] Германов В., Некоторые аспекты моделирования и расчета распределительных сетей кабельного телевидения, Теле-Спутник №8(94), 2003, С.-Петербург, стр. 50-53.

Session

TELECOMMUNICATION NETWORKS

Modeling Service Level Monitoring Processes for QoS Guarantee of Managed Services

Todor Georgiev [1] and Aleksandar Tsenov [2]

Abstract – **Nowadays the Service Level Management became very important as a part of the relationships between the Managed Service Providers and their customers. The Service Level Management (SLM) process includes approaches, tools and rules for monitoring of the network Quality of Service (QoS) parameters. The overall process includes things such Service Level Requirements (SLR), Service Management Objectives (SMO) and Service Level Agreements (SLA). In this work a method for modelling the processes of Service Level Monitoring is introduced as a part of the whole Service Support Framework, according to the ITIL (Information Technologies Infrastructure Library) requirements.**

Keywords – **Service Level Management (SLM), Quality of Services (QoS), Information Technology Infrastructure Library (ITIL)**

I. Introduction and Problem Statement

Increasingly, the relationship between IT departments and their internal customers is that of client–supplier, based on the mechanisms of marketing and competition. IT departments are losing their internal company monopoly and have to compete with services offered on external markets. In this scenario a consistent customer focus of the IT management is of pivotal significance [1; 4; 10]. IT departments are facing the challenge of emerging from a technology-oriented applications developer and infrastructure operator to a client-oriented IT service provider.

Reference models help to reduce the costs and risks inherent in the transformation of organizational processes [2; 3]. This explains IT management's increasing interest in reference models for service oriented IT processes. Within the framework of change mentioned above they ensure systematic structuring of customer focused IT management processes at reduced cost and risk [11]. Within the last few years the IT Infrastructure Library (ITIL) has developed into a de facto standard for IT service management [14]. This is corroborated by the rapid increase in membership of the IT Service Management Forum, which is an interest group enhancing and propagating the ITIL principles [9]. Also the large number of practice-oriented ITIL conferences, publications and training opportunities [7; 8; 9; 14] indicate the growing relevance of ITIL. Recent studies substantiate that the ITIL holds a position of high relevance as well as being utilized extensively in the everyday running of German companies [12]. In spite of its relevance, its wide distribution and a large number of publications, a critical analysis of the ITIL reference model from a formal point of view is lacking. On the one hand existing literature is content to simply describe the areas of IT management as documented in the ITIL [3; 6; 8; 9; 11] and on the other it makes suppositions about the general usefulness of the ITIL in practice [7; 13]. The authors know of only a few publications on ITIL in scientific journals [11; 14]. This results in uncertainty in the execution of ITIL projects and misunderstandings regarding the attainable advantages of adapting ITIL.

The principles for orderly modeling provide criteria with which the construction of models can be evaluated and which permit the identification of deficits and advantages. Regarding this last point, it is especially the principle of economy which would imply a benefit analysis. The model of principles for orderly modeling was chosen from among a large number of model assessment approaches [1; 5] on the basis of its particular suitability for reference information models. The evaluation is based on a detailed analysis of four case studies conducted for this purpose.

The paper is structured as follows: Chapter two contains a brief introduction into the basics of the ITIL reference model. Subsequently, Chapters three and four describes the approach and the evaluated model of the Service Level Monitoring process. Following that, selected trends are described and a conclusion is arrived at in chapter five.

II. ITIL as a Common-Practice Reference Model

The ITIL is an English language set of documents consisting of several volumes of IT management concepts, processes and methods. Originally it was developed by the IT service provider of the British government but currently it is being continuously developed and disseminated by the internationally active IT Service Management Forum [2]. The core of the model consists of IT service management, which deals with the control and monitoring of IT services, based on aspects and principles of classical service provision [4]. Within the ITIL there are two areas of IT service management: service support [14] and service delivery [14].

Apart from IT service management the ITIL also addresses the areas of application management, infrastructure management, business perspective, IT software asset management and security management. Because of their minor practical relevance these areas are not investigated further in this paper. Thus, when the ITIL is mentioned below, what is said refers exclusively to the areas of service support and service delivery, i.e. IT service management. ITIL subdivides service support into the areas of incident management, problem manage-

[1]Todor Georgiev is with the TELELINK EAD, Business Park, Building 13, Sofia E-mail: tgeorgiev@telelink.bg

[2]Aleksandar Tsenov is with Telecom Department at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: akz@tu-sofia.bg

ment, change management, release management and configuration management. Service delivery is subdivided into the areas of service level management, financial management, IT service continuity management, capacity management and availability management. In view of the available literature on the subject [1; 14] a detailed description of these areas is not included here. Within the area of reference modeling, models of common practice and best practice can be distinguished [3]. In view of this, the ITIL can be classified as a common-practice model. The modeling object is IT service management and the language of description is a natural language. The nature of recommendation, which by definition has to exist for reference models [2], originates in the description of a branch standard in the area of IT service management. Innovative insights based on a well-founded theory, a requirement for best-practice models, are not emphasized by ITIL. Therefore it is not a best practice model. Rather, the ITIL is a common practice model possessing the character of a branch standard. This standard is valid for internal as well as external IT service providers. Furthermore, the ITIL's validity excludes neither specific branches, nor certain sizes of companies.

III. Models of Service Level Management as a part of the ITIL – Service Delivery Process

A. Service Delivery Process

The Service Delivery Process an it's components is shown on Figure.1. In this provider view all the elements are of equal worth, while the customer view represents another types of relationships and hierarchy - Figure 2.

B. Purpose of Service Level Agreement

- A communication tool
- A living document
- A conflict-prevention tool
- A way towards a service culture

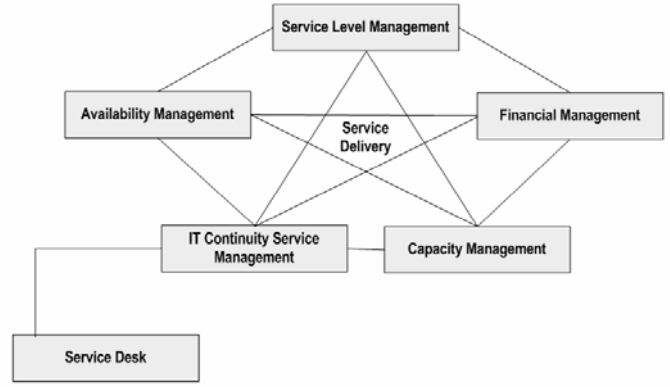


Fig. 1. The Service Delivery Process

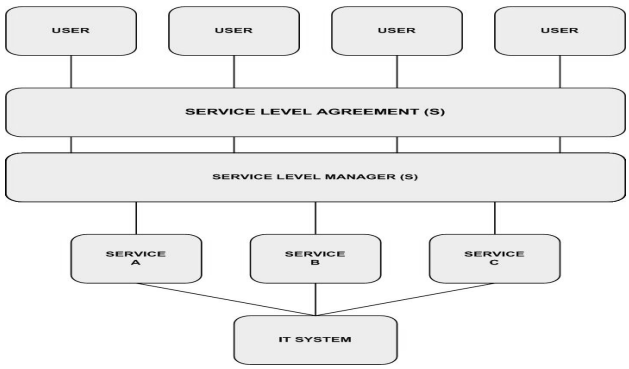


Fig. 2. The Service Level Management Items

- An objective basis for measuring service effectiveness – an SLA ensures that both parties use the same criteria to evaluate service quality (Figure 3).

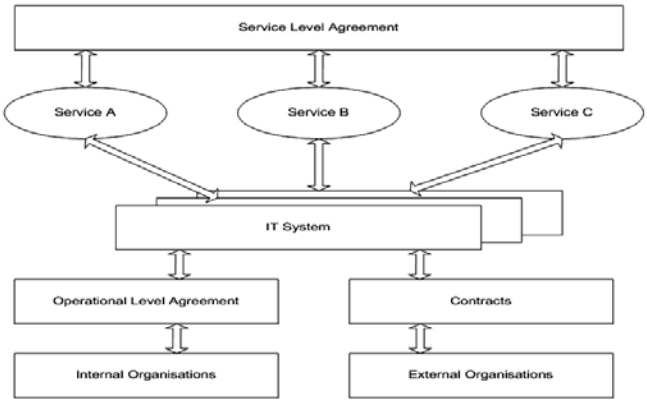


Fig. 3. The SLA Support Structure

C. Modelling the Service Level Management Process

Service Level Management (SLM) should be familiar territory for MSPs. Often, an MSP's contract with the customer is elaborated upon with a number of Service Level Agreements (SLAs). At the very heart of the MSP's relationship with its customer is the setting of expectations or service levels and then meeting them – Figure 4.

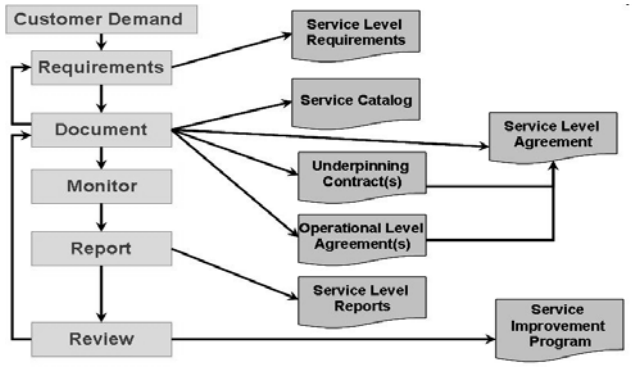


Fig. 4. The Overall Service Level Management Process

In practice, SLAs are negotiated at the beginning of the relationship in the mistaken belief that SLAs are SLM. This is not the case. ITIL indicates that SLM begins with gathering Service Level Requirements from the customer. For an MSP, this typically occurs through the Request For Proposal and Due Diligence processes. Next, a Service Catalogue is developed to identify the services available from IT. Then, internal Operational Level Agreements (OLAs) are identified and negotiated with other IT departments and external Underpinning Contracts (UCs) are established with third party vendors. Finally, an SLA is negotiated with the customer based on the services detailed in the Service Catalogue and the fees the customer is willing to pay. ITIL's SLM approach allows both parties to recognize the customer needs, the provider's capabilities, and the subsequent costs involved. It is this clear definition that moves the MSP onto the right path for supporting agreements and SLAs.

The overall SLM process covers the requirements and characteristics listed below.

Main Requirements

- To check of SLA targets are met
- To react if they are not
- To provide measures
  - For reporting purposes
  - For further review of the service
- Monitor services and customer satisfaction

Measure

- Measure from the customers' viewpoint
- Agree measurement time frames
- Ensure consistency
- Allocate responsibility for measurement

SLM – Report

- Reporting procedures and report must be defined during SLA negotiation
- Reports must be produced as described in the SLA
- Exception reporting for SLA breaches
- Cleary show performance versus targets
- Create impact by using a SLAM (SLA Monitoring) chart…

SLM Review

- Procedures must be defined during initial SLA negotiations
- Regular and Ad-Hoc
- Review achievements versus targets
- Identify corrective actions and improvements
- Review meeting should be held as agreed in SLA

SLM –Renewal/Renegotiation

- Procedures must be defined during initial SLA negotiation
- Fixed date or notice period for renewal
- Renegotiation triggered by
  - Breaches
  - Changes in requirements, technology, volumes

SLM – Service Catalogue

- Provide an accurate picture of the services provided to customers
  - Simplest form is a matrix of services/customers
- Usually includes
  - Customers/users of the services
  - Characteristics of the services
  - Maintainers of the services
- A documentation CI in the CMDB

SLM – Service Improvement Programme

- A key aspect of ITIL is 'continuous improvement'
  - Improve services
  - Decrease cost
- The SIP is the formal approach to improving the Quality of delivered services
  - Planned approach – not driven by failure only
  - Start with existing services and introduce an SIP
- The responsibility for IT Service Quality lies with the Service Level Manager

SLM – SIP – Examples

- Technical improvements
- Process and procedures
- 'Quick wins'
- Training
- Dialogue with customers

SLM – One of 11 disciplines

- SLM define and control the service level targets (the 'what')
- To achieve these targets in an efficient and cost-effective way, SLM must be supported by other ITIL disciplines
- The Service Quality Plan defines 'how' the organization will deliver the agreed service level though Service Management processes

SLM – Content of the SQP

- Management information needed to ensure the services are delivered at the agreed quality level
- Performance targets
  - For the Service Management processes
  - For internal operational teams
  - For external providers
- Monitoring and reporting details

SLM Responsibilities

- Implement SLAs
- Manage SLAs
- Manage UCs and OLAs
- Manage quality of services

Manage customer relationships

IV. The Proposed SLM Concept

An important advantage of the proposed SLM concept is that it provides tight integration across all solutions, permitting them to work together to benefit from - and build upon - the specific capabilities of each (Figure 5).

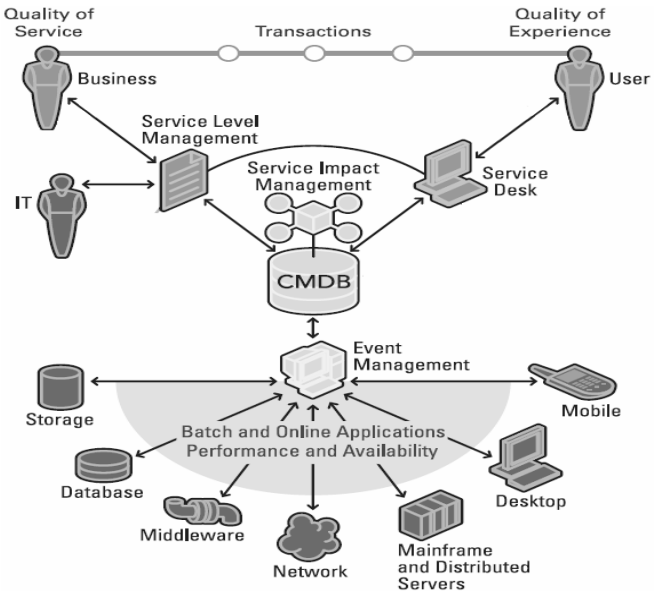


Fig. 5. The proposed Service Level Management concept

In the table below the synergies between the standard Management elements and the functions of the proposed SLM concept are shown.

Standard Management Elements	Proposed SLM Functions
Incident and Problem Management	▪ Tracks help desk response and resolution times and compares them with SLA commitments. ▪ Generates alerts and notifications to the support staff when support SLAs are in danger of being missed. ▪ Provides the help desk with service impact information to assist in determining incident priorities and to facilitate root cause analysis.
Change and Configuration Management	▪ Tracks SLA availability targets to ensure that change tasks and requests are performed in order and on time.
Service Impact and Event Management	▪ Delivers business – aware infor-mation about the real-time state of services.
Capacity Management and Provisioning	▪ Enables the analysis and predic-tive modeling of potential IT configu-ration changes and their effect on ser-vice levels.
Asset Management and Discovery	▪ Measures availability tar-gets for specific assets and ser-vices, and shows the latest cal-culated availability metrics for the specified items in the CMDB to help ensure that cri-tical business assets or ser-vices maintain committed le-vels of availability.
Infrastructure and Appli-cation Management	▪ Uses infrastructure and applica-tion data for service level measuring of both infrastructure and applications.

V. CONCLUSION

The proposed Service Level Management conceptual model is based on the possibility to provide a direct path to Business Service Management by enabling IT to manage the services it delivers from the perspective of the business. It strengthens IT's ability to meet business and user demands and to improve the Quality of Service and the Quality of Experience. The Service Level Management process, introduced in this work, delivers measurable benefits and value as the organization progresses from managing service level agreements from the technical perspective (inside out) to managing service level agreements from the business and the end-user perspective (outside in).

ACKNOWLEDGEMENT

This work is made in connection to the **Project BY-TH 105/2005**.

REFERENCES

[1] The Official Introduction to the ITIL Service Lifecycle, 2007.
[2] Guldentops, E., G. Hardy, J. Heschl, R. Stroud, Aligning COBIT®, ITIL® and ISO 17799 for Business Benefit. A Management Briefing from ITGI and OGC, 2005
[3] Hoekstra, A., N. Conradie, CobiT, ITIL and ISO17799. How to use them in conjunction, 2002
[4] Curtis, B., Integrating CMMI® with COBIT® and ITIL®, 2005
[5] Defining, Modeling & Costing IT Services. Integrating Service Level, Configuration & Financial Management Processes, September 2004
[6] Computer Associates' Delivering. Best Practice Support ITIL, ISO 17799, BS15000, White Paper
[7] Symons, C., IT Strategy Maps: A Tool For Strategic Alignment by BEST PRACTICES November 21, 2005
[8] Delivery of ITIL Best Practices using Computer Associates management solutions
[9] Nabiollahi A., Considering Service Strategy in ITIL V3 as a Framework for IT Governance, The IT Service Management Forum, UK , 2007
[10] Brown A.B., A. Keller, A Best Practice Approach for Automating IT Management Processes, IEEE Publishing, 2006
[11] Jantti1M., A. Eerolal, A Conceptual Model of IT Service Problem Management, IEEE Publishing, 2006
[12] "ITIL Das Munich Institute for IT Service Management – mITSM", available on http://www.mitsm.de/
[13] "ITIL Forum 2008 - Neupositionierung und Wertorientierung", available on http://www.itil-kongress-iir.de/
[14] "ITIL process procedure templates", available on http://www.metocube.com
[15] "OGC Best Management Practice - It Service Management – ITIL" available on http://www.best-management-practice.com

Policy Models for Resource Management

Evelina Pencheva[1] Ivaylo Atanasov[2] and Dora Marinska[3]

Abstract – **In this paper an approach to modeling of policies for resource management is presented. Policy information models are structured according the quality of service mechanisms in Internet Protocol based multimedia networks. The usage of Open Service Access Policy Management interfaces is exemplified to illustrate the way policy information can be managed by 3rd party service providers. Formal description of policy rules is presented.**

Keywords – **IP Multimedia subsystem, quality of service, policy management, open access.**

I. INTRODUCTION

Internet Protocol Multimedia Subsystem (IMS) is standardized service control architecture that provides access-independence and IP connectivity. IMS is built on concepts that offer basic and advanced multimedia services to end-users using common Internet-based protocols. The underlying access and transport networks together with the IMS provide end-to-end Quality of Service (QoS). Via the IMS, the user equipment negotiates its capabilities and expresses its QoS requirements during a Session Initiation Protocol (SIP) session setup or session modification procedure. IP policy control means the capability to authorize and control the usage of multimedia traffic based on the signaling parameters at the IMS session [1]. It provides a way to allocate access and transport network resources, primarily network bandwidth, QoS, and security, according to defined business policies. This requires interaction between the IP connectivity access network and the IMS.

Policy is an ordered combination of policy rules that define how to administer, manage, and control access to resources. Policy defines the access to network resources, the traffic priorities, required bandwidth that has to be allocated to ensure guaranteed delivery and the eligible traffic for discard when the network becomes busy and congested.

Policy-based management systems are usually implemented for enterprise networks [2]. Network administrators apply application- and user-based centralized policy control to optimize resource allocation. Resources include devices that manage network bandwidth, security, IP addresses, storage, processors, and agents, as well as systems that manage services such as billing, accounting, and service mapping, and automated reliable policy deployment.

[1] Evelina Pencheva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: enp@tu-sofia.bg.
[2]Ivaylo Atanasov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: iia@tu-sofia.bg.
[3]Dora Marinska is with Global Communications Nets, 46 St. Kiril i Metodi, 1202 Sofia, Bulgaria E-mail: dmarinska@gcn.bg

The policy is operator specific and depends on network infrastructure and services provisioned [3].

The IMS standards just define a framework for policy control and do not specify in details the structure of QoS management policy. In this paper we suggest an approach to policy modeling in IMS environment. First we describe the IMS policy control architecture and then we model the structure of policy information for resource management using the QoS mechanisms. We exemplify the application of the Policy management service defined for Open Service Access (OSA) to illustrate the way in which policy information can be managed by 3rd party service providers.

II. IMS POLICY CONTROL ARCHITECTURE

To deliver service performance that determines the degree of user satisfaction of the service, an architectural framework for QoS support is defined [4]. The QoS architectural framework is a set of generic network mechanisms for controlling the network service response to a service request, which can be specific to a network element, or for signalling between network elements, or for controlling and administering traffic across a network.

In IMS the Policy and Charging Enforcement Function (PCEF) encompasses service data flow detection, QoS handling, policy enforcement and flow based charging functionalities. This PCEF is located at the media gateway at the IMS User plane. The Policy and Charging Rule Function (PCRF) includes policy control decision and flow based charging control functionalities. It is responsible for finding routes in the network that meet QoS requirements and is located at the IMS Control plane. The PCRF provides network control regarding the service data flow detection, gating, quality of service and flow based charging (except credit management) towards the PCEF [5], [6].

Service requirements are signalled at the Control plane during session establishment and reflected on the underlying IP access and transport networks. Without interaction between User plane and Control plane the operator will not be able to provide the required QoS. The gateway containing PCEF is capable of policing packet flow into the IP network, and restricting the set of IP destinations that may be reached according to a packet classifier. This service-based policy "gate" function has an external control interface that allows the gate to be selectively "opened" or "closed" on the basis of IP destination address and port. When open, the gate allows packets to pass through (to the destination specified in the classifier) and when closed, no packets are allowed to pass through.

In IMS, value-added services are delivered by Application Servers (AS). The OSA Service Capability Server (SCS) is a special type of AS which provides open interfaces to network functions (like call and session control, messaging, user interaction, location etc.) for 3rd party applications. The OSA Application Programming Interfaces (APIs) hide underlying network technology and protocol complexity from application developers. The OSA AS hosts third party applications that use network functions exposed through OSA APIs.

OSA Policy Management APIs have been defined to offer provisioning services [7]. Using the APIs it is possible to create, update or view policy information for any policy enabled service. The APIs facilitate interactions between clients and the policies of any policy enabled service. These include APIs to subscribe to policy events, to request evaluation of policies and to request the generation of policy events. Policy Management clients include both 3rd party applications and network service administration. It is expected that more and more OSA services will use policies to express operational criteria. It is also expected that network providers will host policy-enabled services that have been written by 3rd party application service providers. The network operator can populate the repository with the policy-related conditions and actions that it can support. The PCRF, PCEF and Policy repository form network policy engine as shown in Fig.1.

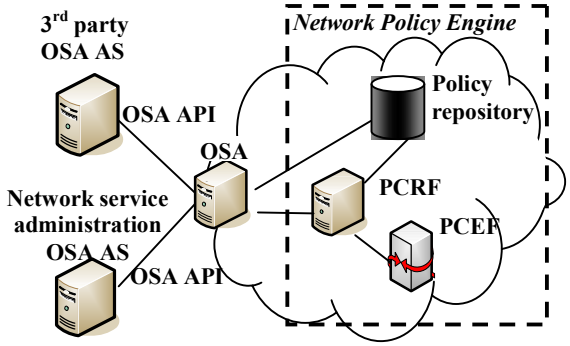


Fig.1 Deployment of OSA Policy Management APIs in IMS environment

III. Resource Management Policy Information Model

Policy domain allows aggregation of policy domains, policy groups, policy rules, or policy event definitions in a single container. Policy group aggregates policy rules or other policy groups. Policy rule is a combination of conditions and actions to be performed if the condition is evaluated as true. The repository is meant to hold unattached conditions and actions.

Considering control functions for IP policy control, we suggest the model the policy domain for Resource management composed of Charging Control Policy domain and Quality of service Provisioning Policy domains as shown in Fig.2. The Quality of Service Provisioning Policy domain includes QoS control domain, QoS data domain and QoS management domain. QoS Control Policy domain defines policy groups that deal with the pathways through which user traffic travels. These policy groups include admission control, QoS routing, and resource reservation. QoS Data Policy domain defines policy groups that deal with the user traffic directly. These policy groups include buffer management, congestion avoidance, packet marking, queuing and scheduling, traffic classification, traffic policing and traffic shaping. QoS Management Policy domain defines policy groups that deal with the operation, administration and management aspects of the network. These policy groups include SLA, traffic restoration, metering and recording.

The admission control policy rules control the traffic to be admitted into the network. Whether traffic is admitted depends on an a priori service level agreement. In addition, the decision can depend on if adequate network resources are available so that newly admitted traffic does not overload the network and degrade service to ongoing traffic. For a service provider, maximal traffic must be admitted while the same level of QoS is maintained for the existing traffic. The admission policy rules are typically parameter or measurement-based. The parameter-based approach derives the worst-case bounds for a set of metrics (e.g., packet loss, delay and jitter) from traffic parameters and is appropriate for providing QoS for real-time services. In contrast, the measurement-based approach uses measurements of existing traffic for making an admission decision. It does not warrant throughput or hard bounds on packet loss, delay or jitter and is appropriate for providing QoS for non real-time services. Admission control can also be used to meet requirements for service reliability/availability over a specified period for the desired transaction types as negotiated in the SLA. Admission control policies give preference to traffic streams (e.g., for emergency communications) deemed to be more critical by a service provider under conditions of congestion [8].

When a QoS resource is modified by the user equipment, such that the requested QoS falls outside of the limits which were authorized, then the PCEF needs to verify the authorization of this QoS resource modification. If the PCEF does not have sufficient information to authorize the QoS resource modification request, the PCEF sends an authorization request to the PCRF. The PCRF authorizes the modified QoS resources based on the current session information.

The resource reservation rules are used to set aside required network resources on demand for delivering desired network performance. Whether a reservation request is granted is closely tied to admission control. In IMS resource reservation is always initiated by the user equipment after successful authorization. The user equipment includes in the resource reservation request the authentication token granted. With request for QoS resource reservation, the PCEF in the gateway needs to assure that the requested resources match to the authorized resources. The PCEF forwards the token, together

with the requested QoS parameters, to the PCRF. The PCRF checks if the corresponding requested QoS resources are within the limit of what was negotiated. The PCRF uses the token as the key to find the stored negotiated session description.

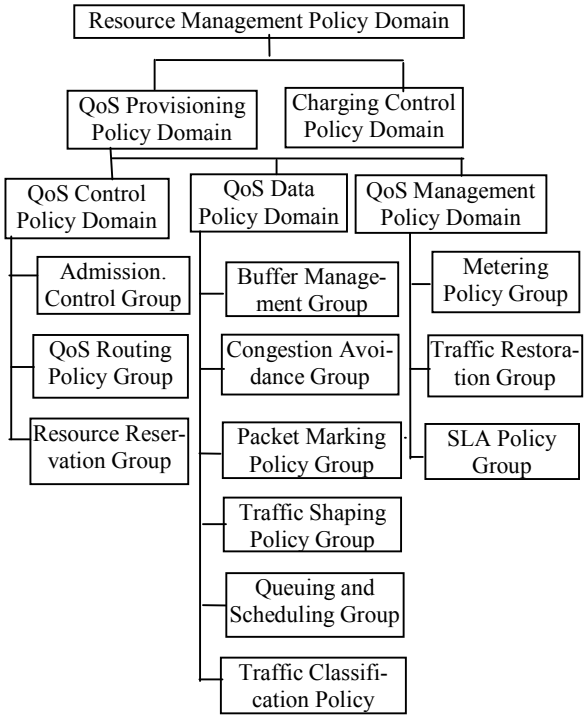


Fig.2 Resource management policy model

QoS routing policy rules concern the selection of a path satisfying the QoS requirements of a flow. Practical QoS routing schemes consider mainly cases for a single QoS metric (e.g., bandwidth or delay) or for dual QoS metrics (e.g., cost-delay, cost-bandwidth, and bandwidth-delay). The path selection process involves evaluation of policy rules having knowledge of the flow's QoS requirements and characteristics and information on the availability of network resources. To guarantee performance on a selected path, QoS routing needs to be used in conjunction with resource reservation to reserve necessary network resources along the path.

Queue or buffer management policy rules deal with which packets, awaiting transmission, to store or drop.

Congestion avoidance policy rules deal with means for keeping the load of the network under its capacity such that it can operate at an acceptable performance level, not experiencing congestion collapse.

Queuing and scheduling policy rule control which packets to select for transmission on an outgoing link. Incoming traffic is held in a queuing system, which is made of, typically, multiple queues and scheduler. Rules governing the queuing system determine the queuing and scheduling discipline it employs.

Packets can be marked according to the specific service classes that they will receive in the network on a per-packet basis. The policy rules for packet marking need to be provisioned or configured dynamically.

Traffic classification can be done at the flow or packet level. At the edge of the network, the policy rules determines the aggregate to which the packet belongs and the respective service level agreement.

Traffic policing rules deal with the determination of whether the traffic being presented is on a hop-by-hop basis compliant with pre-negotiated policies or contracts.

Traffic shaping rules deal with controlling the rate and volume of traffic entering the network.

A Service Level Agreement (SLA) typically represents the agreement between a customer and a provider of a service that specifies the level of availability, serviceability, performance, operation or other attributes of the service. Basic composition of the SLA content, regardless of the services provided includes the business part, service part, technology part and QoS report part [9]. The business part describes the general business information and business procedures related to the service. Policy rules may describe service violation processing and billing information. The SLA information stored in the service part related to negotiated service content and the agreed service level and SLA information in the technological part related to QoS parameters can be used in admission control rules as described above. The QoS report part includes the QoS report information provided to the service customer and the service provider in order to evaluate service level negotiated in the SLA. This information may be used to define policy events concerning QoS monitoring and reporting.

IV. OSA Policy Management Interfaces in Use

The OSA "Policy Management" interfaces allow policies to be provisioned and compliance of service usage with policies to be evaluated. Client (3rd party) applications can use the Policy Management API to express operational criteria in a form of policy. It is possible for an application to manage policy information, control access to it and to request evaluation of policies.

Fig.3 shows the formal definition of a rule for traffic marking based on Single rate Three Colour Marker (SRTCM) [10]. The SRTCM implements two Token Bucket mechanisms. The packet is first tested by the TB(CIR,CBS), where CIR denotes the committed information rate, while CBS denotes the committed burst size. If the packet conforms to the TB1(CIR, CBS) it is marked for low DP (green). Otherwise, it is forwarded to the TB2(CIR, EBS), where EBS denotes the contracted excess burst size. If the packet conforms to this TB, it is marked for medium DP (yellow) otherwise it is marled for high DP (red).

Fig. 4 shows the formal definition of a rule in Admission control policy group. The rule may be used in session authorization considering requested QoS parameters and the allowed QoS parameters as defined in the user profile.

if (TB1(CIR,CBS)-B(CIR,CBS)>=0)
then setConformanceLevel(COLOR-MODE FLAG,"green")
else if (TB2(CIR,EBS)-B(CIR,EBS)>=0)
then setConformanceLevel(COLOR-MODE FLAG,"yellow")
else setConformanceLevel(COLOR-MODE FLAG,"red") end

ConditionAttribute.AttributeName = "TokenBucketSize".
ConditionAttribute.AttributeValue.SimpleValue.StringValue = "MaximumSizeOfTokenBucket > CurrentPacketSize".
conditionList.Condition == <comparison between CIR, CBS parameters of the bucket and the packet >.
conditionList.GroupNumber == 1; indicates how the conditions need to be grouped in DNF or CNF in case more groups of rules exist.
conditionList.Negated == FALSE.
actionList.Action == <marking the packet with COLOR –MODE FLAG>
actionList.SequenceNumber == 1.
IF " Token bucket1(CIR, CBS) >= B(packet)" THEN "green mark == TRUE" , Else CBS -> EBS
IF " Token bucket2(CIR, EBS) >= B(packet)" THEN "yellow mark == TRUE", Else COLOR-MODE FLAG ==’red’

Fig.3 A rule for packet marking based on SRTCM

V. CONCLUSION

In this paper we investigate the IP policy control in IMS and present an approach to modeling of policy information. Policy information is structured according the quality of service mechanisms for controlling the network service response to a service request, which can be specific to a network element, or for signaling between network elements, or for controlling and administering traffic across a network.

The usage of Open Service Access Policy Management interfaces is exemplified. The interfaces can be used to perform administrative tasks on behalf of resource management, e.g. create, update or delete policy information and to invoke evaluation of policies of resource management service. Following the Policy Management specific data definitions we describe in a formal way resource management policy rules. The formal description of policy rules allows reuse of policy-related information in IMS policy engines.

ACKNOWLEDGEMENT

The research is partially supported by the Research project DO-02-135/2008 funded by Bulgarian Ministry of Education and Science.

REFERENCES

[1] Miikka Poikselka, Georg Mayer, Hisham Khartabil, Aki Niemi, The IMS Multimedia Concepts and Services, Wiley, 2008
[2] Michael Massimi, Ursula Wolz, Peer-to-Peer Policy Management System for Wearable Mobile Devices, www.dgp.toronto.edu/~mikem/pubs/MassimiWolz-ISWC03.pdf
[3] W. Lv, J. Kang, and W. Chen, TOOPM: A Telecom Operations-Oriented Policy Management System, Robotics and Applications and Telematics (RA 2007), Proceedings, pp.237-246

if (RVC)-(AVC)>=0)
then setConformanceLevel(video resources,"allowed")
if (RAC)-(AAC)>=0)
then setConformanceLevel(audio resources,"allowed")
if (RQoSclass)-(AQoSclass)>=0)
then setConformanceLevel(bidirectional conversation,"allowed")
if (RVC)-(AVC)>=0) AND (RAC)-(AAC)>=0) AND (RQoSclass)-(AQoSclass)>=0)
then setConformanceLevel(IMS Session,"established")
else setConformanceLevel(IP flows,"unauthorized")
setConformanceLevel(IMS Session,"failure") end

ConditionAttribute.AttributeName = "Session establishment".
ConditionAttribute.AttributeValue.SimpleValue.StringValue = "MediaCharacteristicsOfNetwork > =UserProfileMediaCharasteristics".
ActionAttribute.AttributeName = "NegotiatedSDPParameters".
ActionAttribute.AttributeValue.SimpleValue.StringValue = "AllowedVideo == TRUE".
ActionAttribute.AttributeValue.SimpleValue.StringValue = "AllowedAudio == TRUE".
ActionAttribute.AttributeValue.SimpleValue.StringValue = "QoS class A == TRUE".
conditionList.Condition == < establishing IMS session by comparing allowed SDP parameters with the negotiated ones>
conditionList.Negated == FALSE.
actionList.Action == <authorization of IP flows of the chosen media components by mapping from SDP parameters to authorized IP QoS parameters >.
IF " requestedSDPparametersForVideo >= UserProfileAuthorized " THEN "AllowedVideo == TRUE".
IF " requestedSDPparametersForAudio >= UserProfileAuthorized " THEN "AllowedAudio == TRUE".
IF " requestedSDPparametersForQoSclass >= UserProfileAuthorized " THEN "AllowedQoSclassA == TRUE".
IF "AllowedVideo == TRUE" & "AllowedAudio == TRUE" & "AllowedQoSclassA == TRUE"
THEN “SessionEstablishmentWithRequestedParameters==TRUE”
Else
AuthorizationOfIPflows ==FALSE & SessionEstablishment ==FALSE.

Fig.4 A rule for admission control based on user profile data

[4] ITU-T Y.1291, An architectural framework for support of Quality of Service in packet networks
[5] 3GPP TS 23.203 v8.4.0, Policy and charging control architecture
[6] ETSI TS 183 017 v1.1.1, Resource and Admission Control: DIAMETER protocol for session based policy set-up information exchange between the Application Function (AF) and the Service Policy Decision Function (SPDF); Protocol specification
[7] 3GPP TS 29.198-13 v7.0.0, Open Service Access (OSA); Application Programming Interface (API); Part 13: Policy Management Service Capability Feature (SCF)
[8] ITU-T Y.2171, Next Generation Networks – Quality of Service and Performance, Admission control priority levels in Next Generation Networks
[9] ITU-T, M.3342, Guidelines for the definition of SLA representation templates
[10] J. Heinanen, R. Guerin, A Two Rate Three Color Marker, RFC 2697, 1999.

Study on Open Access to Connectivity Management

Evelina Pencheva[1] and Ivaylo Atanasov[2]

Abstract – **This paper presents an analysis on open access to functions for quality of service management in Internet Protocol based multimedia networks (IMS). The conformity of Open Service Access interfaces for connectivity management to the requirements for quality of service management in IMS is assessed. Improvements to the specifications are suggested.**

Keywords – **IP Multimedia subsystem, quality of service, connectivity management, open access.**

I. Introduction

Ensuring quality of service and policy control are basic requirements for deployment of Internet Protocol based multimedia networks (IMS) [1]. Quality of service (QoS) is seen as one of the aspects that telecom operators will use in order to differentiate their multimedia offerings from those of Internet service providers who in most cases do not have the means to provide QoS [2]. This factor is particularly relevant when using limited-bandwidth access network (e.g. wireless). In order to offer QoS, IMS need to implement QoS mechanisms in conjunction with the access network and the transport network, provide negotiated QoS at signaling level and realized resource reservation in a coordinated manner with session establishment.

One of the main features of IMS is the secured open access to network functions [3]. The open access is through application programming interfaces (APIs) that hide underlying network technology and control protocol complexity from application developers. Open Service Access is defined as service architecture for open access to functions in multimedia networks [4]. OSA Connectivity Manager API is defined to provide access for 3rd party applications to QoS management functions [5]. The OSA Connectivity Manager API is a stable standard defined before the production of the detail specification of IMS mechanisms for QoS and policy control. There exists criticism on capabilities of the API for Resource Admission Control in next generation networks [6], [7]. Resource and Admission Control Subsystem is responsible for elements of policy control, resource reservation and admission control. In IMS, the Policy and Charging Control functions are further clarified and extended.

[1]Evelina Pencheva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: enp@tu-sofia.bg.
[2]Ivaylo Atanasov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: iia@tu-sofia.bg.

In this paper we present an analysis on capabilities of OSA Connectivity Manager APIs and assess their conformity to the IMS requirements for QoS management. We suggest an improvement of the OSA Connectivity Manager.

II. OSA Connectivity Manager Overview

The Connectivity Management is a set of functions that provide configuration and control of both the attributes of IP connectivity and policies governing IP connectivity, within and between IP domains. Such attributes include QoS, security, and routing policy.

The "OSA Connectivity Manager" service capability feature (SCF) is defined to establish QoS parameters for an enterprise network traffic travelling through a provider network. Assuming that the underlying packet network can be configured as a virtual private network (VPN), the Connectivity Manager interfaces provide methods that allow management applications to configure inter-site virtual connections.

The "Connectivity Manager SCF" can be used by a VPN client (enterprise operator subscribed for VPN services) that has entered a relationship with a VPN provider (network operator) to set up a provisioned QoS. Connectivity Manager includes API between VPN client and VPN provider to establish QoS parameters for VPN packets passing through the provider network.

The API requires any specific QoS method to be used neither in the VPN network nor in the operator network. To deliver QoS between networks the differentiated services approach is used which is based on giving preferential treatment to some packets over others in the edge routers. Each packet arriving from the VPN client network into the VPN provider network is marked with a tag called DSCP (differentiated services code point) [8]. Only marked packets can enjoy the QoS service provisioned in the VPN provider network.

The VPN client may be an enterprise operator that owns a number of enterprise sites connected via virtual private network provided by a VPN provider. The VPN provider is available at a number of sites by service access points for the VPN client. An application using "Connectivity Manager" SCF is hosted at the VPN client domain. The telecom operator gives the VPN provider access to the "Connectivity Manager" SCF. The access through an OSA service capability server is subject to the safeguards provided by the OSA Framework [9]. An imaginary VPN configuration is shown in Fig.1.

The VPN provider offers configuration service to the VPN client. Using the "Connectivity Manager" API the VPN client

can create virtual provisioned pipes (VPrP) in the VPN provider network to carry the enterprise traffic and support it with pre-specified QoS. The VPrP defines QoS parameters for traffic flowing through the provider network between two specified enterprise endpoints. The VPN provider offers a set of templates that are used by the VPN client to specify a VPrP.

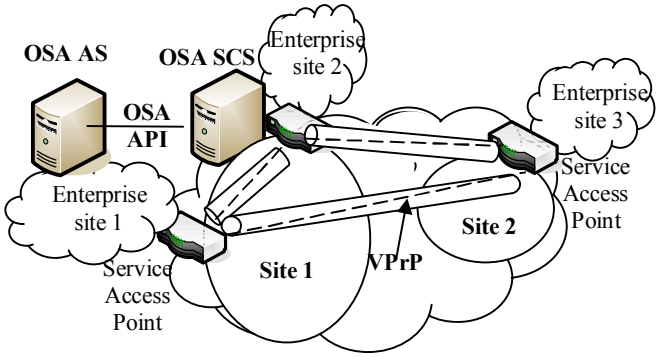


Fig.1 An imaginary Virtual Private Network of an enterprise operator

For instance, the provider may offer templates for video conferencing, audio conferencing, Gold Service, Silver Service, etc. Using these templates the VPN client can select and provision a VPrP with specific QoS attributes. Elements that can be specified for a VPrP include attributes such as packet delay and packet loss, and traffic characteristics such as maximum rate and burst rate. The collection of all the VPrPs, provisioned within the enterprise VPN, constitutes the Virtual Provisioned Network (VPrN).

The Connectivity Manager interface is the entry point to this service. From this interface the client application can get reference to the VPN client network interface and to the QoS menu interface. The QoS menu interface provides a list of templates, each of which specifies the QoS service parameters that are offered by the VPN provider. The service is composed of components that are associated with a provisioned QoS. The client application can use the template interface to specify the service parameters that are offered by the VPN provider, and also, to store the parameters that the VPN client selects temporarily. The VPN client network interface is associated with two components: enterprise sites, and the VPrN that has been already provisioned in the provider network. The Virtual Provisioned Network interface contains references to all the VPrPs already established. The client application can use the QoS Menu to get references to all the QoS templates offered by the provider. Once the VPN client selects the QoS parameters provided in the QoS template, and submits the request to create a new VPrP, the VPN provider validates the information submitted and if the request is approved, the new VPrP is activated.

Fig.2 illustrates the sequence in which VPN client selects service components and creates a new VPrP. In the figure, the client application collects the information required to select a service, then selects service parameters, and finally submits it to the Connectivity manager.

The Enterprise Network interface stores enterprise network information maintained by the provider as it relates to the VPN service and the virtual provisioned network service that the VPN client had already established with the provider network. The VPN client can only retrieve information regarding an existing VPrN, can list the sites connected to the VPN, and can get the handle to a specific site interface that stores information about the site.

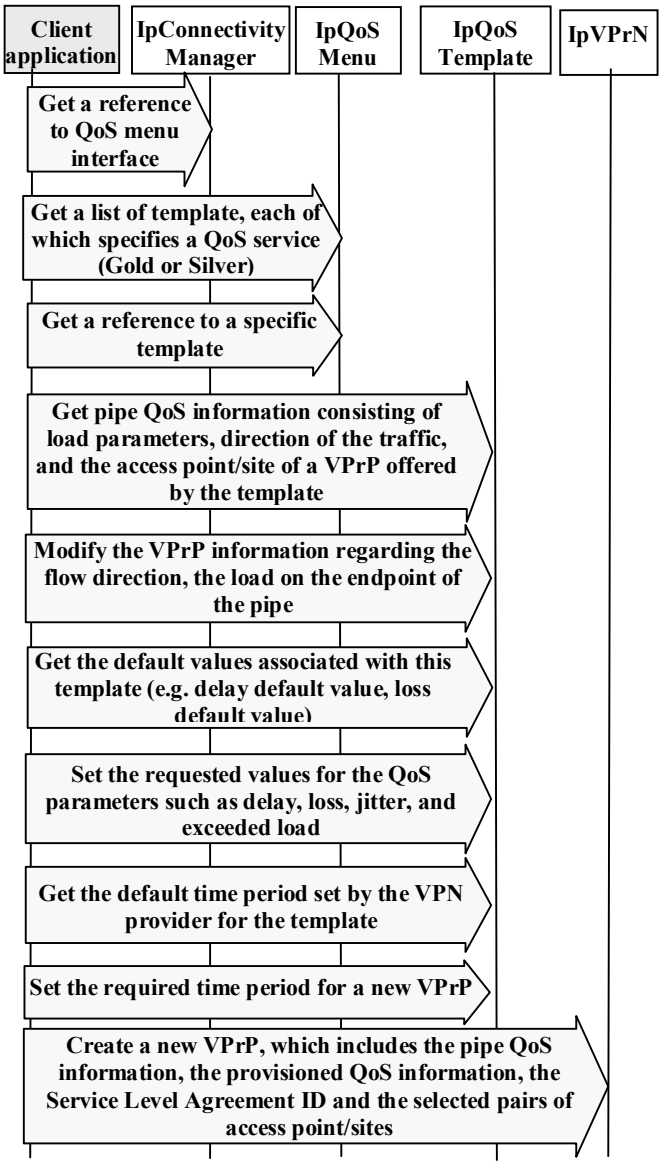


Fig.2 The Application creates a new virtual provisioned pipe

Fig.3 illustrates the way in which a VPN client browses a VPrP and collects information regarding existing VPrP, including all QoS parameters that have been set for this pipe.

The VPrP interface provides status information on a VPrP. The Enterprise Network Site stores site information of the VPN client network. This information is maintained by the VPN provider.

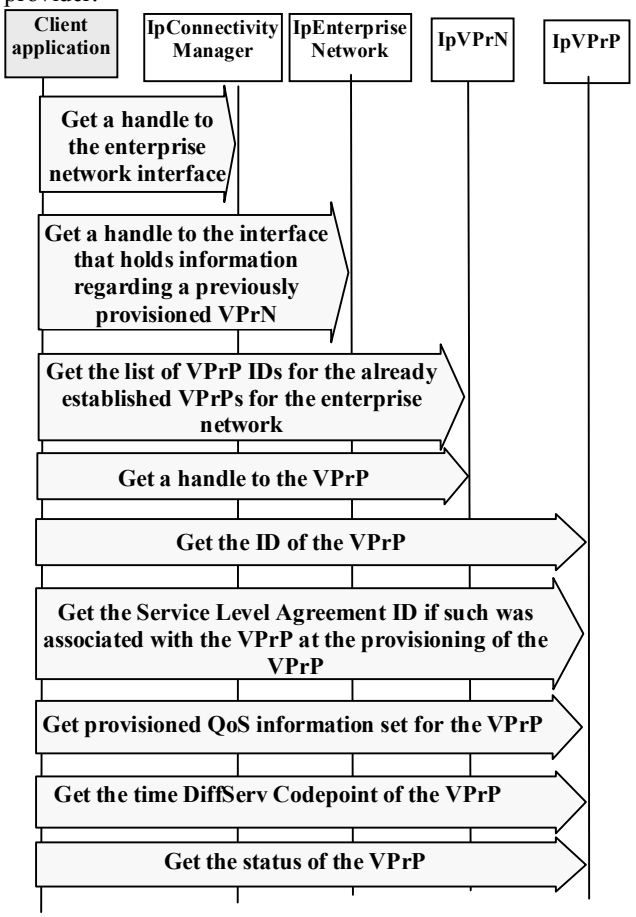


Fig.3 The Application browses a virtual provisioned pipe

III. OSA Connectivity Manager and Policy Control Requirements

In IMS control architecture the functional entities that are involved in policy QoS and policy control include Policy and Charging Enforcement Function (PCEF), Policy and Charging Rule Function (PCRF) and Call Session Control Functions (CSCFs). The PCEF encompasses service data flow detection, policy enforcement and flow based charging functionalities. The PCRF provides network control regarding the service data flow detection, gating, quality of service and flow based charging (except credit management) towards the PCEF. CSCFs manage service control, voice coder negotiation for audio communication, and authentication, authorization and accounting [10].

The requirements of Policy and Charging Control (PCC) in IMS are defined in [11], [12]. The following points include comments about the strengths and weaknesses of the existing OSA/Parlay Connectivity Manager SCF and their support to the PCC requirements.

Gating control applied on a per service data flow basis - To enable the PCRF gating control decisions, the CSCF has to report session events (e.g. session termination, modification) to the PCRF. Applications could access the PCC mechanism via an open API such as OSA/Parlay. The open API would be logically located at the IMS Application plane above the interface between CSCF and application servers. Existing OSA/Parlay Connectivity Manager does not support modification of VPrPs. This needs to be included in the open API requirements. Accordingly, applications could dynamically react to QoS notifications, or modify existing QoS parameters. Existing OSA Connectivity Manager does not support notification of resource changes. Notification can be synchronous or asynchronous. There is a need to formalize how this feedback mechanism is specified.

QoS control per service data flow - The PCEF enforces the authorized QoS for each specific service data flow. Criteria such as the QoS subscription information may be used together with policy rules such as, service-based, subscription-based, or pre-defined PCRF internal policies to derive the authorized QoS to be enforced for a service data flow.

QoS control of QoS reservation procedures (UE-initiated or network-initiated) for IP connectivity access networks (IP-CAN) - QoS service capabilities available at the IP-CAN include Differentiated Service and MPLS-based Traffic Engineering at the edge. Since these are technology-specific examples that may require specific knowledge of the network, the network functions need to be abstracted for the open API to be used at an application level.

QoS Conflict Handling - The PCC architecture must support conflict resolution when the authorized bandwidth associated with multiple PCC rules exceeds the Subscribed Guaranteed bandwidth QoS. If there are multiple requests from CSCFs, then each request can be assigned a priority. In the case of multiple requests, the PCC subsequently reacts to the request with highest priority. In the case of single requests, the PCC doesn't need the CSCFs specified priority.

IV. Improvement of API for Connectivity Management

Existing OSA Connectivity Manager does not support notification of QoS resource changes. Instead, application can only poll the server for network changes.

To meet the requirements for PCC, OSA compatible interfaces for QoS notification service have to be defined. Notification would be assisted by OSA Framework. Notification can be synchronous or asynchronous. There is a need to formalize how this feedback mechanism is specified. We suggest support of five new methods for the Connectivity Manager interface to manage the registration for notifications. The method createNotifications() can be used to enable

notifications so that QoS events can be sent to the application. . This is the first step an application has to do to get initial notifications of QoS events happening in the VPrN. When such an event happens, the application will be informed by reportNotification(). In case the application is interested in QoS events in the context of a particular VPrP it has to use the eventReportReq() method on the VPrP object. The method destroyNotifications() can be used by application to disable call notification. The method changeNotification() can be used by the application to change event criteria introduced with createNotification. The methods enableNotifications() and disableNotifications() can be used to indicate that the application is able to receive notifications which are provisioned from within the network or disable aaccordingly. We also suggest a new application interface IpAppConnectivityManager for receiving notifications from the IpConnectivityManager interface. The new interface supports the method reportNotification() which notifies the application of the arrival of QoS related events. To report QoS events which are specific for a particular VPrP we propose a new application interfaces IpAppVPrP which supports eventReportRes() method. This asynchronous method reports that an QoS event related to a particular VPrP has occurred that was requested to be reported.

Fig.4 shows the usage of the newly defined tools for provisioning notifications for QoS events.

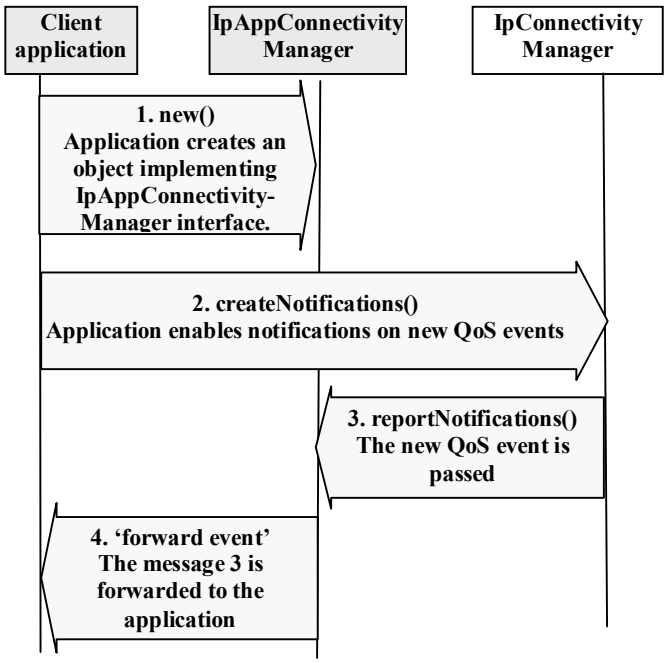


Fig.4 OSA application registers for and receives notification of a QoS event

V. CONCLUSION

In this paper we analyze the functionality for open access to connectivity management in IP-based multimedia networks. The paper focuses on OSA Connectivity Management APIs and analyzes the interface to the requirements to policy control based on recent 3GPP standard Policy and Charging Control Architecture.

We describe the strengths and weaknesses of the existing API specifications and provide suggestions about changes that are needed to improve the specification.

ACKNOWLEDGEMENT

The research is partially supported by the Research project 080910dni-7 funded by the Research and Development Sector of Technical University of Sofia.

REFERENCES

[1] Miikka Poikselka, Georg Mayer, Hisham Khartabil, Aki Niemi, The IMS Multimedia Concepts and Services, Wiley, 2008
[2] ITU-T Y.1291, An architectural framework for support of Quality of Service in packet networks
[3] Hu Hanrahan, Network Convergence, Services, Applications, Transport and Operations Support, Wiley, 2008
[4] 3GPP TS 29.198-1, Open Service Access (OSA); Application Programming Interface (API); Part 10: Connectivity Manager Service Capability Feature (SCF)
[5] 3GPP TS 29.198-10, V8.0.0, Open Service Access (OSA); Application Programming Interface (API); Part 1: Overview.
[6] Samson Lee, John Leaney, Tim O'Neill and Mark Hunter, Open Service Access for QoS Control in Next Generation Networks – Improving the OSA/Parlay Connectivity Manager, Operations and Management in IP-Based Networks, Springer Berlin / Heidelberg Volume 3751/200529-38
[7] J Yan and H Hanrahan, VPN Provisioning using the OSA Parlay Connection Management Interface, http://www.ee.wits.ac.za/comms/Telecomms%20output/output/satnac04/Yan.pdf
[8] S. Blake, D. Black, M. Carlson, E. Davies Z. Wang, W. Weiss, RFC 2475, An Architecture for Differentiated Services
[9] 3GPP TS 29.198-3, Open Service Access (OSA); Application Programming Interface (API); Part 3: Framework
[10] Inge Grønbæk, NGN, IMS and service control – collected information, R&I Research Note N 31/2006
[11] 3GPP TS 23.203 v8.4.0, Policy and charging control architecture
[12] ETSI TS 183 017 v1.1.1, Resource and Admission Control: DIAMETER protocol for session based policy set-up information exchange between the Application Function (AF) and the Service Policy Decision Function (SPDF); Protocol specification

On Cross-layer Design of Wireless Mesh Networks Using Network Coding

Nikolay Petrov[1], Vladimir Poulkov[2] and Georgi Iliev[3]

Abstract – **This paper presents an overview of network coding and its application in cross-layer design. It includes a simple network coding example. A reasoning is given for choosing the cross-layer design approach over a layered design approach. An example of cross-layer design scenario, based on primal-dual interior point method, is briefly reviewed in a simplified and comprehensible manner, using a basic data model representation.**

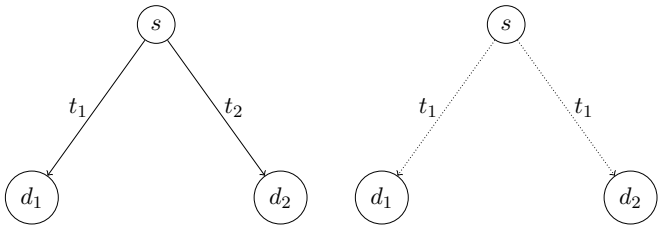
Keywords – **Network coding, Cross-layer design, Wireless mesh networks.**

I. INTRODUCTION

Nowadays, in the majority of existing computer networks, each node functions as a switch in a sense that it either forwards or replicates information from an input link to an output link or to a certain set of output links. Nevertheless, the specifics of the information flow in today's networks and the emergence of new services is a good reason to reconsider and reconstruct the idea of how the network should work.

In 2000, a paper by Ahlswede, Cai, Li and Yeung [1] introduced the idea of network coding. It suggests that the intermediate nodes in a network are allowed not only to route but also to combine incoming data from different nodes with coding operations. The complexity of coding is an important issue, related to additional processing time at some of the intermediate nodes, which are not interested in the transmitted information, in general. Consequently, in practice, the simplest and preferred method for such coding is to apply XOR operations to the input packets at the node that performs the network coding, and to output the result to the destination adjacent nodes.

Network coding has been proven to be effective for multicast applications [2, 3], and according to recent studies [4–6] in many scenarios of multiple unicast applications when such coding is employed in a multi-hop wireless environment. Leading direction of the research on network coding is related to mobile networks. One of the main reasons for that is the broadcasting feature of the wireless channel, which plays an important role in network coding over multiple unicasts. As shown in Fig. 1a in the case of wireline communication the transmission is limited to one destination d_i (one port) at a time, and requires number of channels or time slots (t_1, t_2) equivalent to the number of data sessions ($s \rightarrow d_1$, $s \rightarrow d_2$). On the contrary, the wireless channel Fig. 1b allows concurrent, at the same time t_1, transmission of packets to multiple destinations or sinks, situated in the coverage of the transmit antenna.



(a) Wireline communication (b) Wireless communication

Fig. 1. Advantage of broadcasting in aggregate capacity gain.

II. AN EXAMPLE OF NETWORK CODING

Consider the network topology presented in Fig. 2a. The indices of s and d relate to three independent unicast sessions, and b_1, b_2 and b_3 correspond to the information in each of these sessions. Every node is equipped with one radio; therefore, it can either transmit or receive information at the same time. Using a traditional approach as shown in Fig. 2b it would take 6 time slots in order to transmit the data of all of the three sessions to its destinations.

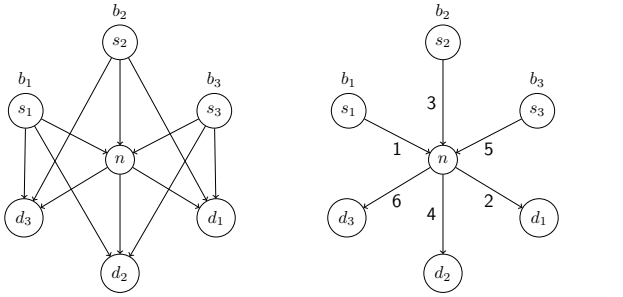
As known, in a network with one source and one sink the maximum information flow is limited by the weakest set of links which cut the source form the sink completely. In this case node n acts as a bottleneck and has to be used three times. Fig. 3 presents the case when network coding is enabled. In this example the network coding is actually the utilization of node n to perform XOR operations and later on decoding the information at the destination nodes by XORing the received data bits.

In details, the communication goes as follows. In the first step (Fig. 3a) the message b_1 sent from node s_1 reaches nodes d_3, d_2, and n. In the second step (Fig. 3b) s_2 sends the message b_2 to nodes d_3, n and d_2. In the third step (Fig. 3c) the nodes n, d_2 and d_1 receive b_3 from the source node s_3. After node n has received b_1, b_2 and b_3 it performs the network coding and broadcasts the message ($b_1 \oplus b_2 \oplus b_3$) to d_3, d_2 and d_1 in the fourth step (Fig. 3d). Now by applying XOR operation to the three messages received at each of the destination nodes the original message intended for the particular node is retrieved. The achieved network throughput gain due to the network coding in this case is 1.5.

[1]Nikolay Petrov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: petrovn@ieee.org.

[2]Vladimir Poulkov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: vkp@tu-sofia.bg.

[3]Georgi Iliev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gli@tu-sofia.bg.



(a) Network topology (b) Sequence of channel occupation

Fig. 2. A three-source/three-destination unicast wireless network and the sequence of channel occupation in the case when network coding is disabled

III. Motivation for Choosing Cross-Layer Design Over Layered Design

The widely used layered architecture models as the OSI model and the Internet Protocol Suite (TCP/IP) facilitate handling the complexity of large communication networks. Each layer fulfills a limited, well-defined purpose and delivers a digest and simpler model of the network to its upper layer. This allows splitting the communication network design into several smaller design problems which are easier to solve.

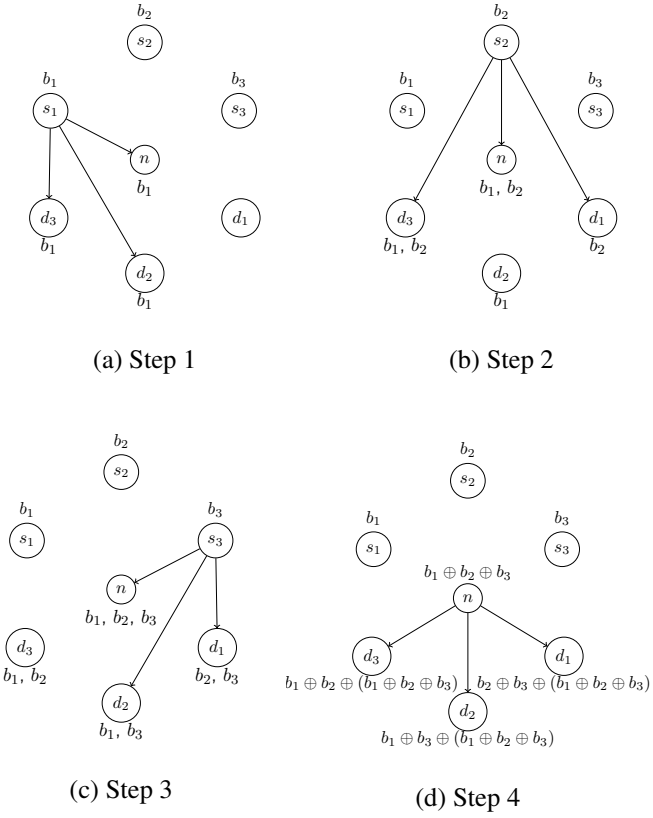
Nevertheless, with the emergence of wireless networks and other new networking technologies in the past decade, environments and circumstances have changed. The characteristics of the wireless networks are quite different from wireline systems. System developers and researchers face different problems and challenges compared to the wireline networks. In this altered situation, a layered design approach can be in general suboptimal to a cross-layer design approach, where several layers are designed jointly.

In a cross-layer design scheme several different aspects of the layered model can be taken into account. Some of these can be congestion control, energy-conservancy, optimal channel use (according to achievable rates for transmission over the wireless medium), resource allocation, interference management, transmission scheduling and media access schemes, delay constrains. Such an algorithm is presented in [7].

IV. Problem Formulation and Data Model

The algorithm presented in [7] is considered. In this method, a pseudobroadcasting technique is taken into account. It complicates the data representation in the means of adding new virtual nodes and corresponding to them links. For the purpose of simplicity and in order to reproduce clearer the essence of the algorithm, pseudobroadcasting will not be explained in this paper.

The network is modeled as a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$. Where $\mathcal{N} = \{1, 2, ..., N\}$ represents the nodes in the network and $\mathcal{L} = \{1, 2, ..., L\}$ the links in the network. $\boldsymbol{I} = \{I_1, I_2, ..., I_N\}$ is defined as the radio distribution vector, where I_i is number of radios at the i^{th} node. The network topology can be represented by an N x L indicator matrix $\boldsymbol{D}$, whose entries $D(n, l)$ satisfy $D(n, l) = 1$ if the n^{th} node is the start node of the l^{th} link, $D(n, l) = -1$, if the n^{th} node is the end node of the l^{th} link, $D(n, l) = 0$, otherwise. There are T independent data sessions in the network and for each one R_t defines the data rate. They form a unicast data rates vector for the T sessions: $\boldsymbol{R} = \{R_1, R_2, ..., R_T\}$. s_t and d_t will denote respectively the source node and the destination node of the t^{th} network session. The aggregate flow rate over the l^{th} link, where $(1 \leq l \leq L)$ is denoted with f_l. The data flows are assumed to be lossless across the links and the traffic flow can be split arbitrarily at nodes as long as the flow conservation law is satisfied. $\mathcal{F}(\boldsymbol{R})$ forms the set of all possible network flow vectors that support $\boldsymbol{R}$. $\mathcal{C}(\boldsymbol{I})$ represents the region containing all achievable rate vectors $\boldsymbol{c} = \{c_1, c_2, ..., c_L\}$. By time-sharing over a large number of time slots of the available links [8], the link capacity set $\mathcal{C}$ can be represented as a convex hull. Therefore, $\mathcal{C}$ is fully determined by its vertices and it can be defined using another, approximate to $\mathcal{C}$, convex hull $\mathcal{C}' \subseteq \mathcal{C}$. Depending on different scenarios for the power used by the transmitter, and the interference, and noise at the receiver, a finite number of transmission rates may be defined. This leads to a finite number K of feasible link-rate vectors $\mathcal{V}' = \{\boldsymbol{c}_k \in \mathcal{C}, k = 1, 2, ..., K\}$. Now,



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

Fig. 3. Example of network coding technique utilized in the network given in Fig. 2 (a).

using the set $\mathcal{V}'$, the corresponding convex hull

$$\mathcal{C}' = \left\{ \boldsymbol{c} | \boldsymbol{c} = \sum_{k=1}^{K} \alpha_k \boldsymbol{c}_k, \text{ s.t.} \sum_{k=1}^{K} \alpha_k = 1, \alpha_k \geq 0, \forall k \right\} \quad (1)$$

is fully defined. Behind the simple definition of the sets $\mathcal{F}(\boldsymbol{R})$ and $\mathcal{C}(I)$ given here underlies a detailed mathematical interpretation. It is discussed (in details) in [7, 8] and shows how resource allocation, data scheduling, routing, and network coding schemes characterize the two sets $\mathcal{F}(\boldsymbol{R})$ and $\mathcal{C}(\boldsymbol{I})$. Given these notations the optimization problem is defined as [7]:

$$\begin{aligned} &\text{maximize } U(\boldsymbol{R}, \boldsymbol{f}) \\ &\text{subject to: } \boldsymbol{f} \in \mathcal{F}(\boldsymbol{R}); \\ &\boldsymbol{R} \succeq 0; \\ &\boldsymbol{c} \in \mathcal{C}(\boldsymbol{I}); \\ &f_l < c_l, \forall l. \end{aligned} \quad (2)$$

The constraints in Eq. (2) have the meaning as follows. The first constraint enforces the dependence between the achievable rates $\boldsymbol{R}$ and the data flows $\boldsymbol{f}$. Rates R_t should be non-negative, as stated in the second constraint. The third constraint indicates the relationship between the achievable link capacity $\boldsymbol{c}$ and radio allocation along with the resource allocation, scheduling, and data routing schemes. Finally, the fourth constraint states that the sum of the flow rate on each link is bounded by the link capacity.

The choice for the utility function $U(\boldsymbol{R}, \boldsymbol{f})$ may vary depending on the design target. It is practical to aim maximization of the throughput ($U(\boldsymbol{R}, \boldsymbol{f}) = \sum_{t=1}^{T} R_t$, i.e. the sum of the data rates in all the sessions) or to pursue fairness, maximizing the minimum end-to-end communication rates ($U(\boldsymbol{R}, \boldsymbol{f}) = \min\{R_t\}$). The utility function is assumed to be concave.

The surveyed algorithm involves joint consideration of the physical-layer, wireless Media Access Control (MAC), and network-layer planning. In the physical layer the achievable rates for transmission over the wireless medium is taken into account, considering issues such as resource allocation, interference management. The wireless MAC layer includes transmission scheduling and media access schemes. The network layer accounts the actual aggregate load over a particular link. Also, in order to prevent adverse effect on the upper layer performance, the rate flow region $\mathcal{F}(\boldsymbol{R})$ includes delay constrains in its formulation.

The discussed algorithm is based on a primal-dual interior point method. This approach offers very good control of the distance to optimality. The key idea here is to start from a rather loose tolerance ϵ and to build as fast as possible the first rough approximation to the original problem. Subsequently, the optimality tolerance required for the solution of the restricted master problem is tightened until a predefined threshold ϵ_0 is achieved or a maximum iteration number is reached. Depending on the choice of ϵ_0, the algorithm may give fast some locally optimal solution for the original problem Eq. (2) or eventually, taking longer, to reach the global optimum.

In this approach the problem described above is relaxed to the following problem:

$$\begin{aligned} &\text{maximize } U(\boldsymbol{R}, \boldsymbol{f}) \\ &\text{subject to: } \boldsymbol{f} \in \mathcal{F}(\boldsymbol{R}); \\ &\boldsymbol{R} \succeq 0; \\ &\boldsymbol{c} = \sum_{k=1}^{K} \alpha_k c_{\boldsymbol{k}}, \sum_{k=1}^{K} \alpha_k = 1; \\ &f_l < c_l, \forall l. \end{aligned} \quad (3)$$

This problem Eq. (3) is referred as the restricted primal problem. The solution of this problem provides a lower bound U_{lower} for the original problem. On the other hand the dual of the original problem provides an upper bound U_{upper} of Eq. (2). In [7] it is defined as:

$$\max_{\boldsymbol{R} \succeq 0, \boldsymbol{f} \in \mathcal{F}(\boldsymbol{R})} \left\{ U(\boldsymbol{R}, \boldsymbol{f}) - \sum_{l=1}^{L} \lambda_l f_l \right\} + \max_{\boldsymbol{c} \in \mathcal{C}(\boldsymbol{I})} \sum_{l=1}^{L} \lambda_l c_l, \quad (4)$$

$$\text{subject to: } \lambda_l \geq 0, l = 1, 2, ..., L$$

Given the solution of these two formulations the gap between the two bounds can be obtained. This gap, as an indication of the accuracy of the current optimization result, is compared to ϵ. From this result, using an iterative algorithm with column generation [9], the distance to optimality can be controlled.

V. RESULTS

Consider the network topology shown in Fig. 2a. It has the following parameters: $N = 7$, $L = 12$, $T = 3$. The source nodes are denoted as s_1, s_2 and s_3. The destination nodes are d_1, d_2 and d_3. The model of the network is given by the graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N} = \{s_1, s_2, s_3, n, d_1, d_2, d_3\}$ and $\mathcal{L} = \{l_1, l_2, ..., l_{12}\}$. For this example the radio distribution vector will be defined as $\boldsymbol{I} = \{1, 1, 1, 4, 3, 3, 3\}$. This indicates that the number of radios at the source nodes remains 1, as in the previous example; the node n and the destination nodes are equipped with multiple radios.

The performance of the algorithm employed in this scenario can be seen from the results obtained in [7] on Fig. 4. A feasible solution of the problem Eq. (3), before the ϵ_0 threshold is reached, corresponds to a valid suboptimal network coding solution. In such a solution the system throughput has the value of the current utility function U_{lower}. Using the discussed iterative column generation interior-point method, the optimal primal solution of Eq. (2) gradually coincides with its dual solution. As it is seen from the figure that the defined tolerance here is $\epsilon_0 = 0$ and the final solution is actually the optimum solution for this particular network. The maximum achieved throughput is $U_{lower} = U_{upper} = 3$. In the case when network coding is disabled and only routing is allowed the optimized system throughput is 2.

VI. CONCLUSIONS

For the short period of its existence the network coding approach has captured the attention of a significant number of researchers. Different approaches have been proposed concerning

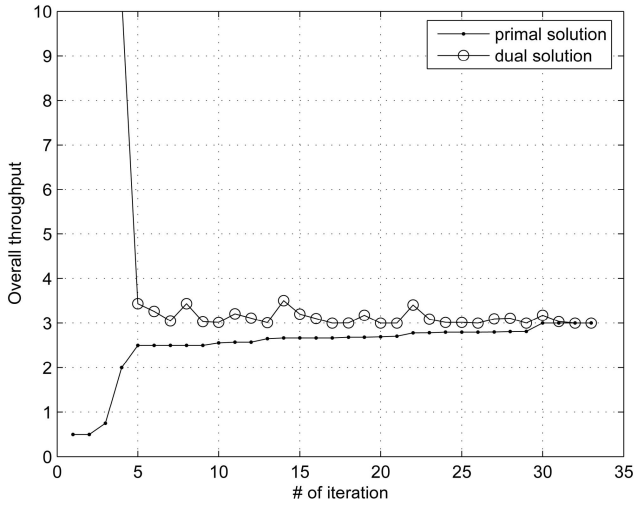


Fig. 4. Performance of the primal-dual algorithm [7] as a function of the increasing number of iterations for a three-source/three-destination network. The utility function of the optimization problem is the total system throughput.

different aspects of the network coding design. This paper has attempted to provide an overview of the cross-layer design for wireless mesh networks employing network coding to support multiple unicast applications. The reviewed approach considers the physical layer, MAC, network layer and sets up delay constraints. Generally, the process of optimization takes some time to converge to an optimal solution. However, the transmission can start before the convergence is done. This makes such an approach practical for application in wireless networks. Investigation in this direction is a future research topic.

ACKNOWLEDGEMENT

This work was supported by the Bulgarian National Science Fund Grant No. ДО-02-135/2008 “Research on Cross Layer Optimization of Telecommunication Resource Allocation”.

REFERENCES

[1] R. Ahlswede, N. Cai, S.Y.R. Li, and R.W. Yeung, “Network Information Flow,” *IEEE Trans. Information Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[2] R. Koetter and M. Médard, “An Algebraic Approach to Network Coding,” *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782-795, Oct. 2003.

[3] S.Y.R. Li, R.W. Yeung, and N. Cai, “Linear Network Coding,” *IEEE Trans. Information Theory*, vol. 49, no. 2, pp. 371-381, Feb. 2003.

[4] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, “XORs in the Air: Practical Wireless Network Coding,” *Proc. ACM SIGCOMM*, 2006.

[5] D. Traskov, N. Ratnakar, D.S. Lun, R. Koetter, and M. Médard, “Network Coding for Multiple Unicasts: An Approach Based on Linear Optimization,” *Proc. IEEE Intl Symp. Information Theory (ISIT 06)*, July 2006.

[6] T. Ho, D. Lun, *Network Coding, An Introduction*, Cambridge University Press, 2008.

[7] K. Li, X. Wang, “Cross-Layer Design of Wireless Mesh Networks with Network Coding,” *IEEE Trans. Mobile. Comp.*, vol. 7, no.11, pp.1363-1373, Nov. 2008.

[8] M. Johansson and X. Lin, “Cross-Layer Optimization of Wireless Networks Using Nonlinear Column Generation,” *IEEE Trans. Wireless Comm.*, vol. 5, no. 2, pp. 435-445, Feb. 2006.

[9] J. Gondzio and R. Sarkissian, “Column Generation with a Primal-Dual Method,” technical report, Logilab, HEC Geneva, Univ. of Geneva, Geneva, 1996.

Competitive Pricing Using Game Theory in the Next Generation Networks

Vesna Radonjić and Vladanka Aćimović Raspopović[1]

Abstract – **In this paper we present the game theory model for achieving Nash equilibrium in pricing Next Generation Networks services. We consider the competition between two Internet Service Providers offering the same service. Their competition is modelled as a simultaneous-play game in which the solution is obtained by Nash equilibrium. The proposed model is verified through numerous simulations performed by software that we developed for that purpose.**

Keywords – **Competitive Pricing, Quality of Service, Quality of Experience, Nash Equilibrium, Next Generation Networks.**

I. Introduction

Pricing is one of the important issues in Next Generation Networks (NGN). Through an appropriate pricing mechanism, a service provider offering NGN services tends to maximize his revenue, while users tend to achieve the highest satisfaction from service usage at the affordable price.

Two main factors influencing the price setting are users' demand and competition among service providers. Price and demands for the same or even similar services are mutually dependent. If the demand is high, a high price can be charged by a service provider and his revenue will be increased. On the other hand, if the demand is low, the price must be reduced to attract more users. Competition among the service providers impacts the price setting. If the services are substitutable users buy a service that provides the highest satisfaction at the lowest price. If one service provider reduces its offered price to attract more users and gain higher revenue, this will impact the revenues of other providers who will try to compete by reducing their offering prices as well.

In this paper we present a model for service competition and pricing in a NGN where two Internet Service Providers (ISPs) compete with each other to detain existing and attract new users in a particular service area. They try to achieve that goal with trade-off between Quality of Service (QoS) and price. In our model price and bandwidth consumption are optimized for users which bandwidth demand is in certain range of interest, named as partially elastic users. We consider the case where both ISPs offer their prices at the same time, i.e. simultaneous-play game. The solution of this competition is given by Nash equilibrium for which both ISPs are satisfied the solution in terms of prices.

The paper is organized in the following way. In Section 2 we explain the meaning of pricing, charging and billing processes and the pricing role in a QoS differentiation especially in NGN. In Section 3 both users' and ISPs optimization problems are presented and the model for solving these problems is proposed. In Section 4 simulation results are presented and analyzed. Conclusion is given in the Section 5.

[1] Vesna Radonjić (v.radonjic@sf.bg.ac.yu) and Vladanka Aćimović Raspopović (v.acimovic@sf.bg.ac.yu) are with the University of Belgrade – the Faculty of Transport and Traffic Engeneering, Vojvode Stepe 305, 11000 Belgrade, Serbia.

II. Pricing for the Next Generation Networks

A. Pricing Issue in the Next Generation Networks

Pricing is the process of determining tariffs, i.e., cost per unit. It is based on particular pricing model and controlled by a pricing policy. Charging combines the tariffs and the results of metering needed for the charge of users. The output of charging process is the charge per party (customer, service provider, content provider). The billing process produces an invoice on the basis of the charge per party [1]. The process can be configured by means of the billing policy, e.g., how often a bill is sent to a user. The payment process results in the actual transfer of money, based on an invoice as input.

QoS differentiation introduces a clear need for incentives to be offered to users and encourage them to choose the service that is most appropriate for their needs. In commercial networks, this can be most effectively achieved through pricing. Price discrimination of services is appropriate for encouraging service differentiation with the associated revenues that should be paid for any needed network expansions. NGN must be flexible enough to enable the use of different pricing models [2], [3]. Pricing model should fulfil a trade-off between providing satisfying user's utility and provider's revenue, still preserving implementation efficiency and feasibility. User's utility can be expressed as a function of available network resource offered to a user which indicates a user's sensitivity to changes in QoS.

It is suggested for NGN that the basic best-effort architecture should be left intact with QoS schemes solely reserved for resource intensive high quality real-time services [4].

B. Description of Quality of Service, Network Performance and Quality of Experience in the NGN

The QoS paradigm requires a network that could carry out service differentiation with packets serviced depending upon their value. QoS is defined in Recommendation E.800 as follows: "Collective effect of service performance which determines the degree of satisfaction of a user of the service". This definition is a wide one encompassing many areas of work, including subjective user satisfaction. However, in [4]

the aspects of QoS are restricted to the identification of parameters that can be directly observed and measured at the point at which the service is accessed by the user. Recommendation I.350 defines Network Performance (NP) as "NP is measured in terms of parameters which are meaningful to the network provider and are used for the purpose of system design, configuration, operation and maintenance. NP is defined independently of terminal performance and user actions". Quality of Experience (QoE) is defined as the overall acceptability of an application or service, as perceived subjectively by the end user. QoE includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc). Overall acceptability may be influenced by user expectations and context [4].

QoS provides a valuable framework for network provider, but it is not necessarily usable in specifying performance requirements for particular network technologies (i.e. IP, MPLS, etc.). Similarly, NP ultimately determines the user observed QoS, but it does not necessarily describe that quality in a way that is meaningful to users. QoE is subjective in nature, i.e. depends upon user actions and subjective opinions. The definition of QoS, NP and QoE should make mapping clear in cases where there is not a simple one-to-one relationship among them. Table 1 shows some of the characteristics which distinguish QoS, NP and QoE.

TABLE 1.
DISTINCTION BETWEEN QOE, QOS AND NP [4]

QoE	QoS	NP
User oriented		Provider oriented
User behaviour attribute	Service attribute	Connection/Flow element attribute
Focus on user-expected effects	Focus on user-observable effects	Focus on planning, development (design), operations and maintenance
User subject	Between (at) service access points	End-to-end or network elements capabilities

NP definition includes transmitting time and response time. Transmitting time is the time interval during which a packet is transmitted between two network nodes. Response time is the time interval between the requirement sending and the receiving of required data. In this paper we did not particularly observed those parameters. The analyzed service model reflects partially elastic users for whom the QoS can be determined solely as a function of the average bandwidth. We defined QoE parameters through positive constants that regulate the sensitivity of users' satisfaction to the QoS/price trade-off.

C. Simultaneous-play Game and Nash equilibrium

Pricing problem can be modelled as a simultaneous-play game between ISPs in which all of them aim to maximize their corresponding objective functions. A strategy profile is the vector containing the strategies of all players. Each strategy profile yields the payoffs to each player. We assume that there are two players in the game. If the payoff function for the *k*th player is $T_k(x)$, where $x=(x_1,x_k)$ is the set of decision variables and x_k is the decision variable of the *k*th player, then each player $k=1,2$ in this game has to maximize his payoff function:

$$\max_{x_k} T_k(x) \qquad (1)$$

The solution of this competition can be achieved as Nash equilibrium [5], [6].

Definition: A pure strategy Nash equilibrium is a strategy profile from which no player has a unilateral incentive to change his strategy.

In other words, Nash equilibrium is a state of the game where no player prefers a different action if the current actions of other players are fixed. Nash equilibrium can be interpreted as the best action that each player can play based on the given set of actions of the other players. Each player cannot profit from changing his action, and because the players are rational, this is a "steady state" [5].

III. SERVICE COMPETITION AND PRICING IN NGN

A. User's Utility and Bandwidth Demand

A utility function which best models user behaviour is a generalization of the logarithmic function employed, tailored for a connection oriented setting. QoS is defined by bandwidth obtained from the ISP. Depending upon the quality of service requested, each user would require a minimum bandwidth γ. Fewer bandwidth than γ on average are of no utility to the user [2]. The law of diminishing marginal utility ensures that the user derives the same amount of satisfaction from any bandwidth more than the maximum π (Figure 1).

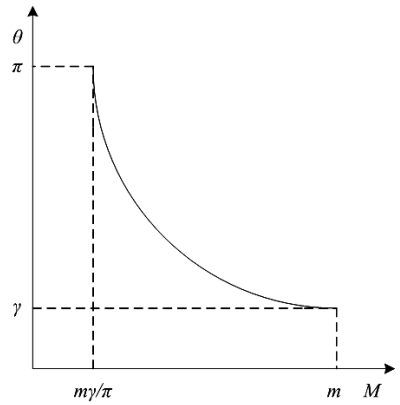


Figure 1: Desired bandwidth

It is considered that the user is willing to pay a maximum m per unit of bandwidth. When the ISP price M equals the maximal price m, the user will desire only the minimum acceptable bandwidth, γ. Any price beyond the maximal price reduces the user's desired bandwidth to zero. Over the interval $m\gamma/\pi \le M \le m$ the desired bandwidth θ decreases

logarithmic with price with π. Dependence of desired bandwidth with price for arbitrarily chosen user is shown in Figure 1.

According to [2], user utility function can be shown to be:

$$U(\theta)=\begin{cases} m\theta, & \text{if } 0\le\theta\le\gamma \\ m\gamma\left(\log\left(\theta/\gamma\right)\right)+1, & \text{if } \gamma<\theta\le\pi \\ m\gamma\left(\log\left(\pi/\gamma\right)\right)+1, & \text{if } \pi<\theta \end{cases} \quad (2)$$

This utility function (Figure 2) is concave and no decreasing. Also, U is strictly increasing on $[0,\infty)$ only when $\pi\to\infty$. The case of strictly elastic users can be obtained by setting $\gamma=0$ and $\pi\to\infty$, thereby rendering U strictly concave in $[0,\infty)$. Therefore, the utility function encompasses a wider spectrum of user behaviour by incorporating the range of user bandwidth requested [8].

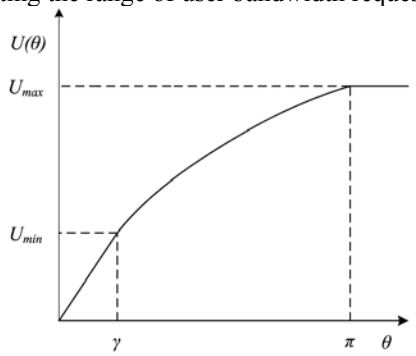


Figure 2: User utility as a function of bandwidth

We suppose that the shape of these functions (shown in Figures 1 and 2) is the same for all users, but parameters γ, π and m are different for different users.

In order to normalize user's utility function, we assume that for less bandwidth than γ user's utility is set to be zero and for more bandwidth than π user's utility is equal to 1. Then, this normalized user's utility function can be defined:

$$U_n=\begin{cases} 0, & \text{if } 0\le\theta\le\gamma \\ \dfrac{m\gamma\left(\log(\theta/\gamma)\right)+1}{m\gamma\left(\log(\pi/\gamma)\right)+1}, & \text{if } \gamma<\theta\le\pi \\ 1, & \text{if } \pi\le\theta \end{cases} \quad (3)$$

The value of θ maximizing utility function (under conditions $0\le M\le m$ and $\gamma\le\theta\le\pi$) is:

$$\theta^*(M)=\frac{m\gamma}{M}. \quad (4)$$

B. Competition model

User's demand D to accept a service is actually its satisfaction probability, which depends on the trade-off between QoS and price. Therefore, it is a function of normalized user utility U_n and price M. It can be defined as [8]:

$$D(M)=1-e^{-kU_n^A M^{-B}}, \quad (5)$$

where k, A and B are positive constants that reflect the sensitivity of users' satisfaction to the QoS/price trade-off: k is a normalization constant, A indicates user's sensitivity to the QoS and B denotes user's sensitivity to the price. For example, increasing A makes the users more sensitive to the QoS, while increasing B does the same to the price. This equation is very general and it points the intuitive behaviour that the satisfaction of a user increases as the quality increases and/or the price decreases.

As distinct from parameters γ and π, which are QoS parameters, k, A and B are QoE parameters.

We consider the game in which two ISPs compete with each other to offer a NGN service to the users. We modelled this problem in a form of Nash game which is simultaneous-play game. We assume that the revenue of each ISPk is given by:

$$T_k=M_k\cdot\sum_{i=1}^{N}D_{ki}(M_1,M_2),\quad k=1,2. \quad (6)$$

where M_1 and M_2 are prices offered by ISP_1 and ISP_2, respectively.

To include both service provider prices M_1 and M_2 in demand functions, we can formulate them in following way:

$$D_{1i}(M_1,M_2)=1-e^{-kU_{ni}^A M_1^{-B}M_2^C} \quad (7)$$

$$\text{and } D_{2i}(M_1,M_2)=1-e^{-kU_{ni}^A M_2^{-2B}M_1^{C/2}} \quad (8)$$

where C is a positive constant which indicates variation of the user's demand for the service offered by one ISP, depending of the price offered by the competing provider. We suppose that ISP_1 is a new provider in the market. Because of that users are more sensitive on change in price offered by ISP_1, comparing with the price offered by ISP_2. This is shown in Equations (7) and (8).

The range of partially elastic users interest for bandwidth is $\gamma\le\theta\le\pi$. Then, revenue functions for provider 1 and 2, respectively are given by:

$$T_1(M_1,M_2)=M_1\cdot\left(1-\sum_{i=1}^{N}e^{-kU_{ni}^A M_1^{-B}M_2^C}\right)=$$
$$M_1\cdot\left(1-\sum_{i=1}^{N}\exp\left(-k\left(\frac{m_i\gamma_i\left(\log(m_i/M_1)\right)+1}{m_i\gamma_i\left(\log(\pi_i/\gamma_i)\right)+1}\right)^A M_1^{-B}M_2^C\right)\right) \quad (9)$$

and

$$T_2(M_1,M_2)=M_2\cdot\left(1-\sum_{i=1}^{N}e^{-kU_{ni}^A M_2^{-B}M_1^C}\right)=$$
$$M_2\cdot\left(1-\sum_{i=1}^{N}\exp\left(-k\left(\frac{m_i\gamma_i\left(\log(m_i/M_2)\right)+1}{m_i\gamma_i\left(\log(\pi_i/\gamma_i)\right)+1}\right)^A M_2^{-2B}M_1^{C/2}\right)\right) \quad (10)$$

The best response of the ISP_1 can be obtained from the optimal price M_1^* for which revenue $T_1(M_1^*,M_2)$ is maximized, given the price M_2 offered by the ISP_2. Similarly, the best response of the ISP_2 is the optimal price M_2^* for which revenue $T_2(M_1,M_2^*)$ is maximized given the price M_1 offered by the ISP_1. This best response is denoted

by $B_k\left(M_p\right)=\arg\ \max_{M_k} T_k\left(M_k,M_p\right)$, where M_p is the price offered by the other ISP. Nash equilibrium gives the set of prices such that none of the service providers can increase the revenue by choosing a different price, with the given price offered by the other service provider. This is the point where $B_1\left(M_2{*}\right)=M_1{*}$ and $B_2\left(M_1{*}\right)=M_2{*}$.

IV. Simulation Results

For the purpose of carrying out simulations of the competitive pricing model, we developed software in C Sharp. In Figure 3, application for determining Nash equilibrium in a simultaneous-play game is presented. Best responses namely, the best prices offered by one ISP to a NGN user with the given prices offered by the other ISP are shown in Figure 4. For model parameters, as shown in Figure 3, Nash equilibrium is obtained for $M_1=0.64$ and $M_2=0.78$. With the given price $M_1=0.64$, ISP_2 cannot increase his revenue by choosing a different price than $M_2=0.78$. The same stands for ISP_1: his best response to price $M_2=0.78$ is $M_1=0.64$, i.e. the price given by Nash equilibrium.

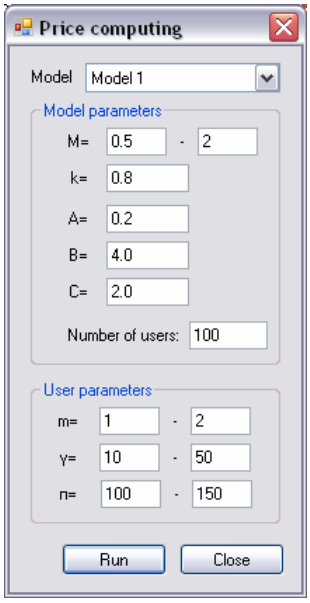


Figure 3: ISP computing prices interface

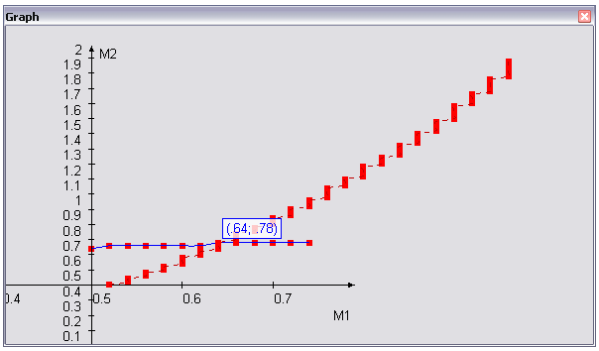


Figure 4: Best responses of both ISPs and Nash equilibrium

V. Conclusion

This paper considers one possibility for modelling the competition between two Internet Service Providers offering the same NGN service. We modelled this problem as a simultaneous-play game, assuming that both ISPs offer their prices at the same time. Both ISPs revenues depend on total users' demand for the service which is a function of acceptable QoS and QoE parameters. Furthermore, important contribution of this paper is defining distinction between QoS and QoE aspects. We supposed that reputations of ISPs are not the same, but the model also gives opportunity of changing market positioning of ISPs. The solution of this competition model is given by Nash equilibrium for which both ISPs are satisfied with the solution related to prices. We verified the proposed model through simulations with software solution especially developed for that purpose.

The important advantage of presented model is in stimulation of each user to choose the amount of available bandwidth to be charged for. At the same time, ISP considers users preferences, defined through QoS and QoE parameters. For solving the competition problem between ISPs, simultaneous-play game with Nash equilibrium proved to be a good scenario.

Acknowledgments

This paper resulted from the researching project supported by Serbian Ministry of Science and Technological Development.

References

[1] V. Aćimović-Raspopović, M. Stojanović, «Accounting in QoS-enabled IP Networks», *Proceedings of the POSTEL*, Belgrade, December pp. 303-312, 2005.

[2] B. M. Ninan, M. Devetsikiotis, „Game-Theoretic Resource Pricing For The Next Generation Internet", *Performance Evaluation and Planning Methods for the Next Generation Internet*, edited by Andre Girard, Brunilde Sanso and Felisa Vazquez Abad, Springer, pp. 141-163, 2005.

[3] V. Radonjić, V. Aćimović-Raspopović, "Responsive Pricing Model with Fixed Bandwidth Usage for the Next Generation Internet", *Proceedings of the ICEST 2008, Volume 2*, Nis, Serbia, June pp. 425-428, 2008.

[4] *ITU-T NGN FG Proceedings Part II*, ITU, 2005, available at: www.itu.int/ITU-T/ngn/files/NGN_FG-book_II.pdf

[5] *Yishay Mansour,* Computational Game Theory, lecture notes available at: http://www.math.tau.ac.il/~mansour/course_games/2006/lecture6.pdf

[6] M. J. Osborne, *An Introduction to Game Theory*, Oxford University Press, 2004.

[7] R. Cocchi *et al.*, "Pricing in Computer Networks: Motivation, Formulation and Example," *IEEE/ACM Trans. Net.*, vol. 1, no. 6, pp. 614–27, 1993.

[8] L. Badia *et al.*, "Pricing VoWLAN Services through a Micro-Economic Framework," *IEEE Wireless Commun.*, vol. 13, no. 1, Feb. 2006, pp. 6–13.

Session

TELETRAFFIC ENGINEERING

Evaluation of traffic self-similarity influence on QoS in local Ethernet network

Evgeniya Gospodinova[1], Nina Sinyagina[2] and Raycho Ilarionov[3]

Abstract - **In the paper is investigated the influence of self-similar traffic on the Quality of Service (QoS) in local Ethernet network in regard to the parameters: percentage of the information packet loss, average delay and jitter of the packets. The obtained results could be used in research of high reliability networks for remote control in online engineering with higher QoS requirements.**

Keywords – **Hurst parameter, Quality of Service, self-similar network traffic, wavelet-based statistical estimator.**

I. INTRODUCTION

Many studies for real network traffic show that it exhibits self-similar property over a wide range of time scales [4]. The properties of self-similar network traffic are very different from traditional models based on Poisson and related processes [1]. The using of traditional models in the networks characterized by self-similar processes can lead to incorrect conclusions about the QoS of analyzed networks [9]. The main properties of self-similar processes include: slowly decaying variance, long-range dependence, 1/f-noise and Hurst effect. These processes are attractive models mainly because they can be characterized by a single parameter, the *Hurst parameter H*, which can be estimated by using the wavelet-based statistical method [2].

The local network design and operation is often required to ensure a certain QoS for the user or to define the network input parameter range, within the limits of which the required QoS level will be sustained in the network parameter metrics.

The paper goals are as follow:

- To present and analyze an algorithm for simulating a self-similar Ethernet traffic;
- To estimate the influence of self-similar traffic over the QoS of local Ethernet network.

II. AN ALGORITHM FOR SIMULATING A SELF-SIMILAR ETHERNET TRAFFIC

Ethernet protocol requires the length of all packets to be between 64 and 1518 bytes. For generating simulated network traffic, an average value, variance, Hurst parameter and the packets' length has to be assigned.

[1]Evgeniya Gospodinova is with the Central Laboratory of Mechatronics and Instrumentation at Bulgarian Academy of Sciences, Sofia, Bulgaria, E-mail: jenigospodinova@abv.bg.

[2]Nina Sinyagina is with the Institute for Parallel Processing at Bulgarian Academy of Sciences, Sofia, Bulgaria, E-mail: nisi@acad.bg.

[3]Raycho Ilarionov is with the Technical University of Gabrovo, Bulgaria, E-mail: ilar@tugab.bg.

The main steps of the mathematical algorithm are:

Step 1 – The input parameters are: traffic length, average length of the packet, minimal and maximal sizes of the packet, Hurst parameter, and relative average error of the Hurst parameter, number of simulated self-similar traffic streams, and the maximal number of tests, based on the wavelet statistical method for determining the degree of self-similarity of the simulated network traffic.

Step 2 - Generating of pseudo-random self-similar sequence with an average value equals zero, variance equals one and a value of the Hurst parameter in the range of 0.5 and 1.

Step 3 - The simulated self-similar sequence is normalized in order to generate only positive numbers. This is accomplished through the transformation formula $Y_i=2^{Xi}$, where X_i is a positive or negative number of the generated self-similar sequence, and Y_i is the normalized positive number, relative to the size of the informational packet of the simulated network traffic.

Step 4 - The simulated self-similar sequence from step three is normalized one more time using the assigned average value and variance.

Step 5 - Using a wavelet-based statistical method, the value of the Hurst parameter of the generated self-similar Ethernet traffic is determined.

Step 6 - The value of the Hurst parameter determined by step five is compared with the input value of the Hurst parameter. If the new value of the Hurst parameter is in the relative average error range, then save the simulated network traffic in a file, otherwise go back to step two.

Step 7 – Check whether the number of the generated network traffics is equal to the maximum number of the simulated self-similar traffic streams. If it is not go to step two, otherwise the process is over.

III. COMPARISON ANALYSIS OF SIMULATED AND REAL ETHERNET TRAFFIC

The validity of the suggested algorithm for simulating self-similar Ethernet traffic is determined by comparing the results of the above algorithm with those published online [10] by Bellcore laboratory's researchers. The results are for Ethernet traffic with self-similarity degree H=0.8.

Fig. 1 shows graphically the simulated Ethernet traffic, using low (H=0.6) and high (H=0.9) degree of self-similarity. It is obvious that the higher degree of self-similarity, the higher level of correlation exists.

Table I demonstrates the average value and the standard deviation of the packet size of the simulated Ethernet traffic, using different values of the Hurst parameter. The average value of the packet size of the simulated Ethernet traffic is

lower than the average value between minimal (64 bytes) and maximal (1518 bytes) packet sizes.

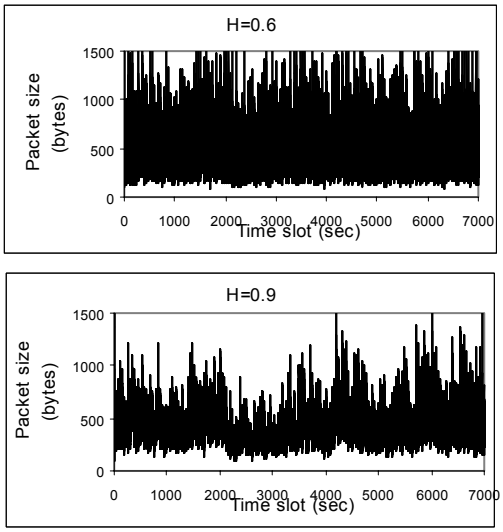


Fig. 1. Sample of synthetic Ethernet traffic

TABLE I
PACKET SIZE MEAN AND STANDART DEVIATION OF SIMULATED ETHERNET TRAFFIC

HURST PARAMETER	MEAN VALUE	STANDART DEVIATION
0.6	486.990	279.530
0.7	479.554	275.554
0.8	463.780	262.309
0.9	445.746	225.977

Fig. 2 displays the autocorrelation functions of the simulated Ethernet network traffic. Three different values of the Hurst parameter are used along with the autocorrelation function of the real Ethernet traffic, when the H=0.8 (BC_pAug89_TL). It can be concluded that the difference between autocorrelation functions of the simulated and the real network traffic it is really minor when the H=0.8.

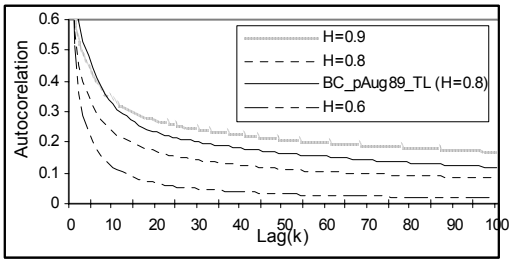


Fig. 2. Autocorrelation function plot of Ethernet traffic

On Fig. 3 are shown the graphical results of the wavelet-based method for determining the Hurst parameter of the simulated and real Ethernet traffic when the H=0.8.

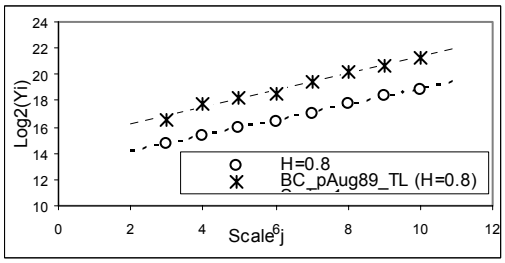


Fig. 3 Wavelet-based plots for Ethernet traffic

Table II demonstrates the determined value of the Hurst parameter and the relative error of the simulated Ethernet traffic. In this case the Hurst parameter is calculated using the wavelet-based statistical method.

TABLE II
PACKET SIZE MEAN AND STANDART DEVIATION OF SIMULATED ETHERNET TRAFFIC

INPUT HURST PARAMETER	SIMULATED ETHERNET DATA		REAL ETHERNET DATA	
	MEAN HURST	RELATIVE ERROR (%)	MEAN HURST	RELATIVE ERROR (%)
0.6	0.600779	0.129806	-	-
0.7	0.703220	0.459953	-	-
0.8	0.800333	0.041686	0.82	2.5%
0.9	0.897661	-0.25988	-	-

The results show that the simulated Ethernet traffic generated by the described above algorithm is very similar to the real Ethernet traffic. A wavelet-based statistical method is used to determine the relative error of the Hurst parameter when the input value H=0.8. The result shows that the error is approximately 2.5%, which is acceptable.

IV. RESEARCH THE INFLUENCE OF THE SELF-SIMILARITY ON QUALITY OF SERVICE IN LOCAL ETHERNET NETWORK

The influence of the self-similarity on the QoS in local Ethernet network is determined in regard to the following parameters: the percentage of the information packet loss, the average delay of the packets in the switch's queue, and the jitter of the packets of multiplexed network traffic (which is a

result of the fusion of self-similar traffic streams at the switch's input ports).

The simplified model of a network node (Fig. 4) is consisted of an Ethernet switch. Consider that, two self-similar Ethernet streams are at the switch's input ports and each of them has a maximal speed of 1 Gbps. Moreover, at the ports there are three more traffic streams with total maximum input speed of 1 Gbps. Thess five input traffic streams form all together one self-similar traffic stream with total maximal speed 3 Gbps.

The speed of servicing the packets is 2 Gbps. The simulation time is 32 768 (2^{15}) seconds. The following sizes of the switch's buffer are analyzed: 1 MB, 25 MB, 50 MB, 75 MB, 100 MB and 125 MB. The bottleneck of the analyzed network node is the connection of the Ethernet switch with the feed network, which is with 2 Gbps capacity.

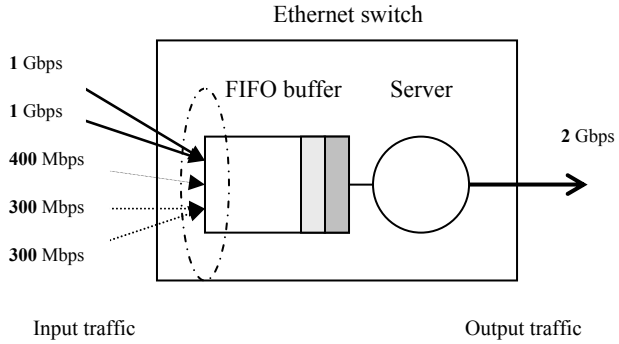


Fig. 4. Network node

Information packets loss

Information packet loss is the ratio of unserviced information packet to the total number of the input packets.

On Figure 5 are shown the graphical results of the information packets loss in respect of the network channel capacity. The traffic streams passing through this channel are with different self-similarity degree. It can be concluded that the highest loss of packets is when H=0.9 and the lowest loss is when H=0.6, 0.7, 0.8 and it is approximately 4 times lower when the channel capacity is 2 Gbps. When the channel capacity is 3 Gbps the packets loss incline to zero. The packet loss of the Poison traffic is 14 times lower than the self-similar traffic stream when H=0.9, and from 2 to 3 times lower when H=0.7 and 0.8 with a channel capacity 2 Gbps.

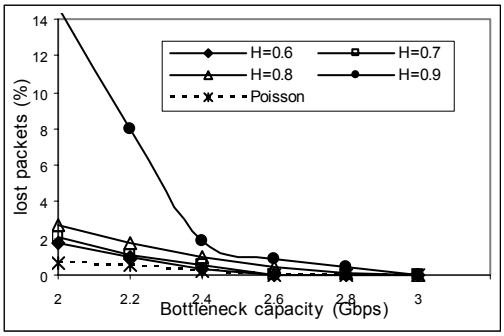


Fig. 5. Estimate of multiplexed self-similar traffic effect on the percentage of lost packets

Network traffic intensity

Traffic intensity is the ratio of the mean rate of packet arrival to the mean rate of service.

The network traffic intensity shows the extent to which the switch capacity is used. If the intensity is closed to one, the switch is overloaded and there is a queue of packets waiting to be served and it is highly possible part of them to be discarded. Contrariwise, if the traffic intensity is closed to zero, then the switched capacity is not entirely used.

On Fig. 6 is shown graphically the relation between traffic streams intensity, Hurst parameter, and network channel capacity. The intensity of the self-similar multiplexed network traffic when H=0.9 and the capacity is between 2.0 and 2.6 Gbps, decreases to 1.0 erl. Then the traffic intensity decreases to 0.9 erl. The traffic intensity of the self-similar multiplexed traffic streams when H=0.6, 0.7 and 0.8 linearly decreases with the network capacity increase. These self-similar traffic streams are with e relatively high traffic intensity. Poison traffic intensity linearly decreases when the channel capacity increases, but it stays less than 0.6 erl for the whole traffic capacity range. The Poison traffic intensity is 0.4 erl lower than the self-similar multiplexed traffic stream intensity when H=0.9, and 0.2 erl lower when H=0.6, 0.7, 0.8.

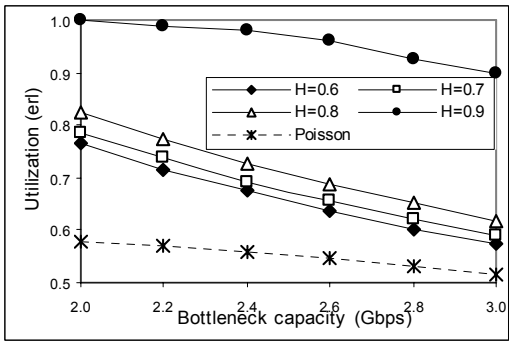


Fig. 6. Estimate of multiplexed self-similar traffic effect on the system utilization factor

Delay

The delay of the information packets in the switch buffer is the ratio of the total time the packets stay in the switch buffer to the number of input packets for a specific period of time.

On Fig. 7 is shown the relation between traffic stream delays, Hurst parameter and channel capacity. The packets delays are highest when H=0.9, and the linearly decrease with capacity channel increase. When the capacity is 2.2 and 2.4 Gbps the self-similar traffic streams delays decrease and they remind unchanged beyond these values. The Poison traffic stream delays are the lowest ones, as they remain almost unchanged with capacity channel increases.

Jitter

Jitter is the variation of delayed packets. It can be determined by the average absolute deviation and is equal to the mathematical expectancy of the absolute value of the subtraction of the packet delay by the average packets delay.

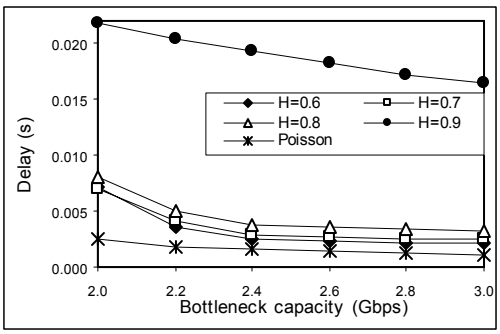


Fig. 7. Estimate of multiplexed self-similar traffic effect on the average delay

On Fig. 8 is shown the relationship between jitter, Hurst parameter, and network traffic capacity. Increasing the capacity values, Jitter decreases for all Hurst parameter values. Poison traffic jitter is from 1.5 to 2 times lower than the self-similar traffic stream jitter within the capacity range 2.2 and 2.4 Gbps.

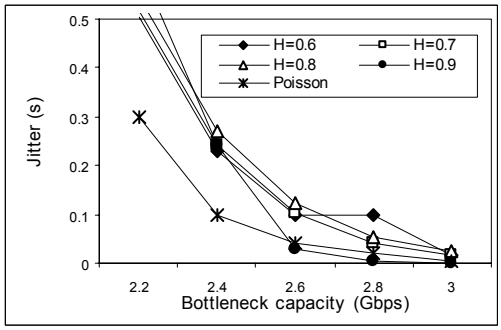


Fig. 8 Estimate of multiplexed self-similar traffic effect on the jitter

V. CONCLUSION

The obtained results in this paper show that self-similar traffic has potentially serious implications on local Ethernet networks. The lost packets, delay and jitter may by significantly higher than predicted by classical models. Poisson traffic models, as opposed to self-similar traffic models, may underestimate Ethernet packet loss by 2 or 14 orders of magnitude. Self-similar traffic models are thus a valuable tool as a real stress test to local Ethernet networks. The results could be used in research of high reliability networks for remote control in telerobotics, teleservice and online engineering with higher QoS requirements.

REFERENCES

[1] Beran, J., R. Sherman. "Long-range dependence in variable-bit-rate video traffic". IEEE Transactions on Communications, vol.43, 1995.

[2] Gospodinov, M., E.Gospodinova. "Generator of fractional Gaussian noise for modeling self-similar network traffic", CompSysTech'2005.

[3] Gospodinov, M., E.Gospodinova. "The graphical methods for estimating Hurst parameter of self-similar network traffic", CompSysTech'2005.

[4] Gospodinova, E., *"New algorithm for modeling of self-similar network traffic, based on fractional Gaussian noise"*. Eighteenth International Conference "ROBOTICS and MECHATRONICS 2008" , Varna, 17-21 September, 2008, pp. 138-142.

[5] Ledesma, S., D.Liu. "Synthesis of fractional Gaussian noise using linear approximation for generating self-similar network traffic". Computer communication, vol. 30, 2000.

[6] Paxson, V. "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic". ACM SIGCOMM, 1997.

[7] Veitch, D., P. Arby. "A wavelet based joint estimator of the parameters of long-range dependence". IEEE Transactions on Information Theory 45, 1999.

[8] Willinger, W., M.S.Taqqu. "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level". IEEE/ACM Transactions on Networking, vol.5, 1993.

[9] Xue, F., S.Yao. "Self-similar traffic shaping at the edge router in optical packet-switched networks", Proc. IEEE ICC 2002, New York, 2002.

[10] ftp://ita.ee.lbl.gov//traces/BC-pAug89.TL.Z

Novel Adaptive Scheduling Scheme for Multimedia Networks with Differentiated Services

Toni Janevski[1] and Dimitar Mladenovski[2]

Abstract – **In this paper we present a new scheduling discipline, which is a designed to be used with diffserv enabled multimedia networks. The main approach is to use Weighted Round Robin that dynamically adapts to the traffic behavior. The goal is to avoid use of weights that are statically assigned. Dynamic weights adjustment is crucial in multimedia networks which provide multimedia services including real-time services such as video streaming and Voice over IP, with implemented admission control. Also, the correlation of the scheduling scheme and the admission control is investigated in the paper.**

Keywords – **Admission Control, DiffServ, Packet Scheduling, Weighted Round Robin.**

I. INTRODUCTION

Today we are facing an explosion of Internet traffic, mostly real-time traffic such as VoIP and video conferencing, which puts a lot of demand on the current network topologies. It complicates the process of planning and dimensioning of such networks in terms of Quality of Service (QoS) support [1]. Network designers need to develop mechanisms to support such growth of services and improve network performance.

There are many technologies, like DiffServ (Differentiated Services), IntServ (Integrated Services), Admission Control and different scheduling schemas etc, which support to real-time services in multimedia networks [2]. DiffServ means Differentiated Services and provides a way to classify traffic for different treatment by the network [3]. This lets us to give different treatment to different types of traffic. The standard proposes three classes of traffic: Expedited forwarding or the premium service, Assured forwarding and best-effort. In addition assured forwarding is divided in four classes and each having three subclasses.

Scheduling schemas along with DiffServ can further improve network performance. Admission Control lets us determine how much traffic we let in the network. If there are available resources in the network a connection is allowed, otherwise is denied.

For DiffServ network on needs a scheduling discipline in each router in the DiffServ domain [4]. Most used scheduling disciplines are priority queuing, Weighted Fair Queuing and Weighted Round Robin. When priority queuing is used the priority class packets are serviced prior to other classes. This may result in poor performance of the other classes, because bandwidth is monopolized by higher priority class.

[1]Dr. Toni Janevski is an Associate Professor at the Faculty of Electrical Engineering, Karpos 2 bb, 1000 Skopje, Macedonia, E-mail: tonij@feit.ukim.edu.mk.

[2]Dimitar Mladenovski is a graduate student at the Faculty of Electrical Engineering, Karpos 2 bb, 1000 Skopje, Macedonia.

When a discipline is based on usage of weight coefficients, then each class receives bandwidth that is proportional to the given weight. In this approach the main problem is how to set the weights. Additionally, admission control is needed for real-time flows such as conversational services [5] (e.g. Voice over IP – VoIP). With admission control and frequent traffic load changes, the static weights approach doesn't give best performance of the network and desired service quality.

In this paper we propose a novel discipline with adaptive weights. Weights are adapted dynamically so that each class receives bandwidth proportional to the assigned weight. Such scheduling schemes are Weighted Round Robin (WRR) and Weighted Fair Queuing (WFQ). We perform analysis of the performance of the proposed scheduling scheme in scenarios with implemented admission control.

In our simulation analysis we use three classes, i.e. for video, voice and Internet traffic, respectively.

II. ADAPTIVE SCHEDULING SCHEME

The general concept of scheduling is shown in figure 1. The weights are associated with the amount of bandwidth that each class gets. The figure shows the scheduling for a DiffServ environment.

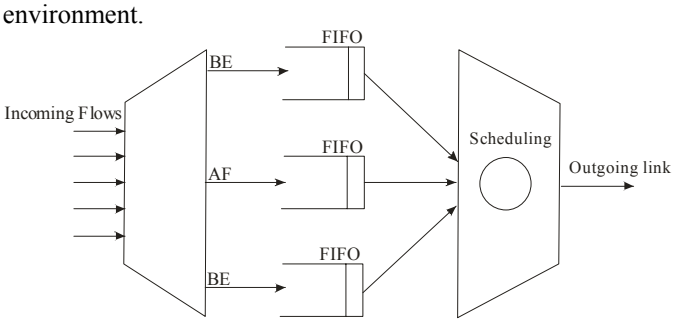


Fig. 1. Scheduling for DiffServ environment

WRR or weighted Round Robin is a discipline in which the packets are served in round robin manner. The service time that each class gets is derived from the weights of that particular class. The serving algorithm first calculates the normalized weights by taking into account the class average packet size i.e. $w_i = w_i / P_i$. Then, it finds the minimum normalized weight. For each nonempty connection every round WRR serves the minimum between the packets that are weighting and the packets to be served.

We propose an adaptive schema that adjusts the weights for each connection. It this schema the normalized weights are given as:

$$w_i = K\rho_i / P_i \quad (1)$$

where K is a adjusting constant, P_i is an average packet size of class i and ρ is:

$$\rho_i = \frac{r_i}{\sum_i r_i} \quad (2)$$

We propose another schema that takes buffer behavior in to account. We use the average buffer state of all classes:

$$avg = \frac{\sum_N avg_i}{N} \quad (3)$$

When the average is calculated the weights are calculated according as follows:

$$w_i = K\rho_i / P + avg_i - avg \quad (4)$$

The average queue size is calculated as in the case with Random Early Detection. The average queue size is calculated by using low-pass filter. In such case, assuming q as instantaneous queue size and f_l for low-pass filter, we obtain:

$$avg \leftarrow (1 - f_l) \cdot avg + f_l \cdot q \quad (5)$$

The adaptive schema according to (4) is shown in Figure 2. The class's average packets in the buffer are checked prior to weights' settings and implementing the WRR.

III. Simulation Results

For the purpose of demonstrating the behavior of the adaptive scheduling we used ns-2 network simulator. We set a topology with DiffServ, Admission Control and two scheduling disciplines defined by (1) and (4). The topology consists of source nodes, access node and destination node. We analyze the queue at the access node. The queue is formed from the packets from the source nodes send to the destination node. We recognize three form of source nodes. Internet traffic is presented with a single node that generates packets with Poisson arrival process and Pareto distributed packet length.

The intensity of the traffic is 0.33 meaning that the incoming traffic is 33% of the outgoing link. The outgoing link is 8.192 Mbps. The packet size is Pareto distributed with mean packet size 128 bytes. The mean arrival time is calculated from the mean packet size and incoming traffic rate. The shape factor is set to two which provides self similarity to the packet size [6], [7].

IP Telephony traffic is presented with audio source nodes and each source node presents a single audio source. We used on/off audio sources.

The mean packet size is 64 bytes and is distributed according to Exponential model. Silent period or idle time is set to be 650 ms and the burst period is 352 ms. During the burst period the source generates traffic with 32 Kbps rate.

For video streaming flows we used MPEG-4 traces [8]. The trace is taken from the movie Jurassic Park. Its parameters are given in Table II.

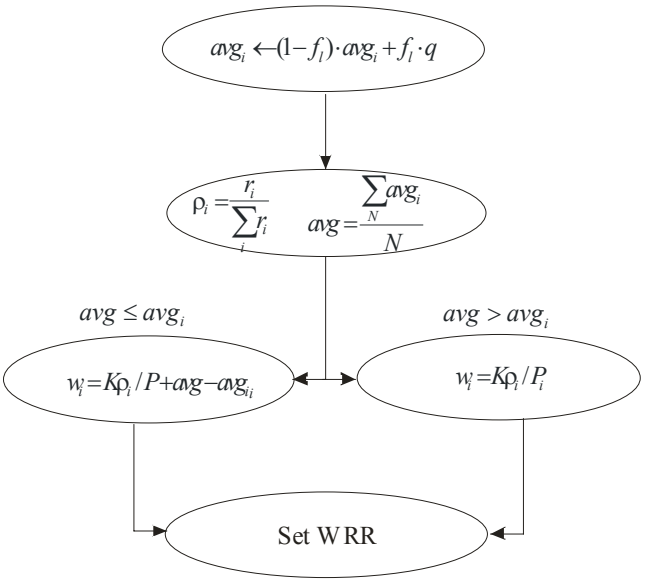


Fig. 2. Adapting the weights

TABLE II
MPEG – 4 SOURCE PARAMETERS

File Size	byte	3.4e+08
Video run time	msec	3.6e+06
# of Frames	-	89998
mean frame size	byte	3.8e+03
var frame size	-	5.1e+06
CoV of frame size	-	0.59
min frame size	byte	72
max frame size	byte	16745
Mean bit rate	bit/sec	7.7e+05
Peak bit rate	bit/sec	3.3e+06
Peak/Mean of bit rate	-	4.37

From the source nodes the traffic is accepted in the access node at which point we do our analysis. At the access node traffic is grouped into three classes and each class is representing different media type. The grouping is available by implementing DiffServ.

The next mechanism implemented is Admission Control. Every source, VoIP or MPEG-4, tries to establish a connection with duration of 20 seconds. If it establishes a connection, at the end of the connection, it tries to establish a new one. This means that connections always are incoming.

Also, connections are allowed if the measured capacity plus the rate of the connection is greater then 90% of outgoing link capacity, a measurement based admission control is implemented.

In this paper we analyze the behavior of the packet loss and the average packets in the buffer. Four scenarios are used. In the first one we use RR, while WRR is used in the second case. When RR is used in each round a packet gets served from each nonempty connection. WRR introduces weights and packets are served according to class weights.

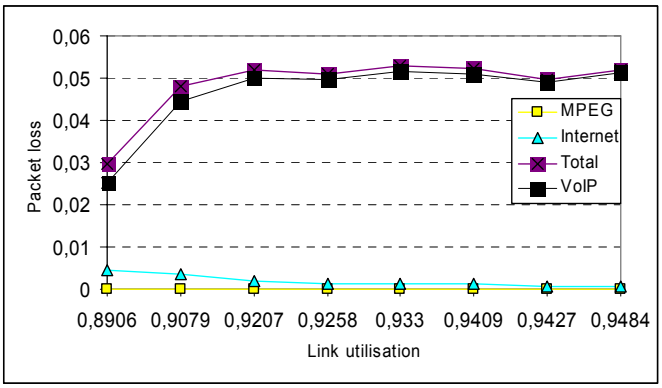


Fig. 3. Packet losses for RR scheduling

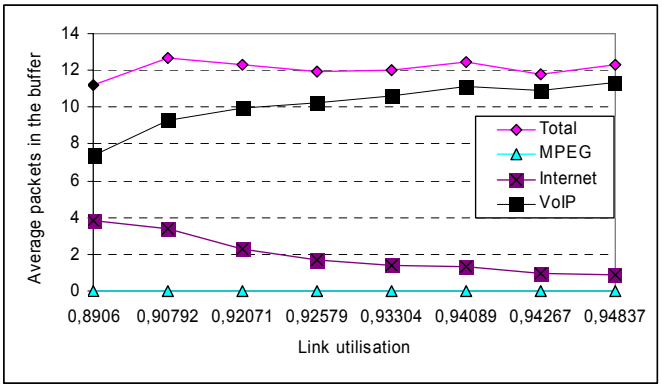


Fig. 4. Average packets in the buffer for RR scheduling

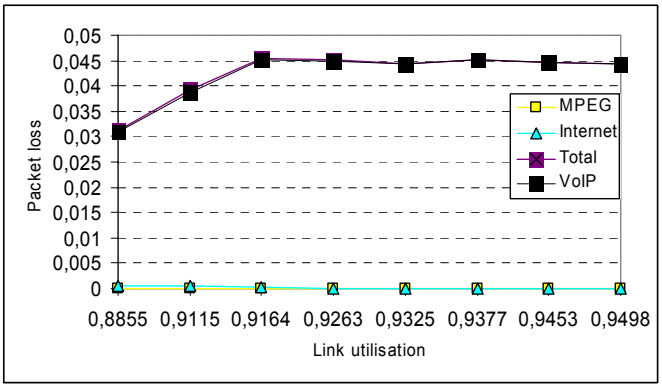


Fig. 5. Packet losses for WRR scheduling

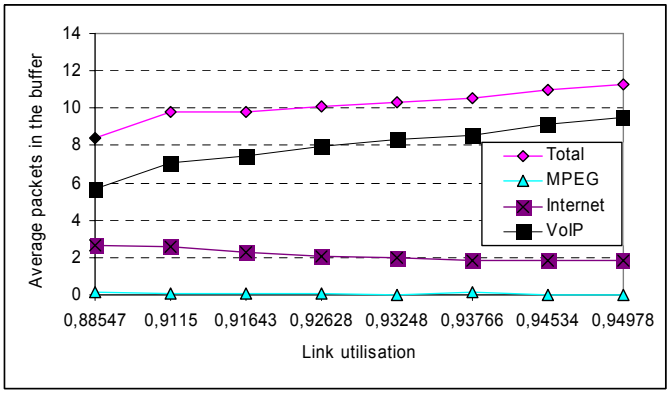


Fig. 6. Average packets in the buffer for WRR scheduling

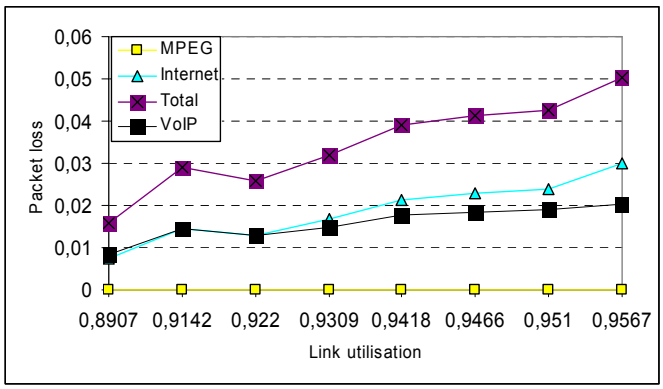


Fig. 7. Packet losses for scheduling according to (1)

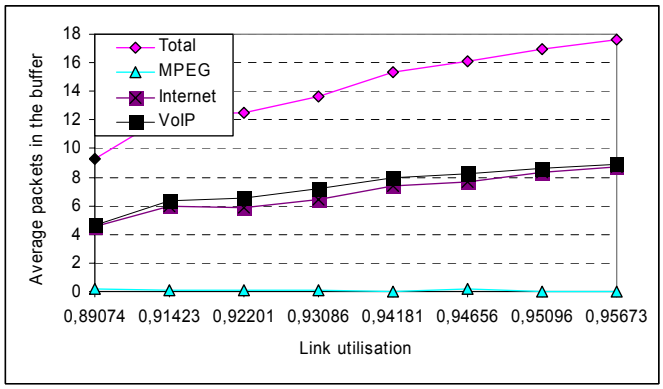


Fig. 8. Average packets in the buffer for scheduling (1)

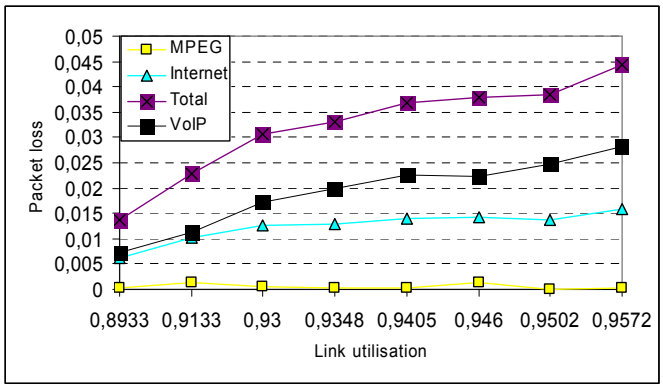


Fig. 9. Packet losses for scheduling according to (4)

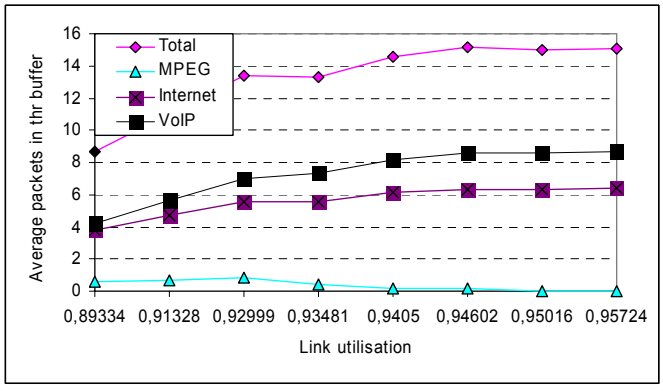


Fig. 10. Average packets in the buffer for scheduling (4)

Such approach solves the problem when classes have different packet sizes. Difficulty appears when WRR is combined with admission control mechanism. Then, this becomes a problem of setting the weights. In our simulation the proposed algorithm sets the weights proportional to the maximum bandwidth that each class demands, and at the same time not calculating the traffic that is to be denied.

We try to avoid the weights problem by using adaptive scheduling according, where the weights are adjusted every time a connection starts or ends, according to (1). Another way to avoid the weights adjusting problem is the fourth scenario where we implement adaptive scheduling that takes into account the behavior of the buffer according to (4). This schema uses the same approach of setting the weights dynamically, but it also uses the buffer behavior to improve the weights settings. The schema finds the class with largest queue length in the buffer and favorites it by giving it larger scheduling weights. This improves performance of the class that has the worst performance and that might have been underestimated by other scheduling disciplines. Results are shown in Figures 3 to 10.

In Figures 3, 5, 7 and 9 are shown packet losses for all disciplines. We can notice that total packet loss varies from 0.045 to 0.05 in the used scenarios. The interesting point is that the RR and WRR discipline show poor performance of the voice class.

When RR is used the voice class gives worst performance because it has smallest packet size. RR serves one packet from each class giving advantage to classes with higher packet size like the class for video traffic. When WRR is used, the lack of knowledge for each class bandwidth usage ends up in bad setup of WRR weights.

Better performance are seen when weights are adopted when traffic changes. Now the total packet loss is split between the voice class and the Internet class. The video class presents traffic with large average packet size. The number of packets that belong to this class is low compared to other classes. This results with low packet loss for the video class.

We can draw the same conclusions when an average packets in the buffer. This can be seen in Figures 4, 6, 8 and 10. When RR and WRR used this class builds up the buffer. We can see that there exist significant difference between this class and other classes. Such difference can be up to 10 packets when RR is used. Adjusting the weights closes the gap between classes.

It is interesting to compare the two disciplines that adjust the weights. When the first adaptive WRR discipline is used the Internet class has the worst performance. Using the second discipline gives less service to the voice class and now it has worst performance. This is expected since the schema favorites the class with worst performance from the buffer point of view. Such performance improvement of the Internet class is followed by degradation in the performances of the video class and the voice class. But, this is significant compared to overall performance. We can notice that overall packet loss is smaller when we use the scheme that uses the buffer state as well.

The same behavior can be seen when the average buffer queue size is analyzed. Figure 4 and 6 show the behavior of the average number of packets in the buffer.

Also there is a large improvement of the Internet class whose value decreases when we use adaptive scheduling scheme with buffer knowledge.

IV. CONCLUSIONS

In this paper we described QoS mechanisms that improve the performance of the real time media. At the beginning we explained the DiffServ mechanism that differentiates the services in to classes so that they get different treatment. This allows us to favor some of the classes. Different scheduling schemas contribute to the goal of improving the performance with the same approach of letting us to favor some class and utilize the service time more efficient. We presented two adaptive scheduling schemes that are based on WRR. The first one adapts the weights to the data rates and the second one adds to the scheduling scheme the behavior of the buffer state. We compared the schemas in an environment where we implemented an admission mechanism that allows us to accept or deny the establishment of new connections according to the available resources in the network. This way we control how much traffic we let in the network and provide an environment where class rates change so often that the difficulty of setting the weights is always present.

Simulations are presented to confirm how these mechanisms contribute to network performance. The first scenario uses the schema where weights are set proportional to the rates of each class. The second schema takes into account the buffer state. We compare the schemas with RR and WRR with statically assigned weights.

The analyses showed that the second adaptive scheduling schema (the one with buffer state) gives better performance. The improvements can be seen in the behavior of the packet loss and the average queue size. Also, these disciplines have better performance compared to RR and WRR.

REFERENCES

[1] Jarmo Siltanen. “Quality of Service and Dynamic Scheduling for Traffic Engineering in Next Generation Networks”, Dissertation, 2007.

[2] Toni Janevski, Dimitar Mladenovski, “The Impact of Real-Time Media Differentiation on the Internet Traffic”, TELSIKS 2007, Nis, Serbia, September 26-28, 2007.

[3] S. Blake et al., “An Architecture for Differentiated Services”, IETF Request For Comment 2475, December 1998.

[4] Stylianos Georgoulas, Panos Trimintzios, George Pavlou and Kin-Hon Ho, “Measurement-based Admission Control for Real-time Traffic in IP Differentiated Services Networks”, Proc. ICT 2005, Cape Town, South Africa, 3-6 May 2005..

[5] Toni Janevski, “Traffic Analysis and Design of Wireless IP Networks”, Artech House Inc, 2003.

[6] A. Tudzarov, D. Temkov, T. Janevski, O. Firfov, “Empirical Modeling of Internet Traffic at Middle-level Burstiness”, IEEE Melecon 2004, Dubrovnik, Croatia, May 12-15, 2004.

[7] Toni Janevski, Dimitar Mladenovski, “Queuing Models with Weibull Distribution for Analysis of Self-similar Traffic”, ICEST 2004, Bitola, Macedonia, June 16-19, 2004.

[8] http://www-tkn.ee.tu-berlin.de/research/trace/trae.html.

Performance of Internet Transport Protocols in 802.11g Wireless Environment

Toni Janevski[1] and Ivan Petrov[2]

Abstract – **Rapid development of wireless communications systems fulfils one of the major challenges in the modern communication era and provides wireless access to Internet. Enabling mobile users to access the vast range of Internet based applications is the next step in wireless communication. Internet connectivity is represented by TCP/IP protocol suite. Performance of the Internet transport protocols may significantly degrade when end to end connection includes wireless links where packets delays and losses are caused by mobility handoffs and transmission errors. In this paper we study the impact of wireless MAC layer design of the native Internet transport protocols during the distribution of multimedia services in realistic outdoor scenario.**

Keywords – **Mobility, Transmission Protocols, Wireless Network.**

I. INTRODUCTION

The need for continuous communication combined with the digital entertainment and the existence of variety multimedia services provided by wireless access communication networks is opening new chapter at the telecommunication market. It is easy to forecast the synergy between wireless access technology and the Internet. This means that remote working, online games, video/music on demand and navigation support are only small part of applications that will be accessible from any palace in any time. Rapid increasing number of active wireless hot spots is providing people to be wireless connected in almost every building and every street they walk and drive in. A large amount of static users are using the IEEE 802.11 g wireless access technology for checking emails, web surfing, video/audio streaming and P2P file shearing.In near future cars will become communication centers enabling us wireless access to Internet. The same scenario will be applicable in the home environment. We know that TCP/IP protocol stack is the most widely used within many applications in the today's computing world. Its parameters were carefully tuned in order to maximize its performance on wired networks where packet delays and losses are caused by congestion [1-6]. In the wireless networks delays and losses are mainly caused by mobility handoffs and transmission errors due to bad wireless channel condition. With the recent developments in mobile wireless networking, the performance of the Internet transport protocols in mobile wireless environment is becoming a topic of interest.

[1]Dr. Toni Janevski is an Associate Professor at the Faculty of Electrical Engineering, Karpos 2 bb, 1000 Skopje, Macedonia, E-mail: tonij@feit.ukim.edu.mk.

[2]Ivan Petrov is with Macedonian Telecom, Key Customers Business Centre, Orce Nikolov bb, 1000 Skopje, Macedonia, E-mail: ivan.petrov@telekom.mk.

We should mention that the protocols for wireless AP's have been designed in order to maximize the utilization of the wireless channel for web browsing and file downloading applications in environment with restricted mobility, which is the main reason why the buffers and the local MAC retransmissions are tuned in a way to maximize the throughput and the reliability for this kind of applications. In this context it is necessary to define and fine tune the technical standards in order to guarantee full interoperability between different digital applications in wireless environment and providing proper wireless/wired interconnection. We focus our attention of the impact of diverse MAC layer and buffer settings of IEEE 802.11g access technology over the Internet native transport protocol during the distribution of multimedia applications in realistic outdoor static and mobile multimedia scenario.

The paper is organized as follows: Section II gives brief overview of the transport protocols, discusses some related work and motivates the need for our approach. It briefly describes the 802.11 MAC protocol. Section III describes our simulation scenario and section IV presents the simulation results. Section V concludes the paper.

II. TRANSPORT PROTOCOLS

Applications can be grouped in two major classes: downloading (TCP) and real-time (UDP). The first class is using reliable data transfer while the second class is based on quick delivery of packets. The performances that are measured by the booth classes of applications are completely different. The first class is measuring the performance in terms of how much time is required to have the whole file transferred that is different from the second one where the performances are measured in terms of percentage of packets that reach the destination within a certain time threshold. We can say that FTP, HTTP, SMTP are applications that for sure are part of the first class and that the interactive on-line games, real-time IP-TV, video/audio chatting, represent examples of applications that are part of the second class. From the statements above we can distinguish the downloading and real-time applications by the employed real time protocol: TCP or UDP. The TCP protocol guarantees the reliable and ordered delivery of every packet sent, to aim this it establishes a session and performs retransmissions of lost packets. We should mention that it employs congestion control functionality. Every TCP flow probes the link with higher and higher data rates eventually filling up the channel. We can be sure that the packets will be queued at the buffer associated with the bottleneck of the link until it overflows causing packet losses. In this situation TCP retransmits the lost packets, and halves its sending rate to diminish the congestion level. Finally, the regular increase of the sending rate is reestablished and so forth. This is not the

case with the UDP transport protocol which is much simpler then TCP because packets are immediately sent toward the receiver with a data rate decided by the sender. UDP does not guarantee reliable and ordered delivery of packets but, its small overhead and lack of retransmissions make it less prone to generate delays in the packets delivery. This is the main reason why the UDP transport protocol is mainly used by real-time applications. The first TCP implementations were using cumulative positive acknowledgements and required a retransmission timer expiration to re send a lost data during the transport. They were following the go-back-n model. Without hesitating we can say that the early TCP implementations did little to maintain network congestion. In order to enable good user throughput and to control network congestion a lot of work has been done in order to improve its characteristics and with time TCP has evolved. Today's TCP implementations contain variety of algorithms that enables to control the network congestion and to maintain good user throughput in the wired network. Several variants of TCP can be found in the today's wired networks. TCP Tahoe, TCP Reno, TCP New Reno, TCP Vegas and TCP Sack are few of them that are going to be used in our simulation scenario. We should mention that most used variant of TCP in the real world is TCP NewReno. Every of these TCP variants have unique congestion and flow control mechanisms that differs them. A problem is defined in coexisting of TCP and UDP traffic in a wireless channel, caused by the TCP congestion control functionality. TCP continuously probes for higher transfer rates, eventually queuing packets on the buffer associated with the bottleneck of the connection. The wireless connection can be shared by several devices and applications; it is evident that the connection level and the queue lengths can increase, thus delaying the delivery of packets stuck in queue and jeopardizing requirements of real-time applications. The situation is more worsened by the wireless nature of the link because the wireless medium allows the transmission of only one packet at a time and is not full-duplex as wired links. This means that packets should wait their turns to be transmitted. Interference, errors, fading, and mobility are causing additional packet losses, the IEEE802.11 MAC layer reacts through local retransmissions which in turn cause subsequent packets to wait in queue until the proceeding ones or their retransmissions eventually reach the receiver. The back off mechanism of the IEEE 802.11 introduces an increasing amount of time before attempting again a retransmission. In the recent years, many researchers have studded the problems that TCP and UDP encounters in a wireless environment and have presented their solutions with aim to overcome them [7-15].

III. Simulation Scenario

The network layout of the simulation scenario that is subject of the conducted analyze is presented at Fig. 1.

We can notice that the network topology is consisting of four wired nodes (A0-A3), two wireless base stations (BS0-BS1) and four wireless nodes (n0-n3).

The wireless stations are configured to work according the IEEE 802.11g Standard. Wired connections are configured as stated in Table I.

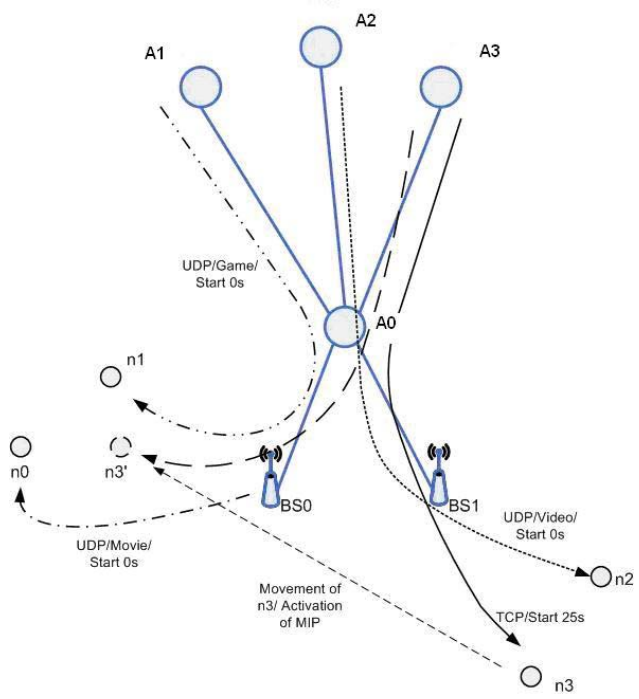


Fig. 1. Simulation Scenario.

TABLE I
CONFIGURATION OF WIRED LINKS SIMULATED AT SCENARIO.

Node 1	Node 2	Delay	Capacity
A1	A0	10ms	100Mbps
A2	A0	20ms	100Mbps
A3	A0	30ms	100Mbps
A0	BS0	10ms	100Mbps
A0	BS1	10ms	100Mbps

TABLE II
TYPES OF APPLICATIONS AND TRAFFIC SIMULATED IN THE PRESENTED SCENARIO.

From	To	Type	Transport Protocol	Start	End
BS0	n0	Movie Stream	UDP	0s	110s
A1	n1	Game Traffic	UDP	10s	110s
n1	A1	Game Traffic	UDP	10.1s	110s
A2	N2	Video Chat	UDP	15s	110s
N2	A2	Video Chat	UDP	15.1s	110s
A3	N3	FTP	TCP	35s	110s

Maximum achievable bandwidth rate is 20Mbps instead of 54Mbps as it is declared for IEEE 802.11g standard. The queue size value used in the simulation comes out by multiplying the longest RTT with the smallest link capacity on the path which is represented by the 20Mbps effectively available over the wireless link. In Table II are presented several applications that are used during the simulation. In the simulation are exploited real trace files for video chat and movie traffic. Two VBR H.263 Lecture Room-Cam are used for the Video chat and

high quality MPEG4 Star Wars IV trace file is used for the movie. These files can be found in [16]. In this simulation the game events have been generated at the client side every 60ms. At the server side updates were transmitted every 50ms toward the client. The payload generated by the client has been set to 42Bytes and the payload generated by the server has been set to 200Bytes. The rest of the packets were set at 512Bytes.

TABLE III
SIMULATION PARAMETERS.

Parameter	Values	Comments
MAC data retransmissions	1,2,3,4	Default value is set at 4
User-BS distance (m)	50,100	Common outdoor environment
MAC queue size (pkts)	25,50,100	Common values
Velocity (m/s)	4.16,7,14	Random choice
TCP Transport protocol	TCP Tahoe, TCP Reno, TCP Newreno, TCP Vegas, TCP Sack	Commonly used types of TCP protocols in wired networks.

The variable parameters used in this scenario are listed in Table III. A number of simulations have been conducted in order to examine the effects generated by different set of parameters defined in the simulation environment. During the simulation the shadowing model is used with one set of parameters. The shadowing deviation (σ_{dB}) was set to 4 while the path loss exponent (β) was set to 2.7. These parameters are common for outdoor soft partition in urban environment. It is known that the signal attenuation grows with the increase of these parameters so it is expected that the packet losses percentage will be higher over the wireless media. MIPv4 is used as a protocol that handles mobility in the defined scenario.

IV. RESULT ANALYSIS

We will observe the simulation scenario presented at Fig.1 with configuration of the links defined as in Table I and applications defined as in Table II. We will study the behavior of the TCP applications and the TCP impact of the real time applications in the defined network when node n3 is mobile. The coordinates of the nodes are given in Table IV. During this analysis we should notice that n3 begins to move 55s after simulation starts.

Like we have stated in Table III the mobile node n3 is moving with three different velocities. We will observe scenarios when n3 is moving with speed of 4.16m/s, 7m/s and 14m/s. At Figs. 2, 3 and 4 is presented the change of the average throughput of the TCP traffic achieved in the simulation environment accordantly. From these figures we can evident that if we increase the number of MAC retransmissions the average throughput will increase too. Best throughput values are achieved when the number of MAC retransmissions in set at value of three. From Figs. 2, 3 and 4 we can notice that best performance has been achieved when IFQ is set at value of 50 pkts and TCP SACK is used as a transport protocol.

TABLE IV
COORDINATES OF WIRELESS NODES IN SIMULATION ENVIRONMENT

Wireless node	X	Y
BS(0)	230	230
BS(1)	330	230
n0	171,41	225,85
n1	172,81	230,97
n2	389,2	227,39
n3	338,13	231
n3'	170,01	230,1

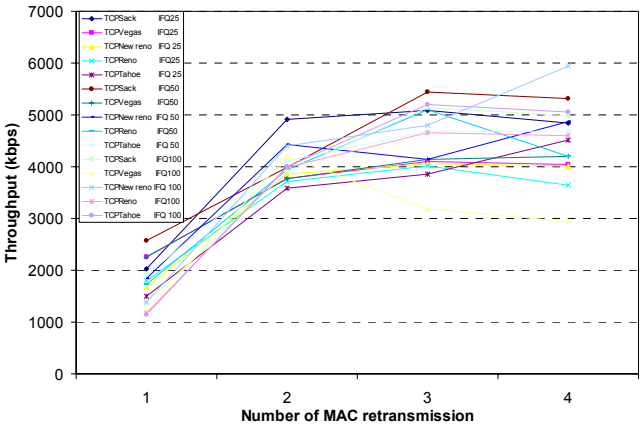


Fig. 2. Average Throughput of the FTP traffic when the node n3 is moving toward BS0 with speed V=4m/s

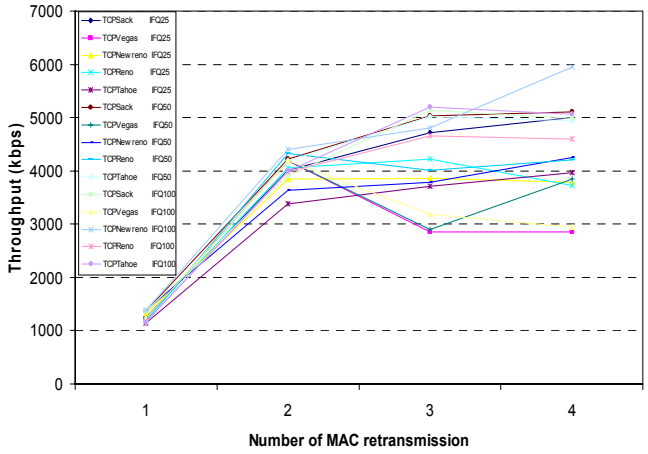


Fig. 3. Average Throughput of the FTP traffic when the node n3 is moving toward BS0 with speed V=7m/s

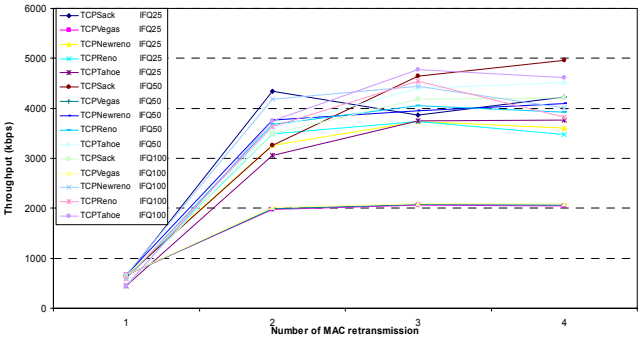


Fig. 4. Average Throughput of the FTP traffic when the node n3 is moving toward BS0 with speed V=14m/s

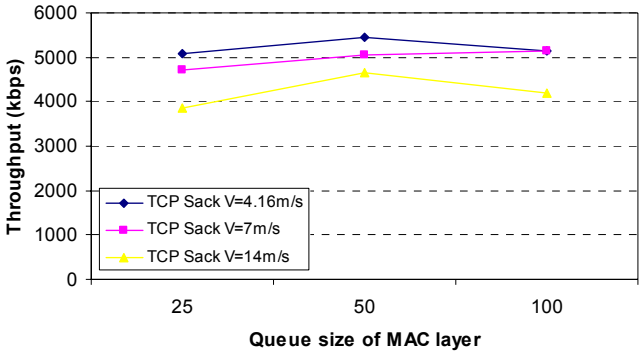


Fig. 5. Average throughput of the FTP flow when the number of MAC retransmissions is set at three and TCP SACK is used as a transport protocol.

We can evident that for smaller speed we are achieving better results when we use queue with size of 50 pkts. This is the case when n3 is moving with speed of 4.16m/s and 7m/s.

In Fig. 5 we have compared the behavior of the average throughput as a function of the MAC queue size when n3 is moving with the defined speed and MAC data retransmissions are set at three. It becomes obvious that the throughput of the flow decreases with increasing of the speed of moving. As TCP transport protocol is used TCP SACK.

V. CONCLUSIONS

In the presented work we have analyzed realistic outdoor wireless scenario in which were considered several everyday applications. Traffic analyze was conducted in order to get more familiar with the traffic behavior of different types of applications in wireless environment. We were considering several types of TCP transport protocols in order to observe the behavior of the FTP traffic in realistic wireless environment. After this analyzes we are able to conclude that: distance between nodes directly impacts the FTP traffic flow and its throughput. The throughput decreases as a function of the distance between the nodes in the wireless environment. The number of the MAC layer retransmissions directly influence of the FTP and UDP traffic flow. MAC layer queue size directly affects the traffic flows in the wireless environment especially for shorter distances between the wireless nodes so it has to be carefully tuned in order to be achieved higher traffic flows.

The choice of the TCP transport protocol affects the throughput in the network.

These simulations have shown that in this scenario is best if we set the max number of MAC layer retransmissions at value of three, the MAC queue size at value of 50 pkts. Mobility directly impacts the throughput. Velocity and MAC Layer Queue Size impact at TCP throughput as a function of MAC Layer Retransmissions. The throughput decreases as a function of the speed. In this simulation environment we are able to conclude that best traffic performances are obtained if we use TCP Sack as a transport protocol. During the conducted analyze we should have in mind that the performances of TCP Vegas are degraded when it is incorporated in scenario with the other TCP protocols.

REFERENCES

[1] W. Richard Stevens, "TCP/IP Illustrated,Volume 1: The Protocols", Addison Wesley,1994.

[2] V. Jacobson, "Congestion Avoidance and Control", SIGCOMM Symposium on Communications Architectures and Protocols, pages 314–329, 1988.

[3] V. Jacobson. "Modified TCP Congestion Avoidance Algorithm", Technical report, 30 Apr. 1990.

[4] K. Fall, S. Floyd, **"**Simulation-based comparisons of Tahoe, Reno and SACK TCP**",** ACM SIGCOMM Computer Communication Review, Volume 26, Issue 3, Pages: 5 - 21, July 1996.

[5] B. Moraru, F. Copaciu, G. Lazar, V. Dobrota, "Practical Analysis of TCP Implementations: Tahoe, Reno, NewReno", 2nd RoEduNet International Conference, 2003.

[6] J. Mo, R.J. La, V. Anantharam, J. Walrand, "Analysis and comparison of TCP Reno and Vegas", IEEE INFOCOM '99, Pages 1556-1563, vol. 3, March 1999.

[7] G. Holland, N. Vaidya, "Analysis of TCP Performance over Mobile Ad Hoc Networks", Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking, Pages 219-230, 1999.

[8] P. V. Nikitin, O. Celebioglu, and V. Kukrer, TCP performance in mobile wireless environment: channel modeling and network simulation, 6th World Multiconference on Systemics, Cybernetics and Informatics, Orlando, FL, July 2002.

[9] M. Gerla, K. Tang, R. Bagrodia, "TCP performance in wireless multi-hop networks", Mobile Computing Systems and Applications Proceedings (WMCSA '99), Pages 41-50, February 1999.

[10] C.F. Chiasserini, M. Garetto, M. Meo, "Improving TCP over wireless by selectively protecting packet transmissions", 4th International Workshop on Mobile and Wireless Communications Network, Pages 316-319, 2002.

[11] Claudio E. Palazzi, Giovanni Pau, Marco Roccetti, Stefano Ferretti, Mario Gerla, "Wireless Home Entertainment Center: Reducing Last Hop Delays for Real-time Applications", ACM International Conference Proceeding Series, vol. 266, article no. 67, 2006.

[12] Martin A.V.D., Mihailovic A., Georganopoulos N., Aghvami A.H., "Adaptation of transport protocols for an IP-micromobility scheme", IEEE International Conference on Communications (ICC 2001), Pages 2462-2466, Helsinki, Finland, 2001.

[13] L. Le, S. Albayrak, M. Elkotob, A.C. Toker, "Improving TCP Goodput in 802.11 Access Networks", ICC '07, IEEE International Conference on Communications, Pages 4494-4499, Glasgow, Scotland, June 2007.

[14] K. Pahlavan, T.H. Probert, M.E. Chase, "Trends in local wireless networks", IEEE Communications Magazine, Pages 88-95, March 1995.

[15] S. Xu, T. Saadawi, "Does the IEEE 802.11 MAC protocol work well in multihop wireless adhoc networks?", IEEE Communications Magazine, Pages 130-137, June 2001.

[16] Movie trace-files available at: http://www-tkn.ee.tu-berlin.de/research/trace/ltvt.html. Last accessed January, 2009.

Analysis of Voice Traffic Performance over OFDM Wireless Access Systems with AMC

Kiril Kassev[1], Boris Tsankov[2]

Abstract – **Orthogonal Frequency Division Multiplexing (OFDM) is an attractive technology developed for future wireless and mobile communications systems.**

In this paper we evaluate the voice traffic performance of a multi-class OFDM wireless access system with specific modulation and coding schemes, employed at the physical layer to match transmission parameters to time-varying channel conditions. Numerical results demonstrate an approach for efficient radio resource management, which can improve the system performance.

Keywords – **performance evaluation, adaptive modulation and coding (AMC), OFDM, VoIP**

I. INTRODUCTION

In order to satisfy customer demands for access to a wide variety of multimedia applications over wireless communication networks, significant research efforts are being put in development of next generation mobile networks (NGMN), also referred to as 4G, based on all-IP environment. This can be achieved by employing the OFDM technique at the air interface, which is immune to intersymbol interference and frequency selective fading, due to the time-varying nature of the wireless channel. As a consequence, OFDM-based systems have become a popular choice, and have been adopted in several standards towards NGMN [1], [2].

In contrast to the traditional channelized multiple access systems, where each user is assigned a fixed amount of bandwidth during the whole connection time, in wireless networks the channel capacity of a wireless link is time-varying, and thus the quality of service (QoS) requirements may not be satisfied, even though a large amount of resource (i.e. bandwidth) is allocated to a certain connection. This is especially true when a mobile station (MS) is located at the cell edge area. In order to maintain a target packet error rate (PER) over wireless links, meeting QoS requirements, adaptive modulation and coding (AMC) schemes have been widely adopted to match the transmission parameters to time-varying channel conditions [3], [4].

The interaction of analytical framework for evaluation of teletraffic performance characteristics of the system under consideration at the data link layer with AMC at the physical layer provides a base for interesting and realistic design work. This topic is covered by a number of excellent papers describing in depth the essence of the underlying problems. Reference [5] investigates the performance of transmission over wireless links, in case an interaction between a finite-length queuing and different AMC schemes are taken into consideration. The Erlang capacity of WiMAX systems with fixed modulation schemes is calculated in [6], considering two traffic classes – streaming and elastic. An interesting approach for evaluation of Erlang capacity of a multi-class TDMA system with AMC by separating the calculation of blocking and outage probabilities is proposed by the authors of [7].

In this paper, we evaluate the teletraffic performance in terms of traffic congestion of multi-rate OFDM wireless access systems with AMC, under specific circumstances, attempting to find the optimal allocation of available resources.

The rest of the paper is organized as follows. We first introduce the system model, which takes into account the OFDM transmission with AMC. Section 3 represents an analytical model of a multi-rate loss system with relevant performance measures. Numerical results illustrating the dependences of system performance on various parameters are given in Section 4. Finally, concluding remarks have been drawn in Section 5.

II. SYSTEM MODEL

We consider an OFDM infrastructure-based wireless access network, where connections are established between a base station (BS) and mobile stations (MSs) inside the cell area. The access to the system resources is based on time-division multiplexing (TDM), and thus at the physical layer, the data stream is organized in frames of fixed length (duration). Since the data transmission is based on OFDM technology, a frame is divided into n time slots (TS). Users can transmit their data in the assigned time slots over all available subchannels. An AMC scheme is employed at the physical layer to match the transmission rate to the time-varying channel conditions as well as to maintain a target PER over wireless link. This results in splitting up the overall cell area on several regions (rings), corresponding to the available modes of AMC scheme, each of which represents a pair of specific modulation format and a forward error correcting (FEC) code.

From network operator's point of view, offered calls from MSs, belonging to the cell area working with the modulation scheme of highest order (closest to the BS), are most profitable, compared to the other ones, since they are allocated the smallest amount of resources (i.e. bandwidth) in order to be served. For this reason, to prioritize this class of calls, a mechanism for reservation of fixed amount of resources can be applied, and it allows the other classes of calls to compete for the remaining bandwidth (class limitation). If n denotes

[1]Kiril Kassev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kmk@tu-sofia.bg.

[2]Boris Tsankov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: bpt@tu-sofia.bg.

the total number of time slots in a frame for data transmission, Δ denotes the number of time slots that can be only reserved for incoming traffic flow of so called "profitable" class.

In this paper, we consider that a voice service is supported by the system. Since the voice is not tolerant to packet delay, it requires constant bit rate, which means that each voice user is allocated a number of time slots per frame, depending on its location in the cell (current modulation and coding scheme used).

III. Performance Analysis

Following the adoption of AMC scheme, we assume that the call attempts, following Erlang case, offered from the MSs belonging to a particular region (ring) of the cell generate a traffic stream. Each traffic stream (class) *i* is characterized by the offered traffic A_a^i (call attempts traffic), maximum number of allocated time slots n_i in the frame (class limitation) and the number of time slots d_i required for establishing one connection (supporting a voice call with a constant bit rate). For the *i*-th arrival process the arrival rate is $\lambda_i(x_i)$, for state $x_i \cdot d_i$, that is, when x_i simultaneous calls of type *i* are being served, and hence, the following restrictions shall be fulfilled:

$$0 \le x_i \cdot d_i \le n_i \le n, \qquad i = 1,..., N \tag{1}$$

and

$$0 \le \sum_{i=1}^{N} x_i \cdot d_i \le n. \tag{2}$$

where number of classes *N* (traffic streams) is equal to the available transmission modes, supported by the AMC scheme.

We conduct the performance analysis of the system, described in Section 2, by applying the classical teletraffic model of multi-rate loss systems, which is based on flexible channel/slot allocation [8]. The call-level characteristic of the system can be modeled by a multi-dimensional Continuous Time Markov Chain (CMTC), which is reversible and has a product form. Due to reversibility, we can apply the local balance equations to calculate the relative state probabilities and to find performance measures of interest. Instead, the numerical evaluations are performed in an efficient way by using the convolution algorithm [8].

The state probabilities of each traffic stream *i* as if it is alone in the system are following Erlang distribution and can be represented as:

$$p_i(j \cdot d_i) = \frac{\frac{(A_a^i)^j}{j!}}{\sum_{k \in \Omega_a^i} \frac{(A_a^i)^k}{k!}} \quad if \quad j \in \Omega_a^i$$
$$p_i(\cdot) = 0 \qquad\qquad else \tag{3}$$

$$\Omega_a^i = \left\{ 0,1,..., \left\lfloor \frac{n_i}{d_i} \right\rfloor \right\} \tag{4}$$

where Ω_a^i are the sets of possible number of users in each service class *i* (*i* = 1,…,N) that can be served by the system and $\lfloor k \rfloor$ denotes the largest integer not exceeding *k*. Eq. (3) takes into account multi-slot traffic ($d_i > 1$) offered to a system.

Based on the state probabilities, we are able to get the performance measure of the system in terms of overall traffic congestion C_{tot}, obtained by using the convolution algorithm. The aggregated state probabilities for the system $Q_{N/i}$ is calculated by successive convolutions of state probabilities vectors of each traffic stream, excepting traffic stream *i*. The convolution of aggregated states probabilities and excluded traffic stream results in a state probabilities vector, which defines the probability of a given number of TS in the frame being occupied by calls offered by all traffic streams:

$$Q_N(j) = \sum_{x=0}^{j} Q_{N/i}(j-x) \cdot p_i(x) = \sum_{x=0}^{j} p_x^i(j), \tag{5}$$

where for $p_x^i(j)$, *i* denotes the traffic stream number, *j* is the total number of busy resources (time slots), and *x* is the number of occupied TS by stream number *i*.

The carried traffic Y_c^i of stream *i* measured in busy channels (TS) is:

$$Y_c^i = \sum_{j=0}^{n_i} \sum_{x=0}^{j} x \cdot p_x^i(j). \tag{6}$$

Since we are interested in determine the carried traffic Y_a^i of stream *i* measured in call attempts, it can be expressed as:

$$Y_a^i = \frac{Y_c^i}{d_i} = A_a^i \cdot (1 - C_i). \tag{7}$$

The traffic congestion C_i is equal to the ratio of offered traffic A_a^i which is blocked. It should be noted that both offered traffic A_a^i and carried traffic Y_a^i of stream *i* are measured in Erlangs (erl).

Due to its attractive performance characteristics, AMC has been adopted at the physical layer of several standards, e.g. 3GPP, 3GPP2, IEEE 802.11, IEEE 802.16 [1], [2], [9]. Based on channel state information (CSI) estimated at the BS, the set of available transmission modes, which are most commonly used, are based on 64-QAM, 16-QAM and QPSK modulation techniques.

As we stated above, the AMC scheme is applied to maintain a target PER over wireless link and therefore both the modulation scheme and coding rate are chosen according to time-varying channel conditions. As a consequence, the number of time slots d_i allocated to users of traffic stream *i* is varying and can be determined in accordance with the

constant bit rate R_V bits per frame, required for voice users as well as the transmission mode i chosen.

$$d_i = \left\{ \left\lceil \frac{R_v}{s \cdot R_i} \right\rceil, \quad i = 1,...., N \right\} \tag{8}$$

where s is the fixed number of OFDM symbols per time slot, and R_i denotes the number of bits carried per OFDM symbol in transmission mode i. Notation $\lceil k \rceil$ denotes the smallest integer larger than k.

We consider the following group of transmission modes i, as they are adopted in [9], [10] (Table I).

TABLE I
TRANSMISSION MODES WITH MODULATION AND CODING SCHEMES

i	Mode 1	Mode 2	Mode 3
Modulation	64-QAM	16-QAM	QPSK
Coding rate	3/4	3/4	3/4
R_i (bits/sym)	4.5	3.0	1.5
Target SNR for 1% PER (dB)	24.4	16.2	8.2

IV. NUMERICAL RESULTS

The current sections deals with numerical results based on the traffic model presented in Section 3.

We consider a single-cell OFDM system with time-division duplex (TDD) operation. The duration of a frame is set to be 1ms. The total number of time slots used for data transmission within the frame is set to be 200, each of which contains 4 OFDM symbols [7]. The modulation order and coding rate in the AMC scheme is determined by the instantaneous SNR of each user in the cell. We follow the AMC schemes shown in Table I, which specifies the minimum SNR required in order to maintain a target packet error rate of 1 %. Thus, the cell area is split up in three non-overlapping areas, at which MSs are served by the BS at a particular modulation and coding scheme. In this reason, the call attempts offered from MSs of specific area can be represented as three independent traffic streams, having an access to the common pool of resources in terms of available time slots n in the frame. The consecutive number of a traffic stream (shortly Stream) corresponds to the transmission mode i, serving data flows offered from a particular cell area.

Since the calls offered from the cell area closest to the BS (Stream 1) are seemed to be most profitable for the network operator, we prioritize them by reservation of a fixed number of time slots Δ in the frame, by restriction of accessibility of the remaining traffic streams to the resources in the frame (a class limitation for Stream 2 and Stream 3 is used).

From traffic engineering point of view, the traffic congestion is the most important measure, and hence the research work is oriented towards finding the optimal allocation of available resources (time slots) for both traffic stream 2 (call attempts served by the BS in transmission mode 2 – 16QAM) and traffic stream 3, so that a minimal value of overall traffic congestion C_{tot} to be obtained.

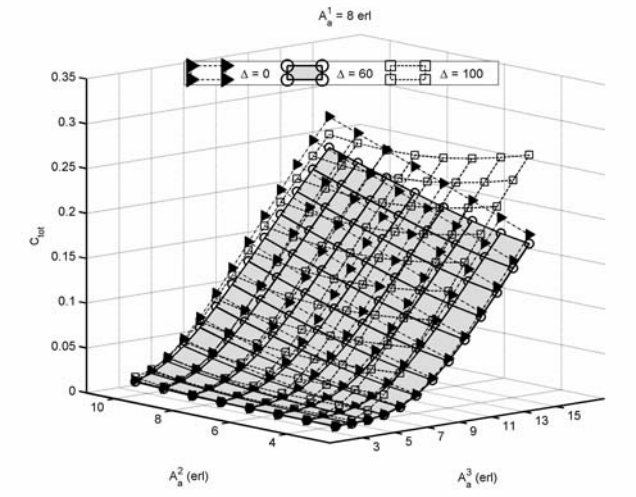


Fig. 1. Overall traffic congestion versus different traffic loads and limited accessibility of both Stream 2 and Stream 3

At the first point, we analyze how various traffic loads of Stream 2 and Stream 3 will affect the overall traffic congestion C_{tot} of the system, assuming a traffic volume of 8 erl. for Stream 1. We have conducted the analysis for three levels of reservation of time slots Δ in the frame for Stream 1 (Δ = 0, 60, 100 TS). Results are depicted on Fig. 1 and demonstrate that minimal levels of the overall traffic congestion, in case of high traffic loads, can be realized by using reservation scheme with Δ = 60 TS, in comparison with results for Δ = 0 (all three traffic flows can access to entire resource space of each frame). It is interesting to notice, that a similar trend can be also observed for reservation mode with Δ = 100 TS, but only if the offered traffic volume of Stream 2 is greater than 8 erl. If this rule is not properly fulfilled, the current reservation scheme that is applied is strongly inefficient (Fig. 1). We can also see that the increase of the traffic load in Stream 3 will increase the traffic congestion more sharply than Stream 2 does, since it requires and occupies more time slots in the frame for establishing a connection than Stream 2 does.

Following the assumptions, stated in Section 2, concerning user distribution into non-overlapping cell areas, at the second point of the analysis, we investigate the overall traffic congestion introduced by the system, versus the class limitation of both Stream 2 and Stream 3. Results depicted on Fig. 2 demonstrate this for different traffic volumes for each traffic stream (A_a^1 = 5 erl, A_a^2 = 8 erl, A_a^3 = 12 erl). We have shown two scenarios, taking into account the case where the traffic volume of Stream 1 is increased as well as the offered traffics of three streams increase their values with approx. 30%. These results are logical, since the class limitation of Stream 2 and Stream 3 leads to reservation of time slots exclusively for traffic Stream 1, resulting in decreasing of its traffic congestion, in case of increased traffic loads generated by the MSs in the cell, being served by the transmission mode employing the available modulation technique of highest order (64-QAM).

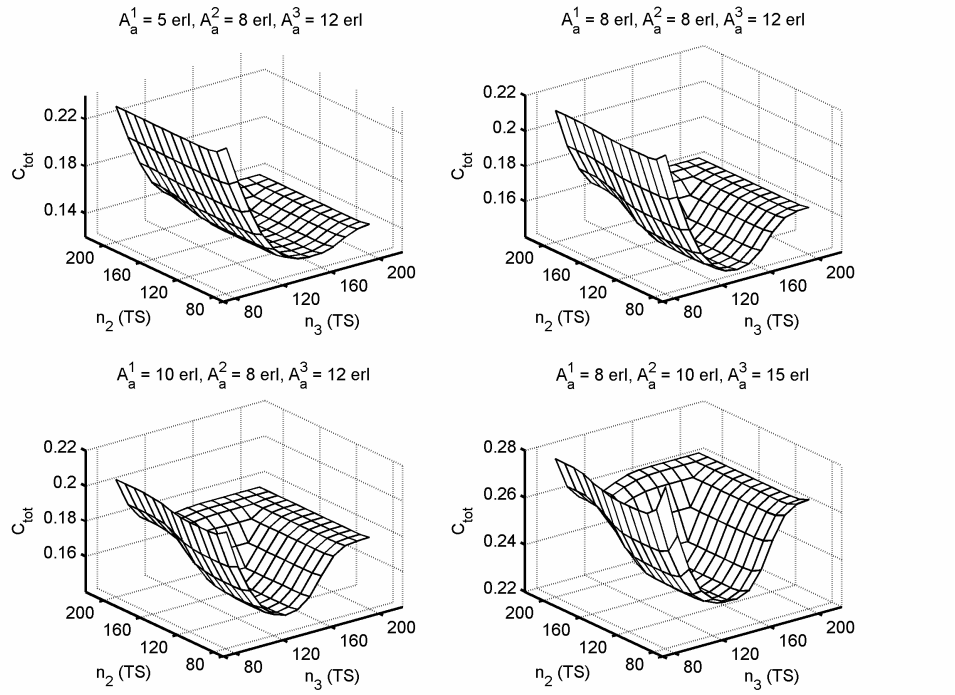


Fig. 2. Overall traffic congestion versus class limitation of Stream 2 and/or Stream 3, over different traffic loads

V. CONCLUSION

In this paper, we have investigated the performance of an OFDM system with AMC that carries voice traffic. Based on the model of multi-rate loss systems, we have determined the overall traffic congestion in an efficient way by applying the convolution algorithm. We have defined several service classes, depending on the AMC scheme used, related to data transmission with constant bit rate from MSs inside a cell area. It has been shown that the traffic congestion of a call attempt of a particular service class depends on the state of the system, bandwidth required as well as the optimal allocation of available resources by applying a class limitation policy (limited accessibility to common resources in the frame of Stream 2 and Stream 3). The analysis, we have conducted, demonstrates that this approach results in finding the minimal overall traffic congestion, which led to improvement in system throughput.

ACKNOWLEDGEMENT

This paper is a part of research work in the context of the following research projects funded by the Bulgarian Ministry of Education and Science: "Optimal telecommunication resource allocation considering cross-layer interaction", (grant DVU01/0109), "Multimedia telecommunication networks planning with quality of service and traffic control", (grant ВУ-ТН-105/2005), as well as a research project funded by the Research and development sector of the Technical University of Sofia with grant 080910dni-7.

REFERENCES

[1] Physical Layer Aspects of UTRA High Speed Downlink Packet Access (Release 4), 3GPP TR 28.848 V4.0.0, 2001.

[2] IEEE Standard 802.16 Working Group, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface fox Fixed Broadband Wireless Access Systems, 2002.

[3] M. S. Alouini, A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels", J. Wireless Commun., vol. 13, no.1-2, pp. 119-143, May 2000.

[4] K. J. Hole, H. Holm, G. E. Oien, "Adaptive multidimensional coded modulation over flat fading channels", IEEE J. Sel. Areas Commun., vol. 18, no. 7, pp.1153-1158, Jul. 2000.

[5] L. Qingwen, Z. Shengli, B. Georgious, "Queuing with Adaptive Modulation and Coding over Wireless Links: Cross-layer Analysis and Design", IEEE Trans. Wireless Commun., vol. 4, no. 3, pp. 1142-1153, May 2005.

[6] C. Tarhini, T. Chahed, "System Capacity in OFDMA-based WiMAX", ICSNC'06, Conference Proceedings, vol. 4, no. 3, pp. 70-74, 2006.

[7] H. Wang, V. Iversen, "Erlang Capacity of Multi-class TDMA Systems with Adaptive Modulation and Coding", ICC'08, Conference Proceedings, pp. 115-119, 2008.

[8] V. Iversen, *Teletraffic Engineering Handbook*, COM department, Technical University of Denmark, 2006.

[9] A. Doufexi, et al., "A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards", IEEE Commun. Mag., vol. 40, no. 5, pp. 172-180, May 2002.

[10] C. Hoymann, "Analysis and performance evaluation of the OFDM-based metropolitan area networks IEEE 802.16", IEEE Computer Networks, vol. 49, no. 3, pp. 341-363, 2005.

A Discrete Time Queueing Model with a Constant Packet Size

Seferin T. Mirtchev[1] and Rossitza Goleva[2]

Abstract – **The importance of IP network will further increase and it will serve as a platform for more and more services, requiring different types and degrees of service quality. Modern architectures and protocols are being standardized, which aims at guaranteeing the quality of service delivered to users. In this paper, we investigate the queueing behaviour found in IP output buffers. This queueing increases because multiple streams of packets with different length are being multiplexed together. To analyze these types of behaviour, we study the discrete-time version of the "classical" queue model M/D/1/k called Geo/D/1/k. This is a discrete-time single server FIFO queue with Bernoulli arrivals and deterministic service times. We develop balance equations for the state of the system, from which we derive packet loss and delay results.**

Keywords – **Queueing system, Queueing analyses, Discrete time queue, IP traffic modelling**

I. INTRODUCTION

The initial motivation for this paper is the necessity of traffic engineering in NGN networks. Many analyses of Internet traffic behaviour require accurate knowledge of the traffic characteristics for purposes ranging from a management of the network quality of service to modelling the effect of new protocols on the existing traffic mix.

Modern architectures and protocols are being standardized, which aims at guaranteeing the quality of service delivered to users. The proper functioning of these protocols requires an increasingly detailed knowledge for statistical characteristics of IP packets. The amount of information flowing through the network also increases, and the challenge is to obtain the accurate information from a huge set of data packets [5,7]. The packet queueing in an IP router arises because multiple streams of packets from different input ports are being multiplexed together over the same output port [9,14].

Many communication systems operate on a discrete-time basis and events can only happen at regularly spaced epochs.

Discrete-time queueing systems have been receiving increased attention in recent years due to their usefulness in modeling and analyzing various types of communication systems. But the technical difficulties which arise in working with a discrete time scale are considerably greater than in dealing with continuous time models although the mathematics behind is more elementary than in the continuous time case. The reason for this is the combinational complexity which appears in the solution procedures for the global balance equations of these systems.

Discrete-time queueing systems have been a research topic for several decades now and there are many reference works on discrete-time queueing theory. Over the years, different methodologies have been developed to assess the performance of queueing systems. The two main analytical approaches are the matrix analytic method and the transform method for discrete and for continuous-time analyses. Many authors have considered the Geo/G/1 discrete-time queueing system [8], [10], [12], [13].

[6] has studied a discrete-time Geo/D/1 and Geo/D/1/n queues. The closed-form expressions for the steady-state distributions of the queue length and of the unfinished work in system (i.e. waiting time) are obtained by the method of analysis, using Lindley's equation.

In [1] a complete study of a discrete-time single-server queue with geometrical arrivals of both positive and negative customers is carried out. Negative arrivals are used as a control mechanism in many telecommunication and computer networks. [2] has concerned with the study of a discrete-time single-server retrial queue with geometrical inter-arrival times and a phase-type service process. An iterative algorithm to calculate the stationary distribution of Markov chain is given.

[11] has proposed a traffic model and a parameter fitting procedure that are capable of achieving accurate prediction of the queuing behaviour for IP traffic exhibiting long-range dependence. The modelling process is a discrete-time batch Markovian arrival process (dBMAP) that jointly characterizes the packet arrival process and the packet size distribution. In the proposed dBMAP, packet arrivals occur according to a discrete-time Markov modulated Poisson process (dMMPP) and each arrival is characterized by a packet size with a general distribution that may depend on the phase of the dMMPP.

[3] has presented an introduction to bandwidth estimation and a solution to the problem of the best-effort traffic for the case when the quality criteria specify negligible packet loss. The solution is a simple statistical model, which is built and validated using queueing theory and extensive empirical study.

It has been shown [4] that in the case of real-time communications, for which small buffers are used for delay reasons, short range dependence dominates the loss process and so the Markov-modulated Poisson process (MMPP) might be a reasonable source model. They have presented an exact mathematical model for the loss process of a MMPP+M/Ek/1/K queue and have concluded that the packet size distribution affects the packet loss process and thus the efficiency of forward error correction.

[1]Seferin Mirtchev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: stm@tu-sofia.bg .

[2]Rossitza Goleva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: rig@tu-sofia.bg .

In this paper, we investigate the basic queueing behaviour of packets found in IP output buffers. The traffic is being generated from the packets of constant size that arrive for transmission on the link. The packets can queue up and loss if their size is bigger than the free positions of the buffer. The quality metrics for the best-effort traffic on the Internet are the packets loss and delay. To analyze these types of behaviour, we study the discrete-time version of the "classical" queue model M/D/1/k called Geo/D/1/k. We developed balance equations for the state of the system, from which we derived packets loss and delay.

II. BALANCE EQUATIONS - GEO/D/1/K QUEUE

We investigate a single server finite queue delay system Geo/D/1/k with a geometric distributed inter-arrival time and a constant packet length. We consider queueing phenomena in discrete-time queueing systems. We assume a fundamental time unit (time slot), the time to transmit an octet (byte), T_b. Customers arrive in the queueing system under consideration during the consecutive slots, but they can only start service at the beginning of slots. That is, service of customers is synchronized with respect to slot boundaries. Further, customer service times are integer multiples of the slot length, which implies that customers leave the system at slot boundaries. During the consecutive slots, packets arrive in the system, are stored in a finite capacity queue and are served by a single server on a first in first out (FIFO) basis (Fig. 1).

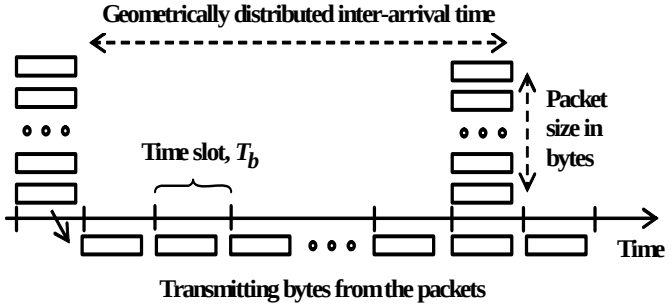


Fig. 1. Timing of events in the Geo/D/1/k queue.

We use a Bernoulli process for the packet arrivals, i.e. a geometrically distributed number of slots between arrivals. Let the probability that a packet arrives in an octet slot is ***p***.

Thus we have a batch arrival process with geometrically distributed inter-arrival times. That is, the number of slots that separate consecutive slots where there are customer arrivals, constitute a series of independent and identically geometrically distributed random variables. The probability of no octets arriving in a time slot is

$$a_0 = 1 - p \tag{1}$$

In this model, we assume a constant packet size with a value ***m***.

The probability that ***m*** octets arrive in a time slot is

$$a_m = p \tag{2}$$

The mean packet service time is the octet transmission time multiplied by the number of octets

$$\tau = T_b\, m,\ \ s \tag{3}$$

The mean arrival rate is

$$\lambda = p/T_b,\ \ \ packets/s \tag{4}$$

Therefore, the offered traffic is given by

$$A = \lambda\tau = p\ m,\ \ Erl \tag{5}$$

The average inter-arrival time of the packets in time slots of the geometric distribution is

$$m_o = 1/p \tag{6}$$

The variance of the inter-arrival time of the packets is

$$v_i = (1-p)/p^2 \tag{7}$$

We define the state probability P_i of being of state ***i***, as the probability that there are ***i*** octets in the system at the end of any time slot. For the system to contain ***i*** bytes at the end of any time slots it could contain any of *0, 1, 2, ..., i+1* at the end of the previous slot. State ***i*** can be reached from any of the states ***0*** up to ***i*** by a precise packet arrival. To move from ***i+1*** to ***i*** there should be no arrivals.

We can write the first equation by considering all the ways in which it is possible to reach the empty state

$$P_0 = (P_0 + P_1)a_0 \tag{8}$$

Similarly, we find a formula for the next state probabilities by writing the balance equations

$$P_i = P_{i+1}a_0,\qquad 1 \le i \le m-1 \tag{9}$$

We continue with this process and take into account that it is possible to enter a packet in a time slot with length ***m*** bytes

$$\begin{aligned}
P_m &= (P_0 + P_1)a_m + P_{m+1}a_0 \\
P_{m+1} &= P_2 a_m + P_{m+2}a_0 \\
P_{m+2} &= P_3 a_m + P_{m+3}a_0 \\
&\ o\ \ o\ \ o \\
P_{k-m+1} &= P_{k-2m+2}a_m + P_{k-m+2}a_0 \\
P_{k-m+2} &= P_{k-2m+3}a_m + P_{k-m+3} \\
&\ o\ \ o\ \ o \\
P_{k-1} &= P_{k-m}a_m + P_k \\
P_k &= P_{k-m+1}a_m + P_{k+1}
\end{aligned} \tag{10}$$

Then using the fact that all the state probabilities must sum to 1 we write the last equation

$$\sum_{i=0}^{k+1} P_i = 1 \qquad (11)$$

We can solve the system Eqs. (8), (9), (10) and (11) and calculate the state probabilities.

III. Performance Measures

The carried traffic is equivalent to the probability that the system is busy

$$A_o = 1 - P_0, \ erl \qquad (12)$$

The packet congestion probability is the ratio of lost traffic (offered minus carried traffic) to offered traffic

$$B = (A - A_o)/A \qquad (13)$$

The mean number of bytes and packets present in the system in steady state by definition is

$$L_b = \sum_{j=1}^{k+1} j P_j, \ bytes; \qquad L_p = L_b/m, \ packets \qquad (14)$$

From the Little formula, we have the normalized mean system time of the bytes (time is measured in time slots)

$$\frac{W_b}{T_b} = \frac{L_b}{T_b \lambda m} = \frac{L_b}{A} \qquad (15)$$

IV. Numerical Results

In this section, we give numerical results obtained by a Pascal program on a personal computer. The described methods were tested on a computer over a wide range of arguments.

Fig. 2 shows the stationary probability distribution in a single server queue *Geo/D/1/k* with 0.95 erl offered traffic, 100 waiting positions and different packet length. We can see that the probability distributions are almost linear decreasing in logarithmic scale and the influence of the packet length distribution kind on the stationary probability is significant.

Fig. 3 compares the packet congestion probability as function of the traffic intensity when the waiting positions are 200 and different packet size. We can see that the influence of the packet length on the packet congestion probability is big.

Fig. 4 presents the normalized mean system time of the bytes (W/T_b) as function of the traffic intensity when the queue length is 200 bytes and different packet size. The influence of the packet size on the mean system time is significant when the offered traffic is smaller then 1 erl.

We compare the discrete-time queue *Geo/D/1/k* with continuous time Poisson arrival queue model *M/D/1/k*. In this model we can accept that the packet length is one byte. When the packets size increases (3, 5 and 10 bytes) the stationary probability distribution is change and the packet congestion probability and mean system time increase vastly.

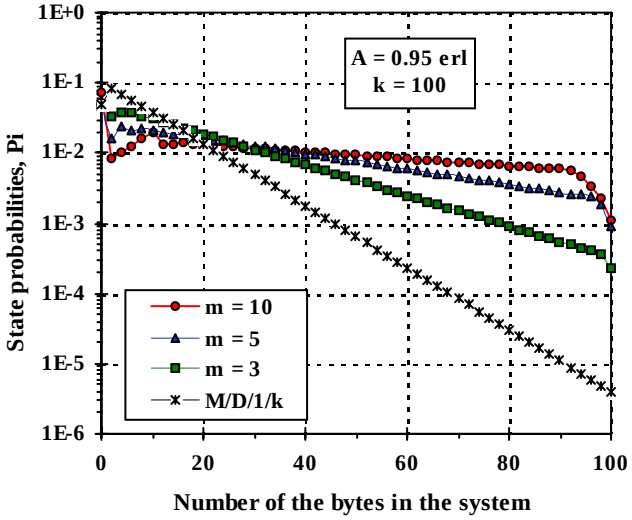


Fig. 2. Stationary probability distribution

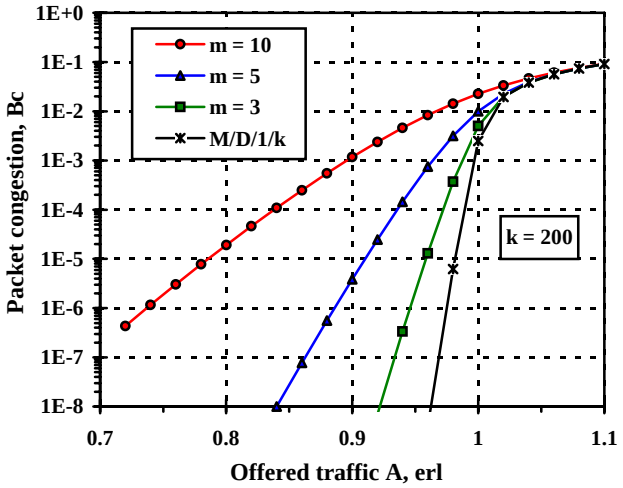


Fig. 3. Stationary probability distribution

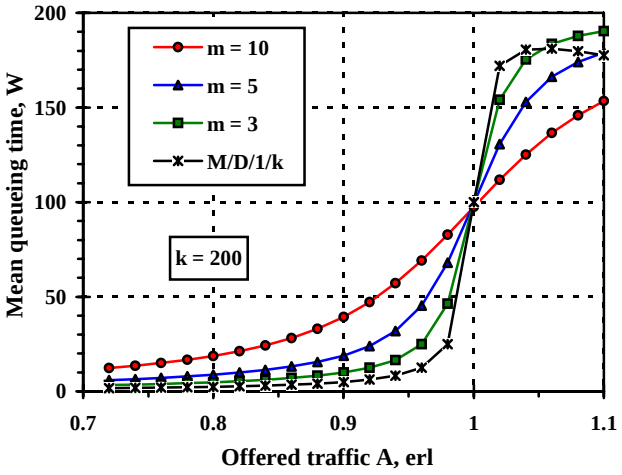


Fig. 4. Stationary probability distribution

V. CONCLUSION

In this paper a basic discrete-time single server teletraffic system Geo/D/1/k with a constant packet length is examined in detail. We present an analysis of this discrete-time queue.

The proposed approach provides a unified framework to model discrete-time single server queue. Numerical results and subsequent experience have shown that this approach is accurate and useful in both analyses and simulations of traffic systems.

The importance of a single server queue in a case of a geometric input stream and constant packet length comes from its ability to describe behaviour that is to be found in more complex real queueing systems. It is the case in a general traffic system, which is an important feature in designing telecommunication networks and systems.

The results presented here add a new aspect to the evaluation of the discrete-time queueing system, and serve as a basis for future research on guaranteeing the quality of service

In conclusion, we believe that the presented analytical model will be useful in practice.

ACKNOWLEDGEMENT

This paper is sponsored by the Ministry of Education and Science of the Republic of Bulgaria in the framework of project ВУ-ТН-105/2005 "Multimedia Telecommunications Networks Planning with Quality of Service and Traffic Management".

REFERENCES

[1] Atencia I. and P. Moreno. A single-server G-queue in discrete-time with geometrical arrival and service process. Perform. Eval. 59: pp. 85-97 (2005)

[2] Atencia I, P. Bocharov and P. Moreno. A discrete-time Geo/PH/1 queueing system with repeated attempts. Информационные процессы, Том 6, N: 3, стр. 272-280 (2006).

[3] Cao J., W. Cleveland and D. Sun. Bandwidth estimation for best-effort Internet traffic. Statist. Sci., Volume 19, Number 3 (2004), pp. 518-543.

[4] Dan G., V. Fodor, and G. Karlsson. "Packet size distribution: an aside?" in Proc. of QoS-IP'05, pp. 75–87, February 2005.

[5] Farber J., S. Bodamer and J. Charzinski. Measurement and Modelling of Internet Traffic at Access Networks. Proceedings of the EUNICE'98,1998,196-203.

[6] Gravey A., J. Louvion and P. Boyer. On the Geo/D/1 and Geo/D/1/n queues. Performance Evaluation, 11, pp. 117 —. 125, 1990.

[7] Janevski T., D. Temkov, A. Tudjarov. Statistical Analysis and Modelling of the Internet Traffic. ICEST Sofia, 2003, pp. 170-173.

[8] Mirtchev S., G. Balabanov and S. Statev. New Teletraffic Models in the IP Networks. National Conference with Foreign Participation, Telecom'2006, Varna, Bulgaria, 2006 (in Bulgarian).

[9] Paxson V. and S. Floyd. Wide Area Traffic: The Failure of Poisson Modelling. IEEE/ACM Transactions on Networking, Vol. 3, no.3, June 1995, pp. 226-244.

[10] Pitts J.and J. Schormans. Introduction to IP and ATM Design and Performance - 2nd Ed. John Wiley & Sons, 2000.

[11] Salvador P., A. Pacheco and R. Valadas. Modelling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs. Computer Networks, Volume 44, Issue 3 , 2004, pp. 335-352.

[12] Vicari N. and P. Tran-Gia. A numerical analysis of the Geo/D/N queueing system. Technical Report 04, COST-257, 1996.

[13] Zhang Z. and N.Tian. Discrete Time Geo/G/1 Queue with Multiple Adaptive Vacations. Queueing Systems, Volume 38, Number 4, August 2001, pp. 419-429.

[14] Zukerman M., T. Neame and R. Addie. Internet Traffic Modelling and Future Technology Implications. IEEE INFOCOM 2003, Vol.22, no.1,March 2003, pp.578-596.

Simulation Modeling of Self-Similar Teletraffic

Dimitar Radev[1], Dragan Stankovski[2], Svetla Radeva3

Abstract — **The paper consider simulation modeling of self-similar teletraffic. Steady-state simulation of self-similar queuing processes was provided with fixed-length sequence generators. A new algorithm for buffer overflow with limited relative error is developed. The implementation of suggested simulation approach for stochastic and long range dependence teletraffic source models is shown. Buffer overflow simulation for finite buffer single server model under self-similar traffic load SSM/M/1/B is considered.**

Keywords — **Communication systems, Modeling, Queuing analysis, Simulation software, Stochastic processes**

I. INTRODUCTION

Apromising method in integration of technologies and services in digital communication networks is the simulation modeling of sources, traffic and their management. The simulation modeling is an abstraction of the real interaction between sources, which leads to simplification and limited performance of the behavior of the elements in real communication network. This is realized with the help of stochastic processes where the generated traffic is based on study of the behavior of certain classes of probability processes like time series analysis, long range dependence stochastic processes, chaotic time series, discrete-time Markov state models, queuing theory [2].

Recent studies of real teletraffic data show that teletraffic exhibits self-similar (or fractal) properties over a wide range of time scales [1]. The self-similar nature of teletrafic (in sense of long-range dependent behaviour is exhibited over a range of time scales: milliseconds, seconds, minutes and hours) can have a significant impact on network performance. The properties of self-similar teletraffic are very different from properties of traditional models based on Poisson, Markov-modulated Poisson, and related processes [2]. The use of traditional models in networks characterized by self-similar processes can lead to incorrect conclusions about the performance of analyzed networks.

The traditional models can lead to over-estimation of the network performance/quality, insufficient allocation of communication and data processing resources, and difficulties in ensuring the quality of service (QoS) expected by network users. Thus, full understanding is that the self-similar nature in teletrafic is an important issue.

Self-similar teletraffic is observed in LAN and WAN, where superposition of strictly independent alternating ON/OFF traffic models whose ON- or OFF-periods have heavy-tailed distributions. In ATM network traffic self-similar traffic arriving at an ATM buffer results in a heavy-tailed buffer occupancy distribution and buffer cell loss probability decreases with the buffer size, not exponentially as in traditional Markovian models, but hyperbolically.

Other implementation of traffic self-similarity is in Internet traffic, where many characteristics of WWW can be moddeled using heavy-tailed distributions to help user requests for documents and the distribution of WWW document sizes. In TCP/IP network traffic the transfer of files or messages show that the reliable transmission and flow control mechanisms of TCP serve to mainain long range dependent structure included by heavy-tailed file size distributions. The relationship between self-similar traffic and network performane is presented in [1].

The self-similarity observed in video traffic give posibility for developing models for VBR video traffic using heavy-tailed distributions. The autocorrelation of the VBR video sequence decay hyperbolically and can be modelled using *fractional autoregressive integrated moving-average* (F-ARIMA) and *fractional Gaussian noise* (FGN) self-similar processes.

The impact of self-similar models on queueing performance is signifficant and the main trends in such findings are connected with (a) permission traffic modelling for high-speed networks, (b) efficient simulation of actual network traffic and (c) analysing queueing models and protocols under realistic traffic scenarios.

II. SELF-SIMILAR PROCESSES

Self-similarity can be classified into two types: deterministic and stochastic. In the first type, deterministic self-similarity, a mathematical object is assumed to be self-similar (or fractal) if it can be decomposed into smaller copies of itself. This work is focused on stochastic self-similarity. In that case, probabilistic properties of self-similar processes remain unchanged or invariant when the process is viewed at different time scales. This is in contrast to Poisson processes that lose their burstiness and flatten out when time scales are changed. One can distinguish two types of stochastic self-similarity. A continuous-time stochastic process Y_t is strictly self-similar with a self-similarily Hurst parameter H ($1/2<H<1$), if Y_{ct} and $c^H Y_t$ (the rescaled process with time scale ct) have identical finite-dimensional probability for any positive time stretching factor c. When the weakly continuous-time self-similar process Y_t has stationary increments, i.e., the finite-dimensional probability distribution of $Y_{t_0+t} - Y_{t_0}$ do

[1] Dimitar Radev is Professor with the Communication Technique and Technology Department, University of Russe, 7017 Russe, e-mail: dradev@abv.bg.

[2] Dragan Stankovski is PhD student with the Communication Technique and Technology Department, University of Russe, 7017 Russe, e-mail: draganstankovski@gmail.com

[3] Svetla Radeva, is Associate Professor at of UACEG Sofia, Bulgaria e-mail: radeva.s@abv.bg.

not depend on t_0, than such process can be constructed as a stationary incremental process $X=\{X_i=Y_{i+1}-Y_i: i=0,1,2\ldots\}$. For second-order self-similar process the self-similarity is observed at the mean, variance and autocorrelation levels. The process X is asymptotically second-order self-similar with self-similarity parameter H ($0.5<H<1$) if for any block size m, the process $\{m^{1-H}X_k^{(m)}: k=1,2\ldots\}$ has the same covariance structure. Modeling and simulation of self-similar traffic can be performed with the generators of synthetic self-similar sequences, divided into two practical classes: the sequential generators and the fixed-length sequence generators.

In this work under consideration are fixed-length sequence generators, which are implemented for buffer overflow at single server models.

III. FIXED – LENGTH SEQUENCE GENERATORS FOR SELF – SIMILAR TELETRAFFIC

Markovian models for self-similar traffic require including several control parameters with a wide range of input values, like size of the sequence, Hurst parameter, scale parameter, vanishing moment, etc. The most appropriate controlling of these values of self-similar processes is realized with fixed-length sequences generators [3]. Here is used a generator of pseudo-random self-similar sequences based on fractional Gaussian noise and Daubechies wavelets (DW) [3], called FGN-DW approach.

Wavelet analysis transforms a sequence onto a time-scale grid, where the term scale is used instead of frequency. The wavelet transform delivers good resolution in both time and scale, as compared to the Fourier transform, which provides only good frequency resolution. The developed algorithm consists of the following steps.

Algorithm for generating of FGN-DW pseudo-random self-similar sequences**:**

Step 1: Given: Hurst parameter H. Start for $i=1$ and continue until $i=n$. Calculate a sequence of values $\{f_1,f_2,\ldots,f_n\}$ using

$$f(\lambda,H)=c_f\,|\,\lambda\,|^{1-2H}+O(|\,\lambda\,|^{\min(3-2H,2)}), \quad (1)$$

where $c_f=\sigma^2(2\pi)^{-1}\sin(\pi H)\Gamma(2H+1)$, $O(\cdot)$ represents the residual error and $f_i=\hat{f}(\pi i/n;H)$, the value of the frequencies f_i corresponds to the spectral density of an FGN process for f_i ranging between $\pi/n \div \pi$.

Step 2: Multiply $\{f_i\}$ by realizations of the independent exponential random variable with the mean of one to obtain $\{\hat{f}_i\}$, because the spectral density estimated for a given frequency is distributed asymptotically as the independent exponential random variable with the mean $f(\lambda,H)$.

Step 3: Generate a sequence $\{Y_1,Y_2,\ldots,Y_n\}$ of complex numbers such that $|\,Y_i\,|=\sqrt{\hat{f}_i}$ and the phase of Y_i is uniformly distributed between 0 and 2π. This random phase technique preserves the spectral density corresponding to $\{\hat{f}_i\}$. It also makes the marginal distribution of the final sequence normal and produces the requirements for FGN.

Step 4: Calculate two synthetic coefficients of orthonormal Daubechies wavelets. The output sequence $\{X_1,X_2,\ldots,X_n\}$ representing approximately self-similar FGN process is obtained by applying the inverse Daubechies wavelets transformation of the sequence $\{Y_1,Y_2,\ldots,Y_n\}$.

The normalized arrivals sequence for $n=10\,000$ observations received after wavelet transform is shown on Fig.1 with FGN-DW algorithm. The sequence is normalized for integer number of arrivals in interval [0, 40] for following parameters: $H=0{,}75$; Scale=4 and Vanishing Moment=6. The reveived results showed that the generator based on FGN-DW method demonstrated a hidh level of accuracy, was fast and can be implemented for long sequences with long-range dependent properties.

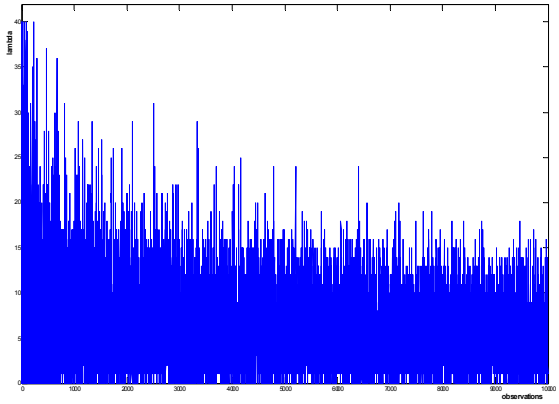


Fig. 1. Arrivals sequence after normalization with wavelet transform.

IV. STEADY – STATE SIMULATION OF SELF – SIMILAR QUEUING SYSTEMS

There exists a significant difference in the queueing performance between traditional models of teletraffic such as Poisson processes and Markovian processes, and those exhibiting self-similar behaviour. More specifically, while tails of queue length distributions in traditional models of teletraffic decrease exponentially, those of self-similar traffic models decrease much slower. Under investigation are potential impacts of traffic characteristics, including the effects of self-similar behaviour on queuing and network performance, protocol analysis, and network congestion controls.

The steady-state simulation of self-similar queueing systems inclide generation of self-similar traffic, simulation of queing process and simulation of overflow probability. Here this is illustated on buffer overflow simulation for SSM/M/1/B queueing systems ($B<\infty$) (i.e. queuing systems with the finite buffer capacity) at self-similar queuing processes. In this case, the difference from M/M/1/B queuing system is that the arrival rate λ_j into SSM/M/1/B queuing system is not a

constant value, where SSM/M/1/B queuing system has exponential service times with constant rates $1/\mu$ as is shown on Fig. 2.

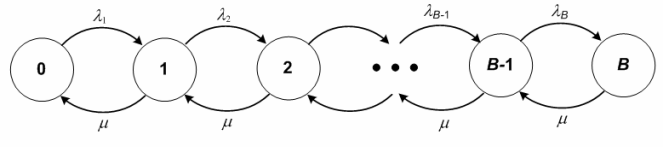


Fig. 2. State transition diagram for a SSM/M/1/B queuing system.

The flow balance equations are

$$\begin{aligned} &\lambda_j = \lambda(i,n,H); && j = 1,2,...,B \\ &\lambda_j = 0; && j \geq B+1 \\ &\mu_j = \mu; && j = 1,2,...,B+1 \end{aligned} \quad (2)$$

This system is stable whether the throughput $\rho = \dfrac{\lambda(i,n,H)}{\mu} < 1$. Let consider two separated cases: ρ=1, and $\rho \neq 1$. For $j = 0,1,2,...,B$ the distribution of number of customers in system is $P_j = \rho^j P_0$, which is determined according to

$$\begin{aligned} &P_j = \frac{\rho^j(1-\rho)}{1-\rho^{B+1}}; && \rho \neq 1 \\ &P_j = \frac{1}{B+1}; && \rho = 1 \end{aligned} \quad (3)$$

Hence, the rate at which the custumers are lost (blocked) is λP_B. The queuing process is described with the steady-state simulation algorithm, presented on Fig.3.

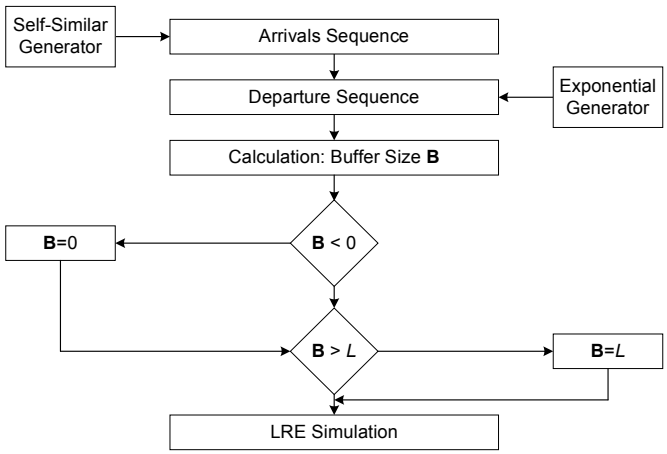


Fig. 3. Algorithm of queuing process.

The simulation is realized with RESTART method [4], which is a variant of splitting where any chain is split by a fixed factor when it hits a level upward, and one of the copies is tagged as *original* for that simulation level [5]. When any of those copies hits that same level downward, if it is the original it just continues its path, otherwise it is killed immediately. This rule applies recursively, and the method is implemented in a depth-first fashion, as follows: whenever there is a split, all the non-original copies are simulated completely, one after the other, then simulation continues for the original chain. The calculation of buffer size for all sequences gives possibilities for determining the overflow probability.

V. LIMITED RELATIVE ERROR SIMULATION OF BUFFER OVERFLOW

The limited relative error (LRE) [5] gives posibility to determine the complementary cummulative function of arrivals at single server buffer queues with Markov processes. For describing the princilpes of LRE for steady-state simulation in disctere-time Markov chains consider homogenious two-node Markov chain, wich is extended to common disctere-time Markov chain, consisting of k+1 nodes with states, respectively $S_0,S_1,\ldots,S_k$. We receive the random generated sequence $x_1,x_2,...,x_t,x_{t+1},\ldots$ for x=0,1,…, k, for wich exists transition for state S_j at the time t, e.g. $x_t=j$ and there are no constraints to the parameters of transition probabilities

$$p_{ij} = \mathrm{P}(j \mid i); \quad (i,j) = 0,1,...,k; \quad \sum_{j=1}^{k} p_{ij} = 1. \quad (4)$$

There are no absorbing states S_i at p_{ii}=1 for all stationary probabilities P_j j=0,1,…,k, which satisfy the constraint condition

$$0 \leq \mathrm{P}_j < 1; \quad \sum_{j=0}^{k} \mathrm{P}_j = 1. \quad (5)$$

The cummulative distribution $F(x)$ can be presented as

$$\left.\begin{aligned} &F(x) = F_i; \quad (i-1) \leq x < i; \quad i = 1,2,...,k+1; \\ &F_i = \sum_{j=0}^{i-1} \mathrm{P}_j; \quad F_0 = 0; \quad F_{k+1} = 1; \end{aligned}\right\}. \quad (6)$$

For simulation of (k+1) nodes Markov chain more significant is complementary cummulative distribution $G(x)$=1-$F(x)$, wich togather with the local correlation coefficient $\rho(x)$ can be determined with the help of limited relative error approach. After determining of two node Markov chain, via changing of the states n times can be received an estimation of the local correlation coefficient $\hat{\rho}(x)$, which can connect the number of transitions througt dividing line $a_i \approx c_i$, with the total number of observated events at left side $l_i=n-d_i$ $(\beta = 0,1,...,i-1)$, and at rigth side d_i $(\beta = i,i+2,...,k)$.

The value of simulated complementary cumulative distribution $\hat{G}_i$ can be determined directly via relative frequency d_i/n, if there are enough number of samples

$$n \geq 10^3; \quad (l_i, d_i \geq 10^2); (a_i, c_i, l_i - a_i, d_i - c_i) \geq 10 \quad (7)$$

The posterior equations can be used for the complementary function $\hat{G}(x)$, for tha average number of generated values of $\hat{\beta}$, for the local correlation coefficient $\hat{\rho}(x)$, for the correlation coefficient Cor[x] and for the relative error RE[x]

$$\hat{G}(x) = \hat{G}_i = d_i / n$$

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{k} d_i$$

$$\hat{\rho}(x) = \hat{\rho}_i = 1 - \frac{c_i / d_i}{1 - d_i / n} \qquad \begin{matrix} i-1 \le x < i \\ i = 1,...,k \end{matrix} \tag{8}$$

$$\mathrm{Cor}[x] = \mathrm{Cor}_i = (1+\hat{\rho}_i)(1-\hat{\rho}_i)$$

$$\mathrm{RE}[x]^2 = \mathrm{RE}_i = \frac{1 - d_i / n}{d_i} \cdot \mathrm{Cor}_i$$

The main advantage of this approach is that the relationships between transitions c_i is reached with routine statistical calculations. The necessary total number of simulation trails n is determined by the maximal relative error $\mathrm{RE}_{\max}[x]^2$ and by the less value of the function $G(x)$, presented as $G_{\min} = \hat{G}_k$ in approximation equation

$$n = \frac{(1 - G_{\min})}{G_{\min}.\mathrm{RE}_{\max}[x]^2} \approx \frac{\mathrm{Cor}_k}{\hat{G}_k.\mathrm{RE}_{\max}[x]^2}; \tag{9}$$

$$\mathrm{Cor}_k = \frac{1+\hat{\rho}_k}{1-\hat{\rho}_k}$$

This procedure is realized with common version of limmited relarive error algorithm for random discrete sequences.

VI. Experimental results

To show the effect of self-similarity on probability of buffer overflow, the received experimental results for SSM/M/1/B queuing system are compared with complementary cumulative distribution for classical single server finite buffer queue M/M/1/B as shown on Fig. 4.

For receiving representative and stable results were used sequences of 10 000 observations. With the suggested LRE algorithm were calculated the values of complementary cumulative function $G(x)$ for different buffer size. The calculations were provided with step m=4. The increasing of Hurst parameter leads to insignificant decreasing of overflow probability. For example, for the value of Hurst parameter H=0,6 the overflow probability is $G(L)$=10,45.10^{-2}, and for H=0,9 –G(L)=5,6.10^{-2}.

VII. Conclusion

Steady-state simulation of self-similar queuing processes was provided with fixed-length sequence generator. The simulation is realized with RESTART method. A new algorithm for simulation of buffer overflow probability was developed for self-similar traffic generation with limited relative error. Buffer overflow simulation for finite buffer single server model under self-similar traffic load with suggested algorithm is considered.

An implementation of suggested simulation approach for the stochastic and long range dependence teletraffic source models is shown. Models of SSM/M/1/40 queuing system with different characteristics of self-similarly process of arrivals and different buffer size are presented.

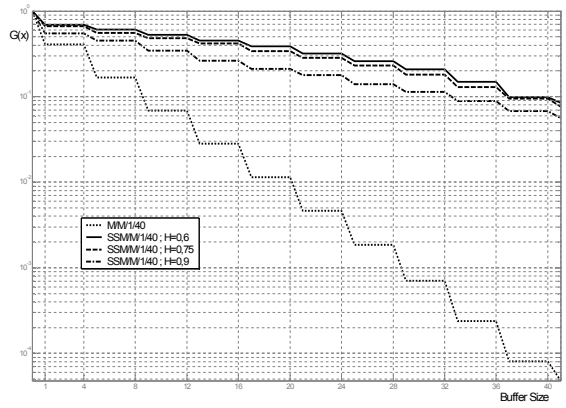


Fig. 4. Simulation of buffer overflow probability (L=41) for SSM/M/1/40 queuing system.

References

[1] J. J. Hae-Duck, “Modelling of Self-Similar Teletraffic for Simulation,” PhD Thesises, University of Centerbury, UK, 2002. p. 269.

[2] D. Radev, “Stochastic process generation for traffic flow simulation.” *Proc. of the IEEE Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services TELSIKS’2005*, vol. 2, Nis, Serbia, 2005, pp. 559-562.

[3] P. L'Ecuyer, “Random Number Generators. - Encyclopedia of Operations Research and Management Science.” Kluwer Academic Publishers, 2001. pp. 695-702.

[4] M. Villen-Altamirano and J. Villen-Altamirano, “On the efficiency of RESTART for multidimensional systems.” *ACM Transactions on Modeling and Computer Simulation*, vol. 16 (3), 2006, pp. 251–279.

[5] C. Gorg and O. Fuss, “Simulating rare event details of ATM delay time distributions with restart/lre”. *Proceedings of the IEEE International Teletrafic Congress, ITC16*, 1999, Elsevier, pp. 777–786.

Session

SIGNAL PROCESSING – 1

Inter/intra-frame Compression of Video Sequences of Dynamic Signatures with Hierarchical Decomposition

R. Kountchev[1], M. Milanova[2], Vl. Todorov[3], R. Kountcheva[4]

Abstract: **In the paper is presented one new method for efficient compression of TV sequences representing dynamic signatures (each signature is a single sequence). The method is based on inter/intra-frame compression of the consecutive TV frames using Hierarchical decomposition. The compression starts with the processing of the reference frame in the sequence. The coarse approximation of the reference frame is first calculated and it is then used as a basis for the whole sequence. The remaining TV frames are coded sequentially one by one, using the approximation for the first frame and correcting it in correspondence with the currently processed frame. The information, which represents the details for every frame is added to the coded video sequence data.**

Key-Words: **Dynamic signatures archiving, intra-frame compression, Hierarchical decomposition**

I. INTRODUCTION

The security threats nowadays require new advanced security concepts to be elaborated and implemented as conceptual and feasible responses to these challenges. The solution of the problem requires more sophisticated and reliable means for information security, tracing of people and goods, and data access based on person's identification or authentication. The visual comparison of an individual's signature has been used for many years as a method of identification and is now popular for electronic comparison in which signature dynamics such as pressure, speed, and pen stroke direction are measured. Signature is not classified as characteristic, pertaining to the individual, because this data may change as the user ages and such systems require regular updates of the used database. One of the widely used solutions for the processing of dynamic signatures is based on the use of famous tools such as table-pads, which capture information about the signatures' dynamic characteristics [1].

One contemporary approach is the use of TV records of signatures, taken in the process of signing. This provides a variety of visual information which permits the extraction of characteristics, concerning the way the pen is hold, the way of signing (with stops or without stops, the way the pen is lifted, etc.), i.e. huge amount of information which else could not be sensed or represented mathematically.

The main problems, related with the security system implementation based on using biometric information, usually result from the large amounts of visual data, which have to be sent, received, stored and processed. Similar problems exist in mobile communications, providing services, which require exact and reliable personalization and authentication (bank transactions, etc.). The usual approach is to archive the related information after compression based on one of the well known video compression standards MPEG-1, MPEG-2, MPEG-4 etc. [2,3,4]. They all ensure high compression and good quality of the restored video. All these standards use intra-frame coding based on DCT (MPEG-1, MPEG-2) or wavelet transforms (MPEG-4), inter-frame prediction and inter-frame interpolation with motion compensation. The main disadvantages of the MPEG standards are the low quality of the fast moving objects (as a result from the inter-frame processing) and the relatively high computational complexity. The Motion JPEG 2000 (M-JPEG) standard [5] is aimed at the intra-frame compression of the consecutive TV frames and is based on wavelet transforms. It offers high quality, but the compression ratio is lower than that of the MPEG standards.

Specific feature of the signatures' TV records is that they are produced using a fixed camera. In this case the motion compensation is not necessary and the video processing could be significantly simplified. For such applications is expected that the inter/intra-frame compression will give relatively good results for lower computational complexity.

In this paper is offered one new approach for compression of video sequences, based on inter/intra-frame processing with 2D orthogonal transforms. The new point is to process the TV records of dynamic signatures with specially developed compression method, based on a Hierarchical decomposition, with Inverse Difference Pyramid (IDP). This decomposition permits the first (coarse) approximation for the lowest decomposition layer of one (reference) frame in a group of TV frames to be used for the processing of the remaining frames in the group and the high-quality approximations for each frame to be then calculated individually. As a result is obtained good quality for the fast moving objects in the restored video sequence.

The paper is arranged as follows: Section 2 introduces in brief the principles of the Hierarchical decomposition; Section 3 presents the new method for inter/intra-frame processing of video sequences; Section 4 gives some experimental results obtained and Section 5 (the Conclusion) focuses on the future development of the method.

II. HIERARCHICAL DECOMPOSITION

The detailed description of the IDP had been presented in earlier publications of the authors [6,7]. The main operations comprising the Hierarchical decomposition (HD) method which is a simplified version of the IDP are presented in brief below.

[1]Roumen K. Kountchev is with the Faculty of Communications and Communications Technologies, Technical University, Kl. Ohridsky 8, 1000 Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg

[2]Mariofanna G. Milanova is with the Department of Computer Science, UALR, USA, E-mail: mgmilanova@ualr.edu

[3]Vladimir T. Todorov is with T&K Engineering, Mladost 3, POB 12, 1712 Sofia, Bulgaria, E-mail: todorov_vl@yahoo.com

[4]Roumiana At. Kountcheva is with T&K Engineering, Mladost 3, POB 12, 1712 Sofia, Bulgaria, E-mail: kountcheva_r@yahoo.com

The initial presumption is that the matrix [X] of the original image is divided into sub-images of size $2^n \times 2^n$ and each is processed with a two-dimensional (2D) orthogonal transform using only the low-frequency spectrum coefficients (i.e. the transform is "truncated"). The values of the transform coefficients are then processed with the corresponding inverse orthogonal transform. In result is obtained the coarse approximation of the processed image. This approximation is subtracted pixel by pixel from the original and in result is obtained the difference image with elements $e_p(i,k)$. The coarse approximation p is defined as:

$$e_p(i,k)=\begin{cases} x(i,k)-\tilde{x}(i,k) & \text{for} \quad p=0; \\ e_{p-1}(i,k)-\tilde{e}_{p-1}(i,k) & \text{for} \ p=1,2,..,P, \end{cases} \qquad (1)$$

where x(i,k) is the pixel (i,k) in the sub-image of size $2^n \times 2^n$ from the input image [X]; $\tilde{x}(i,k)$ and $\tilde{e}_{p-1}(i,k)$ are correspondingly the pixels of the restored and of the difference sub-images in the coarse approximation p. Every difference sub-image is divided into 4 sub-images of size $2^{n-1} \times 2^{n-1}$ and each is then processed with the 2D orthogonal transform again. Using the values of the calculated transform coefficients the image is restored and the second approximation and the corresponding difference image are calculated. The process continues in similar way until the approximation with the needed visual quality is obtained. The processing usually does not require all possible approximations to be calculated, because the needed image quality is obtained earlier, depending on the application.

The approximation models of the input or difference image (p) are represented by the relations:

$$\tilde{x}(i,k)/\tilde{e}_{p-1}(i,k) = IT[y_p(u,v)], \qquad (2)$$
$$y_p(u,v) = T[x(i,k)/e_{p-1}(i,k)]$$

where T[•] is the operator for the "truncated" direct 2D transform applied on the input block of size $2^n \times 2^n$, or on the difference sub-image of size $2^{n-p} \times 2^{n-p}$ for the approximation p=1,2,..,P; IT[•] is the operator for the inverse 2D transform of the spectrum coefficients $y_p(u,v)$ from the approximation p of the "truncated" transform $2^{n-p} \times 2^{n-p}$, obtained in result of the transformation of every ¼ part of the difference sub-image, $e_{p-1}(i,k)$.

The set of transform coefficients, chosen for every approximation, can be different. The coefficients obtained in result of the 2D transform from all consecutive approximations are sorted in accordance with their frequency, scanned sequentially and losslessly compressed. The image restoration (decoding) is performed in reverse order. In the case when this decomposition is used for color images, each of the image components is processed individually. For example, for the YCrCb format 3 pyramids are built.

III. INTER/INTRA-FRAME CODING OF VIDEO SEQUENCES

The HD-based inter/intra-frame coding of video sequences is similar with that for single still image processing, presented above. The difference is that the original TV sequence is divided into groups of consecutive TV frames, processed together. The optimum length of the group is in the range of 12 – 15 frames, which corresponds to the statistical correlation [8] in the TV sequences. The middle or the first frame in the group is used as a reference one and for the whole group is built one common pyramid. For this, the first approximation is calculated for the reference frame only. The difference images for every TV frame in the group are calculated by subtracting the approximation obtained for the reference frame from the corresponding frame. The processing of the sequence continues in similar way for the next decomposition layer. For this, the two parts of the group are divided into smaller groups of 5 TV frames each. The middle frame in the sub-groups is used as a reference one and the next decomposition layer is prepared. The block diagram of the algorithm for 3-layer inter/intra-frame decomposition of 15 consecutive TV frames is given in Fig. 1.

IV. EXPERIMENTAL RESULTS

The experiments were performed with grayscale video sequences (color information in this case is not of high importance). The video camera was fixed. For the experiments were used groups of consecutive TV frames consisting of 15 and 25 frames. The middle frame was used as a reference one. The frames were extracted from DV format with Pinnacle 11. The size of each TV frame was 768 x 576 pixels, 8 bpp. Example single frames are shown in Fig. 2.a,b.

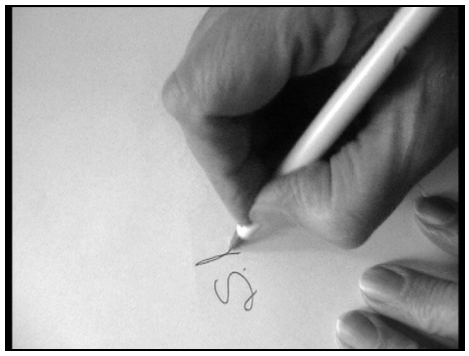


Fig. 2.a. The initial frame in a sequence of 15 TV frames

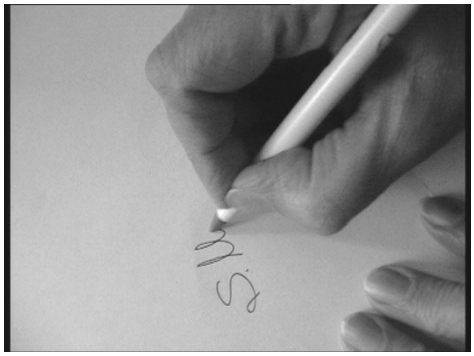


Fig. 2.b. The final frame in a sequence of 15 TV frames

For the experiments was used a 2-layer decomposition. The size of the initial (lower) sub-block was 8x8 pixels and that of the upper-layer sub-block, 4x4 pixels. The coefficients used for the creation of the decomposition approximations were 4 for the lower layer and 3 – for the next one. The results obtained for a sequence of 20 frames are given in Table 1

below. Here is given the information about the compression ratio (CR); the size of the coded approximating images of the consecutive TV frames for layer 2 (Size L2) and the quality of the restored images (PSNR). The size of the approximating file for Layer 1 of the reference TV frame is 12 518 B.

TABLE 1

RESULTS OBTAINED FOR A SEQUENCE OF 20 TV FRAMES

Frame №	Size L2 [B]	File size [B]	PSNR [dB]	CR
10 (ref.)	6 428	442368	38,19	18,35
1	9 659	884736	34,08	20,59
2	10 802	1327104	34,39	25,94
3	10 939	1769472	34,80	29,23
4	12 294	2211840	35,30	28,01
5	12 586	2654208	35,93	33,62
6	12 579	3096576	35,62	39,22
7	12 340	3538944	35,97	44,82
8	13 674	3981312	36,36	50,43
9	13 489	4423680	37,45	56,03
11	14 656	4866048	37,04	61,63
12	16 601	5308416	36,26	67,24
13	16 915	5750784	35,78	72,84
14	17 078	6193152	35,51	78,44
15	17 287	6635520	34,47	84,04
16	17 294	7077888	33,98	89,65
17	17 433	7520256	33,37	95,25
18	17 544	7962624	34,10	100,85
19	17 261	8404992	33,76	106,46
20	17 372	8847360	34,09	112,06
Mean PSNR = 35,3 dB; Coded file size: 225 691 [B]; CR = 112				

The number of participating coefficients in the second layer was defined by taking into account that only ¾ of the coefficients necessary for the presentation of the second-layer approximation are used, as a result from the existing relations between transform coefficients in the consecutive decomposition layers [7]. The column "File size" gives the number of Bytes for the sum of the processed TV frames for the corresponding row in the table. The column "Size L2" gives the size of the approximating file for the second decomposition layer obtained after lossless compression of the coefficients' values. The PSNR (in dB) was calculated individually for each consecutive frame.

Similar investigations were performed for sequences of length 15 and 25 frames. The results obtained show that the compression ratio is lower for sequences longer than 15-18 frames. Naturally, this relation depends on the visual content to a high degree.

The place of the reference frame is of high importance as well. In the example it was chosen to be in the middle of the processed sequence. For the case, when the reference frame was placed in the beginning of the group the results for the compression ratio and the quality of restored images were a little worse. In the first case (reference frame in the middle) the correlation between the consecutive frames was used better and as a result, the size of the approximating images in Layer 2 was smaller. The basic disadvantage is that the process needs the first half of the frames to be saved in a buffer and processed when the reference frame is obtained.

The experimental results for the processed video sequences were compared with those for MJPEG and MPEG-2. The MJPEG is the closest tool, because the compression there is intra-frame. The experimental results show that for similar compression (277 KB for the proposed method and 367 KB for MJPEG) the mean PSNR is correspondingly 35,3dB and 34,2 dB. The comparison results with MPEG are as follows: in general, the bit-rate and the quality of the restored sequence of TV frames are comparable, i.e. for compressed file of size 250 KB, the mean PSNR value for the restored frames for MPEG was 35,9 dB, i.e. the PSNR is a little higher, but as it is known, this is not the best way to evaluate the quality of a moving object. In fact, the visual quality of the fast moving object (the hand) much better for the new method, while the background was reproduced a little better by the MPEG-2 standard. Additional advantage of the presented method it its' lower computational complexity. Further quality enhancement of the restored images in accordance with the new method is possible applying post-processing of the restored TV frames. For this could be used the adaptive fuzzy filter for blocking artifacts reduction, developed by the authors [9]. This filter suits very well the special features of the Hierarchical decomposition and the parameters of the decoded image and ensures mean quality enhancement of 0,4 dB.

V. CONCLUSION

The experiments performed with the new method for intra-frame compression of video sequences, presented in this paper, proved its efficiency. The HD offers some advantages, which still have to be investigated in detail. For example, the relations, existing between some of the 2D orthogonal transform functions permit reduction of the used coefficients number in the decomposition layers [7]. The felicitous use of transform coefficients can ensure quality enhancement with negligible augmentation of the coded data. Instead of evaluating the values of the calculated coefficients, it is possible to prepare some models for certain textures, etc., which will decrease the computational complexity of the method even more.

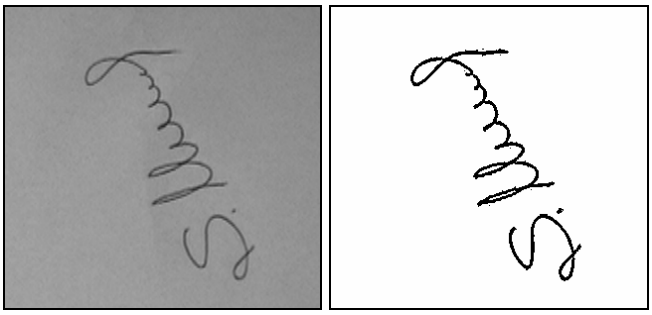
The influence of the decomposition layers number on the compression ratio and the quality of the restored sequence is closely related to the image content and movements speed. Additional investigations will be aimed at these problems.

The HD could be implemented using any kind of orthogonal transform (DCT, Haar, wavelets, etc.), but the Walsh-Hadamard transform has lower computational complexity and together with this ensures the needed quality of the restored images.

The new method is suitable for applications with fixed camera where the motion compensation is not necessary. The new point in this work is that the dynamic signature is saved in the process of signing and after that – coded with inter/intra-frame compression. This method permits together with the traditional similarity comparison to analyze and match a great number of additional parameters and characteristics, which else could not be detected or described. The traditional signature comparison could be done using the same TV sequence of the dynamic signature, because in the last frames the signature is already ready and could be

successfully extracted. Example signature extracted from the TV sequence is shown in Fig. 3.a,b. The high compression ratio, achieved for such kind of video information with the new method for inter/intra-frame coding, permits all the information to be stored as a relatively small file.

The presented method for efficient coding could be used for comparison of video signatures by using content-based information retrieval methods (the compressed document saved in the database is compared with the new signature taken online). Additional possible application areas are the medical imaging, videoconferencing, surveillance, etc.



a) A signature in the TV frame b) the extracted signature

Fig. 3. Part of the TV frame with the finished signature

ACKNOWLEDGEMENT

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science (Contract VU-I 305/2007).

REFERENCES

[1] D. Dessimoz, J. Richiardi, C. Champod, A. Drygajlo. Multimodal biometrics for identity documents: state of the art. Research report, Universite de Lausanne, 2005, pp. 87-94.

[2] B. Haskell, A. Puri and A. Netravali. Digital video: Introduction to MPEG-2. Chapman Hall, NY, 1997

[3] I. Richardson. H.264 and MPEG-4 video compression. J. Wiley & Sons, NY, 2003.

[4] L. Hanzo, P. Cherriman and J. Streit. Video compression and communications: H.261, H.263, H.264, MPEG-4 and HSDPA-style adaptive turbo-transceivers. J. Wiley and IEEE press, 2007.

[5] T. Acharya and P. Tsai. JPEG 2000 standard for image compression: concepts, algorithms and VLSI architectures. J. Wiley & Sons, INC Publication, NJ, 2005.

[6] R. Kountchev, S. Rubin, M. Milanova, Vl. Todorov, R. Kountcheva. Cognitive Image Representation Based on Spectrum Pyramid Decomposition. Proceedings of the WSEAS Intern. Conf. on Mathematical Methods and Computational Techniques in Electrical Engineering (MMACTEE), 2008, pp. 230-235.

[7] R. Kountchev, R. Kountcheva. Image Representation with Reduced Spectrum Pyramid. Book chapter in: New Directions in Intelligent Interactive Multimedia, Eds. G. Tsihrintzis, M. Virvou, R. Howlett, L. Jain, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 275-284.

[8] W. Pratt (Ed.). Image Transmission Techniques. Academic Press, NY, 1979.

[9] R. Kountchev, Vl. Todorov, R. Kountcheva, M. Milanova. Adaptive Fuzzy Filter for Reduction of Blocking Artifacts in Images Compressed with IDP Decomposition. 10th WSEAS Int. Conf. on Computers (CSCC'06), Athens, Greece, July 13-15, 2006, pp. 502-507.

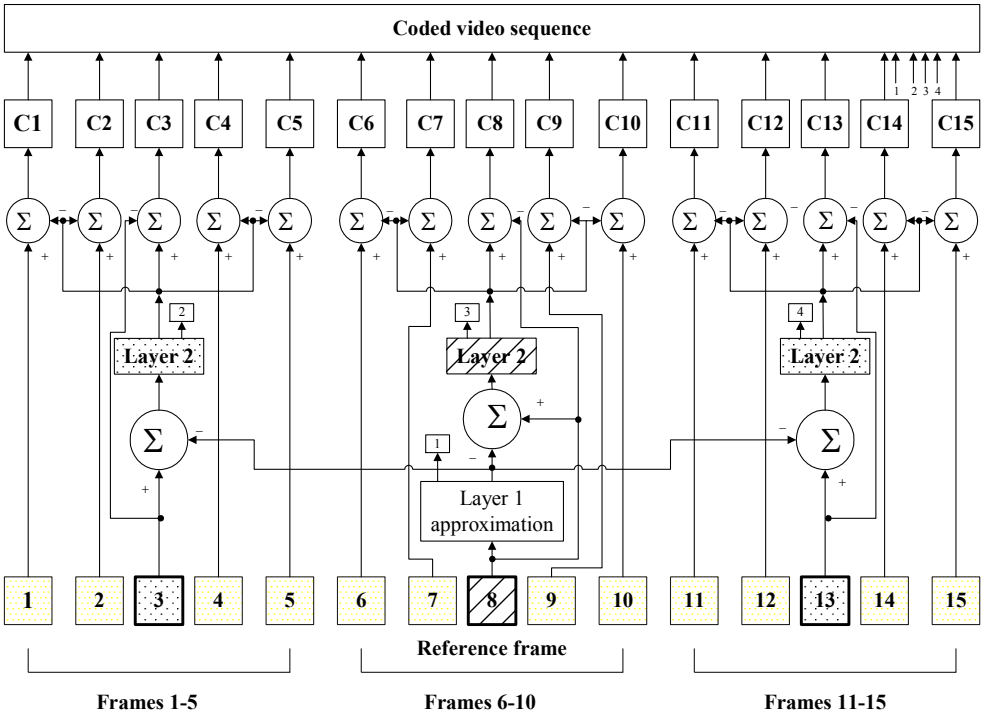


Fig. 1. Block diagram of the inter/intra-frame coding of TV sequences

Abbreviations: 1-15 – original TV frames; C1 – C15 – the coded data for the second-layer approximations of the corresponding TV frames; Σ - block in which is calculated the difference image; Frame 8 – used for the low- layer approximation; Frames 3, 8 and 13 – used for the second layer approximation.

Wavelet Denoising by Using Linear-Phase NPR-QMF Banks

Mitko Kostov and Cvetko Mitrovski

***Abstract* – In this paper we compare the near-perfect reconstruction (NPR) quadrature mirror filter (QMF) design, versus perfect reconstruction (PR)-QMF design in terms of the flexibility of the design and the design filter characteristics, and their efficiency for denoising of the signals based on wavelet shrinkage technique.**

***Keywords* – Near perfect reconstruction, QMF bank, signal-dependent noise, threshold, wavelet domain denoising.**

I. INTRODUCTION

There are many methods, both in space and in a transform domain, for noise removal from the images [1-12]. In cases when the transformation of the original noisy image is adequately chosen, the energy of the signal will be concentrated into a small number of coefficients. One possible choice is to process signals in the discrete wavelet transform (DWT) domain, while other choice is the Fourier transform of a signal which contains energy at all frequencies. Until now, the methods based on wavelet domain filtering use filter banks that satisfy perfect reconstruction (PR) condition.

In this paper we want to show that for denoising purposes a signal can be successfully decomposed and reconstructed by using linear phase near-perfect reconstruction (NPR) QMF banks as an alternative to the well known wavelet filter banks. Even more, using NPR-QMF banks can yield with competitive or even better results. Wavelet shrinkage involves discarding some of the wavelet coefficients and even though wavelet filters satisfy perfect reconstruction condition, the reconstructed signal differs from the original one. Moreover, the transfer characteristic of some of the known wavelets does not have strictly linear phase. However, a linear phase NPR-QMF bank introduces a reconstruction error which is a design parameter in the process of designing the filter bank and hence it can be chosen relatively small. So, if the error introduced by the NPR bank is smaller than the error obtained when known (PR) wavelets are used, the NPR filters can be considered as good filters for denoising purposes.

The paper is organized as follows. The NPR-QMF banks design and wavelet theory are briefly outlined in Sections II and III, respectively. In Section IV we discuss advantages and drawbacks of NPR-QMF filters and commonly used filters. In Section V we compare the filters performances by denoising 1-D and 2-D deterministic signals contaminated with signal-dependent noise. At the end, Section VI concludes the paper.

Mitko Kostov and Cvetko Mitrovski are with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mails: mitko.kostov@uklo.edu.mk, cvetko.mitrovski@uklo.edu.mk

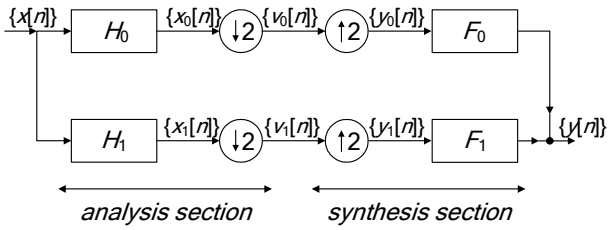


Fig.1. Two-channel filter bank.

II. DESIGN OF TWO-CHANNEL QMF BANKS

The basic block of a two-channel QMF bank consists four even length N linear phase FIR filters, H_0, H_1, F_0, F_1, with impulse responses $h_0[n]$, $h_1[n]$, $f_0[n]$, $f_1[n]$, respectively, out of which H_0, $H_1$are low pass while F_0, F_1 are high pass filters as shown in Fig. 1.

The most commonly used filters for this realization of the filter bank, are the quadrature mirror filters H_0, H_1, F_0, F_1 which satisfy the following property:

$$H_1(e^{j\omega}) = H_0(e^{j(\omega-\pi)}),$$

$$F_0(e^{j\omega}) = H_0(e^{j\omega}),\ F_1(e^{j\omega}) = H_1(e^{j\omega}). \qquad (1)$$

Due to this property, the design problem is simplified, since the coefficients of all the filters are obtained from the lowpass prototype filter coefficients.

In case of PR, the filter coefficients are determined by minimization of the error function between the output and the input of the bank, $e(n) = u_{in}(n) - u_{out}(n)$, in the least-square sense. In our work we propose design of NPR-QMF filter, obtained by minimization of the following error function

$$E = E_r + \alpha E_s, \qquad (2)$$

subject to the filter length, N, and their cut–of frequency ω_s, where: $E_r = \int_0^{\pi} W(\omega)[e_r(\omega)]^2 d\omega$, $E_s = \int_{\omega_s}^{\pi} \left|H_0(e^{j\omega})\right|^2 d\omega$, $e_r(\omega) = \left|H_0(e^{j\omega})\right|^2 + \left|H_0(e^{j(\omega+\pi)})\right|^2 - 1$, $W(\omega)$ is a properly chosen weighting function and α is the relative weight between the measure of the total error signal energy, E_r, and the measure of the error signal energy in the stop band, E_s.

TABLE I
PROTOTYPE FILTER COEFFICIENTS FROM A NPR-QMF BANK

$h_0[0-5]$ =	–0.0095	0.0403	–0.0126	–0.1303
$h_0[11-6]$	0.1407	0.6786		

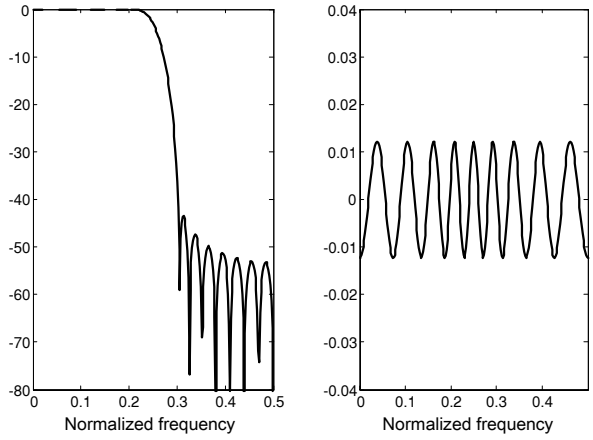


Fig 2. The magnitude responses of the prototype filter and the overall reconstruction error.

III. DISCRETE WAVELET TRANSFORM

In series expansion of discrete-time function f,

$$f(t) = \sum_{j=1}^{J} \sum_{k=1}^{2^{-j}M} d_{jk}\psi_{jk}(t) + \sum_{k=1}^{2^{-J}M} c_{Jk}\phi_{Jk}(t), \qquad (3)$$

ψ_{jk} and ϕ_{jk} denote wavelet and scaling functions

$$\psi_{jk}(t) = 2^{-k/2}\psi(2^{-k}t - n), \qquad (4)$$

a_{Jk} and d_{jk} are approximation and detail coefficients, and j and k are dilatation and translation indexes, respectively. The mother and scaling functions are defined by:

$$\psi(t) = \sum_n 2^{1/2} h_1 \psi(2t - n) \qquad (5)$$

$$\phi(t) = \sum_n 2^{1/2} h_0 \phi(2t - n) \qquad (6)$$

The estimation of d_{jk} and c_{Jk} is carried out through an iterative decomposition algorithm, which uses two QMF filters, h_0 and h_1.

IV. NPR-QMF DESIGN

The idea and motivation for using a NPR-QMF bank as a replacement of a PR-QMF bank in combination with the wavelet shrinkage denoising technique is based on the fact that in both cases the inherent nonlinearities of the wavelet filtrating technique causes reconstruction error of the input signal.

In this section we are presenting characteristics of the NPR-QMF filter designed by using the proposed algorithm in section II, with the design parameters: filter length, $N = 12$ and its stop band cut off frequency, $f_s = 0.6\pi$.

The coefficients of designed low pass filter, h_0, are given in Table I, while its frequency response is given in Fig. 2a. This filter has linear phase, good pass-band and transition band characteristics, but it produces certain overall reconstruction error, given in Fig. 2b.

TABLE II
COMPARISON OF THE SLOPES OF THE TRANSITION BANDS OF THE MOST FREQUENTLY USED FILTERS WITH NPR-QMF FILTERS

	$\omega_p(\times\pi)$ (1)	$\omega_s(\times\pi)$ (2)	$a(\omega_p)$ (3)	$a(\omega_s)$ (4)	slope (dB/rad/s) (5=(4–3)/(2–1))
NPR QMF (N=32)	0.50	0.60	–2.99	–33.23	**–98.54**
NPR QMF (N=12)	0.49	0.66	–2.93	–33.70	–58.99
Db3	0.50	0.88	–2.96	–34.89	–26.28
Db6	0.50	0.79	–3.01	–34.89	–34.54
sym5	0.50	0.82	–2.95	–34.81	–31.86
sym6	0.50	0.79	–3.01	–34.80	–34.54
coif4	0.50	0.75	–3.01	–34.57	–39.87
coif5	0.50	0.73	–3.01	–34.95	–44.49

V. EXPERIMENTAL RESULTS

In this section, we compare the efficiency of the NPR-QMF filter design versus PR-wavelet filter design, on basis of their frequency characteristics, and the efficiency of the both designs in signal denoising by wavelet shrinkage technique.

The NPR-QMF filters are with filter lengths: $N = 12$ and $N = 32$ and stop band cut-off frequency equal to 0.6π

The results of the comparison among the NPR-QMF filters and the PR wavelet filters are expressed in Table II, in terms of their transition band slopes calculated as a slope of their frequency responses among the cut-off frequency and the frequency for which the amplitude response is 35 dB lower than in the pass-band. The advantage of the NPR-QMF is obvious. Further more, NPR-QMF filters have better pass band and stop band characteristics versus biorthogonal 9/7 and Daubechies wavelet filters in terms of their magnitude responses as illustrated in Fig. 3 and Fig. 4. From those figures and from Table II, it is obvious that NPR-QMF filters have better magnitude responses than both the Daubechies and Simlet's family wavelet filters and they do have linear phase while the wavelet filters do not have.

Comparing NPR-QMF filters and the filter that corresponds to the Meyer wavelet, the second is only approximation of FIR filter. Fig. 4b shows its impulse response.

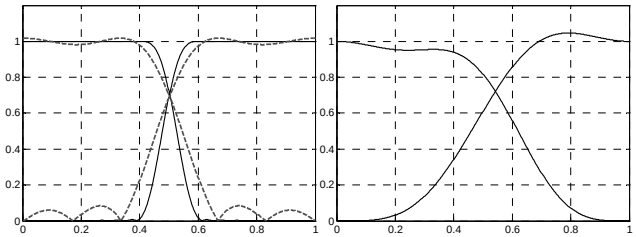


Fig 3. Comparison of magnitude responses of NPR-QMF and biorthogonal filters: (a) NPR-QMF with $\omega_s = 0.6\pi$ and filter lengths $N = 12$ and 32; (b) biorthogonal 9/7 filters.

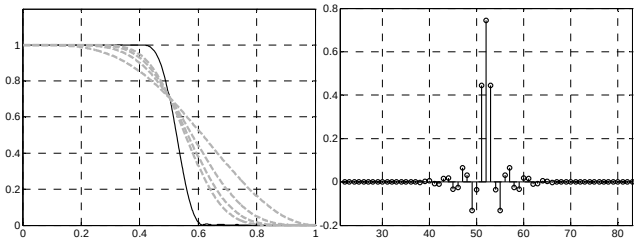


Fig. 4. (a) Magnitude responses of NPR-QMF bank filters ($N = 32$, $\omega_s = 0.6\pi$) (black full line) versus responses of filters that correspond to Daubechies family (db2, db4, db6, db8) (grey lines); b) impulse response of the corresponding filter to the Meyer wavelet.

The frequency responses of the filters that correspond to wavelets from Daubechies and Simlets families, have more zeros at π and the wavelet functions have more vanishing moments (the number is equal to the half of the filter length). Hence, by using such wavelet functions, the signal power is concentrated in small number of wavelet coefficients, which should be kept non-modified by the shrinkage procedure. This stands if the signal is smooth. More zeros at π means that smother signals are better approximated.

In order quantitatively to compare the denoising capabilities of the designed filters we use 1D and 2D test signals shown in Fig. 5-left. The 1D test signal was obtained by adding a Poisson noise to the corresponding artificially generated signal (right), while the 2D test signals (right), were generated from the corresponding phantom images (left) by summing a bell shaped random 2D signals with Gaussian distribution around each pixel, with total intensity proportional to the pixel intensity on the original phantom image. Since the obtained test images have significantly less energy then the original phantom images, we normalized them before the denoising test.

All test signals are denoised by using both: the NPR-QMF bank filter (Table I) and different types of PR wavelets, and the results were compared with the original signals in the average mean square error (MSE) sense.

From the results shown in Figs. 6-9, it is obvious that NPR-QMF banks are concurrent to the PR wavelet filters both for 1D and 2D signal denoising by using wavelet shrinkage technique.

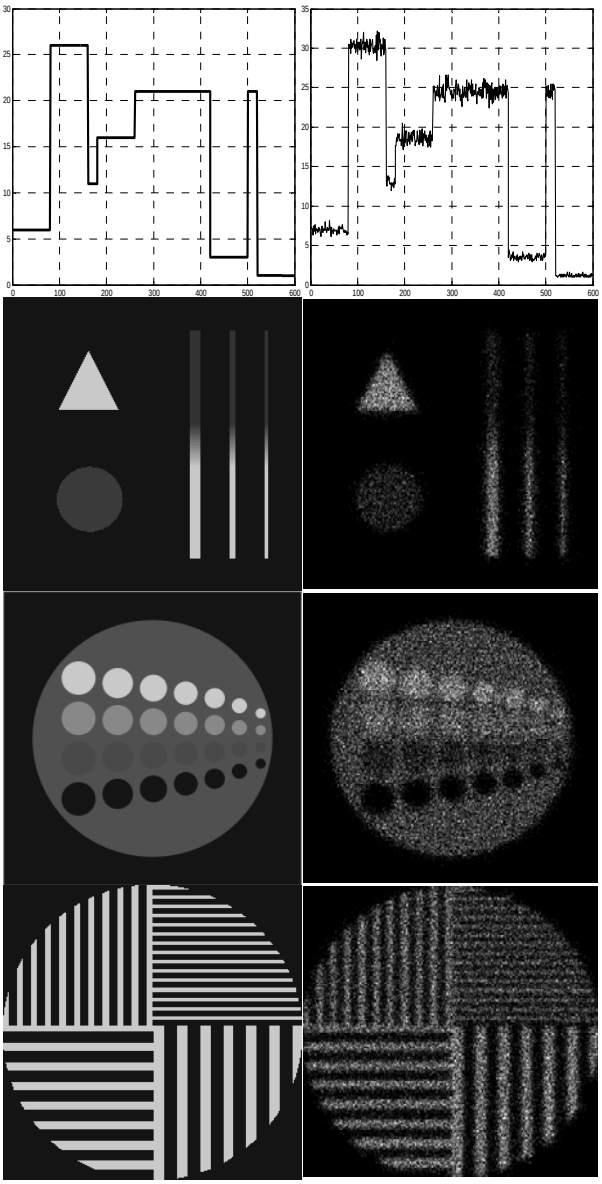


Fig. 5. Deterministic signals and their noisy versions: 1D signal, Phantom, Circles, Bars.

VI. CONCLUSION

In this paper we compare the efficiency of the NPR-QMF filter design versus PR-wavelet filters on basis of test signal denoising. The results show that NPR-QMF banks, designed by choosing low values for α ($\alpha<10^{-2}$), are competitive to the PR wavelets, in terms of design flexibilities and their efficiency on denoising both 1D and 2D signals, by using wavelet shrinkage technique.

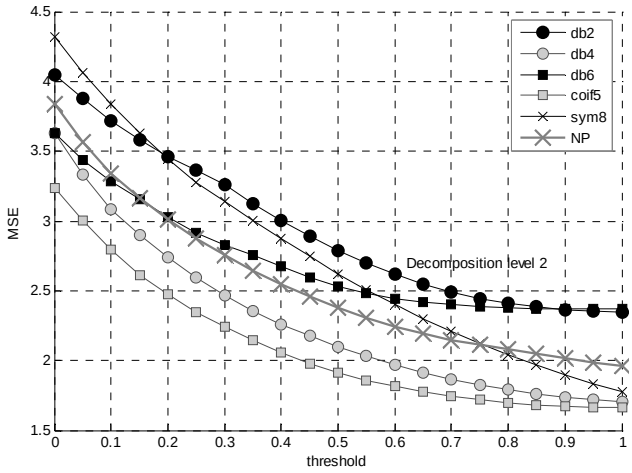


Fig. 6. Comparison of different wavelet filters for the 1D signal and different threshold values.

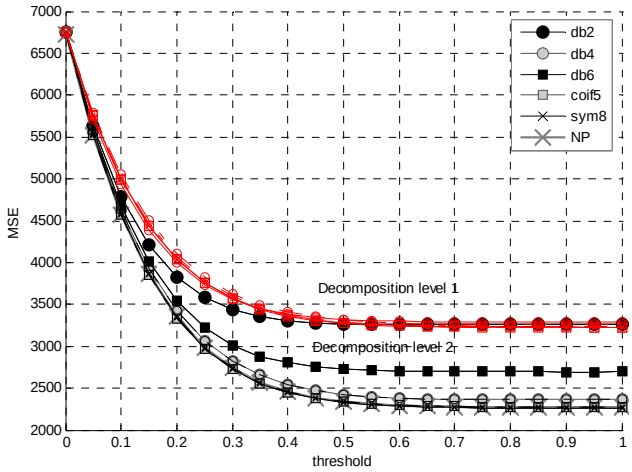


Fig. 7. Comparison of different wavelet filters for the image Bars and different threshold values.

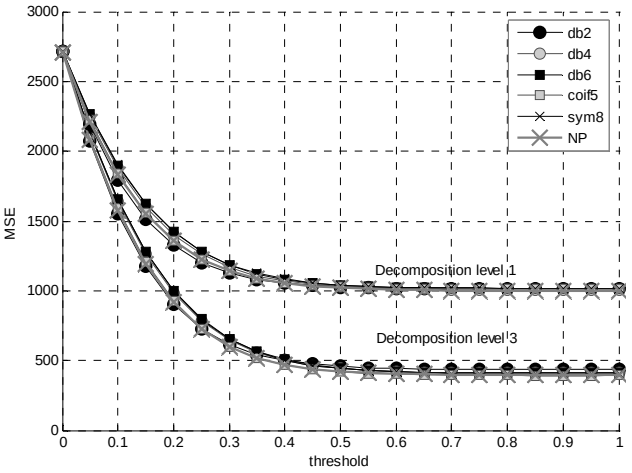


Fig. 8. Comparison of different wavelet filters for the image Circles and different threshold values.

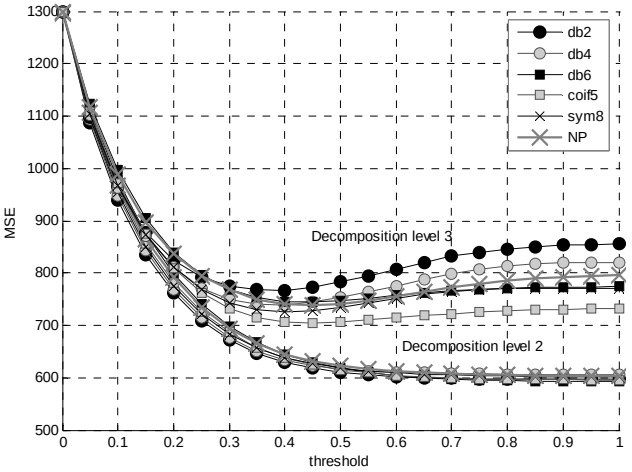


Fig. 9. Comparison of different wavelet filters for the image Phantom and different threshold values.

REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation via Wavelet Thresholding", *Biometrika*, vol. 81, pp. 425-455, 1994;

[2] Y. Xu, J. B. Weaver, D. M. Healy, Jr., and J. Lu "Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique", *IEEE Trans. on Image Processing*, vol. 3, no. 6, pp. 747-758, Nov. 1994;

[3] Robert D. Nowak, Richard G. Baraniuk, "Wavelet-Domain Filtering for Photon Imaging Systems", *IEEE Trans. Image Processing*, vol. 8, Iss. 5, p. 666-678, May 1999;

[4] S. Zhong, V. Cherkassky, "Image Denoising using Wavelet Thresholding and Statistical Learning Theory", *IEEE Trans. on Image Processing*, Feb. 1999.

[5] S.G. Chang, B. Yu, M. Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression", *IEEE Trans. Image Processing*, vol. 9, pp. 1532-1546, Sept. 2000;

[6] S. Zhong, V. Cherkassky, "Image Denoising using Wavelet Thresholding and Model Selection", *Proceedings of Int. Conf. on Image Processing*, vol. 3, pp. 262-265, Vancouver, BC, Canada, Sept. 2000.

[7] L. Sendur, I.W. Selesnick, "Bivariate shrinkage with local variance estimation", *IEEE Signal Process. Letters*, vol. 9, no. 12, pp. 438–441, Dec. 2002;

[8] P. Bao and L. Zhang "Noise Reduction for Magnetic Resonance Images via Adaptive Multiscale Products Thresholding", *IEEE Trans. on Medical Processing*, vol. 22, no. 9, pp. 1089-1099, Sept. 2003;

[9] J. Ge, G. Mirchandani "Softening the Multiscale Product Method for Adaptive Noise Reduction", *Thirty-Seventh Asilomar Conf. on Signals, Systems and Computers*, vol.2, pp. 2124-2128, 9-12 Nov. 2003;

[10] E. J. Balster, Y. F. Zheng and R. L. Ewing "Feature-Based Wavelet Shrinkage Algorithm for Image Denoising", *IEEE Trans. on Image Processing*, vol. 14, no. 12, pp. 2024-2039, Dec. 2005;

[11] A. Pižurica, W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising", *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 645–665, Mar. 2006;

[12] F. Luisier, T. Blu, M. Unser, "A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding", *IEEE Trans. on Image Proc.* vol. 16, pp. 593-606, Mar. 2007.

Architecture of Communication System for Distance Learning of Deaf People

Rumen P. Mironov[1] and Roumen K. Kountchev[2]

Abstract - **Architecture of communication system for distance learning of deaf people is presented. The basic function is to combine standard video information with sign language information, received by speaker. Fundamental block in the system is a special contour detector, which is using for detection and compression of movements of the translator.**

Keywords – **digital image processing, software architectures, communication systems, sign language.**

I. Introduction

Video communication systems for deaf people are limited in terms of quality and performance. The analysis of the visual attention mechanisms for sign language may enable optimization of video coding, transmission, reconstruction and visualization systems for deaf users. Sign language is a complex combination of facial expressions, mouth/lip shapes, hand and body movements, and finger spelling [1].

Visual perception of sign language video requires sufficient spatial and temporal resolution to capture the detailed movements of the sign interpreter. Reasonable visual quality and frame rates can be obtained using contour image compression for video transmission by the Internet Communication System for Distance Learning of Deaf People (CSDLDP), developed in the Video Communication Laboratory, Technical University of Sofia.

The intelligent contours extraction and transmission permit to obtain very efficient lossless compression with high compression ratio and good sign language comprehensibility [2], which is a base for the development of various application tools for the mobile communications and distance learning.

In this paper architecture of the developed Internet communication system for distance learning of deaf people is presented. The basic idea is to use special contour detector, developed in [2], for movements of the sign language translator, make high efficiency compression of extracted objects [3], combine this information with standard video information to present and transmit over Internet [4]. The system is used in science research by the contract VU-MI-104 "Transmission of video information over Internet for distance learning of deaf people".

[1]Rumen P. Mironov is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridsky Str. 8, 1000 Sofia, Bulgaria, E-mail: rmironov@tu-sofia.bg.

[2]Roumen K. Kountchev is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridsky 8, 1000 Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg.

II. Architecture of the System

The architecture of the developed CSDLDP system is presented on Fig.1. It is organized on modular principle and includes the following blocks, worked on the base of the well-known [5], [6], [7] and the new developed methods [8], [9]:

- *Video coder*. In this module functions for compression of real video sequences with different rates and qualities are included.
- *Audio coder*. In this module functions for compression of real audio sequences with different rates and qualities are included. It contain also functions for transferring of audio and sound frequency band into the several frequency strips, which can be understand by the people with partial defection of the hearing.
- *Contour detector*. In this module functions for contours detection of movements of sign language translator are included [2].
- *Contour coder*. In this module functions for high efficiency compression of contour sequences, developed in [3], are included.
- *Video server*. In this server video or TV information can be accepted, managed, processed and distributed over the Internet. He also contains functions for video encryption and saving in special video presentation database. There are interface components for interactive distance learning of students, with or without teachers.
- *Video database*. This is multimedia database in which sign language video sequences, video clips, audio, presentations and text documents are saved [10],[11]. The DBMS can be MS SQL Server 2008, Oracle 11, MySQL 5 or other SQL oriented system.
- *Video streaming server*. This is a special media server for packaging and transmission of presented multimedia data flow over Internet.

The first four modules (video coder, audio coder, contour coder and detector) are incorporated in a new AVC Coder subsystem, which is used to combine the different compressed multimedia data flows into one multimedia audio-video format (MAVF) and transmit it to the video streaming server.

In the receiver side the functional blocks are the same. The video streamers are used for receiving Internet data and unpacking it and transmit to the decoders. The AVC Decoder subsystem, in which are incorporated video decoder, audio decoder and contour decoder, is used to separate the multimedia data flows into audio, video and contours, decompress audio, video and contour data and transmit the developed presentations to the computer classes, presentation halls or individual consumers.

The CSDLDP system is realizing as object-oriented software modules on C++ for operational systems Windows 98/2000/XP/Vista.

For real-time processing and transmission the hardware realization of all sender modules is necessary, which can be accomplished by the time-saving servers, specialized video controllers with TV tuner boards and signal processors from the Texas Instruments, like DSP DM6446 DaVinci.

III. Experimental Results

One of the most complicated blocks in the presented system is the contour detector of sign language interpreter. Using two contour frames from video film of sign language interpreter for deaf people, the visual quality analysis of contour extraction is made. The basic frames of size 320×240 and 24 bpp in BMP format are shown on Fig.2 and represent frame 37 and frame 44 of the signed sentence called "At each other's throat".

Fig. 2. Basic images ateachot37.bmp and ateachot44.bmp (320x240 pixels, 8 bits)

On Fig.3 contour images, obtained from the basic ones by the developed algorithms for contour extraction of size 320×240 and 8 bpp in BMP format are shown.

Fig. 3. Input contour images ateachot37c.bmp and ateachot44c.bmp

Then the extracted frames can be compressed separately by the high efficiency lossless compression algorithm, with compression ratio more then 100, as is shown in [2]. The visual quality of the moving contours is excellently and they are completely enough for understanding of sign language interpreter.

For the other video components of multimedia data standard compression techniques as H.264/AVC can be used. The audio decomposition and compression is in development now.

The efficiency of the developed multimedia database, which application programming interface (API) is shown on Fig.4, is testing by 8 clients for 5 video clips.

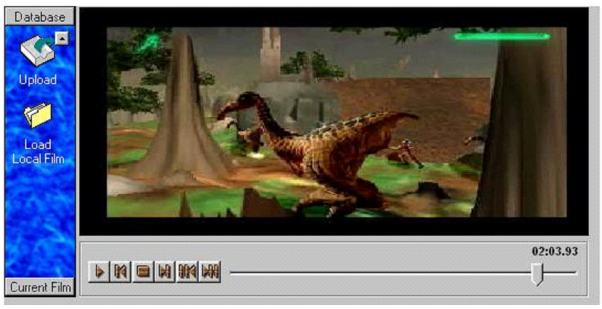


Fig. 4. Multimedia database API

The network tests for video transfer from/to the server with 10 Mb/s and 100 Mb/s in Intranet set are done and the measured times in seconds are given in Tabl.1.

TABLE I
DATABASE PERFORMANSE

Name of video clips	Size of video clips (bytes)	Transfer to the server (sec)	Transfer from the server (sec)
		(10 Mbit/s)	
Transmit.avi	34816	33.71	7.41
Test66.mpg	323761	129.60	12.25
IcVideo14.mpeg	563419	273.71	18.57
Model.avi	1828354	1106.41	79.26
IcVideo43.mpeg	8556548		
Name of video clips	**Size of video clips (bytes)**	**Transfer to the server (sec)**	**Transfer from the server (sec)**
		(100 Mbit/s)	
Transmit.avi	34816	2.57	0.57
Test66.mpg	323761	10.76	0.98
IcVideo14.mpeg	563419	22.17	1.51
Model.avi	1828354	105.41	7.55
IcVideo43.mpeg	8556548	1063.21	180.65

During the transfer time of video data to the server with 10 Mbit/s the CPU performance is average 15.7% and the virtual memory size is increasing from 100 MB to 140 MB. For transferring of video data from the server the CPU performance is average 9.4% and the virtual memory is increased from 100 MB to 105 MB. For the analogical tests with 100 Mbit/s the CPU performance is average 26.2% and for loading the CPU performance is decreasing about 0.8%. The virtual memory is the same as by the 10 Mbit/s tests. These results show that general factor for transfer time of video data is the performance of SQL components.

For testing of client-server performance an asynchronous test with 100 Mbit/s for 5 and 8 simultaneously connected clients is made. The CPU performance is 18.1% for 5 clients and 19.4% for 8 clients and the virtual memory is increased with about 8-9 MB. The received results are averaged for one client and shows increasing of requesting time with about 4%.

We have run also films with large size in order to estimate the developed MM Browser performance. The CPU Usage (in %) for data stream decompression and the used RAM (in Kbytes) are estimated. A comparison is performed with some

software tools widespread on the market. The results are shown in Table II. The experimental files own MPEG, QT and AVI formats (OutCast.mpeg - 38.2 MB, Menace_480.mov - 24.8 MB, ICVIDEO27.AVI - 4.10 MB).

TABLE II
COMPARISION OF MM BROUSER

Program Name	OutCast.mpeg		Menace480.mov		ICVIDEO27.AVI	
	CPU	MEM	CPU	MEM	CPU	MEM
Xing Player	86%	6085	-	-	-	-
QT Movie Player	-	-	53%	8870	-	-
MS Media Player	88%	7496	65%	8432	18%	7532
MM Browser	98%	8860	75%	9483	51%	10606

The developed multimedia application for dealing with MM Data Base through Internet/Intranet allows the following options: maintenance of centralized database for films; saving the widespread video formats – AVI, QuickTime and MPEG; maintenance of extended information about the downloaded films, as well as easy access to them; downloading and playing films from files; Internet/Intranet connection for downloading and saving films from/to database; searching of film using the title; saving the current played film into the local hard disk.

Our application need higher CPU and MEM Usage since it works simultaneously with video data, situated on the local hard disk and video data located on SQL server. Note also that Xing Player and QT Movie Player don't support some of the basic video formats: MOV and AVI; MPEG and AVI respectively (see Table II).

IV. CONCLUSION

The developed architecture has the following advantages:

- communication over the standard IP network;
- module design, based on multimedia processing tasks;
- possibilities to use in presentation and interactive distance learning mode;
- using of advanced image and video processing algorithms for decomposition and compression;
- large possibilities for upgrades and improvements.

The developed Internet communication software system can be used in different areas for distance learning not only of deaf people, but for videoconferences, science research and etc. The system will allow improvements of human life quality and expand the possibilities for learning in different areas - quantitative microscopy, analysis of biomedical images, biotechnologies, robotics, ecological monitoring, visual control in the industry, economics, medicine, science researches and etc.

ACKNOWLEDGEMENT

The authors thank the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science for the financial support by the contract VU-MI-104/2006.

REFERENCES

[1] D. Agrafiotisa, N. Canagarajaha, D.R. Bulla, J. Kyleb, H. Seersb, M. Dyec. "A perceptually optimised video coding system for sign language communication at low bit rates". *Signal Processing: Image Communication*. vol. 21, No. 7, pp. 531-549, 2006.

[2] R. Kountchev, S. Rubin, M. Milanova, Vl. Todorov, R. Kountcheva. "Extraction of Objects Contours for Body Language Study". *The 2008 IEEE Intern. Conference on Information Reuse and Integration (IEEE IRI'08)*, Las Vegas, USA, pp. 220-225, July 13-15, 2008.

[3] A. Cavender, D.K. Barney, R.E. Ladner, R. Vanam, E.A. Riskin. "Mobile ASL: Intelligibility of Sign Language Video as Constrained by Mobile Phone Technology". *Disability and Rehabilitation: Assistive Technologies*, Special Issue ASSETS 2006, vol. 3, pp. 93-105, 2008.

[4] ITU-T, Series H: Application profile – Sign language and lip-reading real-time conversation using low bit-rate video communication. *CCITT Recommendations.* Supplement 1, 05, 1999.

[5] Maher A., Sid-Ahmed, *Image Processing: Theory, Algorithms, and Architectures*. McGaw-Hill, Inc., 1995.

[6] B. Jähne. *Practical Handbook on Image Processing for Scientific and Technical Applications*. 2nd Ed., CRC Press, 2004.

[7] W. K. Pratt. *Digital Image Processing*. 4th Ed., John Wiley & Sons, 2007.

[8] C. Ware. *Information Visualization – Perception for Design*. 2nd Ed., Morgan Kaufmann, 2004.

[9] R. Kountchev, R. Mironov. "System for Image Processing, Analysis and Recognition". *Journal for Automatic and Informatic*, Sofia, Number 4, pp.41-44, 1996, (in Bulgarian).

[10] A. Mironov, R. Mironov. "Using database for presentation of multimedia information". *National Conference with foreign participation "Evolution of Telecommunications Networks and Systems"*, TELECOM'99, October 26-28, pp.642-649, Varna, Bulgaria.(in Bulgarian).

[11] R. Mironov, N. Sirakov, F. Muge. "An Architecture of Virtual Multimedia Library". *Proc. of V Ibero – American Symposium on Pattern Recognition*, SIARP'2000, Lisbon, pp.103-111, September 11-13, 2000.

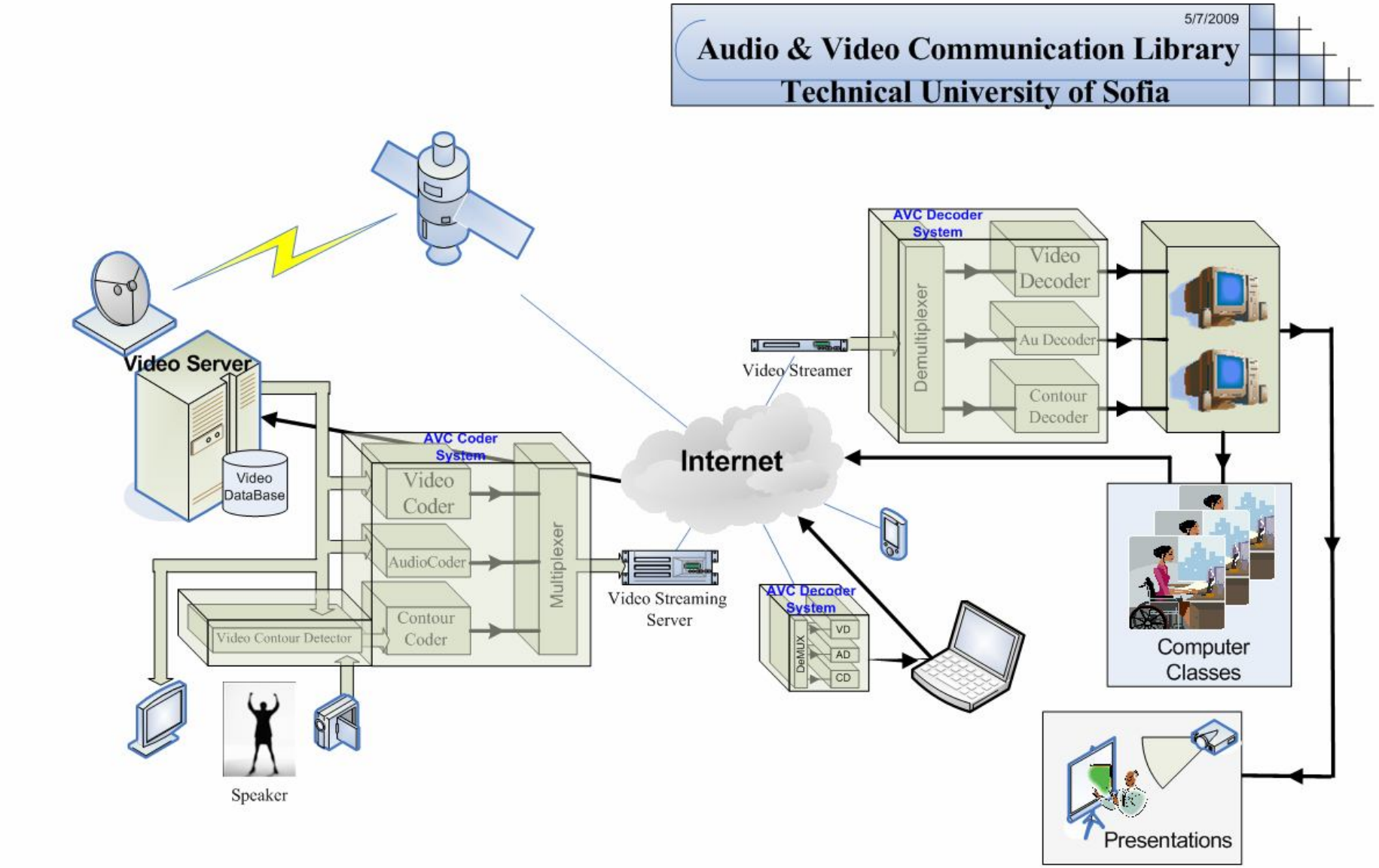


Fig. 1. Architecture of Internet Communication System for Distance Learning of Deaf People

Architecture of Image Processing System for Documents

Rumen P. Mironov[1] and Roumen K. Kountchev[2]

Abstract – **Architecture for document image processing, analysis and archiving is presented. The basic functions include input/output of halftone images, pre- and post-processing, filtration, compression, enhancement, 2D transformations and interpolations.**

Keywords – **image processing, software architectures, document archiving.**

I. Introduction

In connection with the fast growing of advanced calculation techniques the systems for image processing on the base of personal computers are standing topical [1], [2]. They are built modularly and include additional frame grabbers for input and output of digital images, modules for compression, transform, filtration, analysis and visualization. In [3], [4] and [5] some image processing systems are described. All of them include basic software modules for image processing [6] and hardware accelerators for increasing of computation operations speed. They are developed on the basis of common requirements, defined in [7]:

- working on standard computer platforms with different operation systems;
- working with different input/output devices;
- working in interactive mode;
- using of internal system program language;
- using of advanced image processing algorithms;
- large possibilities for upgrades and improvements.

The image processing algorithms in these systems are included as standard libraries and each of them can be used separately by the user or supervisor. It is a function-oriented approach, which isn't suitable for imitation of human visual system working.

In this paper architecture for document image processing, analysis and archiving is presented. This architecture is task-oriented and shows the complex image processing flows, as the natural human visual system is operating. On this base in the Laboratory of Video Communications and Technologies, (Faculty of Telecommunications) an object-oriented software image processing system is developed, which is used in the research work for the different science projects.

II. Architecture of Image Processing System

On Fig.1 a block-scheme of the developed architecture for document image processing system is presented. This architecture is organized as a set of the following image processing units, based on well-known and new developed algorithms [8], [9], [10], [11]:

- *Image Loading Unit (ILoadU)*. In this module functions for loading and converting of compressed and uncompressed basic image file formats (TIFF, BMP, JPEG) and the developed in [11] image compressed file format are included. *ILoadU* also contains functions for image decryption and loading from special image documents database.
- *Image Inputting Unit (IInptU)*. In this module functions for input of document images (text, photos, tables, graphics) from scanners, photo and cameras are included.
- *Image Saving Unit (ISaveU)*. In this module functions for saving of compressed and uncompressed basic image file formats (TIFF, BMP, JPEG) and the developed in [11] image compressed file format are included. *ISaveU* also contain functions for image encryption and saving in special image documents database.
- *Image Pre-processing Unit (IPrepU)*. In this module the following functions for image pre-processing are included – functions for arithmetic and geometric operations, table operations, contrast enhancements, histogram operations, linear and nonlinear noise filtration. They are used for image quality improvement or converting in a form better suited for analysis by human or a machine.
- *Image Presentation Unit (IPresU)*. In this module different image presentation models (positioning, quad-tree, pyramidal, structural) and functions for image converting are included.
- *Image Compression Unit (ICmprU)*. In this module image compression functions, based on classical algorithms for run-length, Huffman, LZW, arithmetic, scalar and vector quantization, JPEG and a new algorithm for lossless compression [12] are included.
- *Image Geometrical Transform Unit (IGeoTU)*. In this module functions for translations, scaling, interpolations, rotations, affine and perspective transformations, spatial warping and geometrical resampling are included.
- *Image Linear Transform Unit (ILinTU)*. In this module functions for 2D linear superpositions and convolutions, correlations, discrete cosine, sine, Hadamard, complex Walsh-Hadamard, Fourier and Karhunen-Loev transforms are included.
- *Image Pseudo-Color Transformation Unit (IPCTU)*. In this module functions for pseudo-color transforms in spatial and frequency area and image halftoning by adaptive and non-adaptive error diffusion are included.

[1]Rumen P. Mironov is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridsky Str. 8, 1000 Sofia, Bulgaria, E-mail: rmironov@tu-sofia.bg.

[2]Roumen K. Kountchev is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridsky 8, 1000 Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg.

- *Image Analysis Unit (IAnlyU)*. In this module functions for image segmentation (contour, brightness, texture) morphological image processing (binary, gray), edge detection and feature extractions are included.
- *Image Post-Processing Unit (IPostU)*. In this module functions for adaptive and non-adaptive, linear and non-linear filtration of gray level or binary images are included.
- *Image Visualization Unit (IVizuU)*. In this module functions and drivers for visualization of document images are included.
- *Image Printing Unit (IPrntU)*. In this module functions and drivers for printing of documents are included.

For properly working of the developed software system the following additional modules are necessary:

- *System Supervisor*. In this module the basic functions for system control, diagnostic and interaction between separated components and operation system are included. In the system supervisor analysis and processing of system error messages and receiving of context help information are also achieved.
- *Graphical User Interface (GUI)*. In this module graphical oriented system with menus, dialogs, windows, icons, buttons, fonts and etc. is included, which can be used for dialog input and output of parameters or images, interactive processing and analysis, visualization of documents, graphical presentation of the results from the analyses.
- *Communication Module (CM)*. In this module functions for connection between the *System Supervisor* and the external modules and functions of system interpreter are included. The usage of *Communication Module* facilitate the building of processing tasks – also including new ones, which are not defined in the system. This is a way to create *External Algorithms*, established by the operator.
- *System Peripheral Module*. In this module the system peripheral drivers are situated.

Te system supports two types of data objects: image data objects and image-related, non-image data objects. A system image data object is a multi-dimensional collection of pixels, whose structure is:

- horizontal space index ($0 \le x \le 1024$);
- vertical space index ($0 \le y \le 1024$);
- depth space index ($0 \le z \le 1024$);
- temporal index ($0 \le t \le T_{max}$);
- color or spectral band index (($0 \le fz \le F_{max}$);

The system utilizes the following pixel data types: Boolean, Unsigned Integer, Signed Integer, Real and Complex. The precision and data storage format of pixel data is implementation dependent.

It supports several image related, non-image data objects. These include:

- *Chain*: an identifier of a sequence of operators;
- *Composite identifier*: an identifier of a structure of image arrays, lists and records;
- *Histogram*: a construction of the counts of pixels with some particular amplitude value;
- *Lookup table*: a structure that contains pairs of entries in which the first entry is an input value to be matched and the second is an output value;
- *Matrix*: a two-dimensional array of elements that is used in vector-space algebra operations;
- *Neighbourhood array*: a multi-dimensional moving window associated with each pixel of an image;
- *Pixel record*: a sequence of across-band pixel values;
- *Region-of-interest*: a general mechanism for pixel-by-pixel processing selection;
- *Static array*: an identifier of the same dimension as an image to which it is related;
- *Value list*: a collection of pairs of elements in which the first element is a pixel coordinate and the second element is an image measurement.

III. Experimental Image Processing Tasks

The functionality of the human visual system can serve as a reliable guide for breaking up the complex image processing tasks. First, the optical system forms an image of the observed documents. A sensor converts this image into a form that is usable for digital processing with a computer system. The first processing step, denoted as low-level image processing, enhances, restores, or reconstructs the image formed. Further processing extracts features from the images that finally lead to the identification and classification of the objects in the images and can be saved in document database. In this way, the circle is closed, leading us from documents that are converted into images and processed back to their recognition and description.

On Fig.2 the basic graphical window of the developed system with one sample test grayscale image (with size 256x256, 8 Bpp on the left side), your histogram and the same image after segmentation with threshold 138 (on the right side) are shown.

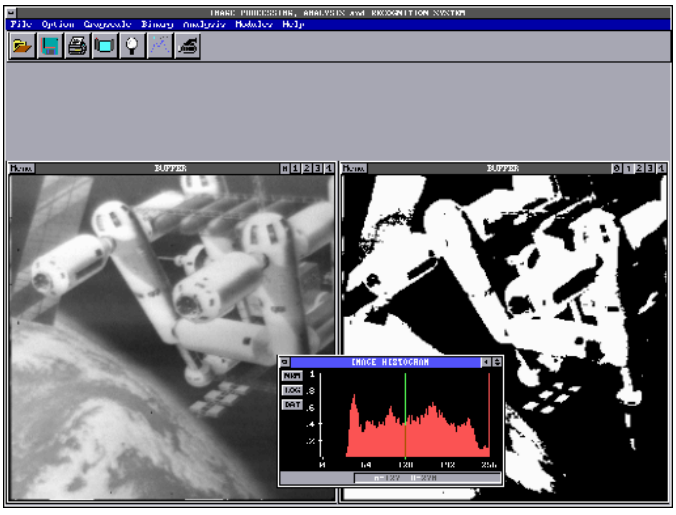


Fig. 2. Basic images ateachot37.bmp and ateachot44.bmp (320x240 pixels, 8 bits)

Using the developed system we can define various image processing tasks, which can be described semantically by the following way:

- *IInptU -> IPrepU -> IPresU -> IPostU -> ISaveU* . Inputting an image from scanner or digital photo camera, enhance image quality by pre-processing, presented in matrix form, post-processing and saving in the database. This semantic chain can be used for preliminary record of uncompressed documents in the database.
- *IInptU -> IPrepU -> IPresU -> ICmprU -> ISaveU* . Inputting an image from scanner or digital photo camera, enhance image quality by pre-processing, presentation in matrix form, compression and saving in the database. This semantic chain can be used for preliminary record of compressed documents in the database.
- *ILoadU -> IPrepU -> IPresU -> IPostU -> IVizuU* . Loading an image from file or database, enhance image quality by pre-processing, presentation in matrix form, post-processing and visualization on the screen. This semantic chain can be used for visualization of saved documents in the database.
- *ILoadU -> IPrepU -> IPresU -> IPostU -> IPrntU* . Loading an image from file or database, enhance image quality by pre-processing, presentation in matrix form, post-processing and printing on the laser or ink-jet printer. This semantic chain can be used for printing of saved documents in the database.

IV. CONCLUSION

The presented architecture, control modules, data objects and interfaces build new task oriented image processing system, which allow this system to imitate a natural human visual system. The developed architecture has the following advantages:

- module design, based on image processing tasks;
- possibilities for working in automatic and interactive mode;
- using of internal system program language;
- using of advanced image processing algorithms;
- large possibilities for upgrades and improvements.

The developed software system can be used in different areas – metallography, quantitative microscopy, analysis of biomedical images, biotechnologies, robotics, ecological monitoring, visual control in the industry, medicine, science researches and etc.

ACKNOWLEDGEMENT

The authors thank the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science for the financial support by the contract VU-I-305/2007.

REFERENCES

[1] Maher A., Sid-Ahmed, *Image Processing: Theory, Algorithms, and Architectures*. McGaw-Hill, Inc., 1995.

[2] P. Stucki. *Advances in Digital Image Processing*. N. J. Plenum Press, 1979.

[3] R. Kountchev, R. Mironov. "Program System for Metallographic Image Processing and Analysis". *XXX Science Session "Communications, Electronics and Computer Systems'95*, Sofia, May 1995, (in Bulgarian).

[4] R. Kountchev, R. Mironov. "System for Image Processing, Analysis and Recognition". *Journal for Automatic and Informatics*, Sofia, pp.41-44, Number 4, 1996 (in Bulgarian).

[5] R. Mironov, N. Sirakov, F. Muge. "An Architecture of Virtual Multimedia Library". *Proc. of V Ibero – American Symposium on Pattern Recognition*, SIARP'2000, Lisbon, pp.103-111, September 11-13, 2000.

[6] H.R. Myler, A.R.Weeks. *Computer Imaging Recipes in C*. Prentice Hall, Englewood Cliffs, N.J, 1993.

[7] Shi-Kuo Chang. *Principles of Pictorial Information Systems Design*. Prentice-Hall, 1989.

[8] C. Ware. *Information Visualization – Perception for Design*. 2nd Ed., Morgan Kaufmann, 2004.

[9] B. Jähne. *Practical Handbook on Image Processing for Scientific and Technical Applications*. 2nd Ed., CRC Press, 2004.

[10] W. K. Pratt. *Digital Image Processing*. 4th Ed., John Wiley & Sons, 2007.

[11] R. Kountchev, V. Todorov. "File Format organization for effective Still Image Transfer with IDP". *37th Intern. Scientific Conf. on Information, Communication and Energy Systems and Technologies*, Proc. Vol. 1, pp. 287-290, Nis, Yugoslavia, Oct., 2002.

[12] R. Kountchev, V. Todorov, R. Kountcheva. "Lossless Image compression with IDP". *Proc. of the Intern. Scientific Conf. on Information, Communication and Energy Systems and Technologies (ICEST'04)*, pp. 127-130, Bitola, Macedonia, 2004.

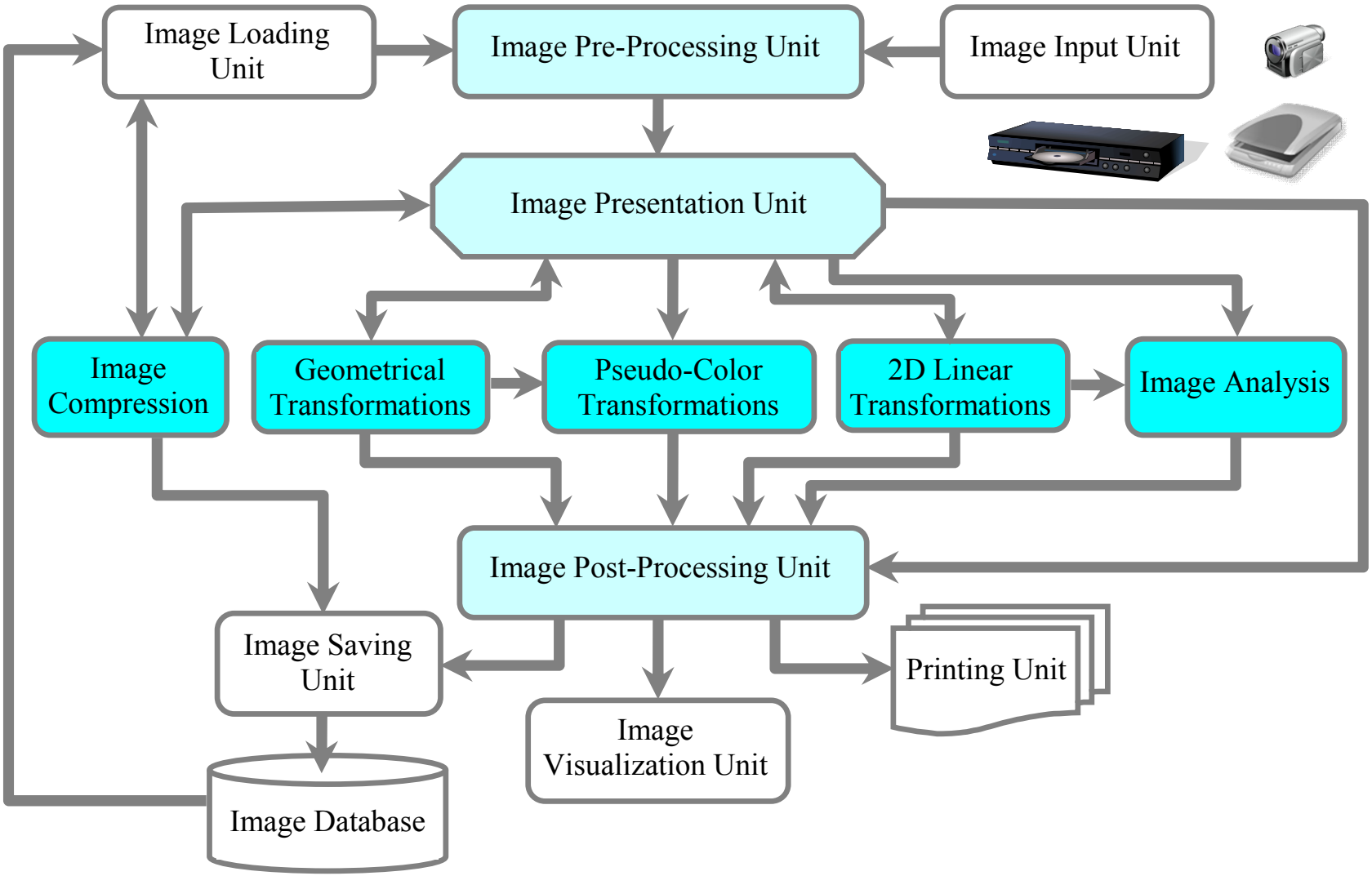


Fig.1. Architecture of Document Image Processing System

Document Protection Based on CHT Coefficients' Phase Modulation

Ivo R. Draganov[1] and Roumen K. Kountchev[2]

***Abstract* – In this paper we present a robust and efficient algorithm for document protection based on digital watermarking in phase domain using all complex coefficients of the two dimensional complex Hadamard transform (2D-CHT). High quality reconstruction and low visibility for the watermark is provided. Also broaden information capacity is guaranteed - up to ¾ of the number of pixels for the original image in bits. Detecting the watermark can be done only on the basis of a simple private/public key.**

***Keywords* – digital watermarking, document protection, complex Hadamard transform, phase modulation**

I. INTRODUCTION

Document protection is an important stage in many processing, transmitting and archival information systems – to detect unauthorized interference over electronic documents or usage without appropriate copyrights is of main importance here [1]. Digital watermarking provides the means for such protection – to prove copyright at one hand and to discover unwanted manipulation over the documents at the other.

Digital watermarking algorithms may be classified in three main groups as for the working domain in which some digital resource is being marked – space, frequency and phase [2]. Previous research suggest that watermarks introduced in space and frequency (amplitude alteration) often are robust enough but their visibility is high or the opposite – aiming low visibility leads to easy deletion of the watermark.

Our previous researches [3, 4] suggest that this contradiction may be suppressed by altering the phase spectrum of an image (or a sound signal). Most manipulations over images (attacks) affect least phases of the harmonics and as addition human eye (ear) is less sensitive to phase changes than to amplitude ones and the watermark is less visible.

To find appropriate phase spectrum a transform should be selected. We consider 2D-CHT as a proper choice since detailed examination of it had been done [5]. Initially we considered the possibility to mark only phases of a harmonics with amplitudes over a preliminary defined threshold but synchronism is easily lost when extracting the watermark so here enhanced form of the algorithm is presented using all 2D-CHT coefficients.

In part two we give detailed algorithm description along with in-depth analysis, in part three – some experimental results are given and then a conclusion is made in part four.

[1]Ivo R. Draganov is with the Faculty of Telecommunications, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: idraganov@abv.bg

[2]Roumen K. Kountchev is with the Faculty of Telecommunications, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg

II. ALGORITHM DESCRIPTION AND ANALYSIS

Inserting the watermark contains the following steps.

Step 1 – getting the input image, in color for the most general case in RGB color space, [*R(i,k), G(i,k), B(i,k)*], where $R,G,B = 0 \div 255$; $i = 0 \div M' - 1$ and $k = 0 \div N' - 1$. Conversion is made then to grayscale (or used by default):

$$I(i,k) = 0.30R(i,k) + 0.59G(i,k) + 0.11B(i,k), \qquad (1)$$

where $I(i,k) = 0 \div 255$.

Step 2 – extension the original image to size $M'' \times N''$ pixels ($M'' = w_M . 2^n$, $M'' = w_N . 2^n$, w_M, w_N – any positive whole number, $n \geq 2$, $M'' \geq M'$, $N'' \geq N'$, by symmetric repetition).

Step 3 – dividing the image into blocks with size NxN pixels where $N = 2^n$.

Step 4 – calculating the coefficients *s(u,v)* for each block using 2D-CHT [5]:

$$s(u,v) = \sum_{i=0}^{N-1}\sum_{k=0}^{N-1} L(i,k) e^{-j\frac{\pi}{2}(ui+vk)} t(u,i)t(v,k), \qquad (2)$$

where $t(p,q) = \begin{cases} 1, & for\ n = 2; \\ (-1)^{\sum_{r=3}^{n} \left\lfloor \frac{p}{2^{r-1}} \right\rfloor \left\lfloor \frac{q}{2^{r-1}} \right\rfloor}, & for\ n = 3, 4, \ldots \end{cases}$ is a sign function, $\left\lfloor \frac{a}{b} \right\rfloor$ - operator for extracting the whole part of a quotient, $u, v = 0, 1, \ldots, N\text{-}1$, $j = \sqrt{-1}$, $p, q = 0, 1, \ldots, 2^n\text{-}1$, $n = \log_2 N$.

Step 5 – calculating the modules and phases of the spectral coefficients:

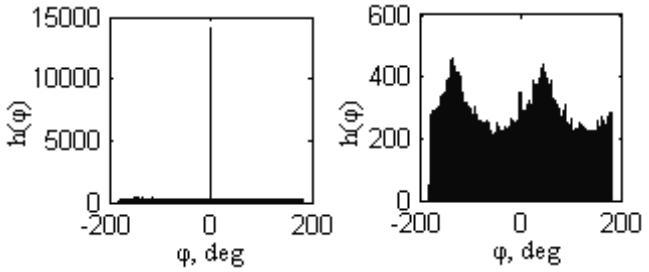
$$M(u,v) = \sqrt{[s_R(u,v)]^2 + [s_I(u,v)]^2}, \quad \varphi(u,v) = -arctg\left[\frac{s_I(u,v)}{s_R(u,v)}\right], \qquad (3)$$

where

$$s_R(u,v) = \sum_{i=0}^{N-1}\sum_{k=0}^{N-1} L(i,k)t(u,i)t(v,k)\cos[\tfrac{\pi}{2}(ui+vk)], \qquad (4)$$

$$s_I(u,v) = \sum_{i=0}^{N-1}\sum_{k=0}^{N-1} L(i,k)t(u,i)t(v,k)\sin[\tfrac{\pi}{2}(ui+vk)]. \qquad (5)$$

The conducted research on the distribution of φ (only for CHT complex coefficients) over the whole image reveals significant difference between natural scene and document images. This is confirmed by the presented averaged histogram $h(\varphi)$ in Fig.1.a for document images and in Fig.1.b – for natural scene images.



a) b)

Fig.1. Phase histograms $h(\varphi)$ for different types of images
a) document images b) natural images

Firstly all the distributions for document images appear very close – a huge peak for $\varphi = 0°$ persists and almost even spread for all other phases in the interval [-180°,+180°] with some small compared to the central peak maximums for $\varphi = -135°$ and 45°. As for the natural scene images nevertheless of their different content their distributions are also very close – the central peak here is absent, but clearly visible maximums for $\varphi = -135°$ and 45° are observed. These results can be explained by investigating (2) for the basic case of $n = 2$. The Hadamard coefficients (R_i for real – ¼$M''xN''$ in number and (C_i, C_i^*) for complex and conjugated complex ones – totally 3/4$M''xN''$) obtained from a block of 4x4 pixels are given in Fig.2.

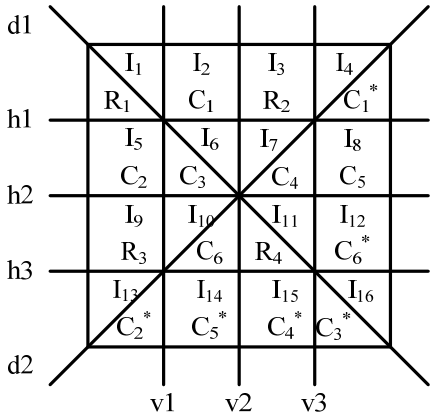


Fig.2. Complex Hadamard coefficients' location for block 4x4

The phase of the complex coefficient C_1 corresponding to $s(1,0)$ is calculated from (2)-(5):

$$\varphi(C_1) = \begin{cases} 0 & ,if \quad \sum_{s=2,6,10,14} I_s = \sum_{s=4,8,12,16} I_s \; and \; \sum_{s=1,5,9,13} I_s = \sum_{s=3,7,11,15} I_s \\ arctg\left(\sum_{s=2,6,10,14} I_s - \sum_{s=4,8,12,16} I_s \Big/ \sum_{s=1,5,9,13} I_s - \sum_{s=3,7,11,15} I_s \right) & - otherwise \end{cases} \quad (6)$$

It is obvious that for a homogenous block $\varphi(C_1) = 0°$, if a vertical transition $v1$ persists $\varphi(C_1) \approx -90°$ or 90° depending on the direction of the brightness gradient, for $v2 - \varphi(C_1) \approx -135°$ or 45°, for $v3 - \varphi(C_1) \approx 0°$ or 180°; for all horizontal transitions $h1$, $h2$, $h3$ $\varphi(C_1) = 0°$; for the diagonal ones $d1$ and $d2 - \varphi(C_1) \approx -135°$ or 45°. The analysis for the others $\varphi(C_i)$, for $i = 1 \div 6$, and for $n > 2$ is analogous.

Now we can explain the results from Fig.1. For document images vast homogenous areas appear and thus a strong peak for $\varphi = 0°$ is obtained. For natural scene images there is no preferable kind of transition (statistically) and because the number of $v1$, $h1$-$h3$, $d1$-$d2$ transitions is dominant over $v1$ and $v3$ maximums for $\varphi = -135°$ and 45° are formed. The latter exist in document images as well but considerably smaller than the central peak.

Step 6 – inserting the current bit w_r of the watermark p into the current phase for the coefficients $s(u,v)$ and $s^*(u,v)$ according to:

$$\varphi_w(u,v) = -\varphi_w^*(u,v) = \begin{cases} \varphi(u,v) + \Delta, & if \quad w_r(p) = 1; \\ \varphi(u,v) - \Delta, & if \quad w_r(p) = 0 \end{cases}, \quad (7)$$

for $r = 1, 2, \ldots, R$, where R is the number of the binary elements $w_r(p)$ of the watermark p and $\varphi_w(u,v)$ and $\varphi_w^*(u,v)$ are the phases of the marked coefficients $s_w(u,v)$ и $s_w^*(u,v)$. Parameter Δ is an angle defining the depth of the watermark and thus its robustness and visibility respectively.

The sequence w_r is obtained using XOR operation over each bit of the watermark and a bit from a randomly generated sequence describing private or public key depending on the cryptographic algorithm used. This sequence should have autocorrelation function close to delta impulse for higher reliability of the extracted watermark. If the current complex coefficient to be marked by phase has zero amplitude, the respective binary value of the watermark is omitted and only the value of the binary random sequence is used without applying XOR. Thus we reduce the probability for introducing errors in the recovered watermark later because in the opposite case the watermark value becomes inextricable – zero amplitude means zero phase even after change. This is highly important for document images where much more complex coefficients are zeros due to the presence of large homogenous areas. Also it is highly important to be no complex coefficient missed in this step because we may get the lack of synchronism at the stage of detecting or recovering the watermark – an improvement made to our previously developed algorithm [3].

Step 7 – calculating the pixel values $L_w(i,k)$ of the current marked block using the inverse two dimensional complex Hadamard transform (2D-ICHT):

$$L_w(i,k) = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} s_w(u,v)\, e^{j\frac{\pi}{2}(ui+vk)}\, t(u,i)\, t(v,k), \quad (8)$$

for $i, k = 0, 1, \ldots, N$-1, where $s_w(u,v) = M(u,v) e^{j\varphi_w(u,v)}$.

Extracting the watermark contains the following steps.

Step 1-5 – repeat the steps 1-5 from the insertion of the watermark over the watermarked image for estimating the coefficients $s(u,v)$.

Step 6 – checking for the presence of p-th watermark from D possible ones calculating the coefficient $C_{m,p}$ which defines the correlation between the p-th and the m-th watermark. p-th is the watermarked originally used for all the blocks of the input image. The cross-correlation coefficient is given by:

$$C_{m,p} = \sum_{r=1}^{R} [\varphi_r + \Delta_r(m)]\,\Delta_r(p) = A_p + B_{m,p}, \qquad (9)$$

for $m, p = 1, 2, \ldots, D$, where $[\varphi_r+\Delta_r(m)] = \varphi_{rw}(m)$ is the phase of the r-th marked coefficient $s_{rw}(u,v)$,

$$A_p = \sum_{r=1}^{R} \varphi_r \Delta_r(p), \quad B_{m,p} = \sum_{r=1}^{R} \Delta_r(m)\,\Delta_r(p), \qquad (10)$$

where φ_r is the absolute value of the phase for the r-th coefficient $s_r(u,v)$ before watermarking and:

$$\Delta_r(p) = (-1)^{w_r(p)}\Delta = \begin{cases} +\Delta & if \;\; w_r(p) = 1; \\ -\Delta & if \;\; w_r(p) = 0. \end{cases} \qquad (11)$$

Here $w_r(p)$ are the binary elements of the random sequence consisting of R bits describing the p-th ciphered watermark.

When R tends to larger values and supposing that $\varphi_r \approx$ const. (in a block 4x4 there may be only one transition present or a homogenous area which leads to this assumption) according to (10) follows:

$$A_p = \varphi_0 \sum_{r=1}^{R} \Delta_r(p) \approx 0. \qquad (12)$$

If all coefficients are non-marked then $C_{0,p} \equiv A_p \approx 0$. For the phase marked spectrum coefficients it is true that:

$$C_{m,p} \approx \begin{cases} \sum_{r=1}^{R} [\Delta_r(m)]^2 = R\Delta^2 & if \;\; m = p; \\ \sum_{r=1}^{R} \Delta_r(m)\Delta_r(p) \approx 0 & if \;\; m \neq p. \end{cases} \qquad (13)$$

Step 7 – making a decision for detecting p_0-th watermark according to the following rule:

$$p_0 = \begin{cases} detected, & if \quad (C_{m,p_0} / R\Delta^2) \geq \theta; \\ absent & - \quad otherwise, \end{cases} \qquad (14)$$

for $m, p_0 = 1, 2, \ldots, D$, where θ is preliminary chosen threshold in the interval $0 < \theta < 1$. The number of errors introduces (false alarms and missed watermark altogether) depend on θ's value.

The extraction of the watermark itself is done by using the original image. It is assumed that the copyright holder is only authorized to do this and thus has access to it. After finding the phase spectrums for the original and the watermarked image a simple subtraction of the respective phases is sufficient to restore the random sequence which then is XOR-ed with the private key to get the watermark.

III. Experimental Results

As experimental dataset we use 10 natural scenes (8 bpp, 512x512 pixels, 72 dpi) and 10 document (containing text, equations, graphics and tables) grayscale images (8 bpp, 2550x3300 pixels, 300 dpi). The watermark is 100x100 pixels binary image.

First of all, the results for natural scene and document images appeared to be very close, so here averaged values are given. In Fig.3 peak signal-to-noise ratio (PSNR) is given between watermarked and original images. It is seen that for wide range of values for Δ the watermarked images preserve high quality not letting PSNR to drop below 35 dB. The size of the window used represented by n has no influence here.

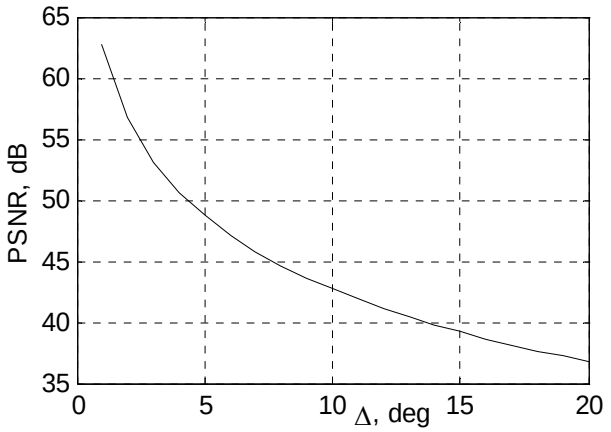


Fig.3. PSNR between watermarked and original images

In Fig.4 averaged results are given for the ratio between squared autocorrelation maximum when using the right private key and the average of the squared values when using other statistically independent random keys (200 in number) for the stage of watermark detection at different Δ. Even for small Δ around 5° this ratio is at least 2 which proves the robust detection for the algorithm proposed.

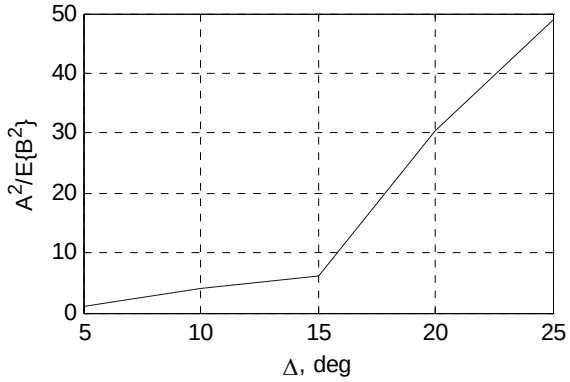


Fig.4. Autocorrelation maximum variation

In Fig.5 mean square error (MSE) is given between the original and extracted watermark when using typical value for Δ = 12° in the case of JPEG compressed (attacked, quality variation between 20 and 80 %) and non-compressed image (100% quality). Its almost constant value in the wide range of 20-80 % compression quality assures high resistance to high frequency attacks represented by the JPEG algorithm here.

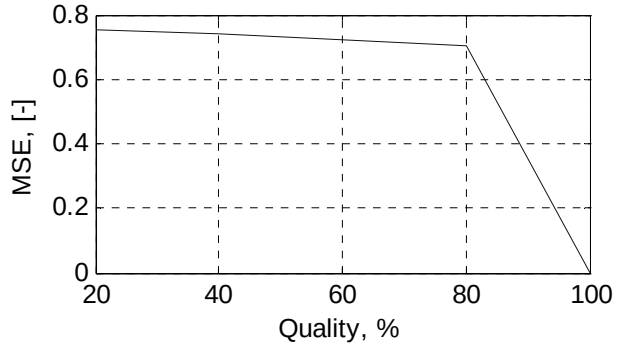


Fig.5 Mean square error for the extracted watermark at different quality levels for the JPEG compressed watermarked image

In Fig.6.a part of the original document containing printed text is given, and in Fig.6.b – the watermarked one for Δ = 12°. The amplified difference of 64 times between these two images is given in Fig.6.c. Slightly changed background is observed which is completely lost (invisible) if documents are stored in compressed form which is the most general case.

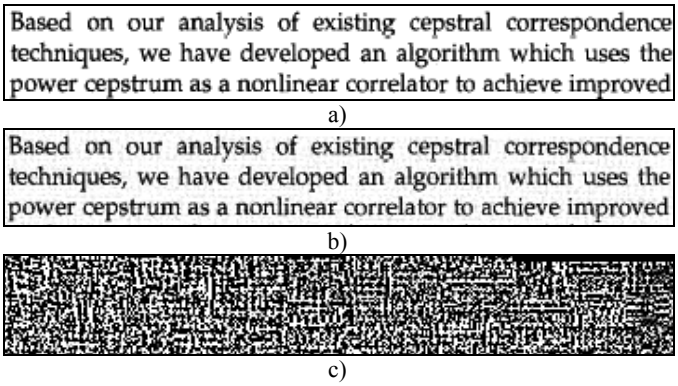


Fig.6. Original and watermarked printed text

In Fig.7.a the original watermark used is given, and in Fig.7.b - the extracted one from a JPEG compressed to 80 % quality image, Fig.7c – from 40 % and in Fig.7.d – from 20 %.

Fig.7. Original and extracted watermark

Fig.7. Original and extracted watermark (continuation)

IV. Conclusion

The proposed algorithm for document protection based on all 2D-CHT coefficients modulation is robust enough to high frequency filtration attack (JPEG compression) which is seen from the almost constant MSE for the extracted watermark. Because of the high PSNR for the watermarked image low visibility for the watermark is guaranteed. Also high capacity for the information inserted is present – ¾ of the total number of pixels in bits, and due to the use of all CHT complex coefficients synchronism is not lost when extracting the watermark. Detecting the presence of watermark is done without the original image. When applied to printed documents last preserve high quality.

Acknowledgement

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science (Contract - ВУ-И-305/2007).

References

[1] S. Mao, A. Rosenfeld, and T. Kanungo, “Document Structure Analysis Algorithms: A Literature Survey”, *In Proceedings of SPIE Electronic Imaging*, Vol. 5010, pp. 197-207, January 2003.

[2] J. Dugelay and S. Roche, “A Survey of Current Watermarking Techniques”, *In Information Techniques for Steganography and Digital Watermarking*, S.C. Katzenbeisser et al., Eds. Northwood, MA Artec House, pp. 121-145, Dec. 1999.

[3] R. Kountchev and V. Mirchev, “Digital Watermarking Using Complex Hadamard Transform and Phase Modulation”, *In ICEST’2005 Conference Proceedings*, Nis, Serbia and Montenegro, Vol. 1, pp. 330-333, June 29 – July 1, 2005.

[4] R. Kountchev, M. Milanova, C. Ford, and S. Rubin. “Multimedia Watermarking with Complex Hadamard Transform in the Inverse Pyramid Decomposition”, *Proc. of the 2003 IEEE Intern. Conf. on Information Reuse and Integration*, Las Vegas, USA, , pp.305-310, October 2003.

[5] R. Mironov and R. Kountchev, “Analysis of Complex Hadamard Transform Properties”, *In ICEST’2006 Conference Proceedings*, Sofia, Bulgaria, pp. 173-176, June 29 – July 1, 2006.

Session

SIGNAL PROCESSING – 2

Investigation of analog neural network used for numbers recognition in images

Liljana Docheva[1] and Aleksander Bekiarski[2]

Abstract – Because of the similarity of analog VLSI Neural Networks signalling to the brain signalling, ones are a preferable implementation for analog signal processing. With the help of the analog neural networks certain computations that are difficult or time-consuming for digital neural network can be done. Nevertheless analog design can be very difficult because of need to compensate the variations in manufacturing, in temperature, etc. In this article has been investigated the parameter variation influence over the behaviour of analogue neural network used for numbers recognition in images.

Keywords – **Analog neural network, VLSI implementation, Image processing.**

I. INTRODUCTION

Analog neural networks have advantages as high speed, low power consumption and compact implementation in comparison with competing digital signal processing approaches. Disadvantages of analog neural networks are their limited accuracy and nonlinear behavior. Variation in the size of discrete transistors and the local mobility will cause random parameter variation [4], [5]. Moreover increase in the precision of any component has as a consequence an increase of its area. Some of these problems: learning algorithm and weight initialization effects on the analog neural networks are investigated and solved [1]. In this paper analog neural network with Backpropagation algorithm used for numbers recognition in images has examined. The aim of this paper is to investigate influence of some analog neural network parameters onto its recognition ability.

II. AN ANALOG NEURAL NETWORK MATHEMATICAL MODEL

The mathematical model of an analog neural network and it VLSI implementation is depicted in [3]. On the base of the neuron and synapse equations described in this implementation are proposed the equations about analog parameter variation over analog neural network behavior and they have been worked out [1]. On this way now in this article is proposed to investigate parameters variation influence over analog neural network behavior.

Equation 1 is the neuron equation including parameter variation.

$$u_{yk}^{l}=\frac{(\alpha_{FC}+\Delta\alpha_{FC})(I_B+\Delta I_B)}{(\beta_{OR}+\Delta\beta_{OR})U_{OR}}\tanh\left(\frac{\frac{g_{mk}+\Delta g_{mk}}{W_0/L_0.U_c}\sum_j W_j/L_j U_{W_{kj}}u_{zj}}{2(\beta_{is}+\Delta\beta_{is})U_{is}U_t}\right), \quad (1)$$

where

u_{yk}^{l} is the neuron output voltage;

k – number of neuron,

l - neural network layer number,

W/L - the MOS resistive circuit multiplier width/length ratios;

α_{FC} - the emitter-collector current gain;

I_B - the bias current;

β - the MOSFET tranconductance parameter;

U_t - the thermal voltage

Δ -a symbol of variable variation [1,2,3].

The following substitution has been done (equations 2 and 3) with the purpose to achieve simplicity of equation:

$$\chi_k^l+\Delta\chi_k^l=\frac{(\alpha_{FC}+\Delta\alpha_{FC})(I_B+\Delta I_B)}{(\beta_{OR}+\Delta\beta_{OR})U_{OR}} \quad (2)$$

$$\xi_k^l+\Delta\xi_k^l=\frac{(g_{mk}+\Delta g_{mk}).W_j/L_j}{2(\beta_{is}+\Delta\beta_{is})U_{is}U_t U_c.W_0/L_0} \quad (3)$$

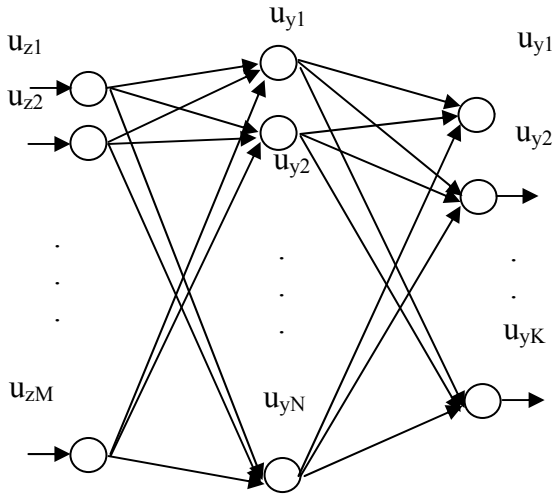


Fig. 1. The structure of neural network that is for numbers recognition in images.

In order to examine the parameter variation influence over the behavior of analog neural network used for numbers recognition in images it is chosen Multi-Layer perceptron (M:N:K) (figure 1), as a well-known feedforward layered

[1] Liljana Docheva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

[2] Aleksander Bekiarski is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

neural network, on which the Backpropagation learning algorithm is implemented. For this neural network architecture equation (1) can be written as:

$$u_{yk}^{2}(t)=\left(\chi_k^2+\Delta\chi_k^2\right)\tanh\left(\left(\xi_k^2+\Delta\xi_k^2\right)\sum_{j=1}^{N}U_{w_{kj}}u_{yj}^{1}(t)\right) \quad (4)$$

The equivalent error of the k[th] neuron of the 2[th] layer δ^2_k is calculated as:

$$\delta_k^2(t)=\left[d_k-\left(\chi_k^2+\Delta\chi_k^2\right)\tanh\left(\left(\xi_k^2+\Delta\xi_k^2\right)\sum_{j=1}^{N}U_{w_{kj}}u_{yj}^{1}(t)\right)\right]. \quad (5)$$

$$\left(\chi_k^2+\Delta\chi_k^2\right)\left(\xi_k^2+\Delta\xi_k^2\right)\left(1-\tanh^2\left(\left(\xi_k^2+\Delta\xi_k^2\right)\sum_{j=1}^{N}U_{w_{kj}}u_{yj}^{1}(t)\right)\right).$$

It can be noted that the terms ($\chi+\Delta\chi$) and ($\xi+\Delta\xi$) directly takes part in equation 5. The weight update for the 2[th] layer is given by:

$$U_{wyk}^{2}(t+1)=U_{wyk}^{2}(t)+(\eta.\delta_k^2(t).u_{yj}^{1}(t)) \quad (6)$$

The neuron output voltage for the 1[th] layer u^1_{yj} is calculated as:

$$u_{yj}^{1}(t)=\left(\chi_j^1+\Delta\chi_j^1\right)\tanh\left(\left(\xi_j^1+\Delta\xi_j^1\right)\sum_{j=1}^{M}U_{w_{ji}}(t)u_{yz}^{1}(t)\right) \quad (7)$$

The equivalent error of the j[th] neuron for the 1[th] layer δ^1_j is calculated as:

$$\delta_j^1(t)=\left(\chi_j^1+\Delta\chi_j^1\right)\left(\xi_j^1+\Delta\xi_j^1\right)\left(1-\tanh^2\left(\left(\xi_j^1+\Delta\xi_j^1\right)\sum_{j=1}^{M}U_{w_{ij}}(t)u_{yz}(t)\right)\right)\sum_{j=1}^{K}U_{w_{kj}}(t)\delta_k^2(t), \quad (8)$$

u_{yz} is the input voltage for the neural network.

The weight update for the 1[th] layer is given by:

$$U_{wyk}^{1}(t+1)=U_{wyk}^{1}(t)+(\eta.\delta_k^1(t).u_{yz}(t)) \quad (9)$$

From equation 4÷9 can be seen that the terms ($\chi+\Delta\chi$) and ($\xi+\Delta\xi$) directly influence over whole (entire) learning phase. This is the reason for the partial neutralization of the parameter variation influence over neural network behavior.

The aim of this paper is to define the boundaries of parameter variation in which numbers recognition in images can be accomplish.

III. Investigation of Analog Neural Network Used for Numbers Recognition in Images

In order to investigate the influence of some analog neural network parameters onto its recognition ability a multilayer perceptron with (M:N:K) structure is used (figure 1), where M=20,N=22 and K=10. The neural network is trained to recognize the numbers shown in figure 2. The numbers are applied to the neural network as a vector with dimension of 20.

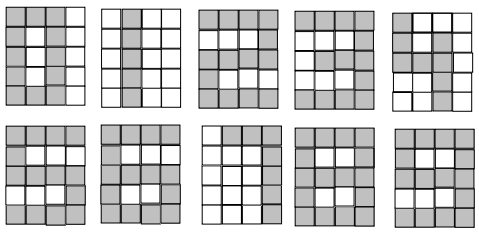


Fig. 2. The patterns.

The desired values are shown in Table I. For the desired values corresponding to the numbers $d_{k,p}$=1, and for desired values at other position are $d_{k,p}$= -1 (equation 10).

TABLE I
THE DESIRED VALUES FOR THE PATTERNS

p / k	0	1	2	3	4	5	6	7	8	9
0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	1	-1	-1	-1	-1	-1	-1	-1
3	-1	-1	-1	1	-1	-1	-1	-1	-1	-1
4	-1	-1	-1	-1	1	-1	-1	-1	-1	-1
5	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
6	-1	-1	-1	-1	-1	-1	1	-1	-1	-1
7	-1	-1	-1	-1	-1	-1	-1	1	-1	-1
8	-1	-1	-1	-1	-1	-1	-1	-1	1	-1
9	-1	-1	-1	-1	-1	-1	-1	-1	-1	1

$$d_{k,p}=\begin{cases}1, for & k=p\\ -1, for & k\neq p\end{cases}, \quad (10)$$

where
k=1,2…K is the number of the output neuron,
p=1,2…P is the number of the pattern that is recognizing.

TABLE II
THE INFLUENCES OF THE PARAMETER χ OVER WEIGHT AND ERROR VALUES OF THE NEURAL NETWORK

χ	w^1	w^2	E
0.7	[-38;38]	[-7;0]	0.174
0.8	[-10;8]	[-5;5]	0.054
0.9	[-1;1]	[-2;2]	0.013
1	[-1;1]	[-1;1]	$9.7.10^{-4}$
1.1	[-8;8]	[-27;22]	1.4

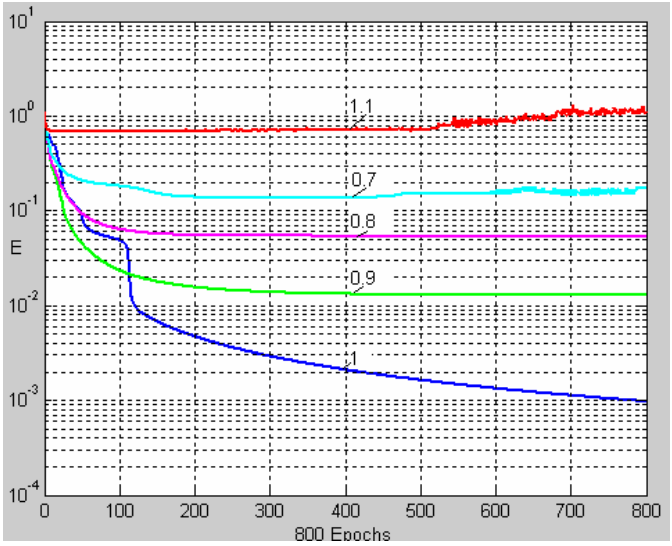


Fig. 3. The influences of the parameter χ variation over output neural network error.

The neural network is trained with backpropagation algorithm, but in the equations the parameters of the real neural networks components take part. The neural network is simulated with Matlab.

Table II shows the influences of the parameter χ over weight and error values of the neural network. It is observed that the parameter χ variation strongly affect to the weight values range.

Figure 3 shows the influences of the parameter χ variation over output error values of the neural network. It san be seen that neural network recognizes correct input patterns in narrow range of parameter χ variation.

Table III shows the influences of the parameter ξ over weight and error values of the neural network.

TABLE III

THE INFLUENCES OF THE PARAMETER ξ OVER WEIGHT AND ERROR VALUES OF THE NEURAL NETWORK

ξ	w^1	w^2	E
0.1	[-1.8;1.8]	[-1.7;1.7]	0.283
0.2	[-1.8;1.8]	[-1.6;1.6]	0.066
0.3	[-1.7;1.3]	[-1.7;1.3]	0.017
0.4	[-1;1]	[-1.4;1.4]	$7.8.10^{-3}$
0.5	[-1.3;1.3]	[-1.2;1.2]	$4.5.10^{-3}$
0.6	[-1;1]	[-1;1]	$2.9.10^{-3}$
0.7	[-1;1]	[-1.3;1.4]	$2.1.10^{-3}$
0.8	[-1;0.8]	[-1.1;1.2]	$1.6.10^{-3}$
0.9	[-0.8;0.8]	[-1;1.1]	$1.2.10^{-3}$
1	[-1;1]	[-1;1]	$9.7.10^{-4}$
1.1	[-0.8;0.8]	[-1.8;1.5]	$7.9.10^{-4}$
1.2	[-0.8;0.8]	[-1;1]	$6.5.10^{-4}$
1.3	[-0.8;0.8]	[-1;1]	$5.6.10^{-4}$
1.4	[-0.8;0.8]	[-0.7;1]	$5.5.10^{-4}$
1.5	[-0.5;0.5]	[-0.5;1]	0.04

Figures 4 and 5 show the influences of the parameter ξ over output error values of the neural network. The range of the ξ variation in which neural network recognizes correct applied input patterns is larger then parameter χ variation.

TABLE IV

THE INFLUENCES OF THE PARAMETER χ OVER WEIGHT AND ERROR VALUES OF THE NEURAL NETWORK FOR NOISY PATTERNS

χ	w^1	w^2	N_e
0.7	[-30;30]	[-7;0]	6
0.8	[-8;8]	[-4;4]	3
0.9	[-1;1]	[-1;1]	1
1	[-1;1]	[-1;1]	0
1.1	[-8;8]	[-22;22]	-

Some noise is added to the original pattern to create test patterns (figure 6). Tables IV and V show the influences of the parameters χ and ξ over weight and error values of the neural network. With N_e is denoted the number of patterns, that aren't recognized. Neural network recognizes correct the input patterns only for χ=1. The variation of the parameter χ

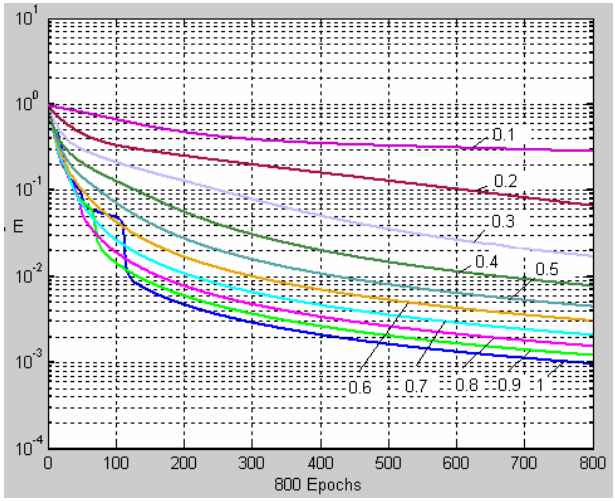


Fig. 4. The influences of the parameter ξ variation over output neural network error for $0.1<\xi<1$.

affect to increasing weight values range and decrease recognition ability of the neural network. For χ=0.9 eight of

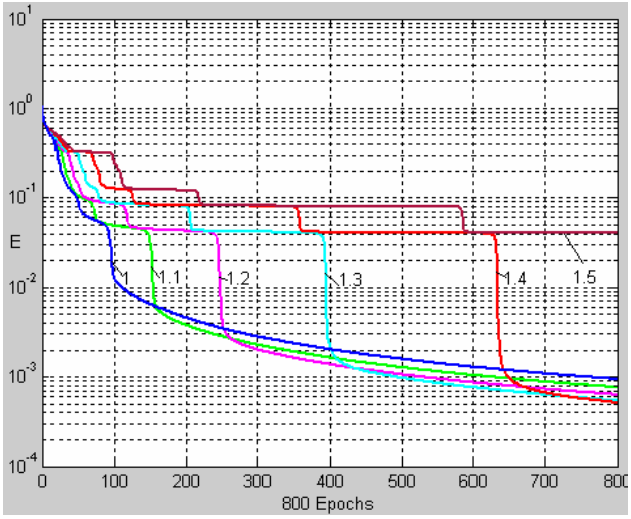


Fig. 5. The influences of the parameter ξ variation over output neural network error for $1<\xi<1.5$.

ten images

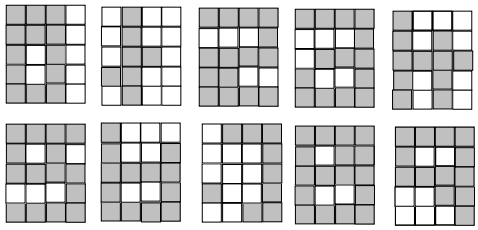


Fig. 6. The noisy patterns.

are recognized for the test patterns from figure 6. If χ=0.8 the recognized images become seven of ten.

The parameter ξ variation have a weak effect over neural network behavior. It can be seen from Table V that the neural network recognizes all patterns for the test sequence for range $0.1<\xi<1.5$. Furthermore the variation of the parameter ξ don't increase weight values range.

TABLE V

THE INFLUENCES OF THE PARAMETER ξ OVER WEIGHT AND ERROR VALUES OF THE NEURAL NETWORK FOR NOISY PATTERNS

ξ	w^1	w^2	N_e
0.1	[-1.7;1.7]	[-1.7;1.7]	-
0.2	[-1.5;1.5]	[-1.5;1.5]	0
0.3	[-1.5;1.5]	[-1.5;1.5]	0
0.4	[-1.5;1.5]	[-1.5;1.5]	0
0.5	[-1.3;1.3]	[-1;1]	0
0.6	[-0.8;0.8]	[-1;1]	0
0.7	[-0.7;0.7]	[-1;1]	0
0.8	[-0.7;0.7]	[-1;1]	0
0.9	[-0.7;0.7]	[-1;1]	0
1	[-1;1]	[-1;1]	0
1.1	[-1;1]	[-1;1]	0
1.2	[-0.7;0.5]	[-0.7;0.6]	0
1.3	[-0.7;0.5]	[-0.7;0.6]	0
1.4	[-0.7;0.5]	[-0.7;0.6]	0
1.5	[-0.5;0.6]	[-0.5;0.5]	1

IV. CONCLUSION

In this paper analog neural network used for pattern recognition has examined. The neural network is trained with backpropagation algorithm, but in the equations the parameters of the real neural networks components take part. There is investigated influence of some analog neural network parameters onto its recognition ability. It is observed that the variation of the parameter χ affect to increasing weight values range and decreases recognition ability of the neural network. The neural network recognizes correct input patterns in narrow range of parameter χ variation. The parameter ξ variation have a weak effect over neural network behavior.

REFERENCES

[1] Alexander Bekiarski, Liliana Docheva, "Influences of weight initialization over analog neural network behavior", E&E, vol : 1, pp. 40-45, 2007.

[2] Liliana Docheva, Alexander Bekiarski ,Ivo Dochev "Analysis of Analog Neural Network Model With CMOS Multipliers", Radioengineering, vol.16 N:3 стр. 103-107, 2007.

[3] LEHMAN T, Hardware Learning in Analog VLSI Neural Networks. Ph.D. thesis, Technical University of Denmark. 1994.

[4] Vittorio Dante, Paolo Del Giudice and Maurizio Mattia "Implementation of neuromorphic systems: from discrete components to analog VLSI chips (testing and communication issues) Ann. Ist. Super. Sanità, vol. 37, n. 2 (2001), pp. 231-239

[5] Elisabetta Chicca, Davide Badoni, Vittorio Dante, Massimo D'Andreagiovanni, Gaetano Salina, Luciana Carota,"A VLSI Recurrent Network of Integrate-and-Fire Neurons Connected by Plastic Synapses With Long-Term Memory", IEEE Transactions on neural, vol. 14, N: 5, 2003, pp. 1297-1307

Fast Anaglyph Retinal Rivalry Reduction Algorithm

Alexander A. Krupev[1] and Antoaneta A. Popova[2]

Abstract – **In this paper is presented an optimized retinal rivalry reduction algorithm for stereoscopic anaglyph image production and 3D visualization. The used color separation of the images for the left and the right eye is by means of Red-Cyan glasses. A rigorous and a fast non-rigorous approaches for reduction of flickering effects, associated with binocular rivalry, are suggested and compared. A correctly reproduced colour gamut is used for this purpose. The results of the retinal rivalry reduction gamut transformation are also compared to the basic anaglyph creation method.**

Keywords – **stereoscopic image, 3D image, anaglyph algorithm, retinal rivalry, uniform vector quantization.**

I. Introduction

In this paper will be discussed the anaglyph stereoscopic method based on color separation of an image by means of Red - R and Cyan - C color filters into two images, intended for the left and the right eye respectively. In order to produce a depth effect, two color layers, representing two slightly different views of a scene, are superimposed.

When there is too much difference between the left and the right eye images, a common artifact, called retinal/binocular rivalry, occurs. Instead of fusing them, the perception alternates between the left and right eye images [4]. A form of rivalry called binocular luster may be observed when the images presented to each eye differ mostly in their luminance and there is a minimal contour difference, as it should be for stereoscopic purposes with correctly chosen stereo basis. The object seems to flicker or shine. Though normal for glossy surfaces, with the color separation based anaglyph method this happens always for objects of certain colors outside of the correctly reproduced color gamut.

The literature on the production of anaglyph images is limited and what exists is empirical mostly. Recent works of E. Dubois [2] and W. Sanders and D. McAllister [1] on the subject are focused mainly on the elimination of ghosting/crosstalk and region merging effects. Our team has published a paper on the subject [6], the used algorithm implementation there is rigorous and precise, but there is a more elegant, simpler and faster approach as we will demonstrate. The works by R. Turnnidge, D. Pizzanelli [5] and others focus on the practical aspects of anaglyph production in Photoshop, but retinal rivalry problem is ignored. Older works focus on monochrome anaglyphs, which lack retinal rivalry. Apart from anaglyphs, retinal/binocular rivalry has been studied as a physiological phenomenon extensively [3], [4].

Our main goal is to create an alternative faster non-rigorous arithmetic implementation of the method from [6] at the minor expense of some precision. In Section II the anaglyph gamut transformation algorithm is explained. In Section III a new non-rigorous implementation is suggested and compared to the rigorous one. The test results and conclusions are shown in Section IV.

II. Retinal Rivalry Reduction Algorithm Using the Anaglyph Gamut

The anaglyph method is used mainly for 3D representation of monochrome images. The monochrome anaglyph is not a monochrome image itself, but it's perceived as such. The set of colours correctly reproduced by the anaglyph method, besides the monochromatic ones, is defined in [6]. These colours form a plane in the RGB-cube - Fig.1.

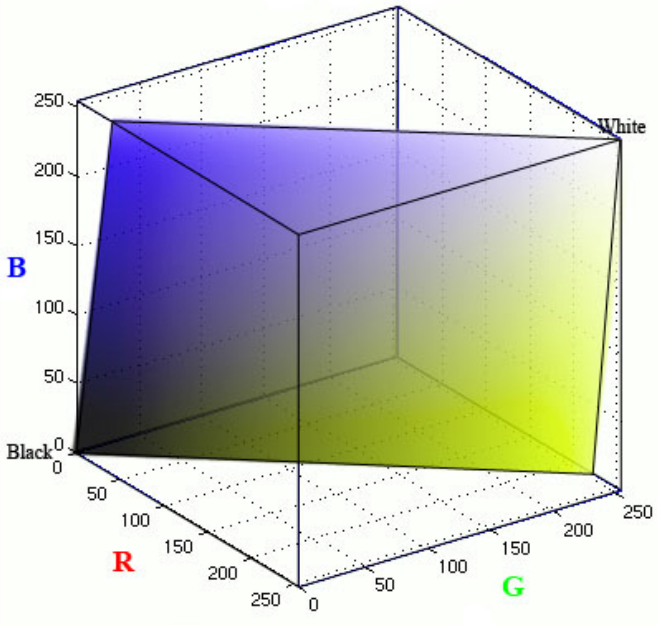


Fig. 1. The plane of colors, forming the defined anaglyph gamut

These colors satisfy the condition:

$$R = \frac{7G + B}{8} \qquad (1)$$

This preserves the luminance ratio between the left and the right channels observed in a monochrome anaglyph. The objects, with colors outside of this plane, show different

[1]Alexander A. Krupev is with the Faculty of Communications, Technical University- Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: asfalot@abv.bg

[2]Antoaneta A. Popova is with the Faculty of Communications, Technical University - Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mail: antoaneta.popova@tu-sofia.bg

degrees of flickering when viewed through the anaglyph glasses, especially objects of bright cyan, red and green colors. The monochrome colors are placed diagonally from the lower left to the upper right corner, varying from black to white.

The purpose of the anaglyph gamut transformation algorithm is to substitute the colors from the stereo pair with colors within the plane from Fig. 1. Each time the closest color from the plane is used. In the next section we will discuss the process of finding this nearest color value.

III. FINDING THE NEAREST DISTANCE COLOUR VALUE FROM THE ANAGLYPH COLOUR PLANE

The previously utilized approach was to divide the plane hierarchically and uniformly into smaller and smaller regions, each represented by its centroid. B increases horizontally from left to right, G-vertically upwards and R, as a function of these two, diagonally. The first and second hierarchical levels are shown on Fig. 2. G and B channels each have 256 colour values ranging from 0 to 255 - 8bit per channel. That's why the segmentation is performed 8 times for maximum precision on increasingly smaller regions from 256x256 to 2x2. The Euclidian distance Eq. (2) is calculated to each one of the four centroids and the nearest one is chosen, thus narrowing the search to its corresponding region. This approach is similar to a uniform vector quantization technique.

$$dist = \sqrt{(R - R_0)^2 + (G - G_0)^2 + (B - B_0)^2} \qquad (2)$$

Here R_0, G_0 and B_0 are the coordinates of a colour from the image and R, G and B - the coordinates of a point from the plane.

The new approach we propose is a numerical one. We will define an equation which calculates the nearest colour from the anaglyph colour plane, instead of choosing from a set of predefined values.

The equation of a plane is given by Eq. (2), d=0 since the beginning of the coordinate system (0, 0, 0) belongs to the plane:

$$\begin{aligned} &aR + bG + cB + d = 0 \\ &d = 0 \end{aligned} \qquad (3)$$

The parametric equations of a line passing through a point (R_0 G_0 B_0) in the RGB color space and perpendicular to the plane from Eq. (3) are:

$$\begin{aligned} R &= R_0 + at \\ G &= G_0 + bt \\ B &= B_0 + ct \end{aligned} \qquad (4)$$

Here t is an independent variable.
The nearest colour to (R_0 G_0 B_0) from the plane is the cross-point between the line and the plane.

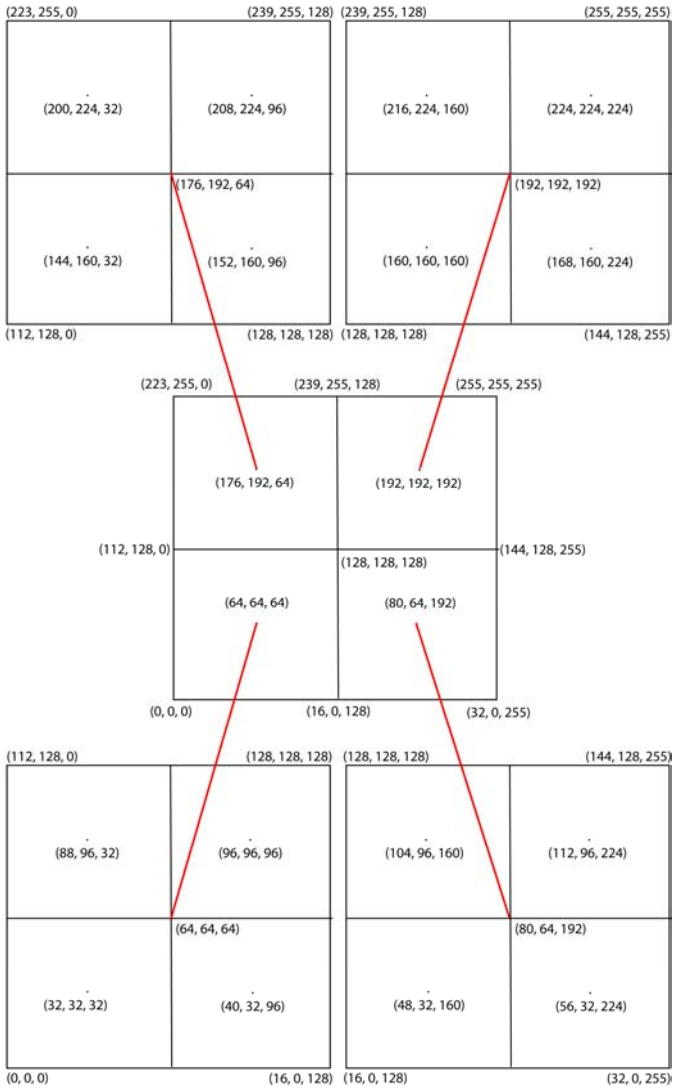


Fig. 2 First and second level hierarchical segmentation of the color plane

We substitute Eq. (4) into Eq. (3):

$$a(R_0 + at) + b(G_0 + bt) + c(B_0 + ct) = 0 \qquad (5)$$

After we solve for t we get:

$$t = -\frac{aR_0 + bG_0 + cB_0}{a^2 + b^2 + c^2} \qquad (6)$$

For this value of t the line crosses the plane.

After we substitute this value of t into Eq. (4) we can calculate the coordinates (R, G, B) of this point:

$$R = \frac{(b^2+c^2)R_0 - bG_0 - cB_0}{a^2+b^2+c^2}$$
$$G = \frac{(a^2+c^2)G_0 - aR_0 - cB_0}{a^2+b^2+c^2} \quad (7)$$
$$B = \frac{(a^2+b^2)B_0 - aR_0 - bG_0}{a^2+b^2+c^2}$$

Or in a matrix form:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{a^2+b^2+c^2} \begin{bmatrix} b^2+c^2 & -b & -c \\ a^2+c^2 & -a & -c \\ a^2+b^2 & -a & -b \end{bmatrix} \begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix} \quad (8)$$

From Eq. (1) we can get the exact a, b and c values for our case:

$$8R - 7G - B = 0$$
$$a = 8;\ \mathrm{b} = -7; c = -1; \quad (9)$$

Using these exact values in Eq. (8) we get the equation (10), used in the alternative implementation of our algorithm:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{114} \begin{bmatrix} 50 & 7 & 1 \\ 65 & -8 & 1 \\ 113 & -8 & 7 \end{bmatrix} \begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix} \quad (10)$$

IV. Experimental Results and Conclusions

All the experiments are implemented in Matlab 7.0.3 working environment. An algorithm detecting flickering objects in stereo pairs/cyclopean images, implemented on Matlab, is used. The implemented algorithm produces maps of the retinal rivalry effect - grayscale images, the brighter colors of which indicate flickering regions.
The formula used for this detection now is:

$$Y = \left| R - R_{expected} \right|$$
$$R_{expected} = \frac{7G + B}{8} \quad (11)$$

R is the Red channel value from the actual image, $R_{expected}$ is the Red channel value, for which there would be no rivalry. Y is the grayscale value from the map of the flickering regions. It ranges from zero (R= $R_{expected}$) to 255 (for pure Red [255, 0, 0] or Cyan [0, 255, 255] colors).

As an example of how does a binocular rivalry map work see Fig. 3. White and gray squares yield black squares on the map, cyan and red – white squares, which indicates that they would cause strong flickering in an anaglyph image [6].

Fig. 3. A test image (left) and its retinal rivalry map (right)

On Fig.4 we see a stereo pair of a brain. The intense red parts appear lighter on the map, while the background and the grayish/whitish parts of the brain appear darker. The retinal rivalry map is not a negative image – if it was, the background would appear white.

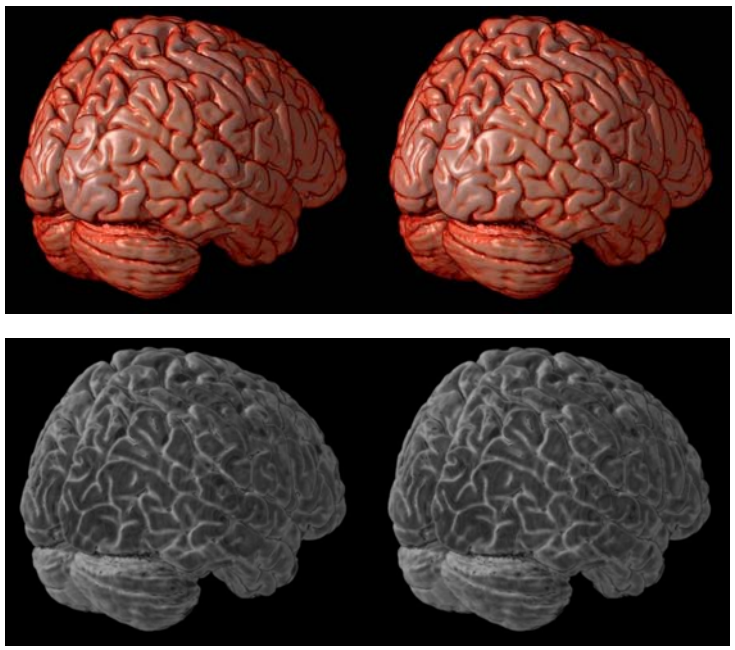


Fig. 4. A stereo pair (Up) and the map of its flickering regions (Down)

On Fig. 6 we see a processed stereo pair with colors from the anaglyph color gamut. The processing is done with the new fast algorithm. There are no noticeable flickering regions on its retinal rivalry map. The resulting anaglyph from Fig. 7 demonstrates no binocular rivalry also as expected.

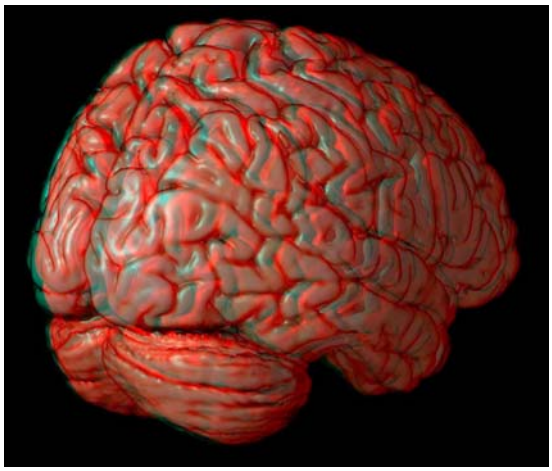


Fig. 5. Anaglyph images of a brain using the stereo pair from Fig. 4

The same cannot be said about the anaglyph from Fig. 5, produced by the stereo pair from Fig. 4 using the basic common PS algorithm, without any further processing. Unnatural shining is most obvious in the cerebellum area, being most reddish in the stereo pair.

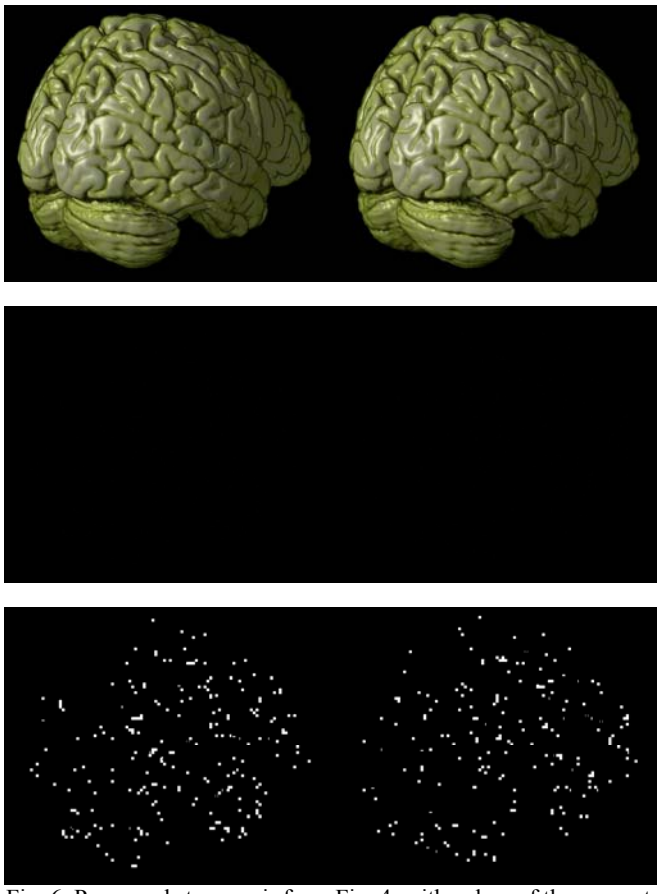


Fig. 6. Processed stereo pair from Fig. 4. with colors of the suggested anaglyph gamut (Up), its binocular rivalry map (Middle) and binarized version of the map (Down)

The retinal rivalry map in Fig. 6 is not perfectly black, though. Due to rounding errors in the fast rivalry reduction algorithm parts of it have values up to 2 (very dark gray). When the map is binarized with a threshold of 1, these parts become obvious (Fig.6-Down).

That's because the new implementation doesn't use predefined color values from the anaglyph color plane to choose from like, but calculates new values. They are close enough to the plane not to introduce visible flickering.

An example: orange color (222, 121, 30) – its nearest color of the anaglyph color plane is (159, 176, 37) selected by the rigorous algorithm, but the new fast algorithm calculates (158, 176, 37) which is quite close. The authors consider the precision acceptable for practical purposes.

Considering that the new method is about 9 times faster (0.5 versus 4.5 seconds for the Brain stereo pair with resolution 1806x768), not very computationally intensive and its results are comparable to the rigorous one, we encourage its use for anaglyph binocular rivalry reduction.

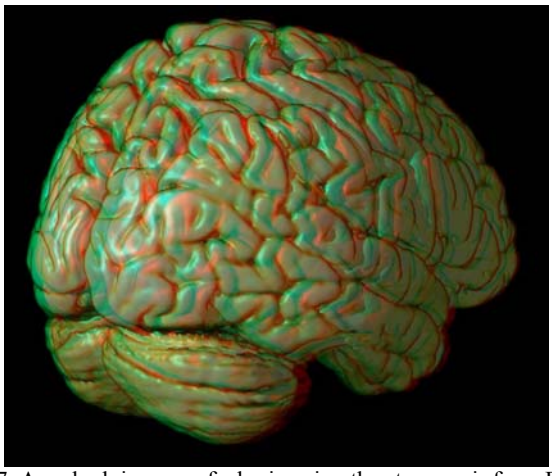


Fig. 7. Anaglyph images of a brain using the stereo pair from Fig. 6

ACKNOWLEDGMENT

This work was supported by Technical University of Sofia under project № 091pd018-07 "Research and development of algorithms for 3D visualization of stereoscopic and multi-view images applying in education".

REFERENCES

[1] W. Sanders and D. McAllister, "Producing Anaglyphs from Synthetic Images", Proc. SPIE Vol. 5006, San Jose, California, pp. 348-358, 2003.

[2] E. Dubois, "A projection method to generate anaglyph stereo images" Acoustics, Speech, and Signal Processing, IEEE International Conference**,** Salt Lake City, UT, USA, 2001

[3] F. Tong and S. A. Engel, "Interocular rivalry revealed in the human cortical blind-spot representation", Nature, 411, 195-199, 2001.

[4] R. Blake, "A natural theory of binocular rivalry", Psychological review, Vol. 96. No. 1, pp. 145-167, Evanston, Ilinois, USA, 1989

[5] R. Turnnidge and D. Pizzanelli, "Methods of previsualizing temporal parallax suitable for making multiplex holograms. Part II: Greyscale and colour anaglyphs made in Photoshop^{TM}," Imaging Science Journal, vol. 45, no. 1, pp. 43–44, 1997.

[6] A. Krupev and A. Popova, "Flickering Reduction Algorithm in Anaglyph Stereoscopic Images", ICEST'08 Proceedings of Papers, Vol. 1, pp.125-128, Nis, Serbia, 2008

Reference Text Line Identification Based on "Water Flow" Algorithm

Darko Brodić[1] and Zoran Milivojević[2]

Abstract – **In this paper a robust and efficient algorithm for detecting reference text line at different skew angles is presented. Algorithm called "water flow" assumes hypothetical water flows under different specified angles of the image frame from left to right and vice versa. As a result of algorithm, unwetted image frames is extracted as area needed for text line extraction. From extracted text line by the least square method, reference text line is identified. Examples using different water flow angles are examined and inspected. Results are analysed, elaborated and summarized. Proposed algorithm showed robustness and should be used for all sorts of skewness.**

Keywords – **Reference text line, Reference text line extraction, OCR**

I. INTRODUCTION

Handwritten text is fully or partially cursive text. It can be composed of discrete characters, broken words or its combination. But handwritten text tend to be multi-oriented and skewed. Appearance of different orientation skewed lines i.e. multi-skewed lines are made the handwritten text to be less readable. Hence, reference text line extraction from optically scanned documents is primary goal of optical character recognition (OCR) [1].

Various reasons exist for appearance of multi-skewed lines in text, but two are most common [1]. Firstly, during scanning process a degrees of misalignment of the document made is unavoidable. But, all text lines in the scanned document are uniformly skewed i.e. reference text line are almost parallel. Secondly, text lines in original document are differently skewed due to specific individual handwriting.

All handwriting text lines are made under different orientation i.e. multi-skewed. To enhance the ability of document analysis system robust algorithm for reference line extraction in multi-skewed text is needed.

In ideal situation extraction of reference text line is simple. A reference text line is described by relation

$$y = ax + b\,. \tag{1}$$

It means parameters a and b define slope and y-intersection, respectively. Due to uniformly skewed text, one reference line extraction led to extraction of other reference lines.

In real situation, all proposed techniques of reference line extraction based on identifying valleys of horizontal pixel density histogram quoted in [1] failed due to multi-skewed text lines.

[1]Darko Brodić is with the Technical Faculty Bor at University of Belgrade, Vojske Jugoslavije 12, Bor 19210, Serbia, E-mail: dbrodic@tf.bor.ac.rs

[2]Zoran Milivojević is with the Technical College Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: zoran.milivojevic@ vtsnis.edu.rs

Method of identifying words contour area as a start of detecting baseline point proposed in [2]. But the assumption made on the definition of elements of word are too specific.

Another method proposed in [3] deal with "simple" multi-skewed text. It uses as a basis simple type of Hough transform for straight lines. This approach is again specific.

Algorithm proposed by [4] model text line detection as an image segmentation problem by enhancing text line structure using a Gaussian window and adopting the level set method to evolve text line boundaries. Authors specified method as robust, but rotating text by an angle of 10° has an significant impact on reference line hit rate.

In this paper proposed method by [1] is modified as in [5] and more examined as well. Actually method [1] hypothetically assumed a flow of water in a particular direction across image frame in a way that it faces obstruction from the characters of the text lines. Definition of unwetted area is corner stone for calculating text line extraction. But in [1] water flow angle is locked up by four values: 45°, 26.6°, 18.4° and 14°. Our modified water flow method works with water flow angle from 0° to 90°. As can be seen different water flow angle region is proposed for the best referent line hit rate. Text under investigation is computer printed sample text rotated around x-axis from -20° to 20°.

Organization of paper is following. In Sect. II. we give an information on proposed algorithm. Every detail of modified algorithm is described. In Sect. III. experimental results are given. Results are analysed and elaborated as well. In Sect. IV. conclusion is made.

II. PROPOSED ALGORITHM

Text line detection procedure consists of three elements as shown in Fig. 1.

Fig. 1. Text line detection procedure

In preprocessing stage, algorithm for line distinction is used. After preprocessing all text lines are separated. During the processing stage, algorithm for identification of skew and reference text line is employed. After all, in postprocessing stage reference text, based on skew and stroke angle, is straightened and repaired.

In this paper we deal with element of processing. Some assumption should be made before defining algorithm. We suppose that there is an element of preprocessing. It had been

made before applying our algorithm. It's job is to split up text lines due to white area between every neighbor text lines. After preprocessing, every text line is separated. So, it initiates distinct entity consists of group of words.

Split up and extracted text line represents digitalized image dimension MxN. Each word in an image contains black points i.e. pixels. Every point is represented by number of coordinate pairs $X_{i,j}$. Algorithm needs to extract bounding areas and define pixel type from situation in Fig.2.

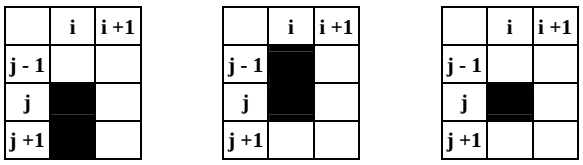


Fig. 2. Boundary pixel type
a. Upper boundary pixel, b. Lower boundary pixel,
c. Candidate for additional investigation of boundary type

Due to making unwetted areas under angle those regions are mark out by lines defined by:

$$y = kx \ , \tag{2}$$

where slope k = tan(α) from Fig.3.

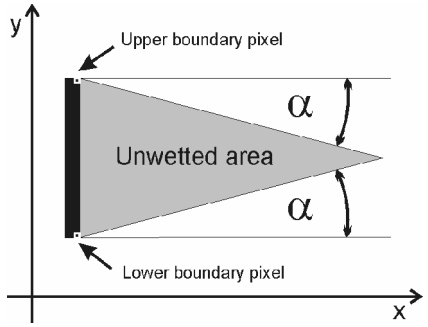


Fig. 3. Unwetted area definition

First, proposed algorithm have to verify boundary pixel type in document image. After verification it makes unwetted areas around the words. Due to pixel type, i.e. upper or lower, slope is α or –α. Lines defined by slope make connection in specific pixel creating unwetted area defined as grey region in Fig. 3.

Pseudo code for pixel type detection (See Fig.2. for reference) can be expressed as follows:

```
begin
   for I = 1 to M step 1
   begin
      for J = 1 to N step 1
if x[i,j]=black and x[i,j+1]=black and x[i,  j-1] =
white and P
      then
         pixel=upper_boundary
         slope=alpha
elseif x[i,j]=black and x[i,j-1]=black x[i,j+1]=
white and P
      then
         pixel=lower_boundary
         slope=-alpha
elseif  x[i,j]=black and x[i,j-1]=white and
x[i,j+1]=white and P
      then
         additional_investigation
   end
end
```

where P is defined as:

```
x[i,j]=white and x[i,j+1]=white and x[i,j-1]=white
```

Additional investigation is made on pixel without complete location. It can be lower, upper or no boundary pixel. It depends on neighbor area of pixels. Apart from [6] and [7] enlarged window RxS pixels is defined as a basis. In this paper R=5 and S=7 is proposed and analysed. Position of window is backwards from pixel candidate for additional investigation. Pixels occurence and position in proposed window led to identification on pixel boundary type.

After additional investigation pixel type is completely located. Throughout previous decision making, algorithm for unwetted areas simply draw area under specified angles. As a result words are bounded by unwetted dark stripes. This situation is given in Fig. 4.

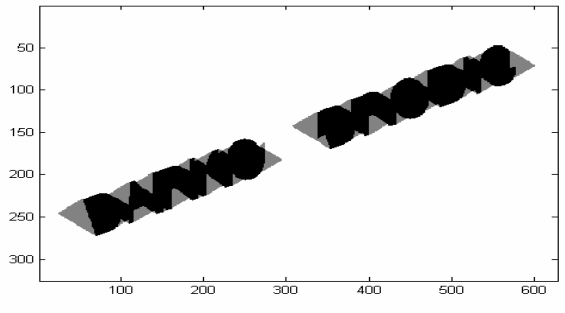


Fig. 4. Document text with unwetted areas

The skew angle detection is based on information obtained from presented algorithm. Defining reference text line means calculating specific average position of every image column. That specific average position is average position of only black pixels in every column of image.

Relation for calculating reference text line is:

$$x_i = \frac{\sum_{j=1}^{L} y_j}{L} \qquad i = 1,\ldots, K \ , \tag{3}$$

where y_j is position of black pixel in column j and L is sum of black pixel in specified column j of an image.

After calculation image matrix with only one black pixel per column is obtained. That black pixel per column defines calculated reference text line and text line skewness. Fig. 5. presented line obtained by reference text line calculation.

Calculated "reference text line" forms discontinuous line partly representing reference text line. To form continuous reference text line from point's collection some numerical method could be used. Candidate methods are interpolation, extrapolation, and combination of interpolation and extrapolation or least squares method [8].

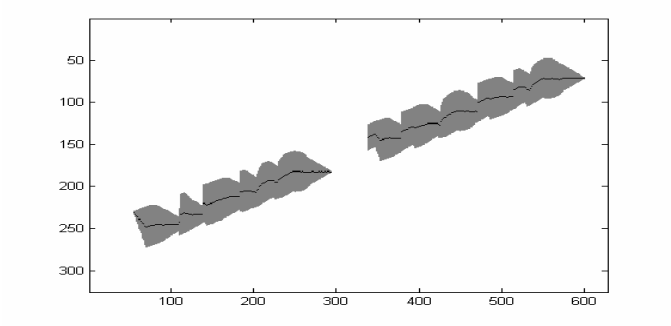


Fig. 5. Calculated "reference text line"

The goal is more common due to linear aspect of reference text line representation from original sample text. But, fundamentally method of fitting data is needed. Least squares method can be interpreted as that. It is an alternative to interpolation for fitting a function to a set of points. Unlike interpolation, it does not require the fitted function to intersect each point. The method of least squares is probably best known for its use in statistical regression. Hence, least square method is used for achieving continuous reference text line from point's collection. The common computational procedure to find a first degree polynomial function approximation in a situation like this is as follows [8]:

1. Use n for the number of data points.
2. Find the four sums:

$$\sum x \ , \ \sum x^2 \ , \sum y \ , \sum xy \ . \tag{4}$$

3. The calculations for the slope, a', and the y-intercept, b', are as follows:

$$a' = \frac{(\sum y)(\sum xy) - n\,(\sum xy)}{(\sum x)^2 - n\,(\sum x^2)} \ , \tag{5}$$

$$b' = \frac{(\sum x)(\sum xy) - (\sum y)(\sum x^2)}{(\sum x)^2 - n\,(\sum x^2)} \ . \tag{6}$$

As a result, continuous reference text line is obtained by:

$$y = a'x + b' \ . \tag{7}$$

III. Experimental results

For the sake of experiment, reference computer sample text no.1 rotated from -20° to +20° by step of 5° around x-axis is used. It is represented in Fig.6.

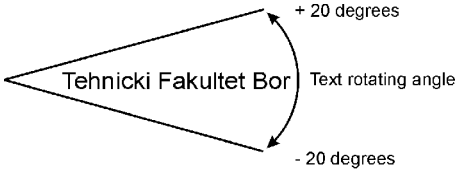


Fig. 6. Sample text rotating from -20° to +20° for the robustness investigation of the algorithm

As a reference, sample text no.2 is introduced to represent handwritten text. It is given in Fig.7.

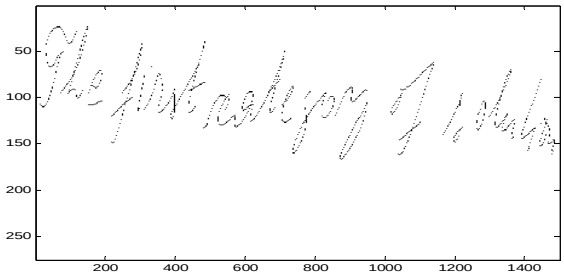


Fig. 7. Sample text no.2 representing handwritten text

Our modified water flow method works with water flow angle from 0° to 90°. Referent line hit rate (RLHR) is defined by:

$$RLHR = 1 - \frac{\beta_{ref} - \beta_{est}}{\beta_{ref}} \ , \tag{8}$$

where β_{ref} = arctan(a) from eq.(1) and β_{est} = arctan(a') from eq.(7).

The best referent line hit rate depends on different water flow angle region. At the beginning, referent line hit rate for "water flow" small angles, i.e. from 0° to 20°, are investigated (see Table I and II). "Water flow" angles smaller than 15° aren't enough robust. Further investigation on varying full range "water flow" angle from 0° to 90° by step 5° gave results presented in Table I and II. In Table II mean value i.e. X, and RMS values are calculated by:

$$\overline{X} = \frac{1}{L}\sum_{i=1}^{L} X_{est} \ , \tag{9}$$

$$RMS = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(X_{ref} - X_{est})^2} \ , \tag{10}$$

where in (9) and (10) L is number of examined text rotating angles range from -20° to +20° by step 5°, X_{ref} is RLHR for β_{est} equal to β_{ref} i.e. due to normalization equal to 1, and X_{est} is RLHR.

Inspecting given results led to conclusion that any "water flow" angle bigger than 20° has enough robustness. This instance is presented in Fig.8.

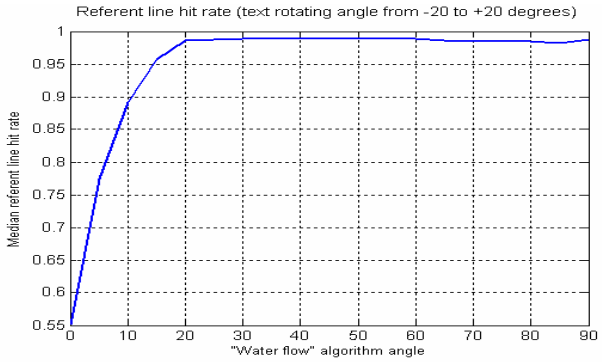


Fig. 8. Referent line hit rate for "water flow" angle from 0 to 90°

Further inspecting Fig.8., the best referent line hit rate is obtained for ”water flow” angle region from 20° to 60°. Specified ”water flow” angle region should be used on further investigation of complicated reference text samples.

Proposed algorithm is suited for printed as well as for handwritten text. Applying algorithm to sample text no.2 proved those claims. Visual results of calculated “reference text line” from sample text no.2 is represented in Fig.9.

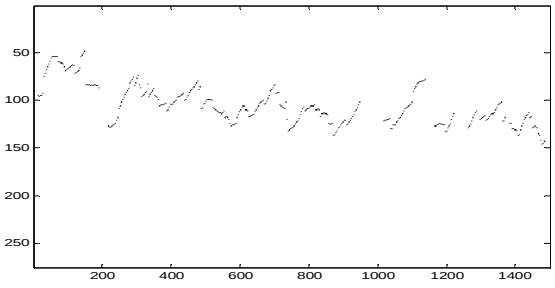


Fig. 9. Calculated ”reference text line” from sample text no.2

As can be seen from Fig.9. algorithm is suited for multi-skew text common in handwritten text. Further investigation will be made on calculating skew rate and reference text line in handwritten text.

IV. Conclusion

In this paper we presented modified algorithm for extraction of reference text line in multi-skewed text image. It assumes hypothetical water flows under different specified angles of the image frame from left to right and vice versa. As a result of algorithm, unwetted image regions defined. Those regions are corner stone needed for reference text line calculation.

Using least square method on calculated reference text line, reference text line is extracted and identified. Robustness of algorithm is examined and validated using computer sample text rotating in region of ±20° around x-axis. Obtained results led to conclusion that ”water flow” angle in region from 20° to 60° is the best suited for high-quality reference line hit rate.

At the end, calculated ”reference text line” from hand-written sample text visually prove that algorithm is suited for multi-skew handwritten text.

References

[1] S. Basu, C. Chaudhuri, M. Kundu, M. Nasipuri, D.K. Basu, "Text Line Extraction from Multi-Skewed Handwritten Documents", Pattern Recognition, Vol.40, pp. 1825-1839, 2006

[2] Jiren Wang, Mazlor K.H. Leung, Siu Cheung Hui, "Cursive Word Reference Line Detection", Pattern Recognition, Vol.30, No.3, pp. 503-511, 1997

[3] G. Louloudis, B.Gatos, I.Pratikakis, C.Halatsis, "Text Line Detection in Handwritten Documents", Pattern Recognition, Vol.41, pp. 3758-3772, 2008

[4] Yi Li, Yefeng Zheng, D. Doermann, S. Jaeger, "A New Algorithm for Detecting Text Line in Handwritten Documents", 18th International Conference on Pattern Recognition, Vol.2, pp. 1030-1033, Hong Kong, 2006

[5] D. Brodić, Z. Milivojević, “Uniform Modified Method for Handwritten Text Reference Line Detection”, MIPRO 2009, CIS Section, paper no.26, Opatija, Croatia, 2009

[6] R.C. Gonzalez, R.E. Woods, “Digital Image Procesing”, 2nd ed., New Jersey: Prentice-Hall, 2002, pp. 67-70

[7] M. Sonka, V. Hlavac, R. Boyle, “Image Processing, Analysis and Machine Vision”, Toronto: Thomson, 2008, pp. 174-177

[8] W.M. Bolstad, “Introduction to Bayesian Statistics”, New Jersey: John Wiley & Sons, 2004, pp. 40-44, 235-240

TABLE I
REFERENCE LINE HIT RATE FOR WATER FLOW ANGLE VARYING FROM 0° TO 90° AND FOR TEXT ROTATING FROM -20° TO 20°

Text	“Water Flow” Angle																		
Rotation	0°	5°	10°	15°	20°	25°	30°	35°	40°	45°	50°	55°	60°	65°	70°	75°	80°	85°	90°
-20°	0.4988	0.6164	0.7362	0.8607	0.9860	0.9879	0.9885	0.9882	0.9879	0.9868	0.9863	0.9893	0.9887	0.9871	0.9865	0.9860	0.9854	0.9860	0.9860
-15°	0.4983	0.6588	0.8223	0.9888	0.9907	0.9907	0.9910	0.9907	0.9899	0.9899	0.9888	0.9862	0.9888	0.9873	0.9851	0.9843	0.9847	0.9851	0.9851
-10°	0.5048	0.7408	0.9870	0.9892	0.9904	0.9915	0.9915	0.9915	0.9909	0.9904	0.9887	0.9858	0.9881	0.9824	0.9796	0.9796	0.9779	0.9779	0.9779
-5°	0.4851	0.9737	0.9771	0.9783	0.9817	0.9840	0.9828	0.9828	0.9828	0.9817	0.9794	0.9817	0.9657	0.9668	0.9634	0.9622	0.9622	0.9565	0.9600
0°	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900
5°	0.4874	0.9760	0.9794	0.9805	0.9817	0.9851	0.9863	0.9874	0.9874	0.9886	0.9931	0.9989	1.0023	1.0023	1.0046	1.0080	1.0092	1.0103	1.0126
10°	0.4980	0.7368	0.9813	0.9841	0.9858	0.9875	0.9887	0.9898	0.9909	0.9926	0.9949	0.9949	1.0000	0.9960	0.9938	0.9960	0.9977	0.9994	1.0011
15°	0.4976	0.6566	0.8163	0.9828	0.9851	0.9858	0.9862	0.9869	0.9877	0.9881	0.9899	0.9903	0.9922	0.9910	0.9881	0.9873	0.9884	0.9903	0.9914
20°	0.4993	0.6169	0.7376	0.8579	0.9868	0.9887	0.9887	0.9893	0.9901	0.9907	0.9901	0.9912	0.9854	0.9780	0.9706	0.9703	0.9722	0.9437	0.9893

TABLE II
REFERENCE LINE HIT RATE MEAN AND RMS FOR WATER FLOW ANGLE VARYING FROM 0° TO 90° AND FOR TEXT ROTATING FROM -20° TO 20°

Text	“Water Flow” Angle																		
Rotation	0°	5°	10°	15°	20°	25°	30°	35°	40°	45°	50°	55°	60°	65°	70°	75°	80°	85°	90°
Mean	0.5510	0.7740	0.8919	0.9569	0.9865	0.9879	0.9882	0.9885	0.9886	0.9887	0.9890	0.9898	0.9890	0.9868	0.9846	0.9849	0.9853	0.9821	0.9881
RMS	0.1647	0.1606	0.1119	0.0555	0.0034	0.0026	0.0027	0.0026	0.0026	0.0031	0.0044	0.0051	0.0104	0.0103	0.0122	0.0134	0.0138	0.0206	0.0146

Improved System for Content Based Image Retrieval Based on Pyramid Decomposition in the Spectrum Domain

Agata Manolova[1], Roumen Kountchev[2], Irina Aleksieva[3]

Abstract **– One of the most important problems concerning the e-client management of large databases is the creation of methods for content-based image retrieval. The goal of this study is to improve the system for fast image retrieval, using pyramid image decomposition in the spectrum domain developed in [1]. The image retrieval is performed on the basis of evaluation of the multi-layer distance between the compared images. This approach permits the use of a recursive algorithm with relatively low computational complexity.**

Keywords – **Content based image retrieval, Inverse Differential Pyramid, Colour image database.**

I. INTRODUCTION

The rapid advancements in multimedia technology have increased the relevance that repositories of digital images are assuming in a wide range of information systems. Effective access to such archives requires that conventional searching techniques based on external textual keywords be complemented by content-based queries addressing appearing visual features of searched data [2], [3]. Content-based image retrieval (CBIR) could be described as a process frame work for efficiently retrieving images from a collection by similarity. The retrieval relies on extracting the appropriate characteristic quantities describing the desired contents of images. In particular, different techniques were identified and experimented with to represent the content of single images according to low-level features, such as color [4], [5], texture [7], [8], shape [9], [10], and structure [11].

The current work is divided in three major steps:

The first part consists of image preprocessing for the needs of the algorithm;

The second stage of the proposed work is the development of the content based image retrieval algorithm based on the Inverse Differential Pyramid (IDP);

The last part includes the implementation of the algorithm on C++ and testing it on databases – small with 301 images divided in four semantic types and extended database with 4468 images divided in 53 semantic types and conclusions.

[1]Agata Manolova and [2]prof. Roumen Kountchev are with the Faculty of Telecommunications at TU - Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: amanolova@tu-sofia.bg, rkountch@tu-sofia.bg .

[3]Irina Aleksieva is with University Joseph Fourier, 621 avenue Centrale , Domaine Universitaire, 38400 St Martin d'Heres, France.

II. DEVELOPMENT OF THE INCREMENTAL ALGORITHM FOR IMAGE COMPARISON WITH IDP

A. Image analysis and preprocessing

Image representation based on transforming the image in another function is a very suitable choice which allows to obtain easily comparable objects and also to do the comparison on a multilevel scale.

There are some necessary conditions for the IDP algorithm depending on the image dimensions and the spectral coefficients as described in [12]. The first condition of the preprocessing is that the image should be in greyscale i.e. there should be a conversion from the RGB or other color space to the YCbCr model where we keep the Y component. The use of the YCbCr color space is not random. It is widely used in image and digital photography systems. The YCbCr color space has the advantage of low correlation between its components. The second important condition is that all the images in the database must have the same height / weight m ($m = 2^n$ and $n \in IN$). Unfortunately few images meet this requirement. There are various ways to handle that problem depending on the database. The simplest way is to resize the image with an interpolation of the pixels in order to obtain the right dimensions. As the standard method for image interpolation is not suitable for this system due to a change in the general forms of the image which perturbs the transform, the proposed algorithm for modified image interpolation can provoke a slight twist in the general forms of the image but usually it is insignificant for the transform. An important perturbation can occur just in cases where the width and the height strongly differ [13]. This algorithm is based on floor and ceiling functions. If the image size is WxH where $W < m$ and $H < m$ (which is mostly the case), the rows and the columns are separately treated. The first step is to compute the scale between the old width (resp. height) and the new one. Supposing the old width is W and the new is w then the ratio is W/w. Then the width coordinate of every pixel from the original image is multiplied by the ratio and the floor function of this multiplication is put in an accumulator. The accumulator receives the values of every pixel of the original image that have obtained the same floor. Computing the value of the new pixel simply demands the mean value of all the pixels in the accumulator as shown on the figure 1. Compared to the standard interpolation approach the results with the modified implementation method present much less distortion. This algorithm is computationally non expensive and easy to implement.

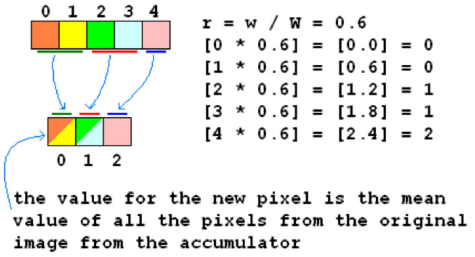


Fig. 1 Modified Interpolation algorithm scheme

B. Image decomposition with Inverse Differential Pyramid

The idea behind this approach is transforming the images from the database to obtain different levels (called layers) providing simplified information about the image. The method used for this decomposition is called Inverse Differential Pyramid (IDP), [1], [12] and it comprises the following steps:

1. The original image [B] with size m^2 ($m=2^n$) pixels is presented as a sum of components, presenting the image decomposition in the spatial domain in correspondence with the relation:

$$\left[B\left(2^n\right)\right]=\left[\tilde{B}_0\left(2^n\right)\right]+\sum_{p=1}^{n-1}\tilde{E}_{p-1}\left(2^n\right)+E_{n-1}\left(2^n\right) \qquad (1)$$

where p = 0, 1, 2, ..., n-1 is the sequence number of the decomposition component.

2. The components of the decomposition (1) are defined as follows:

i. Image decomposition with Inverse Difference Pyramid in (1) for p=0 is presented with the matrix:

$$\left[\tilde{B}_0\left(2^n\right)\right]=\left[T_0\left(2^n\right)\right]^{-1}\left[\tilde{S}_0\left(2^n\right)\right]\left[T_0\left(2^n\right)\right]^{-1} \qquad (2)$$

where $[T_0(2^n)]$ is the matrix, of the used 2D linear orthogonal transform, in this case the Discrete Cosine Transform (DCT). There are infinity of function transform and the use of every transform depends on the approach. Since this study is about image comparison the applied transform is DCT which is a transform commonly used in image manipulation. It is also a non-loss transform meaning that if DCT is applied to an image and then we run over the IDCT (Inverse Discrete Cosine Transform) the result will be the entry image. It is possible to have a slight loss of information due to a computer approximation error but it is insignificant.

$$\left[\tilde{S}_0\left(2^n\right)\right]=m_0(u,v)S_0(u,v)$$
$$m_0(u,v)=\begin{cases}1 \text{ for } (u,v)=(0,0),(0,1),(0,2),(1,0),\\ \qquad (1,1),(1,2),(2,0),(2,1)\\ 0 \text{ elsewhere}\end{cases} \qquad (3)$$

The choice of the binary mask $m_0(u,v)$ is based on previous studies about the IDP algorithm efficiency [12]. S_0 (u, v) are the elements of the spectrum matrix, calculated in accordance with the transform:

$$\left[S_0\left(2^n\right)\right]=\left[T_0\left(2^n\right)\right]\left[B_0\left(2^n\right)\right]\left[T_0\left(2^n\right)\right] \qquad (4)$$

The values of the elements build the lowest layer (p=0) of the spectrum pyramid.

ii. The next components of the decomposition (1) for p = 1, 2, 3, ..., n-1 are represented with the matrix:

$$\left[\tilde{E}_{p-1}\left(2^n\right)\right]=\begin{bmatrix}\left[\tilde{E}^1_{p-1}\left(2^{n-p}\right)\right]\left[\tilde{E}^2_{p-1}\left(2^{n-p}\right)\right] & -- & \left[\tilde{E}^{2^p}_{p-1}\left(2^{n-p}\right)\right]\\ \left[\tilde{E}^{2^p+1}_{p-1}\left(2^{n-p}\right)\right]\left[\tilde{E}^{2^p+2}_{p-1}\left(2^{n-p}\right)\right] & -- & \left[\tilde{E}^{2^p+p}_{p-1}\left(2^{n-p}\right)\right]\\ --- & --- & ---\\ \left[\tilde{E}^{4^p-2^p+1}_{p-1}\left(2^{n-p}\right)\right]\left[\tilde{E}^{4^p-2^p+2}_{p-1}\left(2^{n-p}\right)\right] & -- & \left[\tilde{E}^{4^p+1}_{p-1}\left(2^{n-p}\right)\right]\end{bmatrix} \qquad (5)$$

$$\left[\tilde{E}^{k_p}_{p-1}\left(2^{n-p}\right)\right]=\mathrm{IDCT}\left(\tilde{S}^{k_p}_{p}\left(2^{n-p}\right)\right)$$
$$\text{for } k_p=1,2,\ldots,4^p \qquad (6)$$

$$\left[\tilde{S}^{k_p}_{p}(u,v)\right]=m_p(u,v)S^{k_p}_{p}(u,v)$$
$$m_p(u,v)=\begin{cases}1 \text{ for } (u,v)=(0,0),(0,1),(0,2),(1,0),\\ \qquad (1,1),(1,2),(2,0),(2,1)\\ 0 \text{ elsewhere}\end{cases} \qquad (7)$$

$$\left[E_{p-1}\left(2^{n-p}\right)\right]=\begin{cases}\left[B\left(2^n\right)\right]-\left[\tilde{B}_0\left(2^n\right)\right] \text{ for } p=1\\ \left[E_{p-2}\left(2^{n-p}\right)-\left[\tilde{E}_{p-2}\left(2^{n-p}\right)\right]\right]\\ \text{for } p=2,3,\ldots,n-1\end{cases} \qquad (8)$$

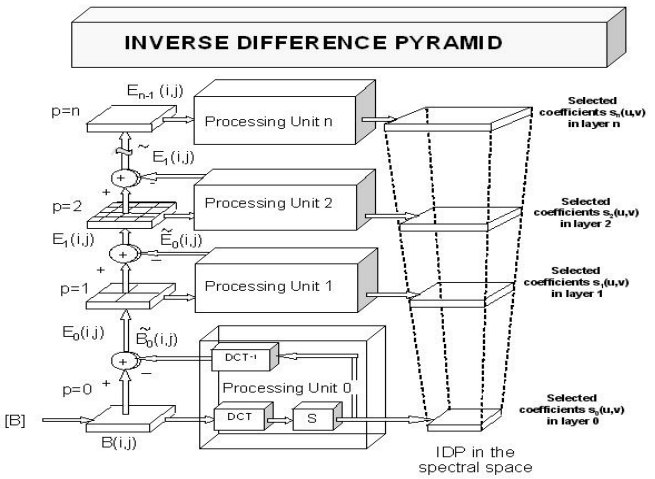


Fig. 2 IDP algorithm

C. Multilevel image distance

In this study the images are compared in the frequency domain (after the application of the DCT). As we know, not all the coefficients in the image-corresponding matrix in the frequency domain bare significance to the object information

i.e. most of them equate to 0. Furthermore we know where are located the 8 most important coefficients (top left corner).
Therefore when we compare two objects on a current pyramid layer comparing the two matrices for all terms is unwise because it would not give us further lore (and it also could provide expensive in terms of time and memory). Thus we could compare only 8 coefficients per entry block. This give us 8 comparisons for level 0, 8 x 4 coefficients for level 1, 8 x 4^2 for level 2, etc to the last level n where we have $8x4^n$. So the total length of the comparator vector is $8\times\frac{4^{n-1}}{3}$ coefficients.
The Multilevel Image Distance computation is done by levels. On each level is computed the difference between the coefficients of that level using the Euclidian distance. The difference of the coefficients of the query image and the current image from the database we compare to is then summed:

$$D_{r,q}\left[\widetilde{E}_{r-1,q}\left(2^{n-r}\right),\widetilde{E}^{k}_{r-1,q}\left(2^{n-r}\right)\right]= \\ =\sum_{u=0}^{2^{n-r}}\sum_{v=0}^{2^{n-r}}\left|m_{r,q}(u,v)\left[S_{r,q}(u,v)-S^{k}_{r,q}(u,v)\right]\right| \quad (10)$$

for q=1,2,…,4^r and r=1,2,…, n-1

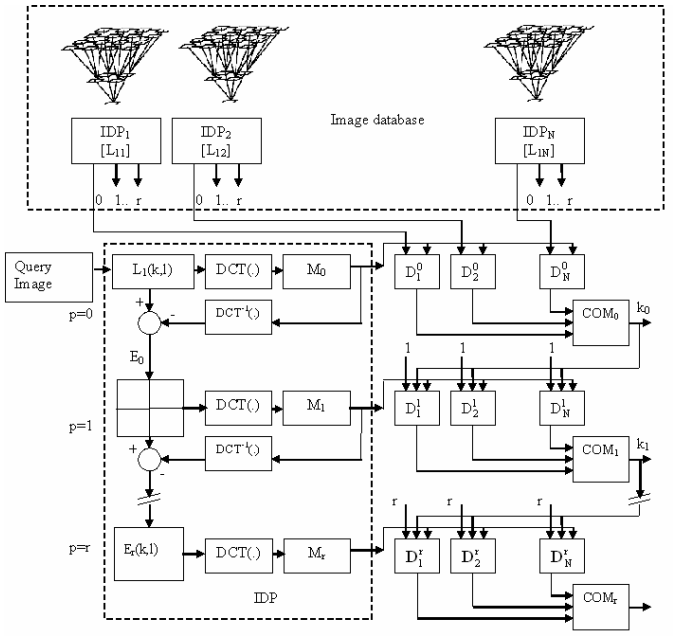


Fig. 3 Multi-layer coarse-to-fine similarity measure

D. CBIR algorithms based on IPD

1. The Incremental Algorithm

Image comparison can be expensive in terms of time and memory. Thus we need a flexible algorithm which focuses on similar images and ignores the rest of the database, enclosing the database to a smaller set at every step/level of the pyramid. Such algorithm can be an incremental algorithm:
If N is the total number of images in the database and we are looking for the 3 to 5 closest images:

a. Compute M_{in1} (a vector containing the N_1 images closest to the query image at level 1 where $N_1 < N$);

b. Compute M_{in2} (a vector containing the N_2 images closest to the query image at level 2 where $N_2 < N_1$);

……

n. Compute M_{inn} (a vector containing the $N^{(p)}$ images closest to the query image at level p-1 where $N^{(p)}< N^{(p-1)}$);
Compute M_{in} (the vector containing the 3 or 5 final images closest to the query image at level p)
The computational complexity of this algorithm is quite low. There are operations $\sum_{i=1}^{p}N^2(i)+2\sum_{i=1}^{p}N(i)$ at all, which is a polynomial algorithm cost.

2. The Prototype search

There are several possible approaches when choosing the prototype. In this study the prototypes were random; hence the database classes were highly homogenous.
Supposing that there are q image classes s.a. [C_1,…, C_q], we could choose couple of representative images par class (the prototypes, j << number images per class) called keys: [$K_1(C_1)$,…, $K_j(C_1)$]; [$K_1(C_2)$,...,$K_j(C_2)$]; …; [$K_1(C_q)$,…,$K_j(C_q)$]. Instead of comparing every query image to the whole database we could just compare them to the keys and keep the r most interesting classes (i.e. where the r << q) and then compare the query image with all the images belonging to the selected classes to find the closest matches [10].

III. Experimental Results

There were two image databases used in the study. The first database consisted in 301 color images divided in 4 classes and a second one - a much larger set of 4468 images grouped in 53 classes. For both the IDP algorithm and the comparison algorithms had to be chosen a suitable platform, the programming language is C++, we have decided to use the Qt4 toolkit and ImageLib for the FT functionalities.

On the big database were experimented different approaches of the incremental algorithm. The image retrieval to the last level (5 levels) proved to be not always necessary and furthermore not constantly the most precise. However this is easily explained by the diversity of the images in the database. The upper levels of the pyramid are far to detailed to give accurate information with a highly heterogeneous collection.
The results in Table I were obtained by testing every image in the database and an evaluation of the retrieval accuracy. Ever image was tested halting the incremental algorithm on five different levels from 0 to 4. The results show that it is not necessary to run the algorithm to the last level.

TABLE I
RETRIEVAL BY PYRAMID LEVELS

Image Level	0	0 and 1	0, 1 and 2
Retrieval accuracy	78 %	81 %	82 %

On Figure 4 some visual results are presented for three different levels. The first image is the query image and the number below the image is the distance between the current image and the query image.

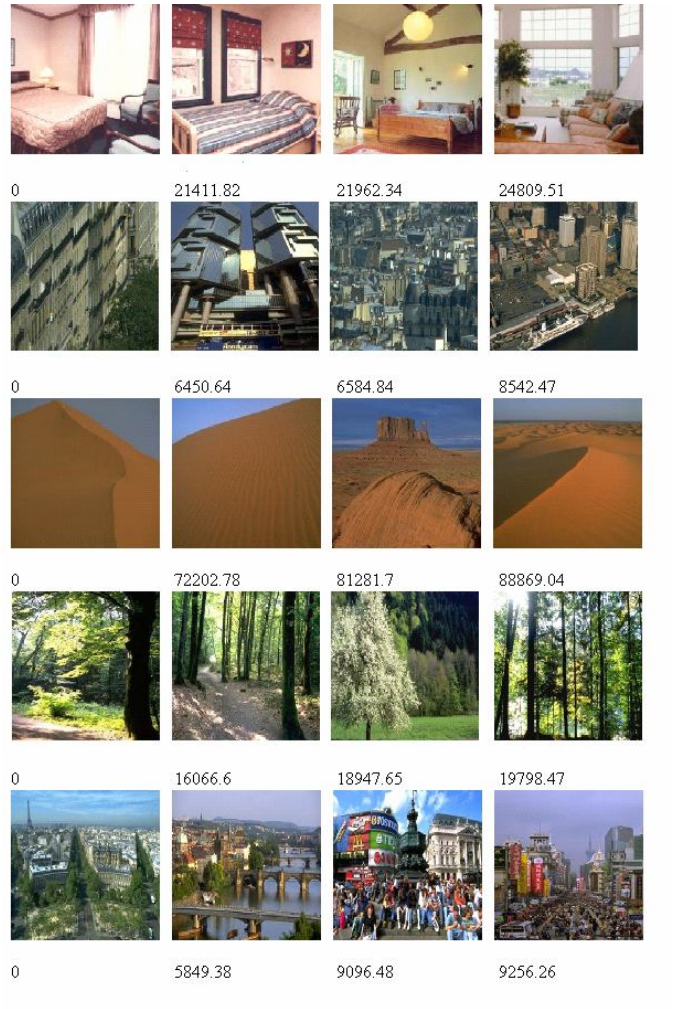


Fig. 4 CBIR results

IV. CONCLUSION

The presented method for efficient image retrieval is based on new image decomposition in the spectrum domain (the Inverse Difference Pyramid). The method permits the evaluation of the images similarity to be performed in the consecutive decomposition layers and to compare parts of the images with increasing resolution, which corresponds to different image scaling. The main advantage of the presented approach is the relatively low computational complexity and the capability to avoid difficulties, due to cropping or other kinds of image editing. All the retrieval computations are done in real time. The time for the retrieval of one query image and displaying the results for the small base without prototypes is 0.07s for 5 levels. The time for the same base but with a prototype search for the same levels is 0.002s. The time for the big base for five levels is 0.9s.

The search speed is one of the most important qualities of the presented method. As a result of the IDP decomposition the search process is significantly accelerated when compared with the well-known Exhaustive Search Algorithm (ESA) [10], for which the search is based on the use of a 3-dimensional color feature vector. For the evaluation were calculated the numbers of operations, necessary for the successful performance of the two algorithms.

Moreover, it proves that the computational complexity can be reduced using the incremental algorithm with a halt condition, or in a case of well-defined classes a prototype search.

All this puts the developed software on a market competitive level and qualifies it as industrially useful.

ACKNOWLEDGEMENT

The authors would like to thank the Gipsa-Lab for the support and the opportunity to make this study. There also should be special gratitude to Anne Guerin-Dugue for the guidance and the ideas. This work was financed by a project grant of the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science by the contract VU-I-305.

REFERENCES

[1] R. Kountchev, A. Manolova, “General Approach for Fast Image Retrieval from Image Database using Consecutive Iterations in the Transform Domain”, Proc. of the Technical University - Sofia, Bulgaria, Vol. 55, 2005, pp. 167-176.

[2] A. Gupta and R. Jain, “Visual information retrieval”, 1997

[3] A. Del Bimbo, *Visual Information Retrieval*, Academic Press, San Francisco, CA, 1999.

[4] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content based image retrieval at the end of the early years”, IEEE Trans. on Patt. Anal. and Machine Intelligence, 22, 1349–1380, 12.2000.

[5] T. Gevers, *Principles of Visual Information Retrieval*, M. Lew, Ed., ch. Color in image search engines. Springer-Verlag, Heidelberg, February 2001, 11–48.

[6] R. Schettini, G. Ciocca, and S. Zuffi, *A Survey of Methods for Color Image Indexing and Retrieval in Image Databases*, ch. Color imaging science: Exploiting digital media, R. Luo and L. Mac Donald, Eds., JohnWiley & Sons, New York, 2001.

[7] N. Sebe and M. Lew, *Texture Features for Content-Based Retrieval*, ch. Principles of visual information Retrieval. Springer-Verlag, Heidelberg, 2001.

[8] D. Hoiem, R. Sukthankar, H. Schneiderman, and L. Huston, “Object-based image retrieval the statistical structure of images”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, IEEE Computer Society, Vol. 2, June 2004, pp. 490–497.

[9] Andrew P. Berman, Linda G. Shapiro, “Efficient Content-Based Image Retrieval: Experimental Results”, IEEE Workshop on Content-Based Access of Image and Video Libraries (1999).

[10] Cawkell Tony, “Image indexing and retrieval by content”, Information Services And Use 20, no. 1 (2000): 49-58.

[11] Roumen Kountchev, Veronique Haese-Coat, Joseph Rosnin, *Inverse Pyramidal Decomposition with multiple DCT*, Signal processing: Image communication 17 (2002) 201-218 Elsevier 2002.

[12] Meijering, Erik, “A chronology of interpolation: from ancient astronomy to modern signal and image processing”, Proceedings of the IEEE 90 (3): 319342, DOI 10.1109/5.993400 (2002).

Automatic Letter Style Recognition of Churchslavic Manuscripts

Mimoza Klekovska[1], Igor Nedelkovski[2]
Vera Stojcevska-Antic[3] and Dragan Mihajlov[4]

Abstract – **In this paper the research on the churchslavic alphabet recognition in old hand-written papers is presented. The known OCR and ICR recognition methods in both printed and hand-written texts are analyzed. Their advantages and disadvantages are calculated by accepting some of their advantages and finally, two new modification methods are suggested. Churchslavic alphabet has a dual nature: it is a hand-written alphabet, but it looks like a printed one.**

Keywords – **Churchslavic alphabet, OCR, ICR, Picture processing, Font, Geometry analyses, ICEEST 2009.**

I. INTRODUCTION

Paperless office, where paper documents are copied onto modern recording media such as CD/DVD, is no doubt an intention in the modern world. Many OCR (Optical Character Recognition), ICR (Intelligent Character Recognition) programs are developed accordingly. There are two basic methods used by those programs: a Matrix Matching and a Feature Extraction. The Matrix Matching compares what the OCR scanner is scanning as a character, comparing it with a library of character matrices or templates. The Feature Extraction is also known as Topological Feature Analysis or ICR. This method varies by how much "computer intelligence" is applied by the particular manufacturer. The computer looks for general features such as open areas, closed shapes, diagonal lines, line intersection etc. The newest versions of the OCR programs support a lot of natural (human) languages (about 180). The churchslavic language or alphabet is not included in those 180 languages. It is a "dead" alphabet not in active use today, but many libraries in the world keep a lot of papers written by this alphabet. There are common situations where two sheets of a particular old book are in a library in, say, Leningrad, 10 sheets in a library in Sofia, in France… The linguists in this field need to exchange information. They use microfilms, photos, descriptions in edited publications etc. It is very difficult to have an access to the originals and normally it is not suitable to operate with such old originals, but it is better for the originals to be scanned and available in electronic format. The number of originals is quiet big and it is much better to keep them in text format. Reprinting old manuscripts in authentic form is a specific task, not very popular among computer engineers. Recognizing the manuscripts and treating them by the actual OCR programs is not popular too. The intention of this work is to initiate a process of including the churchslavic alphabet in the list of OCR programs.

Maybe this is not an urgent problem; the artifacts will be the same in the years to come, but those papers are keeping humanity's previous intelligence for many centuries and should be kept as a guide and a memory for it's past and it's history.

II. THE NATURE OF ALPHABET

The churchslavic alphabet is a handwritten one, but there are many differences in resolving the problems between other handwritten texts and this one. The most important difference is concerning the interconnection of the characters. The manuscript letters (characters) in other handwritten texts are usually joined or written connected (Fig. 1), but in the churchslavic manuscripts the letters (characters) are written separately (standalone), like in the printed texts.

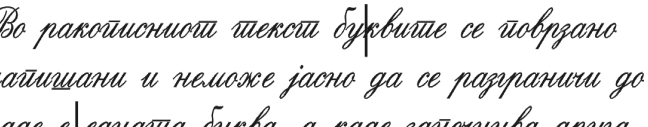


Fig. 1.Connected letters in handwritten manuscripts

So, the churchslavic manuscripts look like a printed document. But there are many additional problems in the handwritten manuscripts than in the printed texts.

Each normal printed text has the same h-high value for the letters in the font. It has a horizontal (or straight) line as the base line. The ascenders, descanters and capitals are always in the equal straight line. In the churchslavic alphabet these rules are not valid.

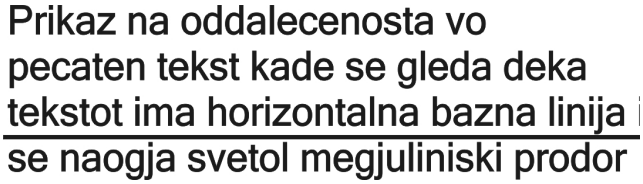


Fig. 2. Horizontal, straight "white" line in the printed texts

[1]Mimoza Klekovska is with the Technical Faculty of Bitola, address: ul. Ivo Ribar Lola bb, 7000 Bitola, Macedonia, e-mail: mimiklek@yahoo.com.mk

[2]Igor Nedelkovski is with the Technical Faculty of Bitola, address: ul. Ivo Ribar Lola, 7000 Bitola, Macedonia, e-mail: igor.nedelkovski@uklo.edu.mk.

[3]Vera Stojcevka-Antic is Professor Emeritus, e-mail: veranino@gmail.com.

[4]Dragan Mihajlov is with the Faculty of Electrical Engineering and Information Technologies of Skopje, address: Karpos bb, 1000 Skopje, Macedonia, e-mail: dragan@feit.ukim.edu.mk

It is possible to find a horizontal or nearly horizontal "white" line between the rows in the printed texts (Fig. 2).

That is not applicable in the churchslavic language (Fig. 3) because of two reasons.

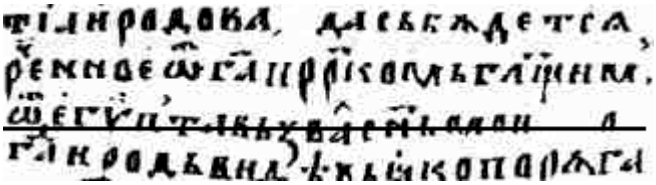


Fig. 3. There is no horizontal straight “white” line

The first one is the orthographic rule to put many upper letters and upper signs (even upper sign above upper letter) in the space between "normal" rows. The second one is because of very frequent use of initials (Fig. 4) or other decorative elements at the beginning of the rows.

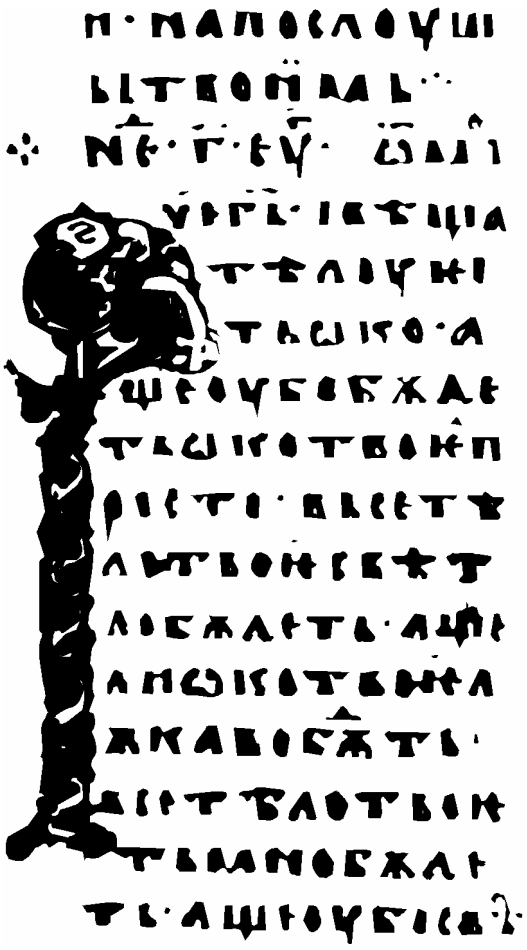


Fig. 4. Initial letter extended over many rows

The decorative elements are extended over many rows (between 5 and 12) and connect the rows in a block.

The decorative element should be cut and deleted to continue the processing intended to find the rows.

In modern printed texts there are spaces between the words; in some cases there is also a thin white space even between the letters in the word (Fig. 5).

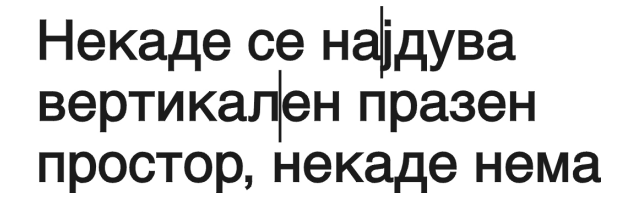


Fig. 5. White space between words, somewhere between letters

The churchslavic rule is "scripta continua", meaning that there is no space between words (Fig. 6). It is not known where the previous word ends and where the next word begins. This produces a problem in searching the whole word in the corresponding dictionary, like other OCR programs do.

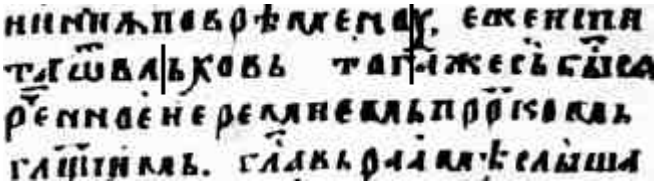


Fig. 6. “Scripta continua” rule in churchslavic manuscripts

But there are still some advantages, used in recognizing printed and handwritten texts that can also be used in processing the churchslavic alphabet. Preprocessing the picture, cleaning it, correcting the slant angle are common problems already treated by the scientific articles of other authors

The aim is to continue in analyzing the geometrical proportions of churchslavic alphabet letters. The decision was made to work in a graphic mode. Monitor's surface is the working area (Fig.7). In that way there is always a visual touch with the evaluation of the author's guiding idea. The Berger's thought: “Seeing comes before speaking” was the driving idea in choosing this operating mode. The pixel is the smallest working measurement. The starting base can be a page of any churchslavic manuscript. The accepted criteria was that the scanned pictures to work with have to be with the same resolution.

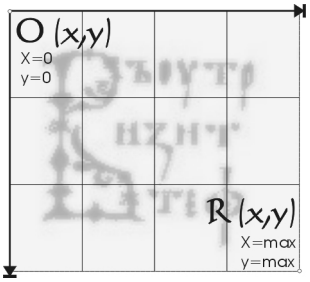


Fig. 7. Rectangular coordinate system over the manuscript

The upper left corner of the manuscript shown on the screen is at the coordinate position (0, 0) in the rectangular coordinate system and the lower right corner at the maximum position (x, y). In determining the average height of the row, the number of rows on the page as a data in the analysis is entered.

The first step was to make a classification of the manuscripts according to their appearance within the

historical period between the 10-Th and the 18-Th century. By some logic, the estimate was that different art styles should have a trace in the written style of the particular century. But it was encountered that there are no dramatic style differences during that period. The next step was to follow the logic of interpolating the manuscripts by the style of scriptor's font (handwriting font). It is similar to the today's light, normal, bold or italic font style. Graphologists would say: "It depends on the writer's personality".

The basic determining measurement is the h-high value. The number of rows on the page based on the value on the y-axis in the letter matrix (in pixels) will direct the activities to calculate h-high of the text row in the range between the minimum and the maximum limit.

The first black pixel on the picture is identified. Then the contour starting with that pixel is followed (tracked down). The number of pixels is counted and their x and y coordinates are compared. The smallest rectangle keeping the form of the sign-letter matrix is determined. High or y_value of the matrix is indicating that either the sign is a valid letter or not. If it is too high there is probably an initial letter or accidentally connected rows. If it is too small, that sign is probably upper letter or upper sign. If it is a valid letter, it is transferred in a new domain to keep it for further analyses of their characteristics; otherwise it is deleted. At the end of this process only the proper signs (letters) on the page (Fig.8) are kept. Initials, other decorative elements, upper letters, upper signs and some undefined small features are removed.

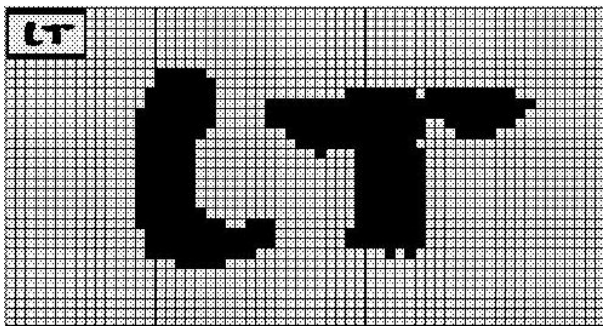


Fig. 8. Proper letters in the matrix are kept

Then each matrix of every particular letter that was preserved (memorized) as a valid sign is analyzed. The fields of interest are:

- Number of matrix rows,
- Number of matrix columns,
- Number of total pixels in the matrix (square surface),
- Number of black pixels and its percentage according to the total pixel numbers,
- X : y ratio,
- Y : x ratio,
- Whether the matrix is symmetrical or not,
- Ratio between contour length and square surface,
- Number of "dictuses" in the letter - primitive features ("dictus" is the trace (or black pixels) from the moment when the pen have touched the paper until the moment when the writer picked up the hand (the pen)),
- Dictus positions (horizontal, vertical, upper-left slanting stroke, upper-right slanting stroke or undefined),
- The weigh (thickness) of the dictus track (line),
- The existence of white holes inside the contour,
- The existence of high percentage black pixels in some specific topological areas in the matrix, such as several left side columns or several upper rows.

38 strings (sets) of signs were produced, because there are 38 letters in the churchslavic alphabet. Every set contains 100 members (elements, samples) of a sign "a" or "b" and so on.

By using the statistical methods the most frequent (common) characteristics of all the samples belonging to the string are defined, keeping those characteristics as a "master letter". Additionally, this "master letter" represents a template in the process of recognizing other letters.

III. LABYRINTH METHOD

By analyzing the style of every particular letter in the churchslavic alphabet the decision criteria for every sign about what letter it is determined. Although each scriptor has his own writing style, there are some rules - the main characteristics of the letter that every scriptor has to respect. For example, the letter "i" will always be with high percentage in black pixels, no matter who wrote it. The most dominant value in geometry analyses of letters is the first criteria in the decision system of the character recognition treated. By the geometry analyses the expectation is to obtain the set of unordered values. To make an order, the values in a range determined during the work should be "normalized" (Fig.9). The H-high is a starting position. All the obtained values are transformed, expressed as a number of pixels, in a system measure that is expressed in a multiple of number 12. The h-high value is either 12 (or a number in a ratio (proportion) of 12, like: 6, 24, 36, ...144 ...) so as to find a proper coefficient in other measurements.

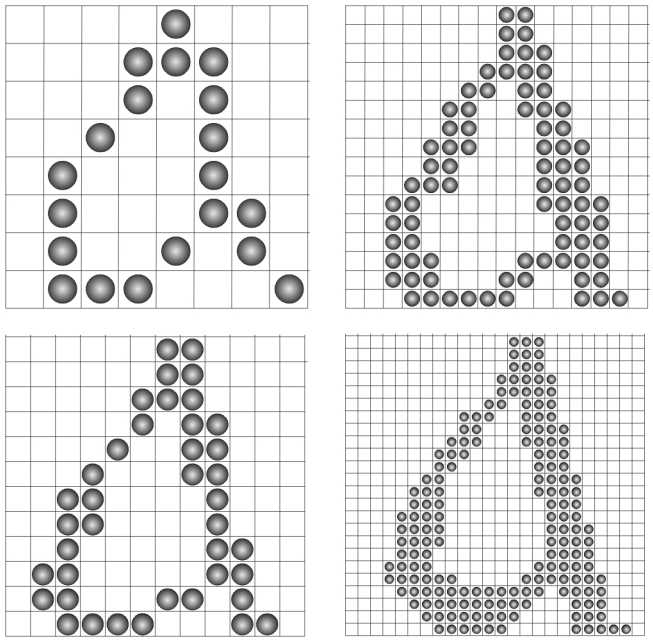


Fig. 9. "Normalizing" the values in a range

The "Labyrinth" method is conceived as a method that can be continually upgraded and refined during it's extended use and testing. It is a system of filters and elimination that, based on the total set of 38 characters in every filter, are decreasing the possible number of correct answers it the subset containing smaller number of members-letters. Every query, every criteria represents one filter determining whether the sign is accepted or rejected from the membership in that set. The sign-letter can in the same time be a member of several subsets. When the tests are made by all the criteria, the result is obtained as a combination or the frequency of "membership" of the particular letter in different subsets. If a situation arises when two or more letters have the same frequency in the subsets by all criteria, the two or more results are recommended.

This is a method that is operating by the matrices and is mostly using a mathematical logic and use of logical operators NOT, OR, AND, XOR, i.e. it's graphical application in the value tables of truth. Mathematically, the logical operators are shown in the table by the values of 0 and 1. The graphic interpretation of those values represents a black or white dot (Fig.10 - First column), i.e. the presence or absence of a pixel.

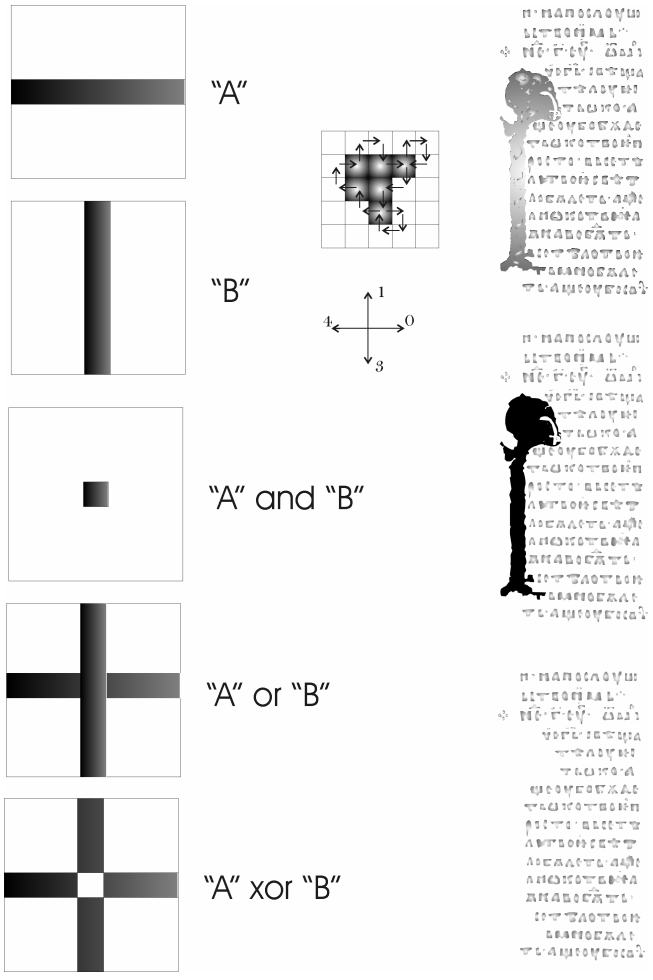


Fig. 10. First column - Graphic result of logical operators application. Second column - An example of contour tracing and an outline of a code . Third column - An outline of the initial elimination.

The programming method of tracing the contour is applied for the sake of limiting (fencing off) the white area from the black area. The graphic interpretation of this method for one approximate black-and-white fenced area is given in the outline (Fig.10 - Second column), together with the selected codes for the direction, the course of movement around the contour.

When one contour will be rimed, the presence of "black dots" in it is counted. Then, the function of filling or replenishment is used. By this function we are sure that all the dots (pixels) that are inside the contour are black. Its recount is done again. If the result (the number of black pixels) in the first and in the second case do not differ dramatically, then there is no presence of white areas or holes in the letter. If there is a real numeric difference (larger than the predefined value of tolerance) in the first and second counting, then we diagnose a presence of white area (hole) in the letter. Besides the counting, we can also make the presence of white holes in the letter by the combined use of logical operators and by filling the contour.

We also use the contour tracing and filling in the graphic elimination of the initial signs of the manuscript, which are not subject of our domain of work (Fig.10 – Third column).

Because the work is in the initial preparation phase, it is hardly possible to talk about satisfactory recognition results. An initial schemata of decision making based on some of the mentioned criteria, can give satisfactory results for one experimentally created font; but for the real world manuscripts additional branching and broadening of this schemata is necessary.

IV. Conclusion

A schematic construction suitable for recognizing the text written by standalone churchslavic characters is presented in this paper. If there are some accidentally connected letters they are treated as initial letter or some decorative element(s) and are extracted as a sign that is not a subject of interest. As a starting point, the effort is to concentrate the research toward 38 main letters in the churchslavic alphabet. The supposition in this work is that there is a clean paper with no slant written letters, because other authors are treating those additional problems. The second groups of additional problems, like upper letters, upper signs, decorative elements or stylized crosses are not currently treated.

References

[1] Chen Tsun Chuang and Lin Yu Tseng, "A Heuristic Algorithm for the Recognition of Printed Chinese Characters", IEEE Transactions on Systems, Man and Cybernetics, vol. 25, no. 4, pp. 710-717, 1995.
[2] David Earls, "Designing Typefaces", Roto Vision, 2002.
[3] Vera Stojcevska Antic, "Makedonska srednovekovna knizevnost", Skopje, 1997.
[4] http://www.dataid.com/aboutocr.htm

Session

SIGNAL PROCESSING – 3

A Genetic-Based Algorithm for the Design of Multiplier-less Halfband IIR Filters

Valentina Markova [1], Juha Yli-Kaakinen [2], and Tapio Saramäki [3]

Abstract – **This paper describes an efficient algorithm for designing multiplierless halfband IIR digital filters. The coefficient optimization is performed in two steps. The proposed procedure gives finite-precision solutions requiring fewer computations than previously reported implementations.**

Keywords – **IIR digital filters, halfband filters, multiplierless filters, genetic algorithms.**

I. INTRODUCTION

Halfband filters are important for applications to multirate digital signal processing. A halfband filter satisfies the passband and stopband symmetry conditions [1], [2]. Both infinite- impulse response (IIR) and finite-impulse response (FIR) filters can be used to realize these filters [3]. For applications requiring exact linear phase, FIR filters are used [4]. However, when the phase requirement is not as strict the IIR filters are the best choice as they require considerably fewer coefficients to meet the given magnitude criteria than their FIR filter counterpart [5], [6].

Odd-order IIR halfband filters can be implemented as a parallel connection of two allpass filters. These filters have turned out to be very efficient for constructing filter banks since all the computations can be performed at the lower sampling rate. Furthermore, this filter class is characterized by low coefficient sensitivity. The importance of such a structure is that if the effect of coefficient value deviation from the ideal value is small, then the short coefficient wordlength can be used with only slightly violating the infinite-precision filter specifications, resulting in a faster, smaller and less expensive hardware [7], [8].

In this paper an algorithm for designing elliptic halfband IIR digital filters with short coefficient wordlength is introduced. This algorithm is based on the following observation: Finding two elliptic halfband filters, one of which has the minimized transition bandwidth and the second one the maximized stopband attenuation such that the given criteria are still met enables one to generate a parameter space including the feasible space where the filter specifications are satisfied. After determining this larger space, all what is needed is to check whether in this space there exist the desired discrete values for the coefficient representations. In order to reduce the computational complexity, a genetic algorithm is applied for finding the solutions meeting the specifications within the parameter space. This strategy is general but particularly efficient for filters implemented as a parallel connection of two allpass filters due to the fact that for these filters only the denominator coefficients of the allpass sections have to be quantized. Furthermore, these coefficients are represented as simple combinations of powers of two, thereby provide a low complexity halfband filter. In this paper, we suggest the multiplierless design with considerably smaller implementation cost of the filter than previously reported realizations.

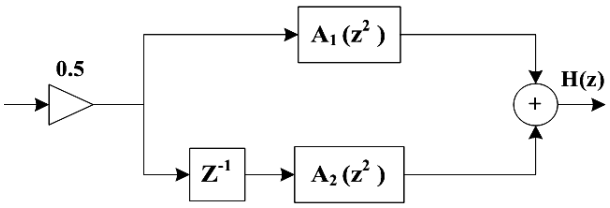


Fig 1. Realization of an odd-order halfband IIR filter.

II. PROPERTIES OF ODD-ORDER HALFBAND ELLIPTIC IIR FILTERS

It is well-known that odd-order lowpass halfband elliptic IIR filters are characterized by the following properties and restrictions (see, e.g., [2]). First, the design criteria for these filters should be stated as

$$\begin{aligned} 1-\delta_p \le \left|H(\Phi, e^{j\omega})\right| \le 1 \quad &\text{for} \quad \omega \in [0, \omega_p] \\ \left|H(\Phi, e^{j\omega})\right| \le \delta_s \quad &\text{for} \quad \omega \in [\omega_s, \pi] \,, \end{aligned} \tag{1}$$

where ω_p and ω_s are restricted to be related through

$$\omega_p + \omega_s = \pi \,, \tag{2}$$

whereas the restrictions between δ_p and δ_s are expressible in terms of $A_p = -\log_{10}(1-(\delta_p)^2)$ and $A_s = -\log_{10}((\delta_p)^2)$, which are called as the passband variation in decibels and the stopband attenuation, respectively, as

$$A_p = 10\log_{10}\left(1+\frac{1}{10^{A_s/10}-1}\right) \tag{3}$$

or

$$A_s = -\log_{10}(1-10^{-A_p/10}) \,. \tag{4}$$

Secondly, due to the fact the poles of the filter are restricted lie on the imaginary axis for a real-valued and stable transfer function, this function can be written as a parallel connection of two all-pass filters, according to Fig. 1, as

[1]Valentina I. Markova is with the Faculty of Telecommunications at Technical University of Varna, 1 Studentska Str, Varna 9010, Bulgaria, E-mail: via@tu-sofia.bg.

[2]Juha Yli-Kaakinen was with the Department of Signal Processing at Tampere University of Technology, Finland He is now with VLSI Solution Oy, FIN-33720 Tampere, Finland, E-mail: juha.ylikaakinen@pp.inet.fi.

[3]Tapio Saramäki is with the Department of Signal Processing at Tampere University of Technology, P.O. Box 553, FIN-33101 Tampere, Finland, E-mail: ts@ cs.tut.fi.

$$H(z)=\frac{1}{2}\left[A_1(z^2)+z^{-1}A_2(z^2)\right], \quad (5)$$

where $A_1(z)$ and $A_2(z)$ can be expressed as

$$A_1(z)=\prod_{\ell=1}^{m}\frac{-c_\ell+z^{-1}}{1-c_\ell z^{-1}} \text{ and } A_2(z)=\prod_{\ell=m+1}^{m+n}\frac{-c_\ell+z^{-1}}{1-c_\ell z^{-1}}. \quad (6)$$

Because this contribution focuses on implementing $A_1(z)$ and $A_2(z)$ as a cascade of Gray-Markel lattice allpass sections [9], they are expressed in the above form. Equally well, the above form suits to be implemented as a lattice wave digital filter [6].

Thirdly, when considering Eqs. (5) and (6), the following attractive observations are made: 1) The second allpass filter $z^{-1}A_2(z^2)$ of $H(z)$ has a first-order section that is a pure delay. 2) Both allpass filters contain second-order terms that are obtained from a first-order sections in Eq. (6) by replacing z^{-1} by z^{-2}. These observations lead to the transfer function of order $2(m+n)+1$, but it requires only $m+n$ multipliers. The orders of the first and second allpass filters are $M=2n$ and $N=2m+1$, which should differ by one. This implies that either $n=m$ or $n=m+1$.

III. Statement of the Problem

Before stating the optimization problem, the transfer function of the halfband filter is denoted for later use by $H(\Phi, z)$, where Φ is the vector containing the adjustable filter parameters c_ℓ for $\ell=1,2,\ldots,m+n$. According to the discussion of Section II, given the minimum stopband attenuation A_s and the stopband edge angle ω_s, the magnitude specifications for the odd-order halfband elliptic filter are given as

$$\begin{aligned} 0 \le -20\log_{10}\left|H(\Phi,e^{j\omega})\right| &\le -10\log_{10}\left(1+\frac{1}{10^{A_s/10}-1}\right) \\ &\text{for } \omega\in[0,\pi-\omega_s] \\ -20\log_{10}\left|H(\Phi,e^{j\omega})\right| &\ge A_s \quad \text{for } \omega\in[\omega_s,\pi]. \end{aligned} \quad (7)$$

This work concentrates on coefficients quantization in fixed-point arithmetic. In VLSI implementations, where general multipliers are very costly, it is attractive to carry out the multiplication of a data sample by a filter coefficient value using a sequence of adds (subtracts) and shifts. For such a purpose, the coefficient values are expressed as

$$\sum_{r=1}^{R} a_r 2^{-P_r}, \quad (8)$$

where each a_r is either 1 or –1 and the P_r's are nonnegative integers in the increasing order. In this case, the aim is to find all the coefficient values so that, first, R, the number of powers-of-two terms, is made as small as possible, and, secondly, P_r, the maximum number of shifts, is made as small as possible.

An estimate for the implementation cost of the filter can be calculated as a sum of the number of the adders and subtracters used to implement all the filter coefficients, that is, the cost is given by

$$\sum_{l=1}^{m+n}\sigma_\ell, \quad (9)$$

where the σ_ℓ's are the number of adders and subtracters required to implement the filter coefficients c_ℓ.

The optimization problem under consideration is the following:

Optimization Problem: Given A_s and ω_s, find m and n, and the adjustable parameter vector Φ such that, first, the criteria of Eq. (7) are met after quantizing the coefficient values corresponding to the parameters included in Φ to achieve the above-mentioned form for their representations and, then, the implementation cost, as given by Eq. (9), is minimized.

IV. Filter Optimization

The solution to the stated optimization problem can be found in the following two-step procedure. In the first step, the stopband edge angle is minimized such that the resulting stopband attenuation achieves just the specified value in Eq. (7) and the stopband attenuation is maximized such that the resulting stopband edge angle achieves just the specified value in Eq. (7). This enables one to find the parameter space of the infinite-precision coefficients including the feasible space where the filter meets the requirements. The second step involves finding the filter parameters in this space using a genetic algorithm such that the resulting filter meets the given criteria with the simplest coefficient representation forms.

A. Optimization of Infinite-Precision Filters

It has been turned out that the desired parameter space for the filter coefficients can be conveniently generated by designing two infinite-precision elliptic halfband filters, which satisfy the specifications as follows:

Design 1: The stopband edge angle is minimized in the criteria of Eq. 7 such that the stopband attenuation reaches just the specified value.

Design 2: The stopband attenuation is maximized in the criteria of Eq. 7 such the stopband edge angle reaches just the specified value.

These designs can be obtained by using simple closed-form algebraic expressions [3]. The parameter vectors containing the optimal infinite-precision filter parameters for Designs 1 and 2 are denoted by $\Phi^{(1)}$ and $\Phi^{(2)}$, respectively, whereas the corresponding sets of the coefficients values for the filter transfer function under consideration are denoted by $c_\ell^{(1)}$'s and $c_\ell^{(2)}$'s. Based on these set of coefficients, the smallest and largest values for the filter coefficients can be determined as

$$c_\ell^{(\min)}=\min\left\{c_\ell^{(1)},\, c_\ell^{(2)}\right\} \text{ and } c_l^{(\max)}=\max\left\{c_l^{(1)},\, c_l^{(2)}\right\} \quad (10)$$

for $\ell=1,2,\ldots,n+m$.

B. Optimization of Finite-Precision Filters

It has been experimentally proved that the parameter space defined above forms a very good approximation for the feasible space where the filter criteria are met. After finding this parameter space, all what is needed is to find in this space the existing combinations of the discrete coefficients values which satisfy the given requirements.

This search can be done in a straightforward manner [6], [7] by first finding the sets of powers-of-two numbers C_l for $\ell = 0, 1, \ldots, n+m$ between the smallest and largest values of each coefficient ($c_l^{(\min)}$ and $c_l^{(\max)}$).

The magnitude response is then estimated for each combination of the $c_l^{(k)}$ for $\ell = 1, 2, \ldots, n+m$ and $k = 0, 1, \ldots, K_\ell$ to check whether the filter meets the given criteria. Here, the number of powers-of-two values between $c_l^{(\min)}$ and $c_l^{(\max)}$ is denoted by K_ℓ, whereas the kth existing discrete value between these smallest and largest values is denoted by $C_l^{(k)}$ for $k = 0, 1, \ldots, K_\ell$. The number of discrete coefficient value combinations is thus given by

$$\prod_{\ell=1}^{n+m} K_\ell \,. \tag{11}$$

The number of discrete coefficient value combinations can be huge. For this reason it is beneficial to use a genetic algorithm for searching those discrete coefficient values with which the specifications are satisfied. This discrete-valued optimization problem can be efficiently solved by utilizing a genetic algorithm as follows: First, the indexes of the power-of-two numbers between the smallest and largest values of the coefficients are represented using a binary code. Furthermore, a lookup table containing the power-of-two values of the corresponding indexes is generated. The next step is to construct the chromosomes by concatenating all these binary strings. At the end, the fitness of the population is evaluated by decoding the chromosomes to their corresponding power-of-two coefficients values using the above-mentioned lookup table.

The fitness function to be maximized is given by

$$f = -\max\left\{\Delta_p / \delta_p, \Delta_s / \delta_s\right\} , \tag{12}$$

where Δ_p and Δ_s are the realized passband and stopband ripples, respectively. The solution meeting the given criteria is obtained when f becomes greater than or equal to minus unity.

V. Design Example

Consider the following criteria for an odd-order IIR halfband filter [8]: $\omega_p = 0.44\pi$ and $A_s = 46$ dB. The rest of the filter criteria are found using Eqs. (2) and (3), which give $\omega_s = 0.56\pi$ and $A_p = 1.1 \cdot 10^{-4}$ dB. These specifications are met by a ninth-order elliptic halfband IIR filter.

The infinite-precision coefficient values of the two elliptic halfband initial filter designs are given in Table 1. In this table, $\Phi^{(1)}$ and $\Phi^{(2)}$ are the optimal coefficient values for Designs 1 and 2, respectively. The corresponding magnitude responses for the initial filters are shown in Fig. 2.

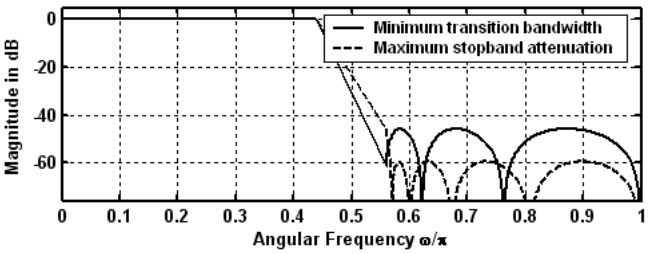


Fig. 2. Magnitude responses for the initial filters.

TABLE I
INFINITE-PRECISION COEFFICIENT VALUES FOR THE INITIAL DESIGNS

	$\Phi^{(1)} = \Phi^{(\min)}$	$\Phi^{(2)} = \Phi^{(\max)}$
c_1	−0.4653	−0.3440
c_2	−0.9204	−0.8686
c_3	−0.1535	−0.1022
c_4	−0.7342	−0.6145

For the corresponding finite-precision filter only three powers-of-two terms ($R = 3$) and eight fractional bits ($P_R = 8$) are required to fulfill the given criteria. The number of bits needed to encode all permissible discrete values between the smallest and largest values of c_ℓ for $\ell = 1, 2, \ldots, 4$ for this coefficient representation form are 5, 3, 4 and 4, respectively, that is, the length of the chromosome is 16. For more details of the above encoding, see [7].

TABLE II
PERFORMANCE OF THE GENETIC ALGORITHM

P_s =0.04				
P_x	P_m	f_{mean}	f_{std}	N_{hit}
0.6	0.04	−1.4871	0.4393	9
0.6	0.05	−1.1253	0.0882	18
0.6	0.06	−1.0929	0.1706	21
0.7	0.04	−1.2308	0.2788	11
0.7	0.05	−1.0834	0.1260	31
0.7	0.06	−1.0659	0.1129	26
0.8	0.04	−1.1977	0.1445	14
0.8	0.05	−1.1042	0.1723	28
0.8	0.06	−1.0820	0.1100	33
P_s = 0.045				
0.6	0.04	−1.2661	0.2876	10
0.6	0.05	−1.0956	0.183	21
0.6	0.06	−1.0674	0.1085	25
0.7	0.04	−1.2401	0.2764	11
0.7	0.05	−1.0705	0.0956	30
0.7	0.06	−1.0568	0.0873	35
0.8	0.04	−1.2035	0.2693	12
0.8	0.05	−1.0861	0.1189	26
0.8	0.06	−1.0603	0.1067	29

TABLE III
OPTIMIZED FINITE-PRECISION COEFFICIENTS VALUES OF THE ELLIPTIC HALFBAND IIR FILTER

$A_1(z)$		$A_2(z)$	
c_1	$-2^{-1}+2^{-3}$	c_3	$-2^{-3}+2^{-7}$
c_2	$-1+2^{-3}-2^{-8}$	c_4	$-2^{-1}-2^{-3}-2^{-6}$

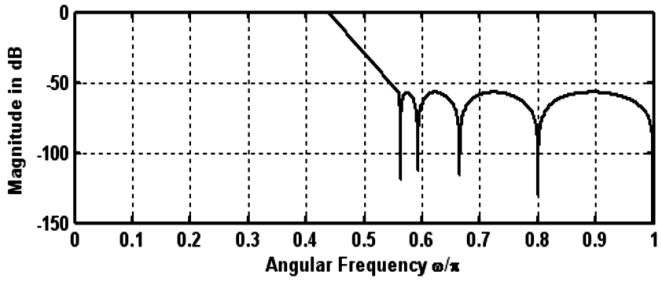


Fig. 3. Magnitude response of the resulting multiplierless halfband IIR filter.

The control parameters of the genetic algorithm such as crossover and mutation rates have been adjusted by, first, running this algorithm 100 times with different parameter settings and, then, selecting the most optimal ones. Some of them are shown in Table II. These results were obtained under the following circumstances. The normalized geometric selection was used as a reproduction operator, the population size was 150, and the number of generations was 250. In this table, P_s denotes the selection probability, P_x and P_m are the crossover and mutation rates, respectively, whereas f_{mean}, f_{std} and N_{hit} give, after these 100 runs, the mean fitness, the standard deviation of the fitness, and the number of solutions meeting the requirements, respectively. The fitness value of the best solution after these 100 runs was always the same that is, f = –0.932. The CPU time required for running one run was approximately 18 seconds when using the genetic algorithm optimization toolbox [10] in MATLAB 6.5 on a 3 GHz Pentium 4. The CPU-time required to evaluate all the possible coefficient value combinations was approximately 12 minutes.

For the optimized finite-precision filter in [8], four powers-of-two terms ($R = 4$) and eight fractional bits ($P_R = 8$) are required to meet the given filter specifications and the number of adders and/or subtracters needed to implement all the coefficients for the Gray-Markel sections is eight. The proposed algorithm results in the optimized filter, which requires only six adders and/or subtracters to implement all the multipliers. The optimized coefficient values are shown in Table III. The magnitude response of the resulting multiplierless halfband IIR filter is displayed in Fig. 3. The corresponding realization is shown in Fig. 4.

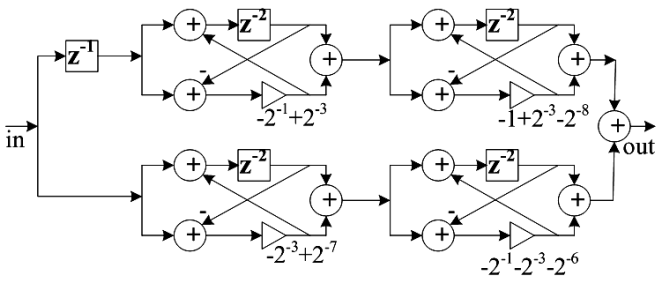


Fig. 4. Realization for the resulting multiplierless halfband IIR filter.

VI. CONCLUSION

A straightforward two-step scheme has been developed for designing multiplierless halfband IIR digital filters. The first step determines a parameter space of the infinite-precision coefficients including the feasible space where the filter meets the given criteria. The second step uses genetic algorithm for finding the coefficients in this space such that the given criteria are met by the simplest representation forms. The efficiency of the proposed procedure has been demonstrated by means of an example.

REFERENCES

[1] M. D. Lutovac and Lj. D. Milić, "Design of Multiplierless Elliptic IIR Halfband Filters and Hilbert Transformers," EUSIPCO'98, Conference Proceedings, pp. 291–294, Rhodes, Greece, Sept. 1998.

[2] M. D. Lutovac, D. V. Tošic, and B. Evans, *Filter Design for Signal Processing using Matlab and Mathematica,* Upperside River, New Jersey: Prentice Hall, 2000.

[3] H. W. Schüβler and P. Steffen, "Halfband Filters and Hilbert Transformers," Circuits, Systems, Signals Processing, vol. 17, no. 2, pp. 137-164, Feb. 1998.

[4] Y. C. Lim, Y. J. Yu, and T. Saramäki, "Optimum Masking Levels and Coefficient Sparseness for Hilbert Transformers and Halfband Filters Designed using the Frequency-Response Masking Technique," IEEE Trans. Circuits. Syst. I: Regular Papers, vol 12, no. 6, pp. 2444–2453, Nov. 2005.

[5] S. A. Samad, A. Hussain and D. Isa, "Wave Digital Filters with Minimum Multiplier for Discrete Hilbert Transformer Realization," Signal Processing Journal, pp. 3761–3768, 2006.

[6] J. Yli-Kaakinen and T. Saramäki, "A Systematic Algorithm for the Design of Multiplierless Lattice Wave Digital Filters," IEEE Trans. Circuits Syst. I: Regular Papers, vol. 54, no. 8, pp. 3761–3768, Aug. 2004.

[7] V. I. Anzova, J. Yli-Kaakinen and T. Saramäki, "An Algorithm for the Design of Multiplierless IIR Filters as a Parallel Connection of Two All-pass Filters" IEEE APCCAS 2006, Conference Proceedings, pp. 744–747, Singapore, 2006.

[8] M. D. Lutovac and Lj. D. Milić, "Design of multiplierless elliptic IIR filters with a small quantization error," IEEE Trans. Signal Processing, vol. 47, pp .469–479, Feb. 1999.

[9] A. H. Gray and J. D. Markel, "Digital Lattice and Ladder Filter Synthesis," IEEE Trans. Audio Electroacoust., vol. AU-21, no. 6, pp. 491–499, Dec. 1973.

[10] C. Houck, J. Joines, and M. Kay, "A Genetic Algorithm for Function Optimization: A Matlab Implementation," NCSU-IE Technical Report 95-09, North Carolina State University, 1995.

Investigations of Variable Fractional Delay Allpass Digital Filters

Kamelia Nikolova[1], Georgi Stoyanov[2] and Valentina Markova[3]

Abstract – **In this paper a comparative study of two approaches to realize variable fractional delay allpass digital filters based on Thiran approximation is carried. After investigations of the structure and design complexity, the worst-case sensitivity, the tuning accuracy in real time and the behavior in a limited wordlength environment it is shown that the method proposed by the authors is considerably better. All the results are verified experimentally**

Keywords – **Digital allpass filter, Fractional phase delay, Variable fractional delay filters, Worst-case sensitivity.**

I. INTRODUCTION

Fractional delay (FD) filters are very useful in numerous applications [1] in digital signal processing and in telecommunications including time delay estimation, timing adjustment in digital modems, precise jitter elimination, asynchronous sample rate conversion and speech processing. In most of these applications the fractional delay value has to be tuned in real time and this is explaining the recent popularity of the variable FD filters [2]-[4].

Variable FD filters can be designed as both FIR and IIR realizations. The Farrow structure is the most popular variable FIR FD filters structure [2]. It allows control of the desired fractional delay by a single parameter and can be successfully used in low frequency applications where due to small coefficients changes the obtained accuracy is sufficient. But, in the most practical cases (high frequency applications) there is a need of much higher transfer function (TF) order to obtain the desired accuracy. The main disadvantage of all the FIR FD filters is that both the magnitude and the phase responses are varying from the desired ones when tuning the fractional delay.

The main advantage of the variable FD filters based on allpass structures is the reduced design complexity (the unity gain for all the frequencies allows concentrating only on phase response characteristics). Still, the TF order of a variable IIR structure satisfying the same phase delay requirements is considerably lower than that of the corresponding variable FIR filter. Although, the variable IIR filters have some disadvantages (complicated design procedure, higher round-off noise, possible instability and worst behavior in a limited wordlength environment) which must be taken into account, they are often preferred because of their lower complexity (less multipliers and adders) and their lower overall delay. The gathering structure [3][4] is the most popular variable allpass FD realization. It is based on Thiran approximation [5]. The design method is simple to use and gives a closed-form solution for the TF coefficients. The structure is derived from the direct form structure with transfer function coefficients represented as polynomials of the fractional delay parameter. The drawbacks of this structure are the long critical path, the large wordlength requirement for VLSI implementation (the transfer function coefficient differ in one or a few order) and the higher sensitivity (as of any direct-form structure).

It is well known that lower sensitivity to the multiplier coefficients will ensure higher accuracy and shorter word-length representation for a given accuracy. High accuracy of the tuning parameters, on the other hand, will speed up the computation and the update of the new tunable TF coefficients in real time. In order to obtain a higher FD time accuracy, we have proposed in [6] a new second order allpass section (called IS) which has lower sensitivity for the specific TF poles positions of Thiran-based allpass FD filters (for the most practical applications around z=0). In [7] we have proposed a new design method for variable allpass-based Thiran FD filters. It is based on truncated Taylor series expansion of the TF coefficients.

In this paper we compare variable FD filters obtained through both design procedures and structures by investigating the worst-case sensitivity, the hardware complexity and the accuracy of realization and tuning in a limited wordlength environment. All theoretical results obtained in this work are verified experimentally.

II. VARIABLE FD WITH GATHERING STRUCTURE

Let an N-th order allpass IIR filter has the following transfer function:

$$H_{AP}(z)=\frac{a_N+a_{N-1}z^{-1}+\ldots+a_1z^{N-1}+a_0z^{-N}}{a_0+a_1z^{-1}+a_2z^{-2}+\ldots+a_Nz^{-N}}. \qquad (1)$$

The coefficients a_N can be determined by Thiran formulas [5] as

$$a_k=(-1)^k\binom{N}{k}\prod_{n=0}^{N}\frac{d+n}{d+k+n}, for\ k=0,1,2\ldots N. \qquad (2)$$

In order to eliminate the division with a time-varying factor which is difficult to implement in hardware or software (due to too many multiplication and division operations) the authors in [3][4] transform the coefficients a_k to a polynomial form:

The first two authors are with the Faculty of Telecommunications, Technical University of Sofia, 8 Kliment Ohridski Blvd, Sofia 1000, Bulgaria.

V. Markova is with the Faculty of Electronics, Technical University of Varna, 1 Studentska Str, Varna 9000, Bulgaria.

E-mails: [1]ksi@tu-sofia.bg; [2]stoyanov@ieee.org [3]via@tu-sofia.bg

$$\hat{a}_k = (-1)^k \binom{N}{k} \frac{\prod_{n=0}^{k-1}(d+n)}{\prod_{n=1}^{k}(d+N+n)} \prod_{n=1}^{N}(d+N+n)$$
$$= (-1)^k \binom{N}{k} \prod_{n=0}^{k-1}(d+n) \prod_{n=k+1\le N}^{N}(d+N+n) \quad (3)$$
$$= \sum_{l=1}^{N} \hat{e}_{lk} d^l, \text{ for } k = 1, 2 \ldots N.$$

Then an N-th order allpass IIR filter can be expressed as

$$H_{AP}(z) = \frac{g(d)[\hat{a}_N + \ldots + \hat{a}_1 z^{-(N-1)}] + z^{-N}}{1 + g(d)[\hat{a}_1 z^{-1} + \ldots + \hat{a}_N z^{-N}]}. \quad (4)$$

The normalized coefficients $g(d)$ are approximated using the truncated Maclaurin series as [3][4]

$$g(d) = \frac{1}{\prod_{n=1}^{N}(d+N+n)} \cong \frac{N!}{(2N)!} \prod_{n=1}^{N}\left[1 + \sum_{k=1}^{I}(-1)^k \left(\frac{d}{N+n}\right)^k\right], \quad (5)$$

where I is the order of the approximating polynomial in terms of d.

The obtained through this method structure (called "gathering structure") in case of second order Thiran-based FD allpass filter ($N = 2$) and second order Maclaurin approximation ($I = 2$) is given in Fig. 1. Third order Maclaurin approximation ($I = 3$) adds to the structure the elements given in the figure with dashed lines. The corresponding TF coefficients are given in Table I.

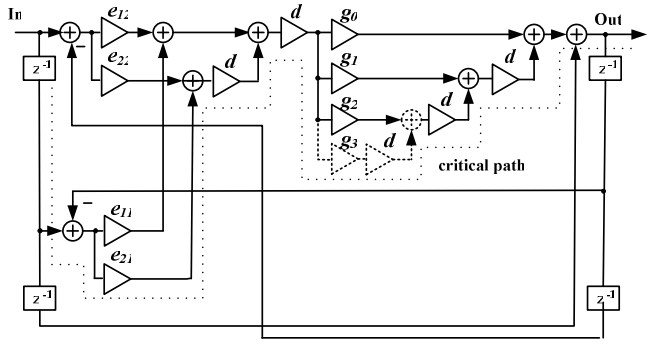


Fig. 1. Gathering structure for the variable FD allpass filter.

TABLE I
TF COEFFICIENTS OF GATHERING STRUCTURE

$\hat{a}_k$		$g(d)$
$\hat{e}_{11} = -8$	$\hat{e}_{12} = 1$	$g_0 = 0.083333$
$\hat{e}_{21} = -2$	$\hat{e}_{22} = 1$	$g_1 = -0.048611$
		$g_2 = 0.021412$
		$g_3 = -0.0084394$

As it can be seen the required wordlength is large because of the significant difference (about 10^2 for $N = 2$, $I = 2$ and about 10^3 for $N = 2$, $I = 3$) between the orders of the coefficients. The higher transfer function order (N) leads to a bigger difference between the smallest and the biggest coefficient (for example: the difference is about 10^5 for $N = 3$, $I = 3$).

III. VARIABLE FD REALIZATION WITH THE PROPOSED ALLPASS STRUCTURE

In order to eliminate the division with a time-varying factor we propose in [7] to express each transfer function coefficients a_k (2) as a truncated Taylor series expansion with respect to d and it was shown that a truncation after the quadratic term was improving considerably the accuracy. Here we expand this method to the following steps:

1. Select a proper order of the transfer function ensuring a required flat part of the delay response.
2. Obtain an allpass FD TF using Thiran approximation.
3. Depending on the TF poles positions, proper allpass sections are selected in a way to ensure minimal sensitivities.
4. Depending on the required tuning accuracy and the initial value of FD, expand each TF coefficient in Taylor series and truncate it after the linear (for adjustment of the phase delay), square or cubic (for tuning of the phase delay) term.
5. Realize all the multiplier coefficients as composite multiplier realizations (see Figs. 3, 4).

The proposed design procedure is simple to use and the obtained structures have no critical path. The method can be applied for an arbitrary TF order and in the cases of first and second order TFs it is possible to implement structures different from direct form and to minimize the sensitivity of the realizations.

The real applications require small FD parameter values (in the range $-0.5 < d < 0.5$) corresponding to TF poles situated in the area around $z = 0$. Our previous investigations for second order allpass FD transfer functions [6] demonstrate that for the poles situated in that area the IS structure proposed by us in [6] and shown in Fig. 2 is the best choice (with the lowest worst-case sensitivity and the best behavior in a limited wordlength environment). Its transfer function is

$$H_{IS}(z) = \frac{b + (-a - 2b + ab)z^{-1} + z^{-2}}{1 + (-a - 2b + ab)z^{-1} + bz^{-2}}. \quad (6)$$

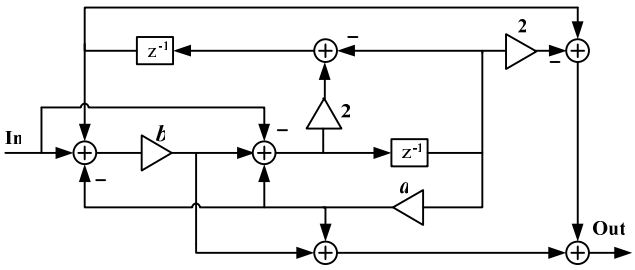


Fig. 2. IS allpass section.

The transfer function coefficients of the IS section realizing fractional delay with maximally flat group delay response can be expressed as

$$a = \frac{d}{d+2} \quad (7) \qquad b = \frac{d(d+1)}{(d+3)(d+4)}. \quad (8)$$

The representations of the coefficients a and b after using second and third order Taylor approximation are correspondingly:

$$a = \frac{1}{2}d - \frac{1}{4}d^2 \quad (9) \qquad b = \frac{1}{12}d + \frac{5}{144}d^2, \quad (10)$$

$$a = \frac{1}{2}d - \frac{1}{4}d^2 + \frac{1}{8}d^3 \quad (11) \quad b = \frac{1}{12}d + \frac{5}{144}d^2 - \frac{47}{1728}d^3. \quad (12)$$

All these coefficients have homogenous structure and can be realized as composite multipliers containing fixed and variable multipliers. The composite multiplier realizations for second and third order Taylor approximation are shown in Fig. 3 and Fig. 4.

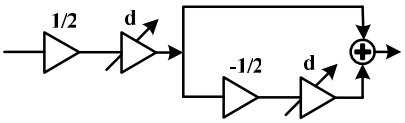


Fig. 3. Composite variable multiplier realization of a (7) after a second order Taylor approximation.

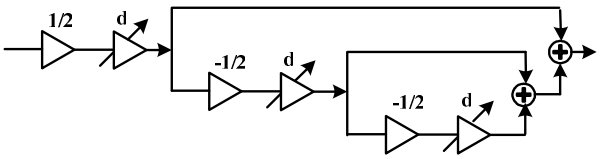


Fig. 4. Composite variable multiplier realization of a (7) after a third order Taylor approximation.

IV. COMPARATIVE STUDIES

We have investigated the phase sensitivity (using the package PANDA [8]) of the gathering and the proposed variable IS structure for first, second and third order Taylor approximation and for different FD parameter values in the range $-0.5 < d < 0.5$. The results for third order Taylor approximation ($I = 3$) and for some FD parameter values are given in Fig. 5. As it can be seen the sensitivity of the IS structure is approximately two times lower than that of the gathering structure for positive values of *d.* The same behavior of the sensitivity (as it is shown in the Fig. 5) is maintained for any of the cases (approximation order and FD parameter values) not given here.

In order to compare the tuning accuracy of both structures we consider the phase delay behavior in a limited wordlength environment for first, second and third order Taylor approximation of the coefficients. It was found that using the first order approximation for the two structures is not applicable because of the quite limited range of tuning.

The phase delay error of the two structures is larger than 0.01 of a single sample time-interval T near dc for any $d > |\pm 0.2|$. The investigations for the second order Taylor approximation show the possibility of tuning in much wider FD range with reasonably good results. The phase delay error is larger near $d = \pm 0.5$. A higher tuning accuracy is obtained by using third order approximation with the price of improvement readily acceptable in many practical cases. The results for the tuning of the phase delay for FD parameter values in the range of $-0.5 < d < 0.5$ in case of third order Taylor approximation and the TF coefficients quantization to 3 significant bits (in CSD code) are shown in Fig. 6. As it can be seen the proposed variable IS structure is behaving better than the gathering structure. The phase delay error of the gathering structure (in the cases shown in Fig. 6) for the worst case $d = 0.5$ and for the TF coefficient quantization to 3 significant bits is -0.0135T at dc while that of the proposed FD allpass filter is only 0.0098T. The difference is more significant for the same case and TF coefficients quantization to 2 significant bits (in CSD code) where the phase delay error of the gathering structure is -0.05T at dc while that of the proposed variable IS structure remains the same 0.0098T.

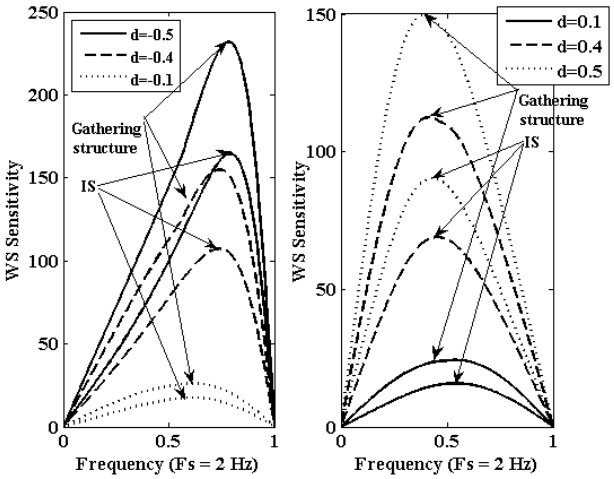
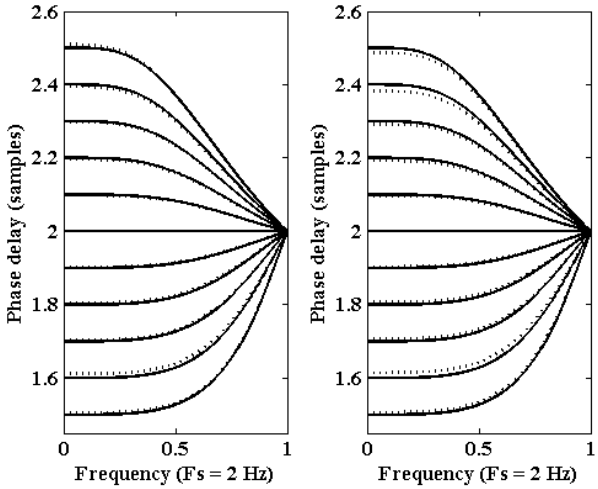


Fig. 5. Worst case sensitivity of the structures in case of third order Taylor approximation ($I = 3$).



a) variable IS structure b) gathering structure

Fig. 6. Tuning of the phase delay in case of third order Taylor approximation: ideal Thiran-based no variable non quantized FD filter (solid line); Thiran-based variable FD filters quantized to 3 bits (dotted line).

TABLE II
COMPARISON OF THE COMPLEXITY OF THE STRUCTURES

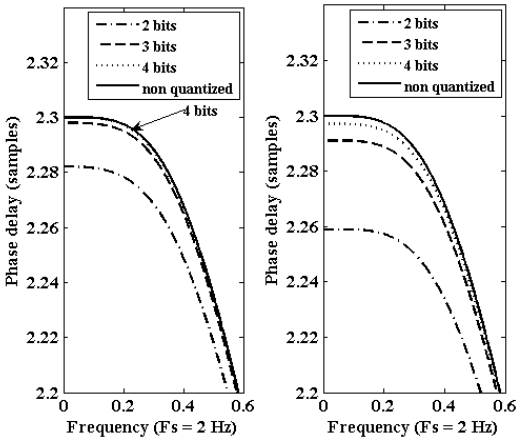
	Variable IS structure			Gathering structure		
	$I=1$	$I=2$	$I=3$	$I=1$	$I=2$	$I=3$
Multiplier	4	8	12	6	11	13
Adder	10	12	14	5	8	9
Delay element	2	2	2	4	4	4

The proposed variable IS FD structure has no critical path while the gathering structure has long critical path requiring 7 consecutive multiplications in case of second order TF and third order TF coefficients approximation. On the other hand, the complexity of the structure of the proposed variable IS FD filter is reduced, as it can be seen from Table II.

All these advantages of the proposed variable IS FD filter are very significant and they allow more precise tuning of the phase delay response in a limited wordlength environment obtained with much less elements (multipliers and delays).

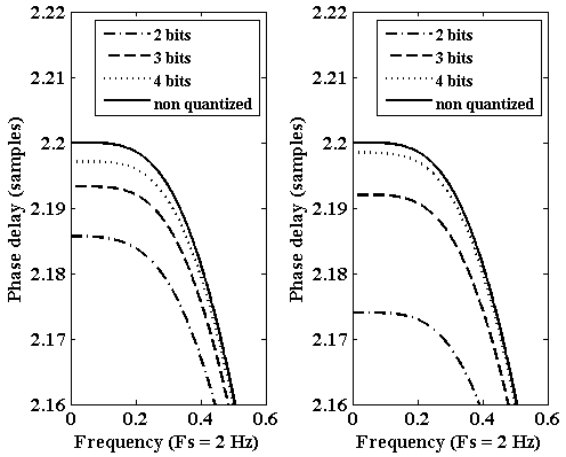
V. EXPERIMENTS

In order to observe and compare the tuning accuracy of the proposed variable IS FD and the gathering structure we have designed two second order allpass FD filters with third order TF coefficients approximation and a given fractional delay parameter value $d = 0.3$. The results are given in Fig. 7. Because of the low sensitivity of the IS structure the tuning accuracy is more precise than that of the gathering structure even when the TF coefficients are quantized to 2 significant bits (in CSD code). As it can be seen the deviations from the desired phase delay (0.3 samples) of variable IS FD filter near dc for 4, 3 and 2 bits are correspondingly smaller than 10^{-5}, -0.002 and -0.0179. The corresponding deviations of gathering structure are -0.0029, -0.009 and -0.289. The same behavior is observed for second order TF coefficients approximation as it is shown in Fig. 8 for a given fractional delay parameter value $d = 0.2$. Because of the lower sensitivities of the IS structure, its accuracy for different word-length is obviously better with exception of the case B=4 bit, due to the influence of the second-order Taylor coefficient approximation.



a) variable IS structure b) gathering structure

Fig. 7. Wordlength dependence of the accuracy of the phase delay for $d = 0.3$ in case of third order Taylor approximation.



a) variable IS structure b) gathering structure

Fig. 8. Wordlength dependence of the accuracy of the phase delay for $d = 0.2$ in case of second order Taylor approximation.

VI. CONCLUSION

A comparison of variable allpass FD filters obtained through two design procedures based on Thiran approximation is proposed in this paper. It was shown that the proposed method can be applied in the case of different from direct form first and second order structures and thus allows minimizing the sensitivity of the variable allpass FD filters. The obtained structure complexity is reduced and moreover, the obtained tuning accuracy is higher even in case of severe coefficients word-length representation.

ACKNOWLEDGEMENT

This work was supported by the Bulgarian National Science Fund under Grant No DO-02-135/15.12.08 and by the Technical University of Sofia (Sofia, Bulgaria) under Grant No 091НИ111-7/2009 of the University Research Foundation.

REFERENCES

[1] T. Laakso et al, "Splitting the unit delay–tools for fractional delay design", IEEE SP Mag., vol. 13, no. 1, pp. 30–60, Jan. 1996.

[2] C. W. Farrow, "A continuously variable digital delay element:" in Proc. 1988 IEEE Int. Symp. Circuits and Systems, Espoo, Finland, vol. 3, pp. 2641 - 2645, June 7-8. 1988.

[3] M. Makundi et al, "Efficient tunable IIR and allpass filter structures," Elect. Lett., vol. 37, no. 6, pp. 344–345, 2001.

[4] M. Makundi et al, "Closed-form design of tunable fractional-delay allpass filter structures," Proc. 2001 IEEE Int. Symp. Circ. and Systems, ISCAS 2001, vol. 4, pp. 434 -437, April, 2001.

[5] J.–P. Thiran, "Recursive digital filter with maximally flat group delay", IEEE Trans. CT, vol. 18, no. 6, pp. 659–664, Nov. 1971.

[6] K. Ivanova and G. Stoyanov, "A new low sensitivity second order allpass section suitable for fractional delay filter realizations", Proc. 8th Intern. Conf. TELSIKS'2007, Nish, Serbia, vol.1, pp. 317-320, Sept. 26-28, 2007.

[7] K. Nikolova and G. Stoyanov, "A new method of design of variable fractional delay digital allpass filters, 43th Int. Conference ICEST'2008, Nish, Serbia, Vol. 1, pp.75–78, June 25-27, 2008.

[8] N. Sugino and A. Nishihara, "Frequency-domain simulator of digital networks from the structural description", Trans. of the IEICE of Japan, Vol. E73, No. 11, pp. 1804-1806, Nov. 1990.

Band-Pass Loudspeaker Systems with Single Vent

Ekaterinoslav S. Sirakov [1]

Abstract – **In this work are presented theoretical analysis of characteristics of fourth order band-pass loudspeaker systems. Mathematical relationships are defined for calculating the electrical and acoustic characteristics of band-pass loudspeaker systems in linear modelling. Using the method of researching, for system is analyzed frequency, pulse and step characteristics.**

Keywords – **Band-pass loudspeaker systems, frequency curve, step and impulse response.**

I. Introduction

The Band-pass loudspeaker systems are known for a long time ago [11]. Band-pass loudspeaker systems are researched by Fincham [5], Geddes [8], Mallory [7], Sutphin [8], Berkhoff [9], Matusiak and Dobrucki [11] and others. Now these systems are used for low-frequency channel of: computer loudspeaker systems, auto audio systems and the soundtrack of TV, video and cinema systems.

Figure 1 presents a band-pass loudspeaker system with single vent.

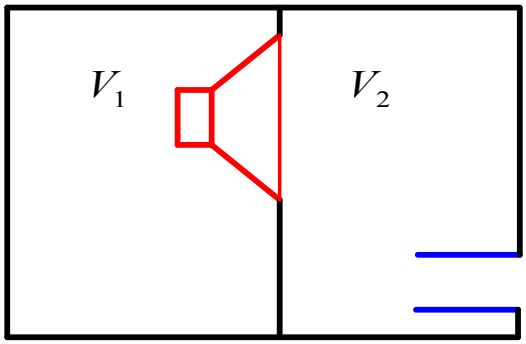


Fig. 1. Sketch of a double cavity single vented loudspeaker system

II. Glossary of Symbols

$B.l$ = *Force factor loudspeaker magnet system* [16], $T.m$

c = velocity of sound in air, $(c = 345)$, m/s [3]

C_{ms} = *mechanical compliance of driver suspension,* m/N [16]

$C_{вн}, C_{MES}$ = electrical capacitance due to driver mass, F $\left(C_{вн} = M_{ms}/(B.l)^2 = M_{as}.S_d^2/(B.l)^2\right)$, [3]

C_{as} = acoustic compliance of driver suspension, $\left(C_{as} = C_{ms}.S_d^2\right)$, m^5/N [3]

C_{ab1or2} = acoustic compliance of air in box V_1 or V_2 $\left(C_{ab1} = V_{1or2}/\rho_0.c^2\right)$, [7] m^5/N

e_g = open circuit output voltage of amplifier, V

i = imaginary factor, $\left(i = \sqrt{-1}\right)$

f = frequency, Hz

$K(s)$ = fourth order band-pass system response (transfer) function, $\left(K(s) = K_1(s).K_2(s)\right)$ [13]

$K_1(s)$ = second order low-pass transfer function

$K_2(s)$ = second order high-pass transfer function

L_e = *loudspeaker electrical inductance*, H [16]

$L_{вн}, L_{CES}$ = electrical inductance due to driver compliance [3], H $\left(L_{вн} = C_{ms}.(B.l)^2 = C_{as}.(B.l)^2/S_d^2\right)$

M_{as} = acoustic mass of driver diaphragm assembly and air load, $\left(M_{as} = M_{ms}/S_d^2\right)$, [3], kg/m^4

M_{ap2} = acoustic mass of air in vent, r - radius vent, L - length of vent, kg/m^4 $\left(M_{ap2} = \rho_0.(L + 0{,}85.r + 0.61.r)/\pi.r^2\right)$, [1, 7]

M_{ms} = *mechanical mass of loudspeaker diaphragm assembly including air load* [16], kg

p_0 = equivalent acoustic pressure, Pa $\left(p_0 = e_g.B.l/(R_g + R_e)S_d\right)$, [3]

Q_1 = box 1 Q at ϖ_S resulting from all losses, [3] $\left(Q_1 = \frac{1}{\varpi_S.C_{as1}.R_{atc}}\right)$, $\left(\frac{1}{C_{as1}} = \frac{1}{C_{as}} + \frac{1}{C_{ab1}}\right)$ [3]

Q_2 = box 2 Q at ϖ_B resulting from all losses, $\left(\frac{1}{Q_2} = \frac{1}{Q_L} + \frac{1}{Q_A} + \frac{1}{Q_P}\right)$ [4]

Q_A = box 2 Q at ϖ_B resulting from absorption losses, $\left(Q_A = \frac{1}{\varpi_B.C_{ab2}.R_{ab2}}\right)$ [4]

Q_L = box 2 Q at ϖ_B resulting from leakage losses, $\left(Q_L = \varpi_B.C_{ab2}.R_{al2}\right)$ [4]

Q_P = box 2 Q at ϖ_B resulting from vent frictional losses, $\left(Q_P = \frac{1}{\varpi_B.C_{ab2}.R_{ap2}}\right)$ [4]

R_e = *loudspeaker voice coil dc resistance* [16], Ω

R_g = amplifier output resistance [3], Ω

R_{ms} = *mechanical mass of loudspeaker suspension losses* [16], $N.s/m$, kg/s, $1N = m.kg.s^{-2}$

[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.

R_{ab1or2} = acoustic resistance of box 1 or 2 losses [7] due to internal energy absorption, $N.s/m^3$

R_{atc} = total series resistance, fig. 7

$\left(R_{atc} = R_{ab1} + R_{at} + R_{as} = R_{ab1} + \frac{R_{ms}}{S_d^2} + \frac{(B.l)^2}{(R_g + R_e).S_d^2}\right)$ [3, 4]

R_{al2} = acoustic resistance of box 2 losses [7] due to leakage, $kg/m^4.s$

R_{ap2} = acoustic resistance of vent losses, $kg/m^4.s$

R_{as} = acoustic losses [3], $N.s/m^5$ $\left(R_{as} = R_{ms}/S_d{}^2 = (B.l)^2/R_{вн}.S_d{}^2\right)$

$R_{вн}, R_{ES}$ = electrical resistance due to driver suspension losses[3], Ω $\left(R_{вн} = (B.l)^2/R_{ms} = (B.l)^2/S_d^2.R_{as}\right)$

s = complex angular frequency, $s = i.\varpi = i.2\pi.f$

S_d = *effective area of drive unit diaphragm* [7], m^2

V_1 = volume of ear in close box [7], m^3

V_1 = volume of ear in box with vent [7], m^3

ρ_0 = density of air (ρ_0 = 1.18 kg/m^3) [1], kg/m^3

ϖ = angular frequency $(\varpi = 2.\pi.f)$, rad/s

ϖ_0 = normalized angular frequency, rad/s

$\left(\omega_0 = \sqrt{\frac{1}{\omega_1.\omega_2}} = \sqrt{\frac{1}{\omega_B.\omega_S}} = \sqrt[4]{\frac{1}{C_{as1}.C_{ab2}.M_{as}.M_{ap2}}}\right)$

ϖ_s = resonance angular frequency of driver in close box [3], rad/s, $\left(\varpi_S = \sqrt{\frac{1}{C_{as1}.M_{as}}}\right)$

ϖ_B = angular frequency for box 2, rad/s $\left(\varpi_B = \sqrt{\frac{1}{C_{ab2}.M_{ap2}}}\right)$ [4]

ϖ_{1or2} = low or high cut of angular frequency, rad/s

Note: In "II. Glossary of symbols" the Thiele-Small loudspeaker parameters [16] presented with *Font style: Italic*.

III. Electrical Characteristics

The analysis of the electrical impedance of the single vented band-pass system of fig. 1 [15] is similar to a vented box system (two maxima and one minimum [11]), see fig. 3.

The input electrical impedance of a loudspeaker in the closed-box is [3]

$$Z_{v1}(\varpi) = R_e + i.\omega.L_e + \frac{(B.l)^2}{R_{ms} + \frac{\frac{1}{C_{ms}} + \frac{S_d^2}{C_{ab1}}}{i.\omega} + i.\omega.M_{ms}}. \quad (1)$$

The electrical impedance of the vented box is

$$Z_{v2}(\varpi) = \left(\frac{B.l}{S_d}\right)^2.\left(i.\varpi.C_{ab2} + \frac{1}{R_{aL}} + \frac{1}{i.\varpi.M_{ap2}}\right). \quad (2)$$

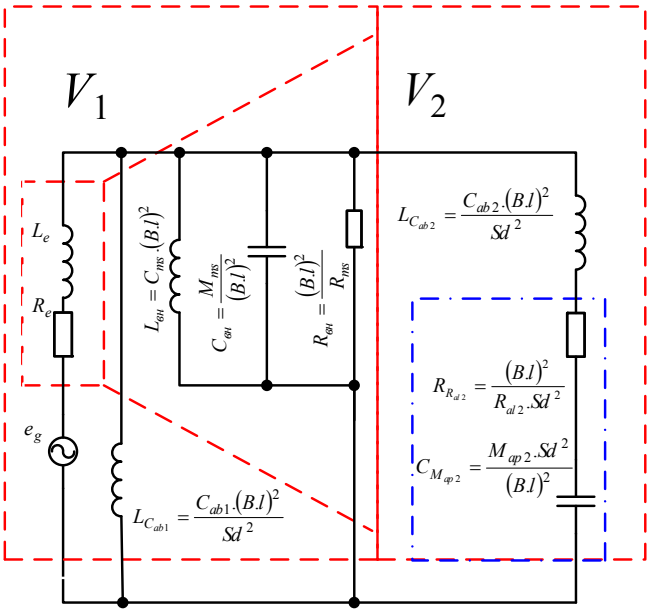


Fig. 2. Simplified electrical equivalent circuit of the double cavity with single vent loudspeaker system [15]

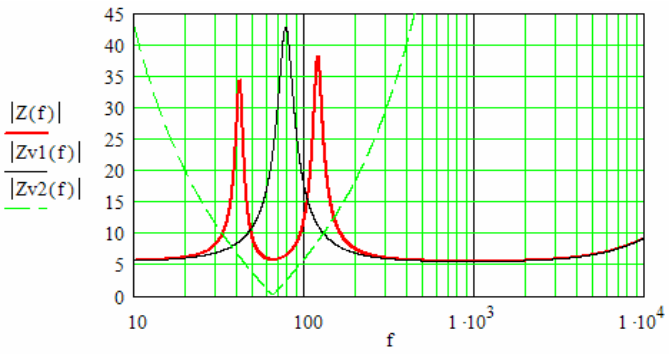


Fig. 3. Magnitude of the Band-pass loudspeaker system impedance

The math expression, Eq. (3), which describes the input electrical impedance, includes the sum of the voice coil electrical impedance $Z_e = R_e + i.\omega.L_e$, the impedance introduced by the mechanical system $Z_{вн}$ and the acoustic volume Z_a (fig.2).

$$Z(\varpi) = R_e + i.\omega.L_e + \frac{(B.l)^2}{R_{ms} + \frac{\frac{1}{C_{ms}} + \frac{S_d^2}{C_{ab1}}}{i.\omega} + i.\omega.M_{ms} + \frac{S_d^2}{i.\varpi.C_{ab2} + \frac{1}{R_{aL}} + \frac{1}{i.\varpi.M_{ap2}}}} \quad (3)$$

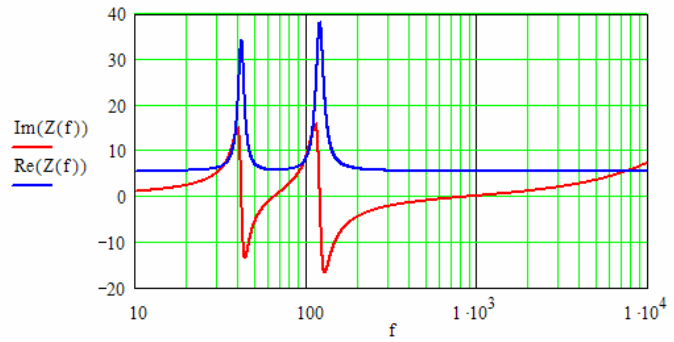


Fig. 4. Real and imaginary part of the Band-pass loudspeaker system impedance

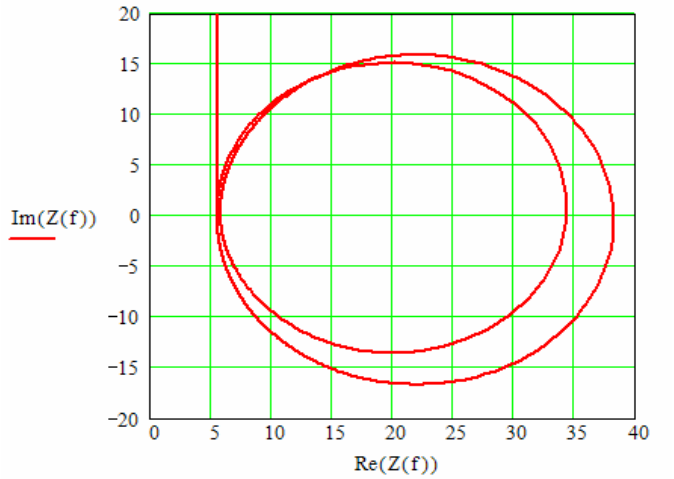


Fig. 5. Complex Impedance (Nyquist plot) of the Band-pass loudspeaker system

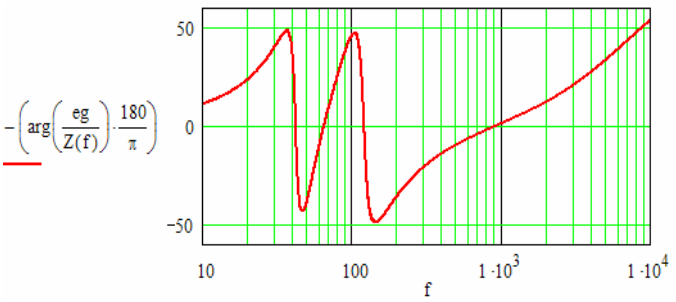


Fig. 6 Phase of the electrical current from the audio amplifier

IV. Acoustical Characteristics

The characteristics of the double cavity with single vent loudspeaker system may be presented with the band-pass system response (transfer) function [13]:

$$K(s)=K_1(s).K_2(s)=\frac{1}{\frac{s^2}{\varpi_1^2}+\frac{s}{\varpi_1.Q_1}+1}.\frac{\frac{s^2}{\varpi_2^2}}{\frac{s^2}{\varpi_2^2}+\frac{s}{\varpi_2.Q_2}+1} \quad (4)$$

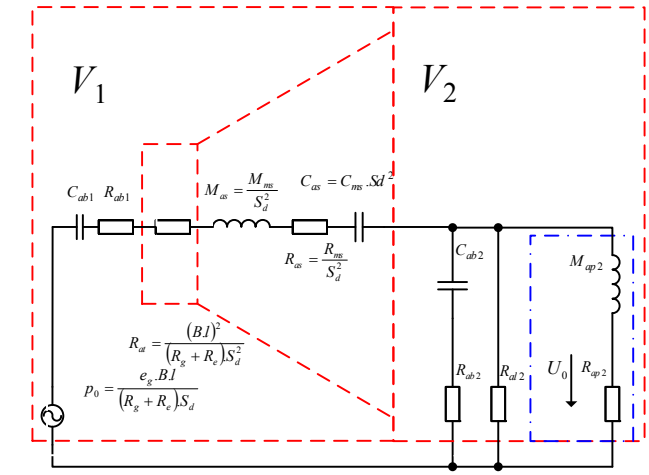


Fig. 7. Simplified acoustical analogous circuit of the double cavity with single vent loudspeaker system [15]

Amplitude – frequency responses are presented in figs. 8÷10.

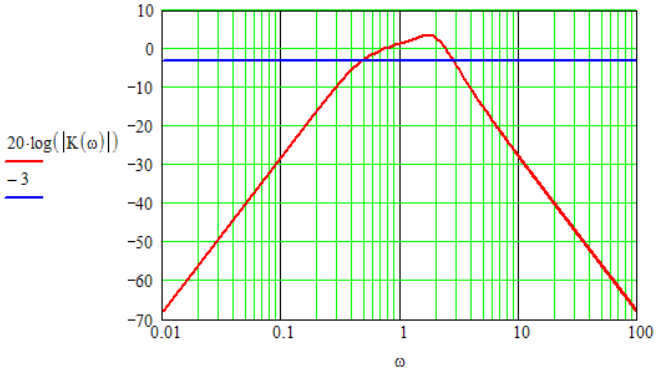


Fig. 8. Normalized amplitude – frequency response of the sound pressure created by the Band-pass loudspeaker system

It is a forth order band-pass filter response (fig. 8) with 12 dB per octave (40 dB per decade) slopes.

The band-pass characteristics will be symmetrical if $\varpi_S = \varpi_B$ [11] and $Q_1 = Q_2$.

The phase of the sound pressure created by the Band-pass loudspeaker system $\arg(K(\varpi))$ is plotted in fig. 9.

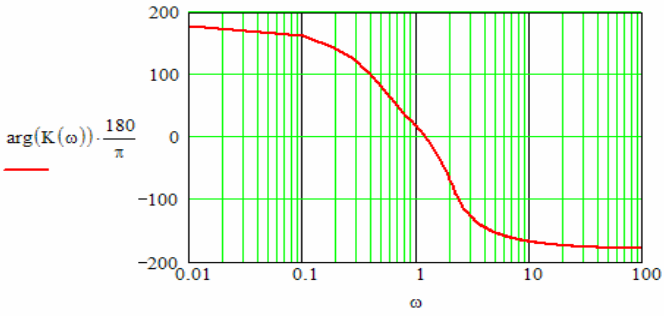


Fig. 9. Normalized frequency response of the phase of the sound pressure created by the Band-pass loudspeaker system

The group time delay $-\frac{d}{d\varpi}\arg(K(\varpi))$ is presented in fig. 10.

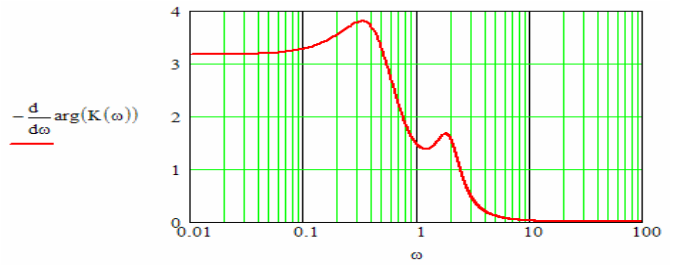


Fig. 10. Normalized frequency response of the group time delay of the sound pressure created by the Band-pass loudspeaker system

For the analysis in the time domain, the impulse response of the loudspeaker $h(t)$ (fig.11) is defined as follows:

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \dot{K}(j\omega)e^{j\omega t}d\omega \qquad (5)$$

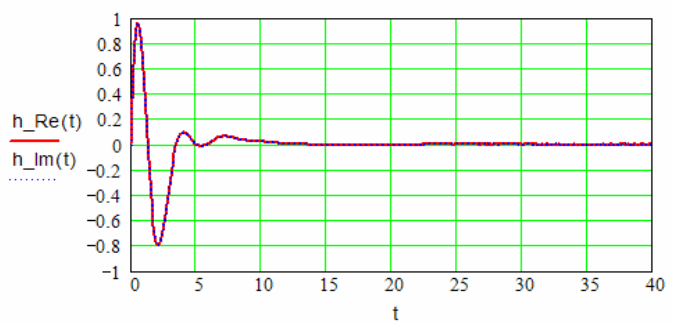


Fig. 11. Normalized impulse response of the sound pressure created by the Band-pass loudspeaker system

The impulse response $h(t)$ is a reaction of $\delta(t)$-Dirac impulse, while the step response $F(t)$ is defined by the integral of $h(t)$:

$$F(t) = \int_0^t h(t)dt \qquad (6)$$

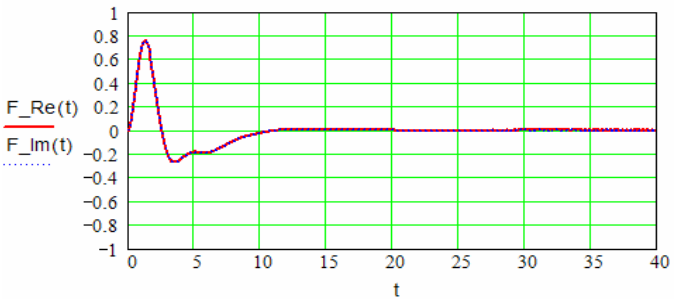


Fig. 12. Normalized step response of the sound pressure created by the Band-pass loudspeaker system

V. Conclusion

The reactance transformation method – Thile-Small [2-4] theory of close and vented system can be applied to the analysis of a band-pass system [5-9, 11, 15].

This work considers researching of the following characteristics of a band-pass loudspeaker system: frequency response of the magnitude (fig. 3), real and imaginer part (fig. 4) of impedance and the complex impedance (fig. 5); electrical phase of the current from audio amplifier (fig. 6); of the amplitude (fig. 8), phase (fig. 9), of the group time delay (fig. 10), impulse (fig. 11) and step (fig. 12), response of the sound pressure created; and in the time domain: impulse (fig. 11) and step (fig. 12) response.

The theoretical analysis of the characteristics and the comparison with the characteristics by computer simulations [15] confirm the proposed ideas.

The results obtained in this paper can be used for theoretical analysis, design and production of band-pass loudspeaker systems.

References

[1] Beranek Leo L., "*ACOUSTICS"*, ISBN: 07-004835-5, 1954÷96

[2] Thiele A. N., *"Loudspeakers in Vented Boxes, Parts I and II"*, J. Audio Eng. Soc., vol. 19, Number 5, pp. 382÷392, 1971 May; pp. 471÷483, 1971 June.

[3] Small R. H., *"Closed-Box Loudspeaker Systems Part I: Analysis"*, J. Audio Eng. Soc.,, vol. 20, Number 10, pp. 798÷808, 1972 Dec, "*Part II: Synthesis"*, J. Audio Eng. Soc., vol. 21, Number 1, pp. 296÷303, 1973.

[4] Small R. H., *"Vented-Box Loudspeaker Systems: Part I: Small-Signal Analysis"*, J. Audio Eng. Soc., vol. 21, n. 5, jun. 1973, pp 363÷372, *"Part IV: Appendices",* J. Audio Eng. Soc., vol.21, n.7, Oct. 1973, pp. 635÷639.

[5] Fincham, L. R., *"A band-pass loudspeaker enclosure",* 63rd Convention Audio Engineering Society, Los Angeles, 1979 may.

[6] Geddes E. R. *"An Introduction to Band-Pass Loudspeaker Systems"*, J. Audio Eng. Soc., vol. 37, № 5, pp. 308÷342, 1989 May.

[7] Mallory Roy, "*Calculating LF response whit PSPICE"*, Speaker Builder, № 3 , 1991, pp. 20÷27

[8] Sutphin E. M., "*An equivalent electrical circuit bandpass enclosure"*, Speaker Builder, № 2, 1992 pp. 10÷13 and 73.

[9] Berkhoff A. P. *"Impedance Analysis of Subwoofer Systems"*, J. Audio Eng. Soc., vol. 42, Number 1/2, pp. 4÷14, 1994 Jan./Feb.

[10] Poularikas Al. D, *The Handbook of Formulas and Tables for Signal Processing,* CRT Press LLC, 1999.

[11] Matusiak G. P., A. B. Dobrucki, "*Fourth-Order Symmetrical Band-Pass Loudspeaker Systems"*, J. Audio Eng. Soc., vol. 50, № 1/2, pp. 12÷18, 2002 Jan./Febr.

[12] Sirakov E. S.**, "***Theoretical Analysis as a Function of the Total Q Factor Loudspeaker Characteristics"*, *XLI ICEST 2006*, 29 June ÷ 01 July, 2006, Sofia, BULGARIA, ISBN-10: 954-9518-37-X, pp. 208÷211.

[13] Sirakov E. S.**, "***Polar Diagram of multiple driver loudspeaker systems"*, *XLII ICEST 2007*, "Sv. Kliment Ohritski", ISBN: 9989-786-06-2, "Mikela" Bitola, 24÷27 June 2007, Ohrid, Macedonia, Volume 2, pp. 317÷320.

[14] Sirakov E. S.**, "***LR-n Models for Voice Coil of the Loudspeaker, Part I: Small Signal Analysis"*, *XLIII ICEST 2008*, 25÷27 June 2008, Printed by: *UNIGRAF*, Nis, Serbia, ISBN: 987-86-85195-61-7, pp. 550÷553.

[15] Сираков Е. С., "Макро модел на нискочестотно озвучително тяло тип Band-pass", *Национална конференция Акустика 1*, 2008 г., Технически Университет, Варна.

[16] www.vissokogovoriteli.plc.bg/, www.madisound.com/, www.audax.fr/, www.dynaudio.com, www.morelhifi.com/, www.scanspeak.com/.

Using the Support Vector Algorithm to Detect Voiced Segments and Eliminate Unvoiced Frames

Damyan Damyanov [1], Zlatka Nikolova [2]

Abstract – **This paper introduces a new algorithm for the detection of voiced speech, which functions as a classifier for frames of recorded speech. In this context, that means that it decides whether or not the frame contains good quality voice data as the frame could contain silence, unvoiced speech, or degraded speech which would be unusable without additional processing. The problem considered in this paper is a very important one and many classical methods exist for solving it. Regardless of the possible advantages of these methods, they all suffer from one major drawback – their efficiency is low, which necessitates some degree of post-processing in order to achieve a high recognition ratio. The Support Vector Algorithm achieves an excellent classification of voiced and other speech frames by the use of nonparametric classification and training.**

Keywords – **support vector machine, speech recognition, classifying of voiced speech frames.**

I. INTRODUCTION

There are two main types of algorithms used in classifiers: parametric ones, in which a priori knowledge of data distributions is assumed, and non-parametric ones, in which no such a priori knowledge is assumed.

Neural networks, fuzzy systems, and support vector machines (SVMs) are typical non-parametric classifiers. Through training using input-output pairs, classifiers acquire decision functions that classify an input datum into one of the given classes [1].

SVM are preferred to neural-networks, because of their better generalization ability in speech processing. Three-layer neural networks are universal classifiers in that they can classify any labeled data correctly if there are no identical data in different classes. In training multilayer neural network classifiers, network weights are usually corrected, so that the sum-of-squares error between the network outputs and the desired output is minimized. However since the decision boundaries between classes acquired by training are not directly determined, classification performance for the unknown data, i.e. the generalization ability, depends on the training method. And it degrades greatly when the number of training data is small and there is no class overlap.

Conversely, in training support vector machines the decision boundaries are determined directly from the training data so that the separating margins of decision boundaries are maximized in the high-dimensional space called feature space. This learning strategy minimizes the classification errors of the training data and the unknown data.

[1] Damyan Damyanov is with the Dept. of Radiocommunications and Video Technology, Technical University of Sofia, Bulgaria, e-mail: ellov@abv.bg

[2] Zlatka Nikolova is with the Dept. of Communication Networks, Technical University of Sofia, Bulgaria, e-mail: zvv@tu-sofia.bg

On the other hand, speech recognition by humans does not pose a problem [2]: One can understand very well that something has been said, even if its information is lost, for example due to loud ambient noise, or when a language, not known by the listener, is spoken. In practical applications this detection of a signal's voiced frames is very useful. It can be employed in speech databases, where the type of a speech should be automatically determined for additional information generation and context-based indexing, or in full automatic speech recognition systems. In the latter case, the room where the acoustic signal propagates is scanned with the aim of recognizing one particular speaker.

Early work in speech/non speech signal classification was carried out, for example, by Hoyt and Wechsler [3]. This research uses a common architecture of signal classifiers: The incoming signal is segmented into frames; discriminative features are extracted and finally classified. Classification of the latter with support vector machines reportedly achieved a recognition rate in excess of 99% [4].

II. SUPPORT VECTOR ALGORITHM

Speech is a phenomenon which constantly changes with time. One approach to registering the non-stationarity of the speech signal is to process it as a temporal sequence of an alphabet of states. The concept beyond the suggested speech detection algorithm is that non-speech signals, which are non-stationary, produce sequences different from those found in speech [5, 6]. This is true, when using the same alphabet for speech and non-speech signals. It is well known [2] that non-speech signals which are stationary are very easy to recognize. Some new efficient methods suggested in [7, 8] can be used.

The parts of speech considered here are voiced frames, the most important components of which are phonemes. In normal speech these are combined to form syllables, words, phrases and more complicated structures. The most important property of voiced-frame detection is the syllabic structure: Normal speech changes constantly between vowels, which are contained in frames with high frame energy, and other phonemes with lower frame energy. Both the vowels and all other types of phonemes possess different spectral features. If a classification algorithm is based on phonemes, then vowels are most conveniently incorporated into a frame model: The support vector machine will discriminate between voiced-like and unvoiced-like frames in certain temporal sequences.

The experiment to derive the speech recognition algorithm was composed of four stages: (1) frame extraction, (2) cepstral Linear Prediction Coefficients (LPC) extraction, (3) segmentation of voiced-unvoiced frames and (4) training and classifying with the support vector algorithm – (Fig.1).

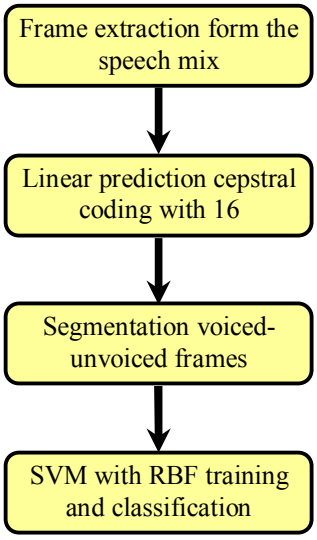


Fig.1: Support Vector Algorithm

III. CLASSIFICATION

In the initial frame extraction step, the signal is recorded at a sampling frequency of 11025 Hz. This is a very good sampling frequency for speech signals, since the frequency spectrum of these seldom exceeds 5 kHz. It also provides good speaker differentiation, compact and de-correlated representation of the signal and computational efficiency.

Next, the signal is windowed into frames of 23 ms or 253 samples. The frames are then multiplied by a rectangular window. Whichever window is taken for frame smoothing, there is always a problem with the samples located on both edges of the frame; their spectrum, and hence their energy, is attenuated by the window. For this reason a half-time overlap of the frames has been chosen. This means approx. 11 ms, or 126 samples of the signal. As classification features, linear prediction cepstral coefficients of order 16 were chosen. This employs the well-known formula:

$$M = F_s + (4 \text{ or } 5) \tag{1}$$

where M is the number of the cepstral LPC coefficients, F_s is the sampling frequency in kHz. The first term in the equation accounts for proper representation of F_s poles from the all-pole of the vocal tract, and the 4 or 5 poles are added for better modelling of the glottal pulse in the voiced case.

The cepstral LPC coefficients have been evaluated, using the recursive formula:

$$y(n,m) = \begin{cases} \log\hat{\Theta}_0(m), & n = 0 \\ a(n;m) + \sum_{k=1}^{n-1}\left(\frac{k}{n}\right) y(k;m)a(n-k;m), & n > 0 \end{cases} \tag{2}$$

where m is the size of the window used, n is the current number of the coefficient. $\hat{\Theta}_0$ is the first sample of the all-pole filter model of the vocal tract, a is the LPC coefficient, and y is the cepstral LPC coefficient.

Next, points of maximal acoustic change (acoustic landmarks) should be evaluated.

As a basis for the detection of acoustic landmarks, a Euclidean distance measure between the cepstral coefficients, $\bar{c}_i$, of frames with a certain time difference, is employed:

$$d_i = \left\| \bar{c}_i - \bar{c}_{i-k} \right\|_2 \tag{3}$$

where k is the time difference between the frames, and i is the current frame. In this paper, the time difference is set to approximately 11 ms, or 126 samples (i.e., k=1). The peaks of d_i can be interpreted as points where the signal spectrum exhibits more significant changes compared to the vicinity, such as phoneme boundaries or instants of rapid spectral change in interferences. Measures like d_i in general produce a huge number of peaks which are not easy to analyse. In this case, smoothing might be preformed: The frame-to-frame Euclidean distance is convolved with a "Mexican hat" function:

$$\Psi(x) = \frac{1}{\left(\sqrt{2\pi}\sigma^3\right)\left(\frac{1-x^2}{\sigma^2}\right)} \exp\left(\frac{-x^2}{2\sigma^2}\right). \tag{4}$$

This is the popular second derivative of a Gaussian, with standard deviation σ set to 3,6 ms. The peaks of the resulting acoustic change function are interpreted as acoustic landmarks and used for segmenting the waveform if they exceed a certain threshold (see Fig.2).

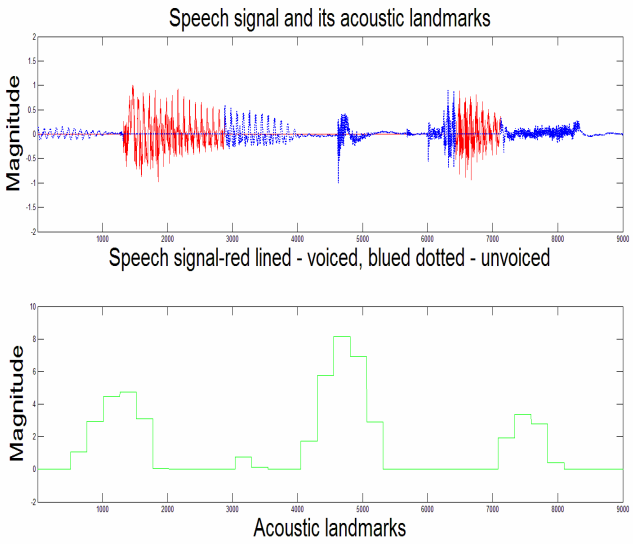


Fig.2: Speech Signal and Acoustic Landmarks

Classification of each frame is performed with a support vector machine. A kernel, based on Radiant Based Function (RBF), has been used for the classification. The analytical expression of the latter is:

$$H(x,x') = \exp(-\gamma\left\|x - x'\right\|^2) \tag{5}$$

where the operator $\| \|$ denotes the Euclidean distance, and x' means in general different variable than x. The parameter γ controls the radius of the function. Its typical values are from $0.9 \div 1$ [1]. In our case it is set to 1. The method for the

evaluation of the separating hyper plane has been set to Least Squares. Thus the SVMs are able to discriminate between classes via topologically complex hyper planes and are computationally efficient at the test stage. The target classes are {voiced, unvoiced}.

IV. Experiments and Discussions

For the experiment a private database has been used. 3000 frames from 30 different speakers have been classified. 300 frames were used for training the support vector machine, and the other 2700 were used in classification.

One utterance from the upper is shown in Fig.3.

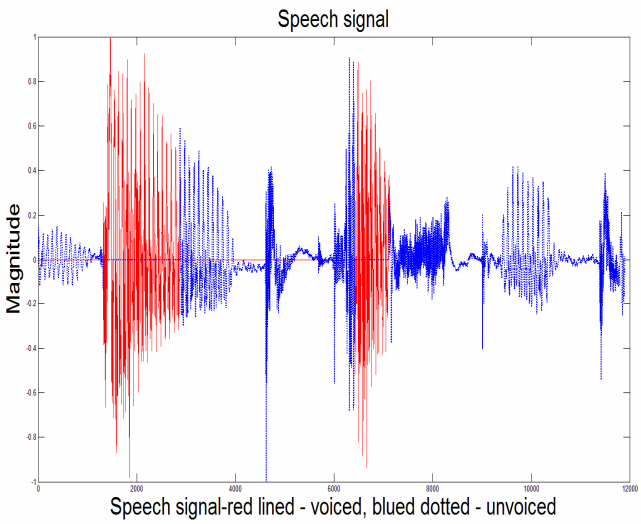


Fig.3: Example for a male utterance. The red straight line represents the voiced frames, and the blue dotted line – the unvoiced ones

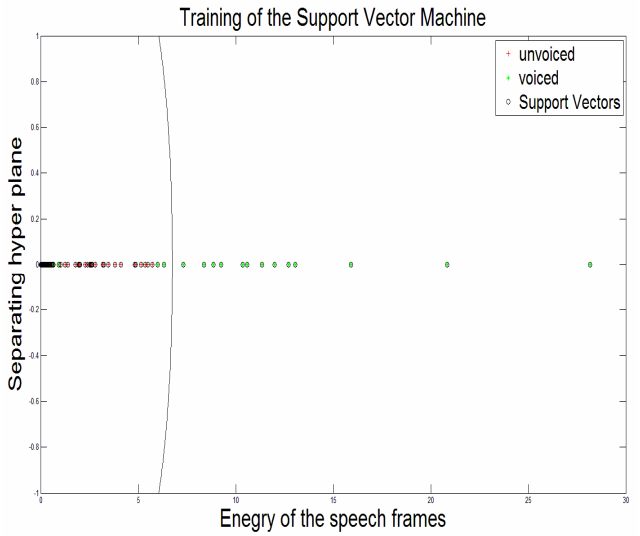


Fig.4: Training of the support vector machine. The red crosses on the left side represent the unvoiced frame energies, and the green crosses on the right side represent the voiced frame energies

The training phase produced the results shown in Fig 4. As can be seen from the latter, this first phase of the classification algorithm has been very effective. The frame energies lie far enough apart from each other to draw a parabola between the two classes and to separate them easily. The classification phase for the utterance from Fig.3 is shown in Fig.5. As expected, the separating parable performs well in separating the two target classes. This can be seen if we show how a classical algorithm performs against the support vector one. As a classical approach, the Spectral Auto-Correlation Peak-to-Valley Ratio (SAVPR) has been chosen, for its simplicity of implementation. It is a parametric approach, meaning that it doesn't need training [5, 6].

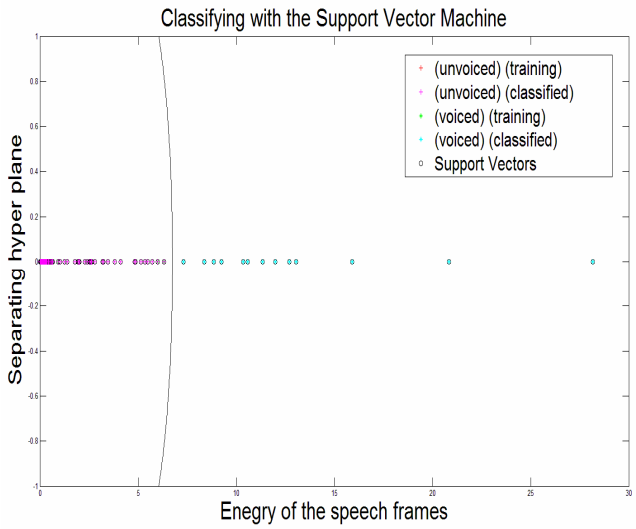


Fig.5: Classifying with the support vector machine. The red and magenta crosses on the left side represent the unvoiced frame energies, and the green and cyan crosses on the right side represent the voiced frame energies

The classification with the SAVPR method for the utterance is shown on Fig. 6, the one with the SVM is shown in Fig. 7, and both are plotted together on Fig. 8. As can be seen, the performance of the support vector algorithm is much better than that shown by the spectral auto-correlation peak-to-valley ratio algorithm. After the classification, using the database mentioned at the beginning of the chapter, the following results were obtained, demonstrating that the overall performance of the SVM is better [6] (Table I).

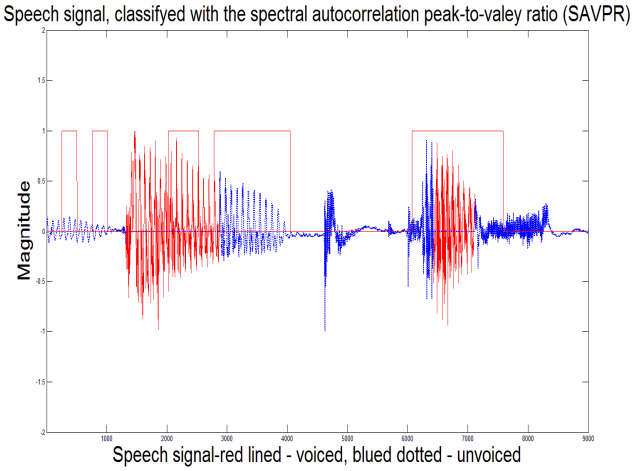


Fig.6: Classification with the spectral auto-correlation peak-to-valley ratio

TABLE I

	Classification with the SAVPR algorithm	Classification with the SVM algorithm
Male speech	61.70 %	92.40 %
Female speech	60.10 %	91.50 %

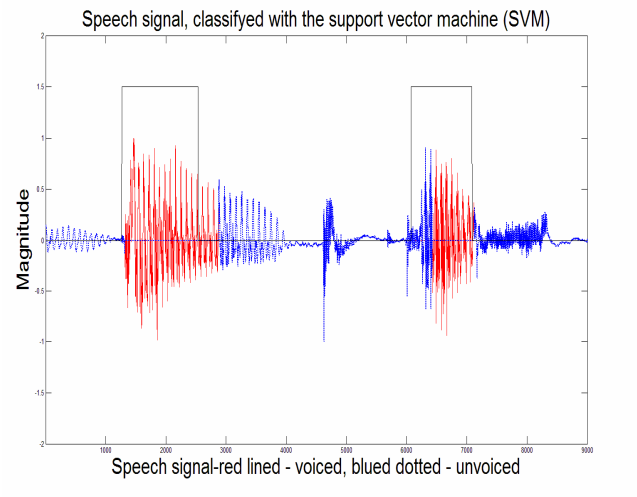


Fig.7: Classification with the spectral SVM

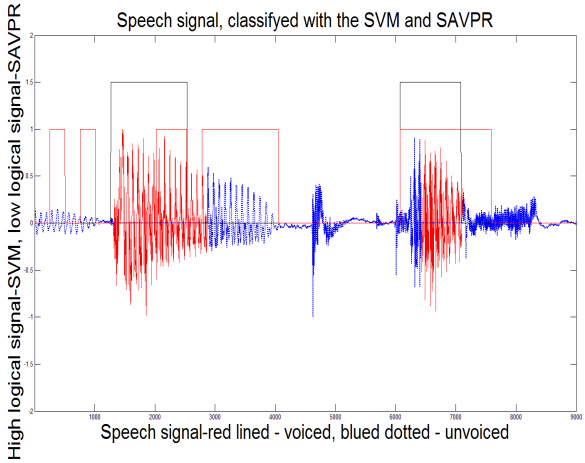


Fig.8: Classification with the SAVPR (low red logical signal) and SVM (high black logical signal)

V. CONCLUSIONS

The support vector algorithm has shown some very good features in the recognition of voiced speech segments – it is non-parametric, requires a small amount of training data and is computationally very effective. In comparison with classical approaches it performs much better, suggesting, that is can be used in "speech overlap detection". This is where two or more people speak simultaneously and recognition of one of the speakers is required. The future work of the authors is to utilize the support vector machine in the field of speech overlap detection.

REFERENCES

[1] Shiego Abe, "Support vector machines for pattern classification", *Springer,* 2005.

[2] John Proakis "Discrete-Time Processing of Speech Signals", *IEEE press*, 2000.

[3] J.D.Hoyt, and H. Wechsler "Detection of human speech using hybrid recognition models", *Proc. ICASSP-94,* Vol.I, pp. 330-333. 1994.

[4] G. Heinrich, "Speech identification using a sequence-based heuristic", *47th International Symposium ELMAR-2005*, 08-10 June 2005, Zadar, Croatia.

[5] Bekiarski Al, and L. Docheva, "Investigation of back Propagation Algorithm Implementation in Analog Neural Networks ", *ICEST,* 2005, Nis Serbia and Montenegro, pp. 637-640.

[6] S. G. Pleshkova-Bekiarska and D. A. Damyanov, "Speech overlap Detection Algorithms Simulation", *ICEST*, Ohrid pp. 495–499, 2007.

[7] Iliev G, Z. Nikolova, V. Poulkov, G. Stoyanov, "Noise Cancellation in OFDM Systems Using Adaptive Complex Narrowband IIR Filtering", *IEEE International Conference On Communications (ICC-2006)*, Istanbul, Turkey, 11-15 June 2006, pp. II-451 - II-455.

[8] Nikolova Z., V. Poulkov, G. Iliev, G. Stoyanov, "Narrowband Interference Cancellation in Multiband OFDM Systems", *3rd Cost 289 Workshop "Enabling Technologies For B3g Systems"*, Aveiro, Portugal, 12-13 July 2006.

Deriving measures for the goodness of 1-D spline interpolator filters

Damyan Damyanov[1] and Snejana Pleshkova[2]

Abstract – **The main topic of the work of the authors was deriving measures for the goodness of modified B-spline interpolators of degrees 5, 3 and 1. They have been optimized, through their degrees of freedom, in order to obtain better interpolator filters. The first measure is minimizing the energy in the stop band of the interpolator, with nonlinear constraints about the ripples in the frequency response. The second measure is minimizing the maximum value of the error kernel of the interpolator from 0 to some α_e.**

Keywords – **spline interpolartor filters, nonlinear constrains,**

I. INTRODUCTION

In The splines theory is well studied during the last years. The modified splines have some very attractive properties – maximum order with minimum support, very easy implementation in discrete structures, designed for processing of continuous data. The frequency responses of the interpolators, based on modified splines, have better attenuation in the stop band; compared to other interpolators, let's say Cubic Lagrange. The errorkernels of the modified spline interpolators are very close to the errorkernel of the sinc interpolator, which is the ideal one. The signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) of the resized and resampled images with spline interpolators are very high, compared to the same case, involving other interpolators.

Some very good studies of splines are given in [1÷4].

II. SPLAIN THEORY

The normalized B-spline functions of degree *M* have *M+2* equally spaced nodes. For M odd the nodes are placed at the integers while for M even they are placed at the half-integers. The B-splines centered at the origin are defined as follows:

$$\beta^M(x) = \sum_{i=0}^{M+1} \frac{(-1)^i}{M!} \binom{M+1}{i} \left(x + \frac{M+1}{2} - i \right)_+^M , \qquad (1)$$

where x_+^n is the truncated power function $\max(0, x)^n$. They belong to the class C^{M+1}, i.e. they are continues and have M-1 continues derivatives. The B-spline functions can be obtained recursively starting with the B-spline of zero degree as follows:

$$\beta^M = \beta^0 * \beta^{M-1}, \qquad (2)$$

where:

$$\beta^0(x) = \begin{cases} 1 \ for \ |x| \le \frac{1}{2} \\ 0 \ for \ |x| > \frac{1}{2} \end{cases} . \qquad (3)$$

The corresponding frequency-domain characteristics are given by:

$$B^M(2\pi f) = \left(\frac{\sin(\pi f)}{\pi f} \right)^{M+1} . \qquad (4)$$

The continuous B-spline dilated by an integer scale factor $m \ge 1$ and of degree *M* is defined by:

$$\beta_m^M(x) = \frac{1}{m} \beta^M \left(\frac{x}{m} \right). \qquad (5)$$

The dilated B-spline can be subsequently sampled:

$$b_m^M[k] = \frac{1}{m} \beta^M \left(\frac{k}{m} \right). \qquad (6)$$

The discrete B-spline of degree M at the integer scale m is given by:

$$u_m^M = \overbrace{u_m^0 * u_m^0 ... u_m^0}^{M+1} , \qquad (7)$$

where:

$$u_m^0 = \frac{1}{m} \overbrace{[1,1,...,1]}^{m} . \qquad (8)$$

The relation between dilated, sampled and discrete B-splines of degree *M* is established through the following so-called *m*-scale relation:

[1]Damyan Damynov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ellow@tu-sofia.bg.

[2]SnejanaPleshkova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: snegpl@tu-sofia.bg.

$$\beta_m^M(x) = \left(u_m^M * \beta_1^M\right)(x)$$
$$b_m^M[k] = (u_m^M * b_1^M)[k]. \quad (9)$$

A form of the modified B-spline, which combines a central B-spline of degree *n* with its derivatives as given by

$$\varphi(x) = \sum_{k=0}^{N} \gamma_k \frac{d^k}{dx^k} \beta^N(x). \quad (10)$$

The support of the modified B-splines (the approximation order as well) up to any number *N* is given by the following modified B-spline basis functions:

$$\beta^{\text{mod}}(x) = \sum_{n=0}^{N} \sum_{m \in Z} \gamma_{nm} \beta^n(x-m). \quad (11)$$

The function β^{mod} is the generating function for certain spline-like space $V(\beta^{\text{mod}})$. The scaling coefficients 'gammas' are actually the so called degrees of freedom, which provide the opportunity to adjust the modified B-spline approximation properties in an application-oriented manner.

In this report we are interested in the combination of 5-th, 3-rd, and
1-st degree B-splines. In this case, the modified B-spline basis function becomes:

$$\begin{aligned}\beta^{\text{mod}}(t) &= \beta^5(t) + \gamma_{30}\beta^3(t) + \\ &+ \gamma_{31}\beta^3(t-1) + \gamma_{31}\beta^3(t+1) + \\ &+ \gamma_{10}\beta^1(t) + \gamma_{11}\beta^1(t-1) + \gamma_{11}\beta^1(t+1) + \\ &+ \gamma_{12}\beta^1(t-2) + \gamma_{12}\beta^1(t+2).\end{aligned} \quad (12)$$

This combination is of particular interest since the quintal term is very smooth with short support and by adding shifted and centered cubic and linear terms the model is improving the sharp signal parts.

Keeping the partition of unity condition gives $\gamma_{30} = -2\gamma_{31} - \gamma_{10} - 2\gamma_{11} - 2\gamma_{12}$. This implies that the latter equation looks like shown above:

$$\begin{aligned}\beta^{\text{mod}}(t) &= \beta^5(t) + \\ &+ \gamma_{31}\left[\beta^3(t-1) - 2\beta^3(t) + \beta^3(t+1)\right] + \\ &+ \gamma_{10}\left[\beta^1(t) - \beta^3(t)\right] + \\ &\gamma_{11}\left[\beta^1(t-1) + \beta^1(t+1) - 2\beta^3(t)\right] + \\ &+ \gamma_{12}\left[\beta^1(t-2) + \beta^1(t+2) - 2\beta^3(t)\right].\end{aligned} \quad (13)$$

That is nothing but a combination of a smoothing function and a difference part. By optimizing in am appropriate manner the parameters $\gamma_{31}, \gamma_{10}, \gamma_{11}, \gamma_{12}$ we can trade between smoothing and sharpening. Figure (1) shows combination between a centered B-spline of 5-rd degree, centered and shifted B-spline of 3-rd degree, and centered and shifted B-splines of 1-st degree.

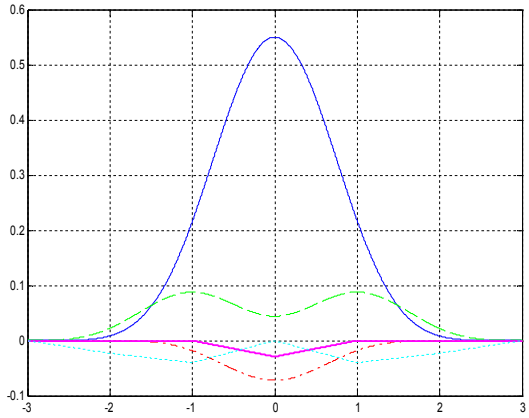


Fig.1. Combination of 5-th, 3-rd and 1-st degrees B-splines

III. INTERPOLATION THEORY

The main algorithm of interpolation due to splines is shown in the following equation:

$$s(x) = \sum_{k \in Z} c(k) \beta^n(x-k), \quad (14)$$

Which involves the integer shifts of the central B-spline of degree n denoted by $\beta^n(x)$; the parameters of the model are the B-spline coefficients c (k), $s(x)$ are the values of the interpolated signal.

To derive this type of signal processing algorithm, we take the signal values $s(x)$, and we want to determine the coefficients c(k) of the B-spline model such that we have a perfect fit at the integers, i.e. $\forall k \in Z$:

$$\sum_{l \in Z} c(l) \beta^n(x-l)\Big|_{x=k} = s(k). \quad (15)$$

Using the discrete B-splines, this constraint can be rewritten in the form of a convolution:

$$s(k) = (b_1^n * c)(k). \quad (16)$$

Defining the inverse convolution operator:

$$\left(b_1^n\right)^{-1}(k) \overset{n}{\leftrightarrow} 1/B_1^n(z). \quad (17)$$

The solution is found by inverse filtering

$$c(k) = \left(b_1^n\right)^{-1} * s(k). \quad (18)$$

Since b_1^n is symmetric FIR (finite impulse response), the so-called direct B-spline filter $\left(b_1^n\right)^{-1}$ is an all-pole system that can be implemented very efficiently using a cascade of first-order causal and anti-causal recursive filters. This algorithm is

stable numerically and is faster and easier to implement than any other numerical technique.

IV. ERROR KERNEL

There are two ways of estimating the properties of the optimized interpolators. The firs way is taking a set of pictures – Barbara, Lena, Hotel, Mandrill, Patterns, and part of the Geometrical image. Then each picture is successively rotated 10 times by 36 grades with the optimized interpolators. Finally, the SNR and the PSNR are computed for each rotated picture with respect to the original one. The latter operations are performed by the program "Spline Explorer", developed by Atanas Gotchev and Grigor Marchokov. The second way is by computing the so called errorkernels for each optimized modified combination of B-splines. Here is a brief explanation:

The main model of image acquisition consists in the following. Consider the one-dimensional function *r(x)* representing a true scene along the x axis. As a result of optical blurring, modeled as a linear convolution with a kernel *s(x)*, we obtain another continues function:

$$g_a(x) = (r * s)(x). \tag{19}$$

For the sake of simplicity we assume that the initial sampling grid is placed on the integer coordinates and generate the discrete sequence $g[k] = g_a(x)$. In continues time it can be modeled by the product $g_a comb(x)$, where:

$$comb(x) = \sum_{k=-\infty}^{\infty} \delta(x-k). \tag{20}$$

If we want to insert an additional parameter *u*, showing the relative position of the sampling device with respect to the function *s(x)*:

$$g_a(x-u) = (r * s)(x-u). \tag{21}$$

Sampling at integers and reconstructing by a continuous interpolation function gives:

$$y_a(x,u) = \sum_{k=-\infty}^{\infty} g_a(x-u)h(x-k). \tag{22}$$

Here, *h(x-k)* represents the impulse response of the interpolator, i.e. of the combination of modified B-splines. The reconstruction function also depends of the sampling phase and it is not simply a function of the difference *x-u*. The error between *g(x-u)* and *y(x, u)* called by the authors sampling and reconstructing (SR) blur, is a function of *u* as well:

$$\varepsilon_{SR}^2(u) = \int_{-\infty}^{\infty} (g_a(x-u) - y(x,u))^2 dx. \tag{23}$$

The average value of the SR blur plays the most important role in the estimation of the properties of the interpolator. It is equal to

$$E\{\varepsilon_{SR}^2\} = \int_{-\infty}^{\infty} \eta^2(2\pi f)|G_a(2\pi f)|^2 df, \tag{24}$$

where:

$$\eta^2(2\pi f) = |1 - H(2\pi f)|^2 + \sum_{k\neq 0} |H(2\pi(f-k))|^2. \tag{25}$$

Here, the function $H(2\pi f)$ represents the frequency response of the interpolator. The non-negative errorkernel η^2 quantifies the amount of errors introduced by the reconstruction kernel. The plotting variable α_e (see Fig.2) is defined by the equation $\alpha_e = \frac{f}{F_s}$, where F_s is the sampling frequency. It is not possible to achieve errorkernel equal to zero, if the sampling frequency is lower than 1 (or equivalently, if the cut-off frequency is higher than 0.5 for fixed unity sampling frequency). For low frequencies, i.e. for the pass band, the reconstruction function must be unity in order to avoid blurring.

A convenient expression of the errorkernel in the case of the modified B-splines can be written as:

$$\eta^2(j2\pi f) = \frac{\left(P\left(e^{j2\pi f}\right)\right)^2 - 2P\left(e^{j2\pi f}\right)\Phi(2\pi f) + \sum_{k=-\infty}^{\infty}\left(\Phi(2\pi(f-k))\right)^2}{\left(P\left(e^{j2\pi f}\right)\right)^2}. \tag{26}$$

Where $\Phi(2\pi f)$ is the Fourier Transform of the continuous kernel, and $P\left(e^{j2\pi f}\right)$ is the Fourier transform of the discrete kernel of the combination of B-splines.

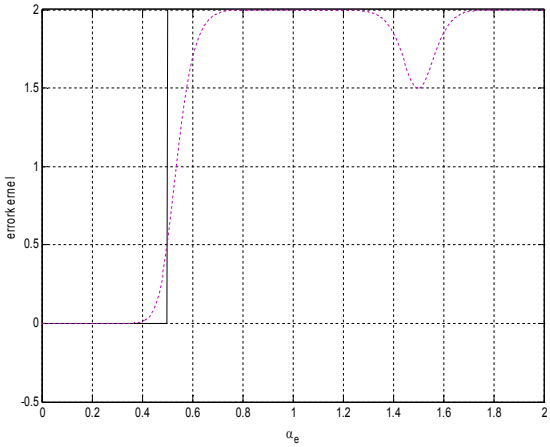


Fig.2. Errorkernels for sinc interpolator (solid line) and for quintal B-spline

The weighting coefficients in (13) are optimized according to two criteria – minimizing the energy in the stop band of the

interpolators, and minimizing the errorkernel from *0* to some α_e .

V. CONCLUSION

Two of the most important unwanted effects, which occur in the interpolated image due to non-ideal interpolator, are aliasing and imaging. Aliasing is the effect of the appearing of unwanted frequencies as a result of the repetition of the original spectrum around multiplies of 2π . Imaging is an effect, which occurs for example by up sampling of the image by some factor *L*. After the up sampling in the pass band of the original spectrum appears together with *L-1* 'images'. The unwanted frequencies can interfere into the pass band during the process of resampling as a result of non-sufficient suppression of the frequency replicas during the previous step of continuous restoration.

It is the stop band of the interpolator, which is responsible for suppressing the frequencies, which cause the above mentioned effects.

In this report, the terms "modified B-splines" and "combination of B-splines" will be used as synonyms.

By the use of the term "interpolator" we mean an interpolator, based on the modified B-splines kernel.

.

ACKNOWLEDGEMENT

For the acknowledgement use the unnumbered section layout.

REFERENCES

[1] C. de Boor – "A Practical Guide to Splines", revised edition, Springer, 2001.
[2] M. Unser and A. Aldroubi, "Polynomial Splines and wavelets – A signal processing perspective," in wavelets – a Tutorial in Theory and Applications, C.K, Chui, Ed., San Diego: Academic Press 1992.
[3] M. Unser and A. Aldroubi, and M.Eden, "Fast B-spline transforms for continuous image representation and interpolation", IEEE transactions Patters Anal. Machine Intell, vol.13 no.3, 1991.
[4] C. de Boor, "On calculating with B-splines", J. Approximation Theory, vol.6, pp. 50-62, 1972.

Diphone Analysis of the Macedonian Language for the Purpose of Text-to-Speech Synthesis

Branislav Gerazov[1] and Zoran Ivanovski[2]

Abstract – **The paper presents the results obtained from a thorough phonetic analysis of the Macedonian language for the purpose of diphone-based Text-to-Speech (TTS) synthesis. A Body of Text containing more than 2300000 words was compiled and analyzed. From it a set of 707 unique diphones was extracted, and is presented in this paper. The set is a necessary starting point for the development of diphone-based TTS for Macedonian. It reduces the man-hour cost involved in building the TTS system's unit inventory by 40%. This is the first paper to present such a thorough analysis of Macedonian diphone characteristics, and is of importance not only for TTS development in Macedonia, but also on the international level for multi-lingual TTS systems.**

Keywords – **Text-to-Speech, diphone, phonetics, Macedonian.**

I. Introduction

There are three paradigms of Text-to-Speech (TTS) synthesis: articulatory, formant and concatenative synthesis. Up until the 1990's, TTS synthesizers predominately used articulatory and formant synthesis. Both of these are based on a model of the speech production process. Because of the cost involved in building the models, TTS synthesizers based on these two paradigms were restricted only to the world's most wide-spoken languages, i.e. English, French, German, Chinese etc.

The third paradigm, concatenative synthesis, abandons modeling speech production altogether. It uses a data base of prerecorded segments of natural speech, which it concatenates one after the other, to generate the requested speech output. Concatenation of prerecorded (natural) segments, also called units, gives the synthetic speech a very natural sound. Various unit lengths are used in concatenative synthesis, the longer the units the more natural the output speech, but the bigger the database. For example a given language may have only 30 – 40 phones, which are the shortest phonetic (acoustic) units of a language; but more than 200000 words. Most systems rely on a compromise between the two – the diphone, i.e. two half-phones including the transition between them. Diphones usually number around 1000 and provide for reasonable speech quality. The most popular diphone based systems are AT&T's diphone based TTS system and Dutoit's MBROLA synthesizer [1, 2].

Owing to the simplicity of concatenative synthesis and the widespread availability of computing power, in 1990's TTS systems began to be developed in languages that lacked world domination. In the Slavic speaking part of the Balkans various systems have appeared, most of them based on diphones. The most prominent of these is the "AlphaNum" unit-selection TTS system for Serbian and later Croatian, [3]. Other systems include "SpeechLab" for Bulgarian and "Govorec" and "Proteus" for Slovenian.

Several attempts have been made to synthesize speech in Macedonian. The first attempts date back to 1997, when a concept solution was found, but never became fully operational, [4]. Later attempts included emulating Macedonian using the Croatian diphone inventory under the MBROLA framework, [5], with a similar approach by the AlfaNum team, this time using a Serbian diphone inventory, [6]. Currently, two high quality systems are under development: TTS-MK, a diphone based system in development at FON University, [7], and "Speak Macedonian", our system, which is based on modified diphone units, [8].

The set of diphones found in a language is an essential starting point for developing a diphone-based TTS system for that particular language. This is highly specific to the language at hand, and calls for research of its phonetic characteristics. Although TTS systems in Macedonian have been in development for some time, results of this kind have not been published yet. This paper presents the results obtained from a thorough phonetic analysis of a large body of text in Macedonian, totaling more than 2300000 words. From the analysis data, the diphone set needed for TTS in Macedonian was extracted and statistically processed. We have found 707 unique diphones which form a practically complete basis for a diphone based concatenative synthesis TTS system. The results presented herein can be of value not only for TTS development efforts in Macedonia, but also for international multi-lingual TTS system development.

II. Diphones

Although the phone is the basic acoustic unit of speech, synthesizing speech by concatenating phones yields bad results. The quality of the synthesized speech is poor, with mediocre intelligibility and no naturalness, [9]. This is because of the great difficulty in simulating the gradual change of acoustic information from one phone to the next, called an interphone transition. Because of this systems rarely rely on phone based units.

The more common approach is to use a unit that includes half of each phone and the transition between them, called a diphone. This way the concatenation point is no longer the transition interval, but rather the middle, somewhat, steady-state of the phone.

Theoretically the number of diphones is given by the number of phones squared. However, due to phonotactic constraints, i.e. restrictions on phone pairing in language, this is not the case. For example, English does not endorse an "sh-t" transition, in IPA (International Phonetic Alphabet) terms /ʃ/-/t/, which is common in Macedonian and other Slavic languages. The actual number of

[1]Branislav Gerazov is with the Faculty of Electrical Engineering and Information Technology at University of Ss. Cyril and Methodius, Skopje, Macedonia, E-mail: gerazov@feit.ukim.edu.mk

[2]Zoran Ivanovski is with the Faculty of Electrical Engineering and Information Technology at University of Ss. Cyril and Methodius, Skopje, Macedonia, E-mail: mars@feit.ukim.edu.mk

diphones varies between languages and is usually far from this theoretical value. For example English has 43 phones and 1162 diphones, as used by AT&T's TTS system. This is far from the theoretical 1849 possible diphone combinations, [1].

III. MACEDONIAN PHONES

Macedonian is comprised of 33 basic phones, 28 of which are represented a unique letter in the alphabet. The letters together with their IPA (International Phonetic Alphabet) equivalents are presented in Table I. Of these, five are vowels (/a/, /ɛ/, /i/, /ɔ/ and /u/), and the rest are consonants. Grouped according to the manner of articulation the consonants are comprised of: 6 plosives (/b/, /p/, /g/, /k/, /d/, /t/), 3 approximants (/j/, /r/), 7 fricatives (/v/, /f/, /z/, /s/, /ʒ/, /ʃ/, /h/), 6 affricates (/dʒ/, /tʃ/, /dz/, /ts/, /ɟ/, /c/) and 3 nasals (/m/, /n/, /ɲ/) [10]. The letter "л" reads /ɬ/ but also /l/ when preceding the front vowels /ɛ/ and /i/, and the approximant /j/. The letter "љ" reads /lj/. The cluster /lj/ was treated as a phone, due to its compactness and the tendency to palatize it in common speech. The phone "r" can become syllabic /r̩ / when enclosed by consonants, as it is in: "првиот" /pr̩viɔt/ (eng. the first), " 'рбет" /r̩bɛt/ (eng. spine) etc. Finally "n" is velarized to /ŋ/ before /k, g/, as in "банка" /baŋka/ or "англиски" /aŋgliski/, [11]. The phones /l/, /r̩ / and /ŋ/ were coded with "q", "w" and "n" in the text, in the phonetic processing step.

TABLE I
PHONE INVENTORY OF MACEDONIAN WITH CORRESPONDING IPA EQUIVALENTS

А	/ a /	И	/ i /	С	/ s /
Б	/ b /	Ј	/ j /	Т	/ t /
В	/ v /	К	/ k /	Ќ	/ c /
Г	/ g /	Л	/ ɬ /, / l /	У	/ u /
Д	/ d /	Љ	/ lj /	Ф	/ f /
Ѓ	/ ɟ /	М	/ m /	Х	/ x/
Е	/ ɛ /	Н	/ n /, / ŋ /	Ц	/ ʦ /
Ж	/ ʒ /	Њ	/ ɲ /	Ч	/ ʧ /
З	/ z /	О	/ ɔ /	Џ	/ ʤ /
S	/ ʣ/	П	/ p /	Ш	/ ʃ /
		Р	/ r /, / r̩ /		

IV. MACEDONIAN DIPHONE ANALYSIS

With a phone count of 31 + /r̩ / + / l / + / ŋ / + /silence/ = 35, the theoretical maximum of diphones for Macedonian is 1225. Again due to Macedonian phonotactics the diphone inventory size is significantly less that the theoretical maximum. To find the exact number of diphones in Macedonian, a large body of text was assembled from various sources. This text was phonetically processed and analyzed in two main process lines shown in Fig. 1. They will be further elaborated in the paper.

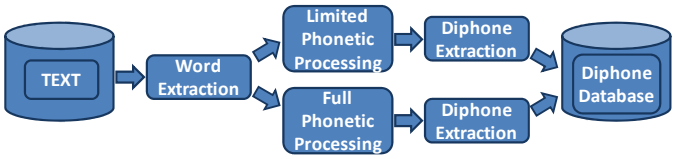


Fig. 1. Two processing lines for Macedonian diphone set extraction

A. Body of Text

The body of text analyzed in our work has a total of more than 2300000 words. It consists of three major parts. The first part, termed the *Reference Body*, was compiled from works of classic and contemporary literature in Macedonian, that were readily accessible on-line. Special care was taken to use a representative cross section of both standard and contemporary trends in the Macedonian language. In addition the Macedonian Constitution as well as various acts of law were included.

The second part of the body was assembled from the Old and New Testament. It's word-count is over 800000. This text was processed separately because the Bible contains names of archaic places and people not in general use today. These distinctive words were later necessarily excluded from the diphone analysis.

The third part of the body comprises a 260000 word Macedonian Dictionary made available by *OpenOffice.org*. Due to the lack of valid word-counts in it, a relevant statistical distribution of diphones could not be extracted. However, the Dictionary was invaluable due to its wide coverage of words in Macedonian. This kind of coverage would not be possible even by making the body of text become arbitrarily large.

B. Word Extraction

The starting point of the diphone analysis process was word extraction from the bodies of text. In total, from the approximately 1300000-word Reference Body, close to 80000 unique words were extracted. Of these, about 44000 occurred only once in the entire body of text. The 404 most common words accounted for 50% of the total word count. The analysis of the Bible text yielded around 40000 unique words, of which 17062 appeared only once, and 94 most common accounted for 50% of words in the text. This data is summarized in Table II.

Here we can see why the Dictionary was invaluable to our analysis. Of about 2100000 words from the Reference Body and the Bible, only 120000 are unique and 55000 of these appeared only once in the whole text. This means that no matter how large the assembled body of text is, it can never cover all the words in a given language. The inclusion of the Dictionary, with its wide coverage of words, reduces this problem.

Here we can see why the Dictionary was invaluable to our analysis. Of about 2100000 words from the Reference Body and the Bible, only 120000 are unique and 55000 of these appeared only once in the whole text. This means that no matter how large the assembled body of text is, it can never cover all the words in a given language. The inclusion of the Dictionary, with its wide coverage of words, reduces this problem.

Table III gives the 10 most common words in the Reference Body and the Bible. It can be seen that the two line up well, i.e. 8 of the words are the same, with the first 5 being almost

equal. This data can be useful should the diphone database be expanded with whole words.

TABLE II
WORD STATISTICS FOR THE THREE PARTS OF THE BODY OF TEXT USED IN THE DIPHONE ANALYSIS

	Word count	Unique	Appeared once	Acc. for 50%	Acc. for 90%
Reference Body	1287513	79065	38593	404	21972
Bible	813878	38572	17062	94	5795
Dictionary	261460	n/a	n/a	n/a	n/a

TABLE III
10 MOST COMMON WORDS IN THE REFERENCE BODY AND THE BIBLE

	Reference Body			Bible		
	Word	Occurences	%	Word	Occurences	%
1	на	59953	4,66	и	45340	5,57
2	и	44937	3,49	на	29195	3,59
3	се	33420	2,60	се	18081	2,22
4	да	33020	2,56	да	17127	2,10
5	во	26959	2,09	во	16306	2,00
6	од	24400	1,90	го	15902	1,95
7	за	21733	1,69	од	15264	1,88
8	со	19494	1,51	ќе	14437	1,77
9	го	17508	1,36	за	11188	1,37
10	не	17496	1,36	што	10957	1,35

A statistical curve, in log-log scale, of the distribution of word appearance against word rank in the frequency distribution table is given in Fig. 2. From the linearity of the plot, we can see that it clearly abides to Zipf's power law, i.e. the frequency of a word is inversely proportional to its rank in the occurrence table, [12].

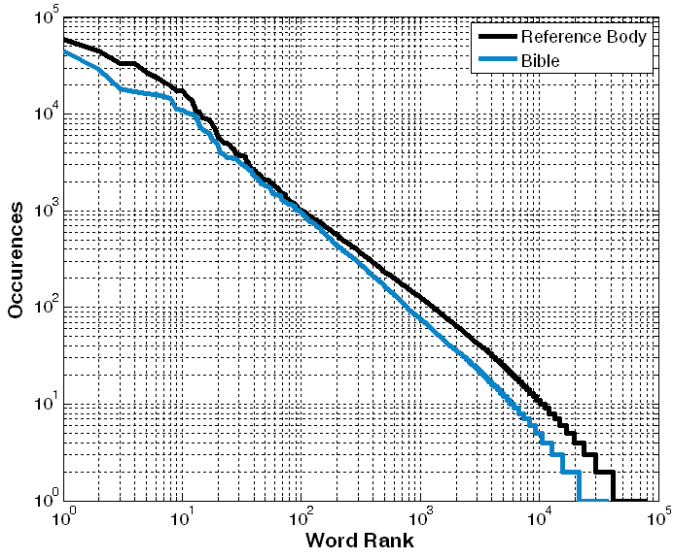


Fig. 2. Word distribution plotted in log-log scale in the Reference Body and the Bible

C. Phonetic Processing

The phonetic processing step compensates for the difference in the rules of orthography (standard writing) vs. those of orthoepy (standard speaking). It was done on two levels: a *limited* ("light") and a *expanded phonetic processing level*.

- *limited phonetic processing level* – singles out the appearance of / r̩ /, / l / and / ŋ /, and codes (substitutes) them with "q", "w" and "n", respectively. The cluster /lj/ represented by "љ", was left intact.
- *expanded phonetic processing level* – builds on the previous level and additionally implements an expanded set of Macedonian orthoepy rules, comprising: voicing assimilation in consonant pairs, reduction of consonants due to assimilation, and devoicing of voiced consonants at the end of words, [10]. This is expected to decrease the total diphone count.

D. Diphone Extraction

Diphone extraction is then carried out on the words that have been phonetically processed. All diphones are extracted and a diphone data base is compiled together with their number of occurrences.

V. RESULTS

This section summarizes the results obtained from the diphone analysis of Macedonian. The total number of diphones from the Reference Body, the Bible, and the Dictionary are given in Table IV. Due to the uniqueness of the Bible texts, two statistics were drawn out, one from the complete word-count, and another from a reduced word-count that excluded the diphones unique to biblical places/people.

Three of the texts: the Reference Body, the Bible Reduced and the Dictionary, were used to generate a single consolidated list of diphones. This *Master List* gives the total count of 728 unique diphones found in the Macedonian language. This is 59,4% of the theoretical maximum of 1225, for the 35 phones used. Its statistics are also given in Table IV. The second processing line, which implemented expanded phonetic processing, yielded a smaller orthoepy based diphone set totaling 707 diphones, or 57.7% of the theoretical maximum. This set is invaluable to the task of building a unit inventory for a diphone-based TTS system for Macedonian, because it provides for a considerable 40% decrease of the man-hour cost involved in the process!

TABLE IV
DIPHONE STATISTICS FOR THE THREE PARTS OF THE ANALYZED TEXT

	Diphone count	Unique	Appeared in a single word	Acc. for 50%	Acc. for 90%
Reference Body	7508399	708	16	41	192
Bible	4407721	680	41	39	170
Bible Reduced	4402611	622	23	39	170
Dictionary	2665939	705	16	47	213
Master List	14582781	728	6	43	195

It should be mentioned that there is a possibility that words in the Macedonian language have not appeared at all in the Body of Text used for the analysis. There is a chance some of these words to contain a diphone not found in this analysis. However, the authors consider this chance as being minute.

Table V shows the 10 most common diphones in the three texts. It can be seen that there is large correlation between the Reference Body and the Reduced Bible texts. The Dictionary most common diphone set differs, because this text contains every word only once (even the most common), while the first two contain natural language word distributions. Even so, a small degree of correlation can still be found.

TABLE V
MOST COMMON DIPHONES IN THE ANALYZED TEXTS

	Reference Body			Bible Reduced			Dictionary		
	Diph.	Occ.	%	Diph.	Occ.	%	Diph.	Occ.	%
1	а#	358709	4,78	а#	195037	4,43	а#	67485	2,53
2	е#	221411	2,95	е#	148312	3,37	ва	67411	2,53
3	о#	197631	2,63	и#	136874	3,11	е#	58687	2,20
4	и#	185966	2,48	о#	110728	2,52	на	51552	1,93
5	#с	154224	2,05	#с	93976	2,13	#п	48735	1,83
6	#н	152863	2,04	#н	91613	2,08	ан	46544	1,75
7	на	152414	2,03	на	73459	1,67	ув	45749	1,72
8	#п	108475	1,44	#и	72787	1,65	ра	41678	1,56
9	#д	97997	1,31	#п	61978	1,41	ни	41006	1,54
10	то	89878	1,20	то	54317	1,23	о#	38590	1,45

Fig. 3 gives the log-log diphone distribution against diphone rank for the three texts, and the consolidated Master List. It can be seen they only partially follow Zipf's law with a sharp roll-off near the end.

The diphone set extracted is given in matrix form in Table VI. Phones are given in Cyrillic as well as IPA characters. The rows represent the first phone in the diphone, while the columns represent the second. Diphones that appeared in both the texts with limited and expanded phonetic processing were marked with an "x". Diphones unique for the text with limited phonetic processing were marked with a "g" (orthography), and the ones unique to the text with expanded phonetic processing with an "e" (orthoepy). The total number of diphones is 732, 728 of these abide orthography and 707 orthoepy.

VI. CONCLUSION

We have presented the results of a phonetic analysis of the structure of the Macedonian language in the respect of diphones. These are the first results from phonetic diphone analysis of Macedonian, for the purpose of TTS synthesis.

The results are of great value for the development of TTS systems in this language, whether in Macedonia or on an international level. The set of 707 diphones acquired through this analysis is a necessary starting point for building unit databases for diphone-based TTS systems in Macedonian. Using the set cuts the man-hour cost of this process down by 40 %.

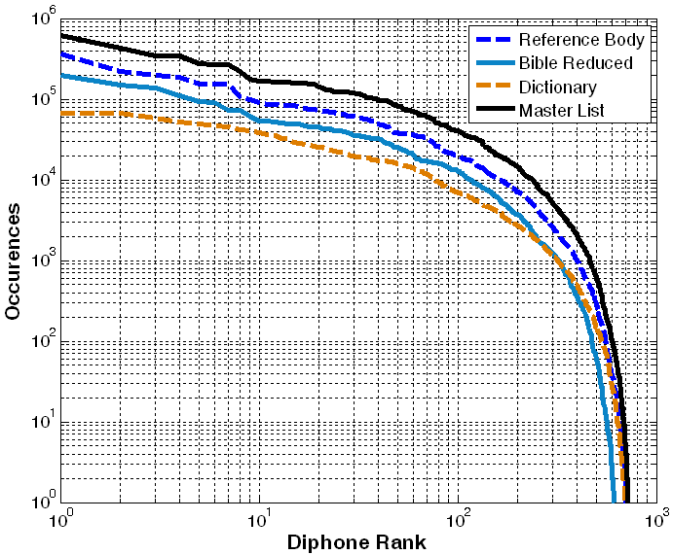


Fig. 3. Log-log scale graph of diphone occurrences vs. rank in the distribution table in the three analyzed texts

REFERENCES

[1] J. Olive, J. van Santen, B. Möbius and C. Shih. "Synthesis". In R. Sproat (Ed.), *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, pp. 191-228, Kluwer, 1998.

[2] T. Dutoit, V. Pagel, N. Pierret, F. Bataille and O. Vrecken, "The MBROLA Project: Towards a Set of High Quality Speech Synthesizers Free of Use for Non-Commercial Purposes", Proceedings of ICSLP 96, 1996.

[3] M. Sečujski, R. Obradović, D. Pekar, Lj. Jovanov and V. Delić, "AlfaNum System for Speech Synthesis in Serbian Language", In Proc. of the 5th Conf. Text, Speech and Dialogue, Brno, 2002.

[4] L. Josifovski, D. Mihajlov and D. Gorgevik, "Speech Synthesizer Based on Time Domain Syllable Concatenation", SPECOM '97 Cluj-Napoca, Oct. 27-30, 1997.

[5] М. Зрмановска, *Синтеза на македонски говор врз база на текст со прозодија*, Магистерска Теза, Електротехнички Факултет, Скопје, 2005.

[6] V. Delić, M. Sečujski, D. Pekar, N. Jakovljević and D. Mishković, "A Review of AlfaNum Speech Technologies for Serbian, Croatian and Macedonian", IS-LTC 06, Ljubljana, Slovenia, 9. - 10 october, 2006

[7] S. Chungurski, I. Kraljevski, D. Mihajlov and S. Arsenovski, "Concatenative Speech Synthesizers and Speech Corpus for Macedonian Language", 30th International Conference ITI Cavtat/Dubrovnik, Croatia, Jun. 23-26, 2008.

[8] B. Gerazov, G. Shutinoski and G. Arsov, "A Novel Quasi-Diphone Inventory Approach to Text-To-Speech Synthesis", MELECON '08, Ajaccio, France, May 5-7, 2008

[9] Б. Геразов и Г. Шутиноски, "Еден Пристап кон Преобразба на Текст во Говор на Македонски", ЕТАИ 2007, Охрид, Македонија, 19 – 21 сеп., 2007.

[10] С. Бојковска, Л. Минова-Ѓуркова, Д. Пандев и Ж. Цветковски, *Опита Граматика на Македонскиот Јазик*, Просветно Дело, 2008.

[11] V. Friedman, "Macedonian", SEELRC, 2001.

[12] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, p. 23, MIT Press, 1999.

TABLE VI
DIPHONE MATRIX FOR MACEDONIAN
DIPHONES FOUND IN: BOTH ORTHOGRAPHY AND ORTHOEPY (x) ; ORTHOGRAPHY ONLY (g) ; ORTHOEPY ONLY (e)

Cyrillic		а	б	в	г	д	ѓ	е	ж	з	ѕ	и	ј	к	л	љ	м	н	њ	о	п	р	с	т	ќ	у	ф	х	ц	ч	џ	ш	#	q	w	n
	IPA	a	b	v	g	d	ɟ	ɛ	ʒ	z	ʣ	i	j	k	ɫ	lj	m	n	ɲ	ɔ	p	r	s	t	c	u	f	x	ʦ	ʧ	ʤ	ʃ	#	r,	l	ŋ
а	a	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
б	b	x		x	x	x		x		x	x	x	x		x	x	x	x		x		x		g		x				g			g	x	x	
в	v	x	x	x	x	x		x	x	x	x	x	x	g	x	x	x	x		x	g	x	g	g		x			g	g	x	g	g	x	x	
г	g	x	x	x		x		x		x		x			x		x	x	x	x	g	x	g	x		x				g		g	g	x	x	
д	d	x	x	x	x	x		x	x	x	x	x	x		x	x	x	x	x	x	g	x	g	g		x		g	g		x		g	x	x	
ѓ	ɟ	x						x				x						x		x						x							g			
е	ɛ	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
ж	ʒ	x	x	x	x	x		x				x	x		x		x	x		x		x		x		x							g	x	x	
з	z	x	x	x	x	x		x	x	x	x	x	x		x	x	x	x		x		x		x		x					x		g	x	x	
ѕ	ʣ	x		x				x				x										x				x								x		
и	i	x	x	x	x	x	x	x	x	x		x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
ј	j	x	x	x	x	x		x	x	x		x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	
к	k	x		x				x				x			x		x	x	x	x	e	x	x	x		x			x	x		x	x	x	x	
л	ɫ	x	x	x	x	x			x	x				x			x	x	x	x	x		x	x		x	x	x	x	x	x	x	x			x
љ	lj	x										x								x				x		x											
м	m	x	x	x	x	x		x				x	x	x	x		x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	
н	n	x	x	x		x	x	x	x	x	x	x			x					x	x	x	x	x		x	x	x	x	x	x	x	x		x	
њ	ɲ	x						x				x								x			x			x				x				x			
о	ɔ	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
п	p	x						x				x	x	x	x	x		x		x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	
р	r	x	x	x	x	x	x	x	x	x		x	x	x	x		x	x		x	x		x	x		x	x	x	x	x	x	x	x		x	
с	s	x		x		x		x				x	x	x	x		x	x		x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	
т	t	x		x	x	x		x				x	x	x	x	x	x	x		x	x	x	x	x		x	x	x	x	x		x	x	x	x	
ќ	c	x		x				x				x		x				x		x				x		x								x			
у	u	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
ф	f	x						x				x		x	x			x		x	e	x	x	x		x			x	x		e	x	x	x	
х	x	x		x				x				x			x		x	x		x		x		x		x		x		x			x	x		
ц	ʦ	x		x				x				x		x	x		x	x		x		x				x							x	x	x	
ч	ʧ	x	x	x				x				x	x	x	x		x	x	x	x	x			x		x							x	x	x	
џ	ʤ	x		x	x			x				x					x			x						x							g			
ш	ʃ	x		x				x				x	x	x	x		x	x	x	x	x	x		x		x			x	x		x	x		x	
#	#	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	
q	r,		x	x	x	x	x		x	x				x	x	x	x	x	x		x	x	x	x			e	x	x	x	x	x			x	x
w	l							x				x	x																							
n	ŋ				x									x																						

Session

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES – 1

The realization of the OO database in the specific CAD/CAM applications

Dejan S. Aleksić[1], Dragan S. Janković[2]

Abstract – **This paper describes the realization of the object-oriented database in the specific CAD/CAM software package supporting the design and manufacturing of the facade carpentry. The application of this software package creates a number of limitations and problems for whose solution a class group Attr API was created. Attr API class allow us to generating the various application variants, starting from the usual desktop application, to the client/server and Internet variant application.**

Keywords – **CAD/CAM database, object-relational mapping, OODB, design and manufacturing of the facade carpentry.**

I. INTRODUCTION

Specialized CAD/CAM programs commonly solve a small number of problems. They are justified as long as they meet the user's needs and solve the problem they were created for. On the other hand, narrowing of the application range does not necessarily lead to decreasing the number of solutions general enough to be used in all other/future applications within the specific area.

This paper illustrates the realization of the object-oriented database in the CAD/CAM software package supporting the design and manufacturing of the facade carpentry [1].

Tha facade carpentry is made out of a number of profiles and items together with the corresponding parts needed to join and assemble them. Profiles may be made of various materials such as aluminium, PVC, wood, iron, or their combinations (aluminium-wood, PVC-aluminium). There is a great number of profile manufacturers, each of them producing lots of profile systems. All profile systems are supposed to abide to the general rules and standards, and yet, each one of them has its own specific characteristics which make it unique. It is precisely because of this that the software packages specialized in giving support in that area should, beside the obviously narrowed application area, offer the solutions general enough to accept and process all the specific characteristics of a profile system. A particular difficulty arises from the fact that it is almost impossible to foresee all the specific solutions in the profile systems which are likely to appear in the future and which must undoubtedly be supported by such software paskages.

[1]Dejan S. Aleksić is with the Faculty of Sciences and Mathematics, University of Nis, Visegradska 3, 18000 Nis, Serbia, e-mail: dejan_aleksic@yahoo.com.

[2]Dragan S. Janković is with the Faculty of Electronics, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, e-mail: dragan.jankovic@elfak.ni.ac.yu.

II. THE CHARACTERISTICS OF APPLICATION

The application of this kind of CAD/CAM as well as its users creates quite a number of characteristics and limitations [2]. We will deal with the most characteristic ones in this paper.

The information concerning the profile systems must undoubtedly be kept in the database; yet, writing in the base occurs rarely whereas reading of the same data is quite intensive, so that the base is Read Only DB in most of the cases. Besides the need for intensive reading of the relatively small amount of data, the critical factor is the reading speed which directly affects the speed at which the main evaluation module works. The factor of the data access speed to the profile system is one of the most significant ones which should be considered with the utmost care in the realization of such an application [3].

It has already been mentioned that there are a great number of profile systems having a lot of common characteristics and rules which will certainly be valid for all future profile systems. Apart from having the common characteristics, each profile system possesses some specific traits which make it different/unique compared to other systems. It is almost impossible to foresee which characteristics some future profile system will have. The second important factor is the possibility to modify the existing application so that it could accept the specific characteristics of the new systems.

Besides the profile systems data, the manipulation with the particular projects data should be attained. The project data may be kept either in the database or on the disk as an independent file [4]. The project data are usually stored in the base in the server/client software variant, but in case of physically separate representatives or operations, the variant of storing the projects into the data bank is used. Each project contains some data concerning the profile system which are practically the copy versions of these data stored in the description base of the profile systems. It is precisely this doubling of the data in the base and projects with the aim of maintaining them (the data changes by number and type, the content changes or the version improvement) and the synchronization after the changes that represent the problem we encounter next.

We are rather inhibited concerning the database choice by the fact that these applications are mainly used by small firms with an insufficient knowledge regarding the database and system maintainance, and by the fact that those applications are supposed to be of a reasonable price.

In accordance with what has been mentioned previously, it is clear that the description of the profile systems and the rules within them will be performed (with rare exception) not by the users themselves, but that the application will have to be

delivered with the already 'filled' bases for the particular profile systems. Such a 'centralized' maintainance of the profile system description and rules database requires a very sophisticated version improvement system of both the database and the application itself.

The multilingual support of the UI application characterizes every resourceful software package and there are quite a number of 'ready-made' solutions for its realization. The problem we encounter arises from the fact that the data stored in the profile system description base have to be translated. The substantial complexity of the system descriptions, their great number and the necessity to support a rather large number of languages make the simple creation and maintainance of the separate base for each profile system and each language impractical and almost impossible.

Concerning the great number of systems, the complexity of the project data processing and the complex report system, it is clear that a hierarchical organization of the data within both the system description base and the project itself is needed.

III. The System Description using Attr API

Regarding the mentioned facts, we chose the open source relation database FireBird. Since this database is available to lots of OS platforms the easy migration from the single user variant to the client/server variant was created. Thus all the requirements concerning the availability, price, easy installation and maintainance were fulfilled.

Unfortunately, such a choice brought about some limitations especially those regarding the fixed number of the fields within the tables and the beforehand defined data types, as well as the data reading speed. Our soulution was based on the creation of a separate group of objects within the application itself that gave us flexibility concerning the number and type of the entity descriptions data, but with the realized mechanism for the data type check. The basis of this approach is the usage of the so called attribute, i.e. the base class tAttribute with the group of its descending classes. Each entity can be described using the arbitrary number of attributes, and each attribute has its type and the momentary value from the list of possible values. Such an approach created a number of advantages, such as the easy realization of some small and yet big enough differences in the profile system description, the multilingual concept support within the base, the easy improvement of the existing database version, the automatic copying of the database into the memory structures and the disk and vice versa [5], the quick and easy project data recording and input onto/from the disk, the automatic data improvement in the base, the possibility to realize the automatic improvement of the recorded project data in accordance with the data changes in the base concerning their value, number and type, the realization of all the advantages created by the object-oriented data approach (data abstraction, inheriting, attribute overburdening) over the data stored in the base, memory or disk.

All the manipulations concerning the input, changes and reading of the attributes were realized within the tAttributes class and tAttributeList.

IV. The Attr API Realization

The *Attr API* itself contains a number of classes but its total functionality can be presented through the three main classes - *tAttribute*, *tAttributeList* and *tXalNode* with their descending classes. The basic functionality when working with attributes was realized in the base class *tAttribut*e. Precisely speaking, class *tAttribute*, being the base class, does not contain much of the functionality but it only defines the basic variables and methods (virtual and abstract) that will be copied and realized in one of the descending classes *tAttrString*, *tAttrInteger*, *tAttrReal* and *tAttrBoolean* (see Fig. 1.). Initially, the support for the four basic data types was realized, although it is possible to realize the support for some arbitrary data type through the new hereditary *tAttribute* class.

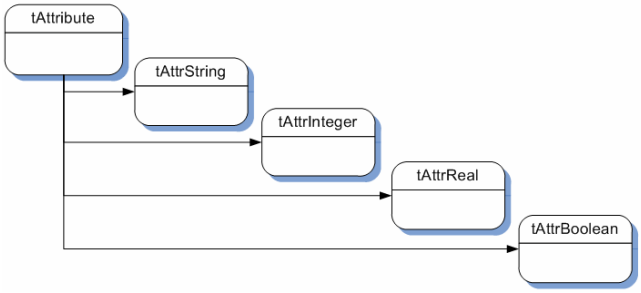


Fig. 1. Inherites tree of base class tAttribute

The second basic class is *tAttributeList* by which the keeping and manipulation of the attribute list was realized, more precisely of the instances of the base class *tAttribute,* i.e., its descending classes. The number of attributes, their type and value may be completely arbitrary.

The third crucial class of *Attr API* is *tXalNode* as the hereditary class descending from the class *tList* (see. Fig. 2.). Its primary task is the formation and manipulation of the basic hierarchical structures. Each instance of this class represents the knot of the hierarchical structure. Each knot of that structural tree has its own list of attributes which describe its state [see Fig. 2.].

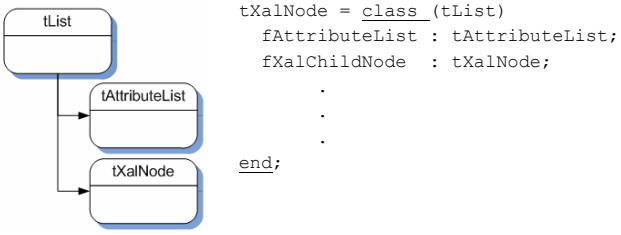


Fig. 2. *tXalNode* - crucial class of *Attr API*

The *Attr API* classes make possible the hierarchical organization and processing of the data stored in the memory as a group of objects – instances of those classes. The data access is quite quick in that case, but the need for a permanent keeping of those data still remains. The input/read process of the hierarchical structure of the *Attr API* objects in/from the base was completely realized within the basic classes, which is completely transparent for the class' users. All the data, together with their hierarchichal structure, no matter how complex it may be or what type or number of attributes are stored in it, are put in only three tables within the relation

database (see Fig. 3.). Furthermore, if lots of hierarchical structures are defined in one application (using *Attr API* classes), then all of them are placed in the same database in the already mentioned three tables.

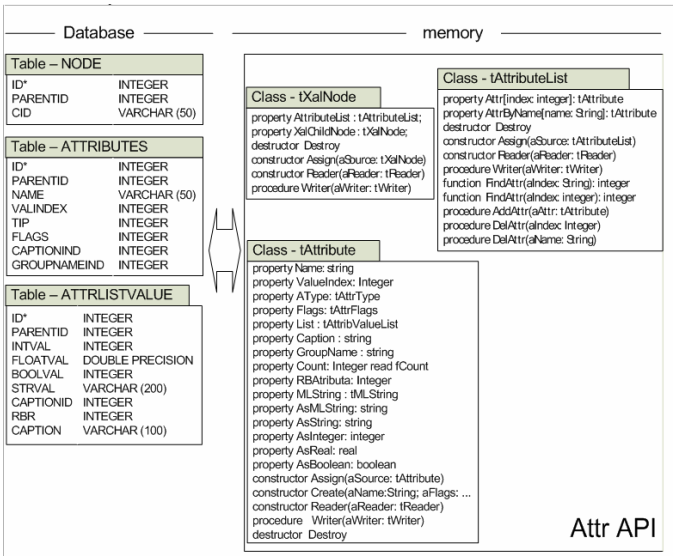


Fig. 3. Relationship between data in relation database and memory (in class of *Attr API*)

It is important to emphasize the possibility of the easy creation of the complex hierarchical structures by using the classes from *Attr API*. The user creates the new classes by commonly inheriting the class *tXalNode*. The additional functionality is realized by copying the methods from the parent class and by creating the new methods. The data in those newly defined classes 'are kept' in the attribute list (class *tAttributeList*) which makes the input/read process of the data in/from the base completely transparent for the user because it has already been realized in the base classes. On the other hand, the user has an absolute freedom in its own class while creating the attributes concerning both their number and their type or values.

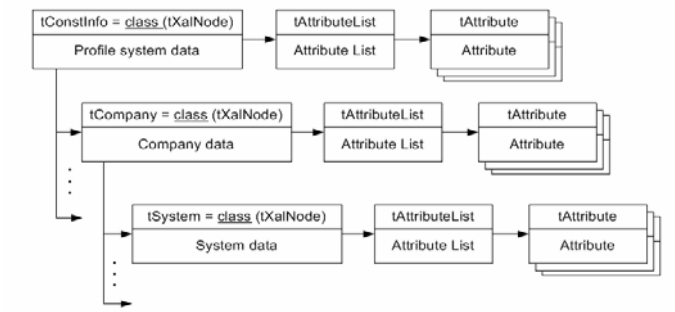


Fig. 4. The profile system description data

The Fig. 4. shows the hierarchical structure created for the manipulation with the profile system description data. The data from this structure are typically edited with a special program (*ConstInfo Manager*) and are put into the database. The main application only reads the data from that structure so that they are quite often cashed because of the data reading speed.

The second example is the hierarchical structure which manipulates the project descripition data (see. Fig. 5.).

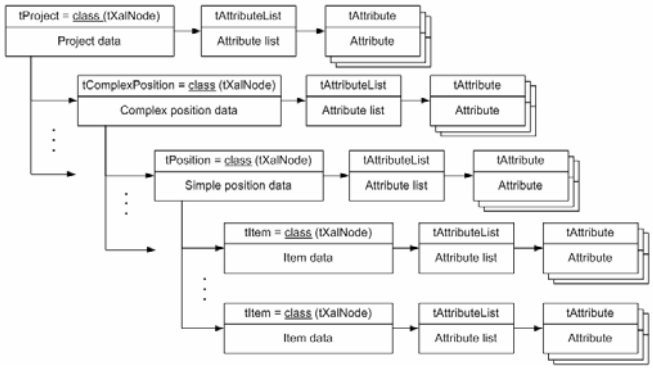


Fig. 5. The hierarchical structure for the project descripition data

Depending on the configuration of the main application, this structure is placed in the file on a disk (common/classical application) or in the database (client/user application). The total functionality has already been realized in both cases in the base classes of *Attr API,* while the user himself has to decide where the data will be stored.

Generally speaking, the hierarchical structures of the Attr API objects can be recorded/input into the various forms and formats. The first of them is the usage of some of the relation databases, the choice of database not depending on the functionality. It is possible to read from one relation database within the same application and to write the processed data into the relation database of some other manufacturer. These operations are completely transparent for the Attr API class users. The working speed makes possible the updating of the data of a particular part of the hierarchical structure from the memory into the database without writing in all the data of that structure. Using the file on a disk for storing the object hierarchical structures is another way of permanently keeping the data from the Attr API objects.

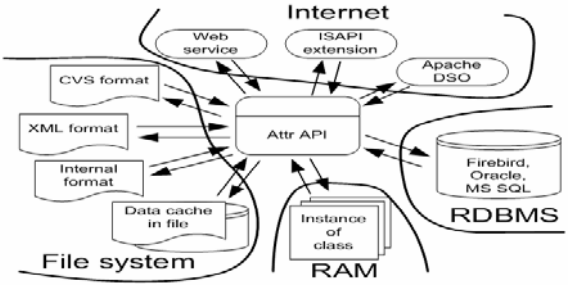


Fig. 6. Attr API objects can be recorded/input into the various forms

Several write-in data formats in the file on a disk are supported (see Fig. 6.) depending on the users' needs. Each one of the formats has its own advantages and disadvantages concerning the crucial parameters, such as the speed of write/read (internal format), the uncomplicated exchange of data with other general programs (CVS format) or in self-desc format (XML). A possibility of the automatic data copying creation from the database into the file on a disk should be particularly emphasized, since it is frequently used for cashing of data on the client side.

The data availability on the Internet with the maximum possible protection is realized by the special modules so that client can exchange data with the server side. The premise of the transparency of the way of keeping and accumulating data

is completely maintained and it is realized without any additional actions on the part of the users.

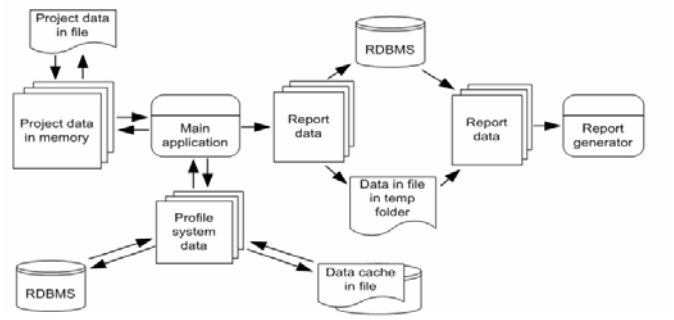


Fig. 7. One common application

This fact allows us the easy realization of the various configuration applications. The Fig. 7. illustrates one common application which uses *Attr API* for keeping and manipulating with the data. There are three different hierarchical structures – the project data, the profile systems data and the report generating data.

In this case, the project data are stored in the file on a disk while the profile systems data (read only) are usually automatically cashed from the database into the file on a disk.

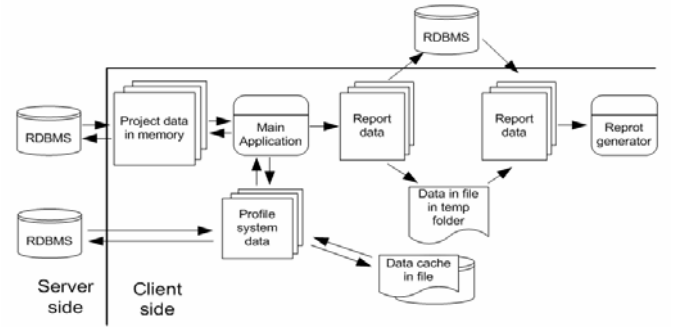


Fig. 8. The client/server variant of application

The data needed for the report generating are stored in either temporary file on a disk or the database itself, depending on the fact whether the user wants to have the reports available later or not. The possibility of the module realization for the report generating not depending on the main application version appears to be an additional advantage. Since the data format is not fixed, and concerning the fact that it is possible to detect the existance of the particular attribute in the hierarchical structure by using the program, then the report generating of various forms becomes quite easy.

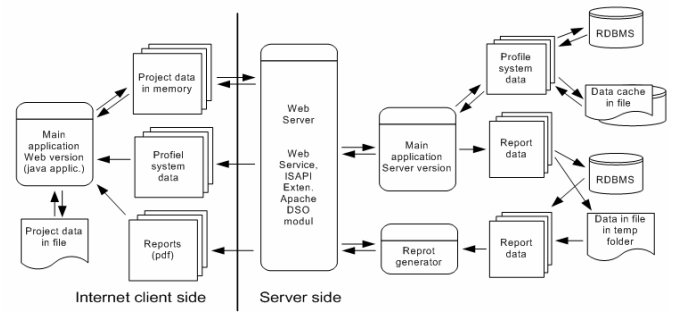


Fig. 9. The web client/server variant of the application

The advantages of using Attr API are particularly evident if we want to realize the client/server variant of our application (see Fig. 8). All we need to do is to define the place and way of storing the data in the configuration Attr API. It is important to point out that it is not necessary to change the remaining part of the application which manipulates the hierarchichally organized data. Following the given logic, the realization of the web client/server variant of the application is not too complicated. The server page has to undergo minimal changes in the logic, whereas the situation is quite similar on the client page, apart from the fact that the main application in the web client variant and Attr API are realized in the Java program language.

V. CONCLUSION

This paper illustrates the realization of the object-oriented database in the CAD/CAM software package for the support of the design and manufacturing of the facade carpentry. The application of this software package creates a number of limitations and problems for whose solution a class group *Attr API* was created. It gives us a necessary flexibility concerning the entity description data number and type with the possibility of checking the data types. Each entity may be described using an arbitrary number of attributes which have a defined type and value from the list of the possible values. This attribute concept offers a number of advantages concerning the modelling of small but considerable differences in the description of profile systems. The classes of Attr *API* allow for the hierarchical organization of the attributes for the profile systems or project description. The process of write/read hierarchical structure of the Attr API objects in/from the database is completely transparent for the *Attr API* users. Generally speaking, the hierarchical structures of Attr API objects can be save/load in various forms and formats (RDBMS, memory, file system, Internet, ...).

All these facts allow for a relatively easy generating of the various application variants, starting from the usual desktop application, to the client/server and Internet variant application.

BIBLIOGRAPHY

[1] Q. Zhang, "Object-oriented database systems in manufacturing: selection and applications", Industrial Management & Data Systems, vol. 101, no. 3, pp. 97-105, 2001.

[2] M.L. Brodie, B. Blanstein, U. Dayal, F. Manola, A. Rosenthal, "CAD/CAM Database Management", IEEE Database Engineering, Vol.7, No.2, pp. 12-20, June 1984.

[3] D. Maier, "Making database systems fast enough for CAD applications", Object-oriented concepts, databases, and applications, ACM Press, New York, NY, 1989.

[4] Ying-Kuei Yang, "An enhanced data model for CAD/CAM database systems", Proceedings of the 25th ACM/IEEE conference on Design automation, pp. 263 - 268, 1988.

[5] M. Blaha, W. Premerlani, H. Shen, "Converting OO Models Into RDBMS Schema" IEEE Software, vol. 11, no. 3, pp. 28-39, May/June 1994.

The use of scripts in a CAD/CAM database

Dejan S. Aleksić[1], Dragan S. Janković[2]

Abstract – **Definition and the execution of combining profiles in profile systems represent one of the key problems in a specific CAD/CAM application for design and manufacture of facade carpentry support. General rules which are common for all the systems are executed within the application. Main mechanism of accomplishing specific rules of combining is comprised of executing scripts affiliated to elements on the lowest hierarchy level of profile system within the database. The script language is specially created for this use and concepts enable us to relatively easy model the rules of combining of some new system with all the specifics it carries with it.**

Keywords – **Scripts , OODB, CAD/CAM, Object-oriented computing, facade carpentry**

I. INTRODUCTION

Specialized CAD/CAM programs usually serve as a solution for only a narrow specter of problems. To justify its existence, they must be maximally adjusted to the user and optimized for the problem they are used for. On the other hand, narrowing its area of usage does not necessarily mean that such programs do not need general enough solutions so that they can be applied to all other/future uses from that area.

Throughout this work, we will illustrate one object-oriented database in a CAD/CAM software package for the project support and the manufacture of the facade carpentry.

During the process of the facade carpentry creation a cluster of profiles and fillings are used, together with the associated parts for their connection and assembly. Profiles must be made of different kinds of materials such as aluminum, PVC, wood, iron or their combination (aluminum-wood or PVC-aluminum). There are lots of profile manufacturers and each of them has a number of profile systems that are used and manufactured. All the profile systems must follow certain general rules and standards, but, on the other hand, each one of them has some specific characteristics which make them unique. This is the reason why software packages specialized for the support in that area, beside beyond doubt narrowed field of application, must have enough general solutions to acquire and process all of the specifics of a profile system. Special difficulty lies in the fact that it is almost impossible to foresee all the specific solutions in profile systems which can appear in the future and which must certainly be supported by these software packages.

[1]Dejan S. Aleksić is with the Faculty of Sciences and Mathematics, University of Nis, Visegradska 3, 18000 Nis, Serbia, e-mail: dejan_aleksic@yahoo.com.

[2]Dragan S. Janković is with the Faculty of Electronics, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia,
e-mail: dragan.jankovic@elfak.ni.ac.yu.

II. THE PRINCIPLE OF ONE APPLICATION – MULTIPLE BASE

By analyzing qualities of the elements and rules of combining, certain laws might be noted which govern all the profile systems. However, each of the profile systems has certain specifics which make it different compared to a competition's product. Systems basically never differ more than 5% to 20% depending on the material, area of use and the manufacturer.

In order for a CAD/CAM application to be applied to a design process and the manufacture of windows and facades, all the elements of the system must be described, and all the rules for their combining must be carried out. The storage and organization of that data is the first problem that needs to be solved.

Being unable to generalize all the rules within the profile system presents the main problem. That problem has been solved in multiple ways. One way is the existence of multiple application versions – one application for each profile system; or that the general rules are executed within the application itself and all the specifics of certain systems are being accomplished through external additional modules. Both suggested solutions posses obvious flaws – firstly in terms of application maintenance, ease of new profile system data input and the crucial one – defining new rules of profile combining.

In our case. general rules are executed within the application itself as well, but all the certain profile system specifics are recorded in the database. This is mainly accomplished by the strict hierarchy in the element organization, by introducing attributes for the element description and by defining and executing scripts (they are being kept in the database and executed within the application itself). So, during description of a new profile system only data within the database is being altered while the main application remains unchanged for all the systems, which leads to the principle of one application – multiple base.

III. DESCRIBING A SYSTEM USING ATTRIBUTES

All the data for the description of elements, as well as rules, are kept in a standard relational database. We have chosen an open-source database – Firebird to satisfy the needs in terms of availability, price, ease of installing and maintenance. Unfortunately, this kind of choice has certain limits such as: fixed number of fields inside tables, data types defined in advance and there are also limits in speed of reading the data [1], [2]. The solution we have applied in this case is based on creating a separate cluster of objects inside the application itself, which enables us the flexibility in terms of number and types of data. The base of this approach is the use of so called

attributes i.e. the base class tAttribute along with its subsequent class clusters. Each entity is described with a random number of attributes and each attribute has its type, current value from the list of possible values [3], [4]. This approach brought a number of benefits regarding ease of execution of small, but also large enough differences in profile system description, Multilanguage terms support within the database, ease of upgrading the existing database version, automatic data copy from the database into the memory or disk structures and vice versa, quick and easy writing and reading of data about projects on/from the disk, possibility of execution of automatic written project data upgrade in accordance with data changes in value, as well as in number and type, accomplishment of all advantages which object-oriented data access brings (data abstraction, inheritance, overload of attributes) over data located inside the database, memory or on disk.

All the actions concerning writing, editing and reading attributes inside the database are executed inside tAttribues class [5].

IV. HIERARCHY SYSTEM DESCRIPTION

It has already been mentioned that each profile system, besides its elements, is defined by the rules of combining those elements. All general rules of combining are executed in the main application code while the other, specific rules of combining are executed within the database itself. In order to model the profile system rules as efficiently as possible, all the system elements data is organized in a hierarchal tree in multiple fixed levels (see Fig. 1.).

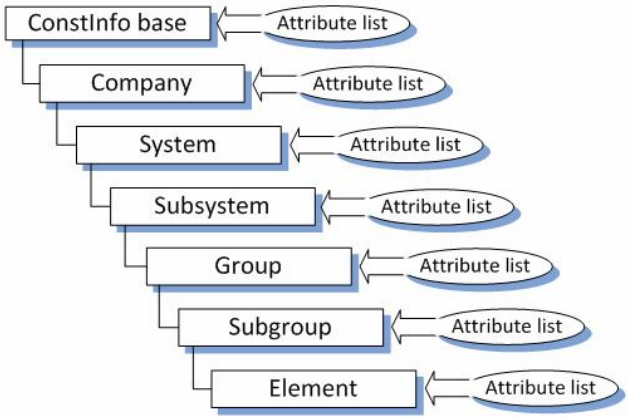


Fig. 1. Schematic display of hierarchal system description

The exception are elements of the lowest hierarchy level which can have their "children" i.e. elements that are added when adding parent-element itself. Process of adding "children"-elements is done by executing scripts which are allocated to each "child"-element and/or using conditional "child" adding system (see Fig. 2.).

V. DEFINING THE RULES OF COMBINING

Main mechanism of accomplishing specific rules of combining is comprised of executing scripts affiliated to

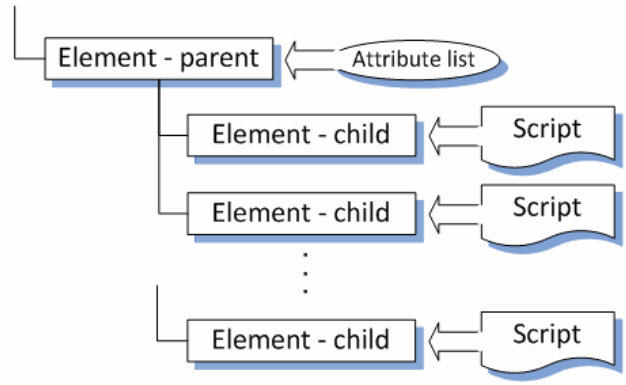


Fig. 2. Parent – child relation

elements on the lowest hierarchy level of profile system within the database. As mentioned before, one parent-element can have unlimited number of child-elements which are conditionally added after the parent-element is added first.

Process of adding one parent-element begins with forming an instance of tItem class inside the memory along with the list of attributes written in the tAttributeList class. After that, the list of children of that element is ran through and for each child-element a process of adding a child is executed. The whole process is done inside the script-executing module which takes an appropriate script from the database and acquires its input parameters (see Fig. 3.).

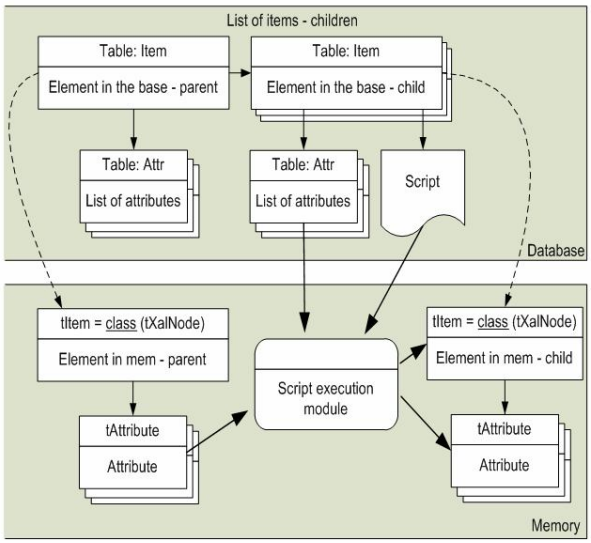


Fig. 3. Generating a new child-element by executing a script

Typically, script input parameters are two lists of attributes. The first one is a list of parents, and the second is a child-element attribute list. Inside the module, script commands are executed in order to get a list of attributes which are allocated to the newly formed instance of the tItem class for the child-element as a result (see Fig. 3.).

The script language (see Fig. 4.) alone has the following features: it follows the Pascal programming language syntax, there is a list of input and output parameters, beside basic types, types defined within the main application can be used (tItem, tAttribute, tAttributeList), possibility of defining local variables, possibility of defining local procedures and functions, access and editing child and parent attributes, access to data structures from the main application during

script execution, syntax script analysis during defining and the possibility of compiling scripts before writing to the database.

```
script  =
 script-heading  block "."
script =
 script ident. "(" ident.-list ")" ","
block =
 declaration-part  statement-part
declaration-part =
 [ type-definition-part ]
 [ variable-declaration-part ]
 proc-and-funct-declaration
proc-and-funct-declaration =
 {(proc-declarat|funct-declarat)";"}
proc-declarat =
 procedure-heading ";" body |
 procedure-heading ";" directive |
 procedure-identification ";" body
funct-declarat =
 function-heading ";" body |
 function-heading ";" directive |
 function-identification ";" body
procedure-heading =
 procedure ident.[formal-parameter-list]
function-heading =
 function ident.[formal-parameter-list]
":" result-type
statement-part =
 begin statement-sequence end
statement-sequence =
 statement { ";" statement }
assignment-statement =
 (variable|function-ident)":="expression
procedure-statement =
 procedure-ident.[actual-parameter-list]
while-statement =
 while expression do statement
repeat-statement =
 repeat statement-seq. until expression
for-statement =
 for variable-ident. ":=" initial-
expression (to | downto) final-
expression do statement
if-statement =
 if expression then statement [ else
statement ]
```

Fig. 4. Part of EBNF definition the script language

VI. CONDITIONAL ELEMENT ADDING

Conditional element adding represents yet another execution mechanism of specific combining rules. Its function represents the expansion of previously described child-element adding mechanism by executing scripts. Namely, one parent-element can contain the list of pairs criteria=values. Each criterion must be registered and corresponds to a certain attribute from the parent-element attribute list. List of criteria contains the list of child-elements with their scripts. When adding a parent-element a special module from the main application is activated which, based on parent-elements attribute values and the values of criteria from the lists, decides if and which list of children will be passed over to the script executing module (see Fig. 5.).

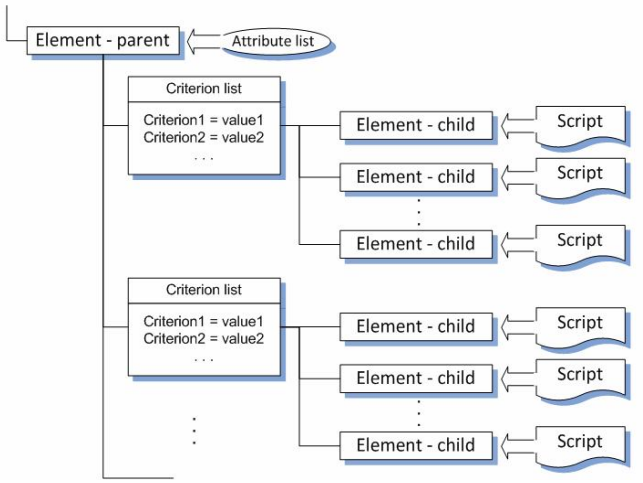


Fig. 5. Parent – child relation

Using this mechanism we emulate the complex if-then command within the very database. Its use could have been accomplished using special conditional commands within the scripts of each child-element as well, but the one we chose is much more efficient and clear especially concerning adding of child-elements.

VII. ONE EXAPMLE

Mechanisms described in previous section are illustrated in this section. We will used PVC profile system in this example. PVC profile is not strong enough and needs to insert Fe profiles inside PVC profile.

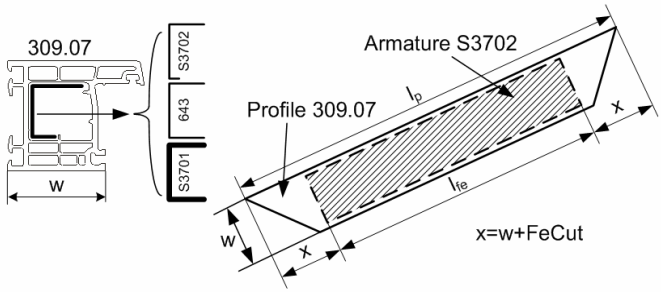


Fig. 6. One PVC profile and Fe armature profile

In one PVC profile can be inserted different Fe profiles depending of length of PVC profile, intensity of wind in this area, atc. Typically, angle of cutting Fe profile is 90° - 90° and his length must be less then length of parent PVC profiles (see Fig. 6).

Assumed that one instance of tItem class and his appropriate attribute list was made for PVC profile with catalogue name 309.07. Values of some typical attribute for this profile was show on Fig. 8 (left side).

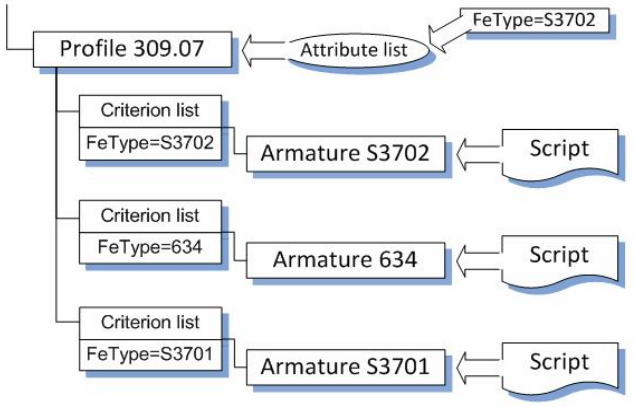


Fig. 7. PCV profile 309.07 (parent) with his "children"

In next step, conditional element adding mechanism is activated and some element - children of PVC profile will be adding.

```
CatName  = "309.07"
Width    = 88.0
Len      = 1206.00
U1       = 135.00
U2       = 45.00
Qty      = 4
FeType   = "S3702"
FeCut    = 5.0
minFeLen = 600
     . . .
```

```
CatName = " S3702"
Len     = 0.00
U1      = 0.00
U2      = 0.00
Qty     = 0
     . . .
```

Fig. 8. Attribute value of PVC profile (left) and Fe profile (right)

In this case, three diferrent Fe profile (S3702, 634 and S33701) can be added depending of value of FeType attribute in attribute list of PVC profile 309.07 (see Fig. 7.). S3702 Fe profile was selected and his attribute list will be loading from hierarchal system description data structure with default attribute value (see Fig. 1.).

```
script AddFe (p:tItem; var c : tItem)
begin
  c.Len := p.Len-(p.w+p.FeCut)*2;
  c.U1 := 90.00;
  c.U2 := 90.00;

  if p.Len > p.minFeLen
  then c.Qty := p.Qty
  else c.Qty := 0
end.
```

Fig. 9. Script for Fe profile adding

After that, appropriate script will be assign to S3702 Fe profile (see Fig. 7.). Attribute list of PVC profile 309.07 (as parent), attribute list of Fe profile S3702 (as child) and appropriate script will be sent to script-execution module (see Fig. 3.). Value of some children attributes will be changed during script executions and lenght and cutting angle of Fe profile will be define. In most of PVC profile systems govern follow rule: if lenght of PVC profile is less then some critical value appropriate Fe profile not need to add. This rule is modeling with if-then-else command on the end of script (see Fig. 9.). On the end of this process, instance of tItem class

```
CatName = "S3702"
Len = 1020.00
U1  = 90.00
U2  = 90.00
Qty = 4
     . . .
```

Fig. 10. Final value of typical attribute of new Fe profile S3702

who represents S33702 Fe profile will be create if value of attribute c.qty is great then 0. Final value of typical attribute of new Fe profile S3702 is shown on Fig. 10.

VIII. CONCLUSION

Definition and the execution of combining profiles in profile systems represent one of the key problems in a specific CAD/CAM application for design and manufacture of facade carpentry support. General rules which are common for all the systems are executed within the application, and all the specifics that certain systems possess are executed within the database. When describing new profile system only information in the database is altered while the main application remains the same for all systems, which brings us to one application – multiple base principle. Information on system elements has hierarchal organization in multiple levels, where lowest level elements can have “children”. Process of child-element adding is accomplished by executing scripts allocated to each child-element and/or with the conditional child adding system. The script language is specially created for this use, follows the Pascal programming language syntax, acquires the list of input and output parameters; beside basic types, types defined within the main application can be used. There is a possibility of defining local variables, procedures and functions, access to and editing of parent/child attributes as well as access to data structures from the main application during the script execution. Before writing the script into the database, a syntax script analysis is performed and there is a possibility of compiling scripts which significantly increases their execution.

All these concepts enable us to relatively easy and within a short period of time model the rules of combining of some new system with all the specifics it carries with it.

REFERENCES

[1] Q. Zhang, "Object-oriented database systems in manufacturing: selection and applications", Industrial Management & Data Systems, vol. 101, no. 3, pp. 97-105, 2001.

[2] D. Maier, "Making database systems fast enough for CAD applications", Object-oriented concepts, databases, and applications, ACM Press, New York, NY, 1989.

[3] M. Blaha, W. Premerlani, H. Shen, "Converting OO Models Into RDBMS Schema", IEEE Software, vol. 11, no. 3, pp. 28-39, May/June 1994.

[4] Y. Yang, "An enhanced data model for CAD/CAM database systems", Proceedings of the 25th ACM/IEEE conference on Design automation, pp. 263 - 268, 1988.
P. Buneman, M. Atkinson, "Inheritance and persistence in database programming languages", Proceedings of the 1986 ACM SIGMOD international conference on Management of data, Pages: 4 – 15, 1986.

Sensor Web Architecture for Crisis Management

Nataša Marković[1], Leonid Stoimenov[2]

Abstract – **This paper presents architecture of a system for crisis management, based on the sensor web technology and Geo-Information Systems (GIS). Rapid development of new technologies, especially Sensor Web, brings many ways for environmental protection and maintenance. Combining GIS, as a platform for data visualization, with Sensor Web, enables us to create crisis management systems for environmental monitoring and protection.**

Keywords – **Sensor Web, Crisis Management, GIS**

I. INTRODUCTION

Geo-Information Systems (GIS) are being widely used for more than forty years. They have found their purpose in environmental monitoring, transportation management, public safety, facility security, disaster management, etc. GIS enables us capturing, storing, analyzing, and displaying geographically referenced information. It allows us to view, understand, query, interpret, and visualize data in a way that is quickly understood and easily shared. GIS technology can be used for scientific research, resource management, and development planning.

Sensor Web, on the other hand is a relatively new technology with wide usability. It enables tracking sensors, obtaining their data and making it available through the Web. According to [1], Sensor Web is a special type of Web-centric information infrastructure for collecting, modeling, storing, retrieving, sharing, manipulating, analyzing, and visualizing information about sensors and sensor observations of phenomena.

The point where these two separate technologies meet is a system where GIS provides maps and objects, and Sensor Web provides interface to sensor data. A general overview of one such system, where geo-location of sensors is interpreted on map, where sensors can be queried and their data represented, is further described in this paper.

The rest of the paper is organized in the following way: Section 2 describes existing technology standards, specifications and practical examples of GIS and Sensor Web systems. Section 3 describes the architecture of GIS and Sensor Web based Crisis Management system (GINISSENSECM). Activity flow and the activity diagram of the GINISSENSECM system is described with details in Section 4. At the end, conclusion and the review of literature is given.

[1]Nataša Marković is with the Faculty of Electronic Engineering at University of Nis, 14 A. Medvedeva, Nis 18000, Serbia, E-mail: natasa.markovic@elfak.ni.ac.yu

[2]Leonid Stoimenov is with the Faculty of Electronic Engineering at University of Nis, 14 A. Medvedeva, Nis 18000, Serbia, E-mail: leni@elfak.ni.ac.yu

II. RELATED WORK

A. Standards and Specifications

In order to use, share and present data received from sensors we need to use a standardized set of protocols and semantics.

The Open Geospatial Consortium (OGC), as a leading organization in the development of standards for geospatial and location based services, has introduced a set of standards called Sensor Web Enablement (SWE). SWE consists of three Markup Language specifications including SensorML [2], Observation and Measurement (O&M) [3], Transducer Markup Language (TML) [4], and four Web Services specifications based on the assumption that all sensors are connected to the web, including Sensor Planning Service [5], Sensor Observation Service [6], Sensor Alert Service [7] and Web Notification Service [8]. This set of services and specifications contributes to exploiting Web-connected sensors and sensor systems of all types: flood gauges, air pollution monitors, stress gauges on bridges, mobile heart monitors, Webcams, satellite-borne earth imaging devices and countless other sensors and sensor systems. OGC specifications also include protocols for data access, such as Web Map Services (WMS), Web Feature Services (WFS), and Web Coverage Services.

B. Practical Examples

Technologies, specifications and protocols mentioned above allow us to create various crisis management applications (disaster management applications, environmental protection and management applications, etc). We will describe in this section two environmental monitoring applications.

An application for detection and monitoring the spread of wild fires [9] uses Sensor Web technology for wild fire detection in Africa. Their aim is to use the Sensor Web to observe specific fire-related phenomena described in the wild fire ontology and employ machine reasoning to determine fire risk. In that way it is possible to predict the occurrence and spreading of wild fire, in order to inform the responsible services so that they can act accordingly and prevent possible disaster.

The South East Alaska MOnitoring Network for Science, Telecommunications, Education and Research (SEAMONSTER) [10] is a smart sensor web project designed to support collaborative environmental science with near-real-time recovery of environmental data. Their work is focused on the Lemon Creek watershed and understanding both physical and biological as a collection of interconnected systems. This project is an effort to implement a sensor web available for

science, education, and sensor web technology evaluation and advancement.

These and many other applications relay on different, specialized architectures. What we have created is a general architecture for crisis management, which could be easily adopted for concrete purposes.

III. GINISSENSECM ARCHITECTURE

The general architecture of the GIS and Sensor Web Crisis Management system (GINISSENSECM) is given in Fig. 1. It consists of the following components: System Operator located at the Crisis Management Center (CMC), Graphical user interface (GUI), Databases, the Decision Making Agent (DMA) and Data Access Layer.

Graphical User Interface is a WebGIS [11]. The user interacts with the GUI, which visualizes data received from different data sources (sensors or community services) as well as objects of interest.

Crisis Management Center (CMC) is a center for resolving numerous problems related to environmental protection. CMC activities are gathered around making decisions on the basis of gathered information about the environment and acting in order to prevent disasters or threats related to the observing environment.

Decision Making Agent (DMA) is a component used for comparing and analyzing data obtained from different sources, making action plans, executing automated spatial queries, running on demand or automated actions and proposing action plans to CMC operator. The DMA component is responsible for executing automated spatial queries, data acquisition from sensors and Community services, and decisioning and acting based on the set of operator demands, or programming logics. DMA component is still in the development phase. So far it is planed as a component consisting of the following subcomponents: knowledge database (support for rules), decision component, interface to user/other component, interface to data sources.

DMA Agent can be configured to act according to operator commands. For now, DMA can act upon two types of events: **on user input** and **on measured values**. From the configuration interface dialog, operator chooses type of event, after which new dialog appears depending on chosen event. In the user input event dialog, operator chooses few keywords from a defined set of metadata, for example ***dump+river***. After choosing the set of keywords, user then configures DMA actions upon the following event: when user input has as metadata keywords **dump+river**, do the following. An action configuration dialog has for now following possible actions: Inform operator on event by **sms/mail**, Alert registered users upon event and Repeat Measurements (for actions based on measured values). So after final step in the DMA configuration the sentence describing acting upon event would be: *When user input has as metadata keywords dump+river inform operator by sms.*

We gave here the example of DMA configuration, upon acting on user input. Different dialog would be for acting upon measured values. In the measured values event dialog, operator configures which values to be monitored and in what range. After choosing range and values of measured parameters, DMA actions dialog appears, same as in the previous paragraph. After final step in the DMA configuration the sentence describing acting upon measured values would be: *When measured values of ph+T is exactly 6,3+34,8 repeat measurements AND alert registered users.*

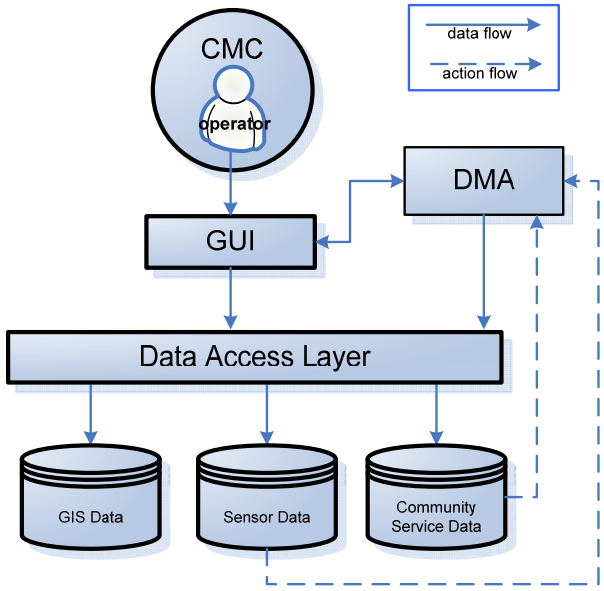


Fig. 1. GINISSENSECM Architecture

Data Access Layer (DAL) represent implementation of different services which enable access to different data sources and are responsible for information flow in the system. DAL is the intermediary between an operator and a data collection management environment.

Databases store data received from sensors, as well as data received from community services and GIS data.

IV. GINISSENSECM ACTIVITY FLOW

The activity flow of the GINISSENSECM system is shown in Fig. 2. The central component of the activity flow is the Decision Making Agent (DMA). The DMA component, implemented using Web services technology, has two major roles. First role is to process automatic measurements based on operator demands. The second role is to decide if there has been major changes in measuring parameters, and offer some actions to the operator or act accordingly (if operator allowed this during DMA configuration). Instead of using only values received from different types of sensors in the process of deciding whether there is a possible threat in the observing environment, CMC operators are provided also with data received from community services. Gathered data is compared and analyzed by DMA and on that basis conclusions are presented to the operator with greater reliability. Activity flow also includes three Web services: Sensor Observation Service, Web Notification Service and Sensor Alert Service, which are responsible for collecting data from sensors, managing sensor actions in real time and delivering notifications to end users.

These services are implemented on the basis of OGC SWE specifications.

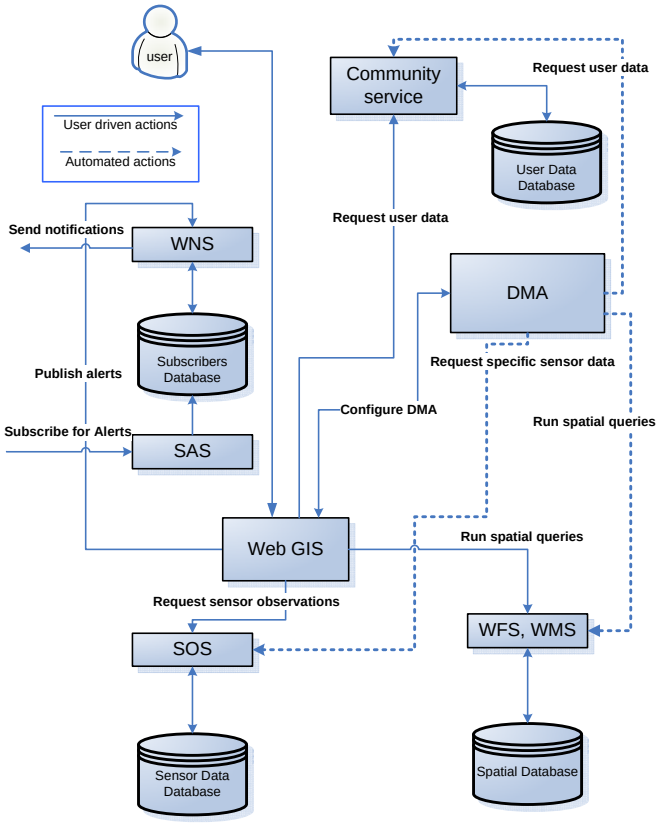


Fig. 2. GINISSENSECM Activity flow

The Community service is a Web service for collecting data from users concerning the environment. It allows users to send information in various ways about their observations of the environment. For example, a certain user could send a picture of dump, fire, etc, which would be useful when decision is made about the various sources and causes of pollution, fire or any other environmental threat or disaster. Community services should register within system in order to be able to interact with the system and send collected user data in various formats (e.g. pictures with descriptions, textual messages …).

GIS data represents a collection of geographical data, maps, vector and raster objects. This component represents different data depending on the system purpose. For example if the final implementation of the system is used for water management, GIS needs to incorporate data about rivers. GIS data is stored in spatial database over which can be executed operator driven or DMA automated spatial queries. There is a predefined set of spatial queries among which the operator can choose one to execute, which is the most appropriate for the current situation.

Operator can choose a type of query executed by DMA agent that will be triggered by some event on the sensor. An example query would be to request objects to be drawn on the map if they are classified objects of threat for the environment) (as illustrated in Fig. 3.a).

a) **SELECT** {objects} FROM {SpatialData D} **WHERE** {object.property **IN** {polluters}}

b) **SELECT** {parameter} **FROM** {SensorData D1, SensorData D2} **WHERE** {D1.loc IN (R1) **AND** D2.loc **IN** (R2)}

c) **SELECT** {metadata} **FROM** {UserDatabase U} **WHERE** {U.loc IN (R1)}

Fig. 3. Example data queries

Sensor Database stores the data obtained through Sensor Observation Service (SOS). SOS is a standard Web service interface for requesting, filtering, and retrieving observations and sensor system information. Operator driven or DMA automated queries are executed over sensor database. For example, user can request the measuring parameters received from the sensors located in two different regions (as illustrated in Fig. 3.b).

User Data Database stores user collected data concerning environmental protection. For example someone can send a picture of environment (waste water, wild dump, object in fire, etc), with a small description of what is in the picture and its geo location. User data can be filtered and queried in different ways. One possible query could be for finding the metadata (user description) for some location (as illustrated in Fig. 3.c).

The activity diagram (shown in Fig. 4) shows basic activities performed in the system by user or configured DMA component. There are two types of actions that can be performed in the system, user driven actions and automated actions.

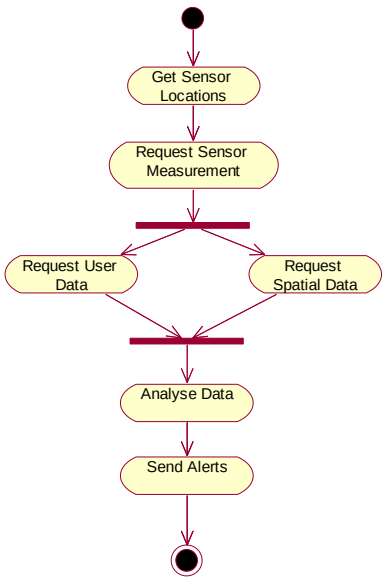


Fig. 4. GINISSENSECM activity diagram

A. User driven actions

User driven actions are associated either with sensors, GIS data and community services or with the DMA component.

The user (operator – further in the text) located at the CMC *Request Sensor Measurements*, in order to check measuring parameters. At the same time, he can also perform two more activities *Request User Data* and *Request Spatial Data*, in order to have a more detailed view on the observing environment. On the basis of received data, operator can analyze data and decide about the actions to be performed. If a possible threat is detected, operator can *Publish Alerts* in order to inform objects of interest. If the operator gets some conclusions about who is responsible for the threat, he can inform them as well. Operator can ask for sensor observations through Sensor Observation Service. He can send request to WMS and WFS services if he needs to see where the sensors are geographically situated. Operator can check the community service data, and finally, when it is needed he can publish alerts and notifications using Web Notification Service.

B. Automated actions

The second types of actions that are possible in the CM system are automated actions. Automated actions are performed by DMA on the basis of operator configuration. Operator configures DMA in order to define a set of rules and actions to be performed automatically in some period of time. For example operator can fill a simple form to initiate specific measurements to SOS, which will be saved into database. DMA can run on-demand sensor observations continuously in certain period in time, and cross-reference the given values with data obtained from community service database. On the basis of predefined criteria, DMA can decide whether the vital parameters are beneath or above certain level, and propose actions to the operator.

V. CONCLUSION

Sensor Web has provided infrastructure for collecting and processing data from distributed and heterogeneous sensors. This set of technologies has found various implementations, especially in the area of environmental monitoring. The GIS and Sensor Web architecture for Crisis Management (GINISSENSECM), described in this paper, provides active monitoring of measured parameters and timely responses in cases of environmental disasters. The Crisis Management System based on Sensor Web and GIS enables access, control and management of desired environment. The GINISSENSECM system is general and it can be personalized and used for concrete purpose. For example if the sensors are situated in river water, the GINISSENSECM system can be used for tracking and managing water pollution, publishing alerts and prevention of the possible disasters (in the cases of releasing toxic substances in water) [12]. Another example of implemened GINISSENSECM architecture is the GinisED DistSense system [13] for tracking parameters of power supply network in real time. The GINISSENSECM system can also be used for fire prevention in buildings as well as for open spaces. The GINISSENSECM system is easily adjustable for management and prevention of any other environmental disaster or threat. There are many ways of upgrading this system and making it more efficient and completely automated. The future work includes modifications and improvements of the current system architecture and its implementation. The DMA component can be further improved by incorporating a knowledge-based engine, in order to run independently and automatically report alerts to the operator in the crisis management center.

ACKNOWLEDGEMENT

Research presented in this paper were partially funded by the Ministry of Science of the Republic of Serbia and PD Jugoistok Niš, within the project in the field of technological development "Intelligent integration of geo-, business and technical information on the company level," No. 13003.

REFERENCES

[1] N. Gross, *Business Week,* Aug. 1999, www.businessweek.com/1999/99_35/b3644024.htm
[2] M. Botts, (Ed.), OpenGIS® Sensor Model Language Implementation Specification (Version 1.0). OGC Document Number: 05-086r2, 117pp., 2006.
[3] S. Cox, (Ed.), OGC™ Observation and Measurement (Version 0.13.0). OGC Document Number: 05-087r3, 136pp., 2006
[4] S. Havens, (Ed.), OpenGIS® Transducer Markup Language Implementation Specification (Version 1.0. 0). OGC Document Number: 06-010r2, 136pp., 2006
[5] I. Simonis, (Ed.), OpenGIS® Sensor Planning Service Implementation Specification (Version 0.0.30). OGC Document Number: 05-089r1, 152pp., 2005
[6] A. Na, M. Priest, (Eds.), OpenGIS® Sensor Observation Service Implementation Specification (Version 0.1.5). OGC Document Number: 06-009r1, 187pp., 2006
[7] I.Simonis, J. Echterhoff, (Eds.), OGC® Sensor Alert Service Implementation Specification (version 0.9.0). OGC Document Number: 06-028r5,144pp., 2006
[8] I. Simonis, A. Wytzisk, (Eds.), Web Notification Service (Version 0.1.0). OGC Document Number: 03-008r2, 46pp., 2003
[9] A. Terhorst, D. Moodley, I. Simonis, G. McFerren, F. Van den Bergh, "Using the Sensor Web to Detect and Monitor the Spread of Vegetation Fires", In: Advances in Geosensor Networks, Springer Lecture Notes in Computer Science Series, 2008.
[10] MJ Heavner, DR. Fatland, E. Hood, C. Connor, "SEAMONSTER: A Sensor Web Technology Implementation and Testbed in Southeast Alaska," presented at NASA Science and Technology Conference, College Park, MD, June 19, 2007.
[11] M. Bogdanovic, N. Davidovic, I. Antolovic, A. Stanimirovic, D. Stojanovic and L. Stoimenov, "WebGIS application for viewing and analysis of electric power supply network geodata", YUINFO 2008, Kopaonik, (in Serbian)
[12] N. Markovic, A. Stanimirovic and l. Stoimenov, "Sensor Web for River Water Pollution Monitoring and Alert System", AGILE 2009, (accepted for publishing)
[13] M. Bogdanovic, N. Markovic, A.Stanimirovic and L. Stoimenov, "GinisEd DistSense – Sensor Web inside GinisED system", Etran 2008, Palic, (in Serbian)

Architectural Enhancement of HASIS 3D Hail Suppression Information System

Vladan Mihajlović[1], Bratislav Predić[1], Dejan Rančić[1], Slobodanka Djordjević–Kajan[1], Igor Antolović[1], Zoran Babić[2]

Abstract – **Previous system HASIS 3D is a high quality tool for tracking the clouds and measurement of some hail cell parameters, providing good visualization and automatic calculation of seeding parameters. Improvement of the system is proposed to provide visualization of different radar data types, at the same time, integration of scalar data gathered from different radars and reconstruction of wind vector gathered from three sources. That requests redesign of HASIS 3D architecture. The new architecture is proposed to be modular, flexible and extensible.**

Keywords – **software architecture, design patterns, data visualization, hail suppression**

I. INTRODUCTION

Modern meteorology is based on information systems, from simple ones that manage meteorological measurements to complex systems that process huge amount of data such as weather conditions monitoring systems [1]. Weather modification is field of meteorology that made great progress based on usage of information systems. Satellites and radars are two main sources for acquisition of data about atmosphere. However, information systems involved in hail suppression mainly use radars for gathering data [2, 3]. Hail suppression systems can be classified in three groups whether they use stationary radars, portable radars or both kinds [4]. Second classification is based on methodology used for seeding the cloud with hail potential. First group of systems use rocket and second use planes.

HASIS 3D (Three-dimensional Hail Suppression Information System) is specific information system developed to support hail suppression activities in Hydrometeorological Service of Republic of Serbia (RHMSS) [5, 6, 7]. The system uses stationary radars for data acquisition and rockets for seeding the clouds with hail probability with chemical reagents. HASIS 3D does not control the radar itself, but use data from other systems, particularly HASIS for Mitsubishi radars and Rainbow [7] for Gematronik radars. However, HASIS 3D is complete system that provide monitoring and tracking of cloud systems, measurement of hail cell parameters, extraction of cloud region that needs to be seeded, calculation of ballistic parameters of rocket and storing data about the whole process.

HASIS 3D system uses data from single radar for weather conditions monitoring. Problem is that topmost elevation of radar limits the part of space from which data can be acquired. Using data from other radar, this lack of data can be exceeded. System can provide more precise visualization of clouds using data from several radars. Also, integration of data from different radars can provide information about situation above the whole country. HASIS 3D system can visualize different types of data about atmosphere, reflectivity, wind speed and spectral width. Better insight into probability of hail can be obtained monitoring all this data types at same time. The nature of scanning of atmosphere using radars is limited to gather data about wind speed only in direction from radar through the point of interest. This is just a projection of real wind vector. To gather the real information about wind direction and speed it is needed to use the data from at least three radars. In this way it is possible to reconstruct the real vector and to create 3D grid with wind data. This grid can be used for visualization to provide new information about processes inside the cloud. Also, this data can be used to determine the real speed of the cloud with hail potential, which can be gathered easy in other way. This requests imposed the enhancement of the system HASIS 3D.

Extension of HASIS 3D architecture is described in following three sections. The second section describes methodology for cloud seeding that is supported by HASIS 3D system. The enhancement of system architecture is presented in third section. The final section summarizes the features of the proposed architecture and specifies the topics for further research.

II. HASIS 3D SYSTEM ORGANIZATION

The main tasks that hail suppression system based on cloud seeding using rockets needs to support are: detecting the processes in atmosphere that can generate a cloud with hail potential, measuring the parameters of the hail cells, automatic verification of criteria for seeding the cloud and automatic creation of cloud regions to be seeded, automatic calculation of ballistic parameters for launching the rockets and storing the data about the seeding process and creating different type of reports.

The configuration of HASIS 3D system [8, 9] is defined to support hail suppression methodology defined by meteorologist from RHMSS and approved in previously conducted researches [5, 6, 7]. According to this, the system consists of three subsystems (fig. 1): Main workstation, GIS workstation and Database workstation.

[1]Vladan Mihajlović, Bratislav Predić, Dejan Rančić, Slobodanka Djordjević-Kajan and Igor Antolović are with the Faculty of Electrical Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: wlada@elfak.ni.ac.rs.

[2]Zoran Babić is with the Republic hydro meteorological service of Serbia, Kneza Višeslava 66, 11000 Beograd, Serbia, E-mail: zbabic2003@yahoo.com.

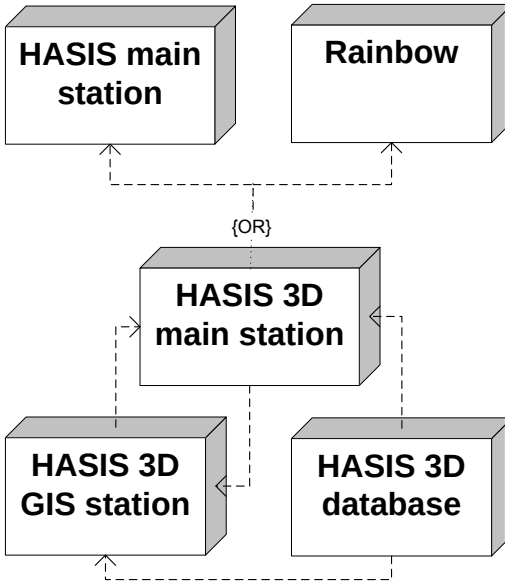


Fig. 1. HASIS 3D system

HASIS 3D Main station's main functionality is to represent situation in atmosphere. The radar data for visualization are transferred as being created on system that controls radar. The system represents the situation through two different views, 2D and 3D. The 2D representation is achieved using horizontal and vertical cross sections (fig. 2). In this view, it is possible to explore cloud configuration, size and structure using contour extraction. Different measurements of cloud parameters can be performed during 2D analysis. In 3D view, situation in atmosphere is represented using multiple isosurfaces. The value for calculation the surface can be dynamically changed providing better insight of cloud shape and dimensions. This way operator has complete support to analyze cloud and estimate cloud hail potential. System, also, helps providing automatic verification of seeding criteria.

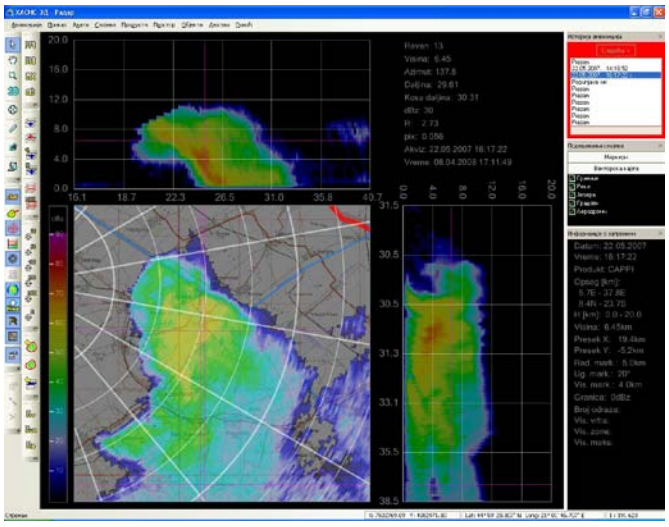


Fig. 2. 2D view on main HASIS 3D subsystem

The GIS workstation displays situation in the atmosphere in 2D and 3D view, but 2D view is limited to horizontal cross sections. The arrangement of launching station on the ground is visualized with their prohibitited zones. This subsystem provides possibility to define square regions where is prohibited to perform actions. According to type of the cloud with hail potential and zones of prohibition, the system automatically determines region in the cloud which need to be seeded. Using location of this region and positions of the launching stations system automatically calculates which stations will perform action and ballistic parameters of rockets for them (fig. 3). Seeding efficiency is, also the input parameter in the process of calculation.

The Database station provides support for storing data about weather conditions, process of cloud monitoring and seeding. These include data about ballistic parameters of rockets, location and status of the launching station and the types and amount of rockets in store. Data about cloud measurement are stored for further analysis. Ballistic parameters calculated by GIS subsystem are available to operator on this workstation, so he can issue command to perform action using radio link.

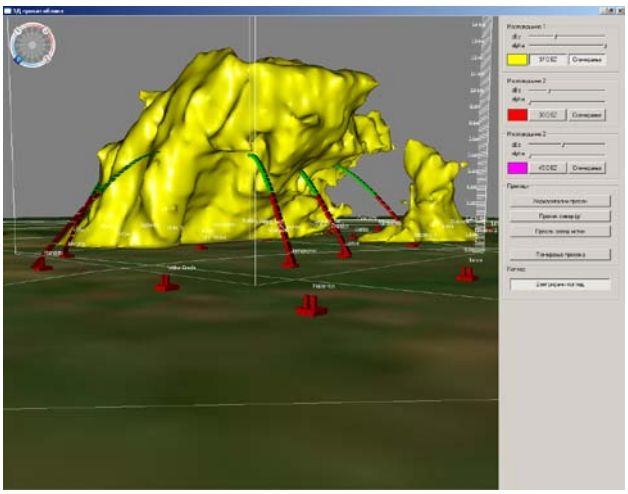


Fig. 3. 3D view on GIS HASIS 3D subsystem

III. ENHANCEMENT OF HASIS 3D ARCHITECTURE

New ideas mentioned earlier influenced enhancement of existing architecture of HASIS 3D and some reconstruction. The main reasons that demanded modifications are visualization of different radar data at the same time, using data from several sources and visualization of vector data types. The most important classes of the new system are *DataLayer*, *DataIntegrator*, *Renderer*, *Observer* and *Manager* (fig. 4).

DataLayer is basic class that manages products that radar collects. There are three types of products: reflectivity, wind speed and spectral width. The basic functionalitie of this class is to prepare data for visualization. Also it can perform different types of data analysis. The radar data are stored in cylindrical coordinate system.

The major task of *DataIntegrator* is to process the data gathered from two or more radars. This class uses the data stored in several *DataLayer* classes. The processing of data is quite different for raster and vector data which would be implemented as *ScalarIntegrator* and *VectorIntegrator*. The scalar data are complete and their integration is performed by

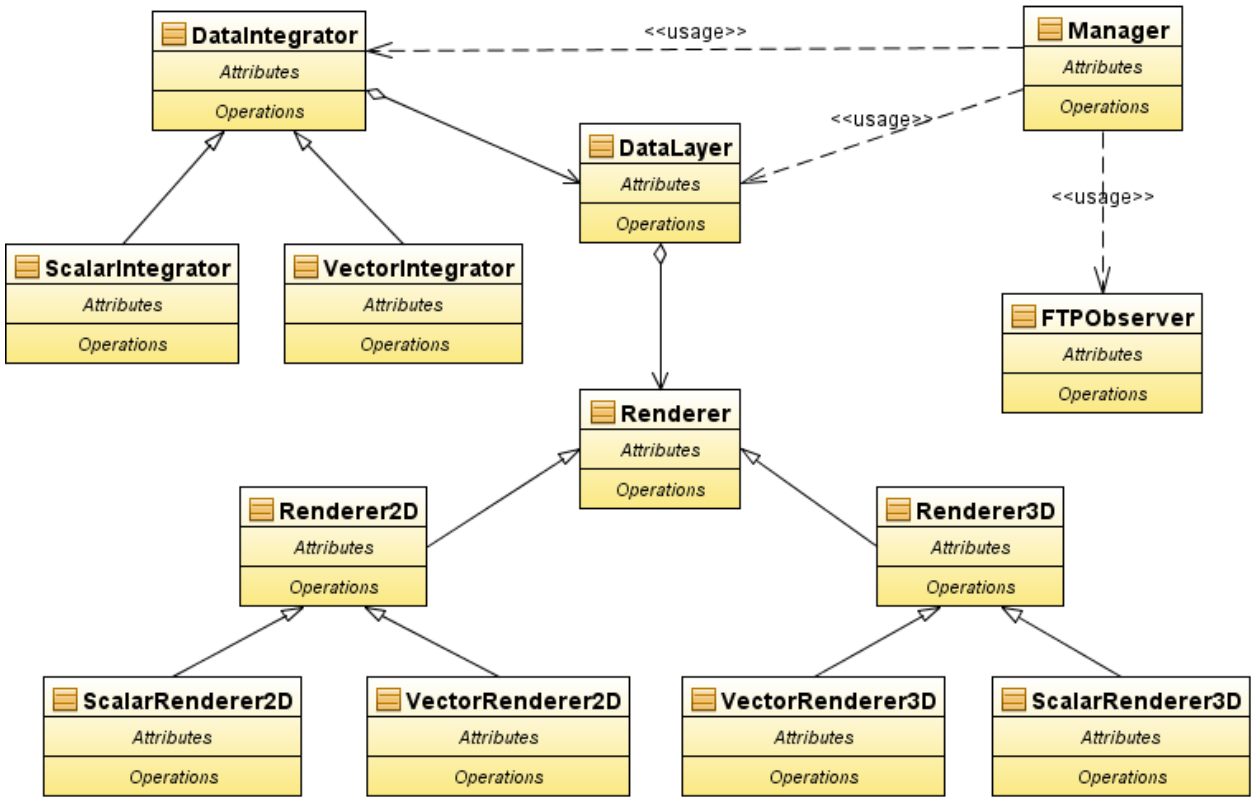


Fig. 4. Class diagram of HASIS 3D enhancement

choosing the most recent or most precise data at particular point in data grid or using some method for averaging. On the other side, vector data gathered from single radar are incomplete because they represent only the projection of real vector on the imaginary line connecting the radar with the particular point in space. Therefore, for calculation of real vector it is required to have data from three radars. However, because radars are positioned on larger distances the good approximation on vector in horizontal plane can be achieved using two radars. The vector reconstruction has inherent error and it is expected that it would need different algorithms which will differently threat this problem. This possibility of using different methods for scalar and vector data integration is imposed Decorator design pattern [10] as solution for *DataIntegrator* to provide simple modification and flexibility.

Quality visualization is primary requirement for cloud tracking and good estimation of situation in atmosphere. *Renderer* is the basic class that performs this task. Two inherited classes visualize radar data in 2D and 3D view. First class has methods for representing the data on horizontal and vertical cross section. The class *ScalarRenderer2D* provides the basic method visualization using color spectra to represent scalar values. Beside this method scalar data can be visualized using contours for specific values. This offers a possibility to combine different radar products in the same picture (cross section). *VectorRenderer2D* and *VectorRenderer3D* have purpose to represent vector data in 2D and 3D view. The simple method to do this is to use regular data grid with vectors in its points. These classes provide vector visualization using streamlines. *ScalarRenderer3D* represents cloud throw multiple isosurfaces for particular values. The values that can be dynamically adjusted to give better insight into cloud parameters. To support simple replacement of method used for visualization, *Renderer* classes are designed using Strategy design pattern with *DataLayer* as a context class.

Significant modification is performed on *Observer* component. The requirement to monitor and gather data from several sources require creation of a new class *ObserverManager*. This class creates and controls several instances of *FTPObserver* components. Each component is configured independently to connect to the workstation which controls the radar. To isolate influence of data acquisition these components work in different threads. *ObserverManager* can add or remove data sources while the system is running.

Besides handling data acquisition the class *Manager* controls data processing and visualization (fig. 5). The class *LayerManager* defines which data will be displayed and which method will be used. This component also specifies how to combine different data layers in the view. Handling data integration is, also, task that this component performs. *AcqusitionManager* is the part of the system that cooperates with *ObserverManager* to properly manage the incoming data. This component allocates different *DataLayer* to accept data and update the history of arrived radar products. The layers that use data from different sources must register with *AcqusitionManager*, so it can inform them whenever new data form source they use are transferred. This way layer that performs data integration can refresh its contents and preserve representation correctness. This functionality is designed using Observer design pattern.

IV. CONCLUSION

Requirement to get better insight into situation of atmosphere invoke improvement of HASIS 3D system. The purpose is to provide simultaneous monitoring several parameters of clouds and better estimation of its hail potential. This will be provided using:

1. Simultaneous visualization of different types of radar data: reflectivity, wind speed and spectral width either in 2D or 3D view.
2. Processing data from several data sources to the gaps in part of the space that cannot be reached by single radar and to improve the accuracy in the part that are covered with several radars.
3. Reconstruction of real vector of wind speed, which will bring new information about processes in the cloud and more accurate calculation of speed cloud motion.

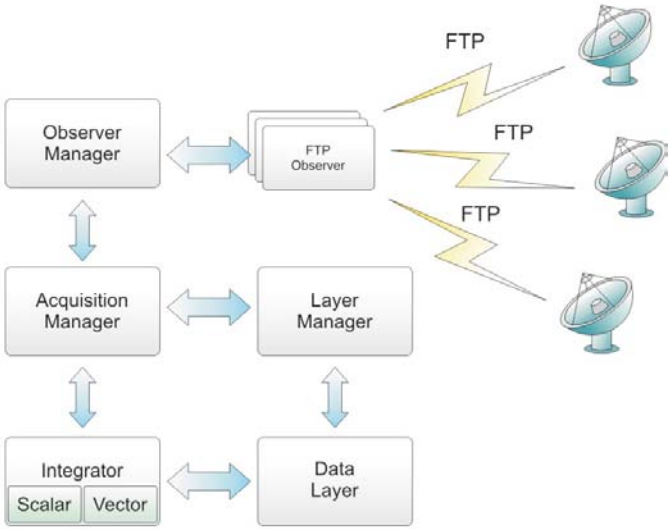


Fig. 5. Radar data management

Enhancement of HASIS 3D system is designed to maintain existing framework. The extension is developed to clearly separate component that control, process and visualize the data. The new part of architecture provides changeability and extensibility of the system. This will enable addition of different methods for data processing, integration and visualizations. To achieve this, different design patterns are used in development of extension HASIS 3D architecture.

The extension of HASIS 3D will provide the base for data integration and visualization. It is expected that improvement of basic methods will invoke development of new methods that better suites in particular cases. Beside, this system use data converted in Descartes coordinate system which contains inherent error. The improvement of precision of data analysis and visualization processes could be provided using data in spherical coordinate system, which is native to radar data acquisition.

ACKNOWLEDGEMENT

Research presented in this paper is partially funded by Serbian ministry of science and technological development and Republic hydrometeorological service of Serbia as part of the project “Integrated hail suppression information system – HASIS 3DI”, ref. no. TR-12003.

REFERENCES

[1] L. A. Treinish, “Visual data fusion for applications of high-resolution numerical weather prediction”, IEEE Proceedings of the conference on Visualization '00, 2000, pp. 477 - 480.

[2] J. Johnson, et al., “The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm”, Weather and Forecasting 13, 1998, pp. 263-276.

[3] T. Johnson, C. Barrere, M. D. Eilts, N. Kuhnert, M. Mathis, D. Axisa, “Weather modification operations with NEXRAD level-II data and products”, Proceedings of 16th Conf. Inadvertent and Planned Weather Modification, San Diego, USA, 2005

[4] TITAN (Thunderstorm Identification. Tracking, Analysis, and Nowcasting) radar processing software, http://www.rap.ucar.edu/staff/vidal/titan/

[5] S. Djordjević-Kajan, B. Milovanović, D. Rančić, A. Kostić, Z. Stanić, L. Stoimenov, "Hail Suppression Information System of a Radar Centre", Proceeding of the Conference GIS/LIS'96 Central Europe, Market Driven Developments, Budapest, Hungary, June 12-14, 1996, pp.102-111

[6] D. Rančić, M. Smiljanić, S. Đorđević-Kajan, A. Kostić, P. Eferica, P. Vuković, Z. Vučinić, “Radar data processing for cloud seeding in hail suppression information system”, RADME 98 - Theoretical, Experimental and Operational Aspects of Radarmeteorology, Rome, Italy, 1998, pp. 137-149.

[7] Z. Babić, D. Rančić, V. Mihajlović, I. Antolović, B. Predić, P. Eferica, “Three-dimensional Clouds Modeling for Hail Suppression Information System”, Ninth WMO scientific conference on weather modification, Antalya, Turkey, October 2007.

[8] Rainbow 5.0 application software, http://www.gematronik.com/website/selex/pdf/datasheets/rainbow.pdf

[9] D. Rančić, V. Mihajlović, B. Predić, I. Antolović, P. Eferica, A. Milosavljević, “Clouds Visualization for Hail Suppression Purposes”, Proceedings of the Sixth IASTED International Conference on Visualization, Imaging and Image Processing, Palma de Mallorca, Spain, 2006, pp.116-121.

[10] V. Mihajlović, S. Djordjević-Kajan, D. Rančić, B. Predić, I. Antolović, P. Eferica, Z. Babić, “Architecture of HASIS-3D System Designed for Hail Suppression Purposes”, Proceedings of the XLII ICEST Conference, Ohrid, Macedonia, June 2007, pp.403-406.

[11] E. Gamma, R. Helm, R. Johnson, J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison Wesley, 1995

Image Based Virtual Tour through Bitola

Igor I. Nedelkovski[1], Svetlana P.Vrskova[2],

Abstract – **A virtual tour (or panoramic tour) is a simulation of an existing location, usually composed of panoramic images, a sequence of hyperlinked still or video images, and/or virtual models of the real location. For creating virtual tour, first step is creating panoramas and then use them for creating virtual tour.**

This paper gives analysis of methods for creating panoramas and method of creating virtual tour.

Keywords – **Virtual tour, image stitching, panoramas, Virtual Walk-Through.**

I. INTRODUCTION

A virtual tour (or panoramic tour) is a simulation of an existing location, usually composed of panoramic images, a sequence of hyperlinked still or video images, and/or virtual models of the real location. They also may use other multimedia elements such as sound effects, music, narration, and text. As opposed to actual tourism, a virtual tour is typically accessed on a personal computer or an interactive kiosk. The phrases "panoramic tour" and "virtual tour" are often used to describe a variety of video and photographic-based media.

The word "panorama" indicates an unbroken view, since a panorama can be either a series of photographs or panning video footage. However, the phrases "panoramic tour" and "virtual tour" have mostly been associated with virtual tours created using still cameras. Such virtual tours are made up of a number of shots taken from a single vantage point. The camera and lens are rotated around what is referred to as a nodal point (the exact point at the back of the lens where the light converges).

A video tour is a full motion video of a location. Unlike the virtual tour's static wrap-around feel, a video tour is as if you were walking through a location. Using a video camera, the location is filmed while moving from place to place. Video tours are continuous movement taken at a walking pace.

Virtual tour today has wide area of implementation like:

1. Real Estate: People from estates make virtual tours of their estates, so that they can create beautiful and vivid effect of the houses to the public. That is quite eye-catching and greatly welcome by customers.

2. Tourism and Resorts: They make virtual tours of resorts, so that tourists can enjoy the beauty and attraction of those resorts beforehand. This has brought them more and more tourists and thus more and more money.

3. Web Design: Web designers embed virtual tours in their website, so that their webpage looks much more lively. Further, their websites are visited more frequently.

4. Exhibition: Some exhibits are made into virtual tour so as to achieve much more attractive effects.

5. Facility Management: Use panoramas rather than normal pictures to documents and manage your project or facilities.

Depending for what purpose is virtual tour, you can use different equipment for creating panoramas (the economical method is stitch photographs taken from the same point in space but of varying pitch and yaw and arranged them in spherical projection). In this paper we demonstrate low cost technology (with use traditional digital camera) for producing continuous 360-degree spherical virtual tour on long-range path (~ 1,5 km), instead traditional use of this technology for spot virtual tours.

II. METHODS OF CREATION PANORAMAS

There are many methods of creation panoramas. You can produce any type of panoramas with any type of camera, digital or analog. Depending on the lens you usually stitch 8-16 images for a 360º panorama.

Wide angle lenses with coverage of 60 degrees or more are usually used. The images are taken with 20%-50% overlapping. Important for all panoramas:

1. Use manual mode with same exposure for all images.

2. If possible do not use automatic white balance.

3. Use a tripod with a special panohead which rotates the camera around the nodalpoint.

Fisheye lenses are popular for spherical panoramas as we can cover a full view with only 3-8 images.

To be able to make a full spherical with 3 images we need a full circle of min 180º. This can be done by using a 35mm film camera or a fullframe digital and an 8 mm fisheye.

Fullframe Fisheye lenses usually 15 - 16 mm has been used for many years for panoramas with 35 mm cameras. The quality is much better than with the 8 mm fisheyes but they need 6-8 images around + top and bottom images.

One shot 360º panoramic solutions are parabolic mirrors which can be used as an ad on to many digital cameras. The resolution is low and the quality is also affected by the quality of the mirror.

Digital panoramic scanning cameras can make very high quality panoramas but the price is also high.

The method we used for creation of panoramas is to stitch photographs taken from the same point in space but of varying pitch and yaw (Fig.1) in order to create a panoramic image

[1]Igor Nedelkovski is with the Faculty of Technical Sciences, St.Kliment Ohridski University - Bitola, address: ul. Ivo Ribar Lola, 7000 Bitola, Macedonia, e-mail: igor.nedelkovski@uklo.edu.mk.

[2]Svetlana Vrshkova is with the Faculty of Technical Sciences, St.Kliment Ohridski University - Bitola, address: ul. Ivo Ribar Lola bb, 7000 Bitola, Macedonia, e-mail: svetlanavrskova@yahoo.com.

which, when rendered using appropriate software (web browser flash plugin), allows the end user to pan and zoom at will.

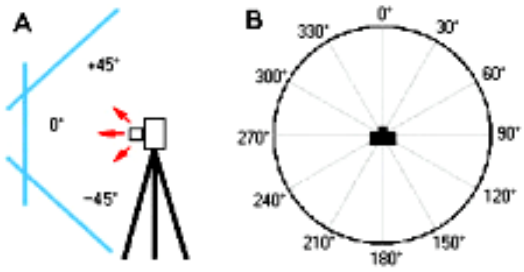


Fig.1 Method of taking photographs for creating panoramas

The benefit of this method is that it does not require any specialized equipment to capture the images, although using such can greatly speed up the process and render a higher quality result. This photographic technique results in a limited depth of field, meaning that the space may appear warped.

The next generation of this technology is the Virtual Walk-Through. This technology eliminates the limitations involving depth of field by enabling a user to travel tangentially throughout a space, in addition to rotating the point-of-view in any direction.

Image stitching or photo stitching is the process of combining multiple photographic images with overlapping fields of view to produce a segmented panorama or high-resolution image. For image segments that have been taken from the same point in space, stitched images can be arranged using one of these graphical projections:

- Rectilinear projection, where the stitched image is viewed on a 2D plane.
- Cylindrical projection, where the stitched image shows a 360° horizontal field of view and a limited vertical field of view. Panoramas in this projection are meant to be viewed as though the image is wrapped into a cylinder and viewed from within. When viewed on a 2D plane, horizontal lines appear curved while vertical lines remain straight.
- Spherical projection, where the stitched image shows a 360° horizontal by 180° vertical field of view. Panoramas in this projection are meant to be viewed as though the image is wrapped into a sphere and viewed from within. When viewed on a 2D plane, horizontal lines appear curved as in a cylindrical projection, while vertical lines curve as they get closer to the poles of the sphere.

The workflow of this method of creation panoramas with stitching photographs taken from the same point in space but of varying pitch and yaw includes:

1. Browsing folder that contains the photos taken from the same point in space (Fig.2)
2. Detecting panoramas (Fig.3)
3. Editing panoramas (cropping) (Fig.4)
4. Save and render panoramas (Fig.5)

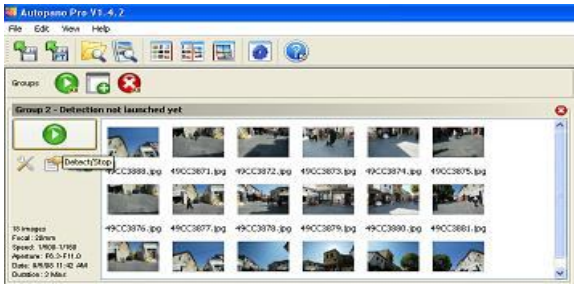


Fig.2 Browse photos

Fig.3 Detecting panorama

Fig.4 Cropping panorama

Fig.5 Render panorama

III. METHOD OF CREATION VIRTUAL TOUR

The first step required to create a virtual tour is to capture the virtual tour images. Virtual tour images are created by capturing from 6-18 images from the center point of view of the scene. This is done by placing a specially equipped digital or film camera on a panoramic tripod head at the center point of the scene and rotating the camera after the image is captured.

Generally 20%-50% overlap is sufficient for stitching together the virtual tour.

The second step required to create a virtual tour is to process and then retouch your stitched image. You must have appropriate virtual tour software for doing this step. There are many companies that provide software for creating virtual tours. Some leaders vendors you can find on this site. Virtual tour software comes in variety configurations from automated to complex, from free to expensive. Depending on your needs and quality your results can vary dramatically.

The workflow of the method of creation virtual tour includes:

1. Make virtual tour skin to get a unique virtual tour.

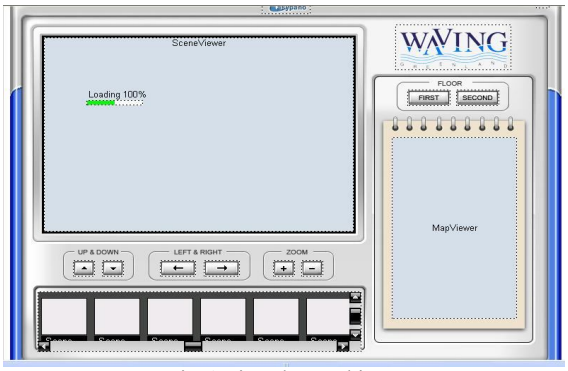


Fig.6 Virtual tour skin

2. Add panoramas and map.

Feature Description		
Add Spherical(S)...	Add Spherical	Import spherical panorama
Add Cylindrical(C)...	Add Cylindrical	Import Cylindrical panorama
Add Still Image(I)...	Add Still Image	Import Still image
Add Kaidan One Shot(K)...	Add Kaidan One Shot	Import Kaidan One Shot
Add Remote Reality One Shot(R)...	Add 0-360 One Shot	Import 0-360 One Shot
Add 0-360 One Shot(O)...	Add Remote Reality One Shot	Import Remote Reality One Shot image
Add Single Fisheye(F)...	Add Single Fisheye	Import single fisheye image
Add Cubic(B)...	Add Cubic	Import cubic panorama

Fig.7 Add panoramas

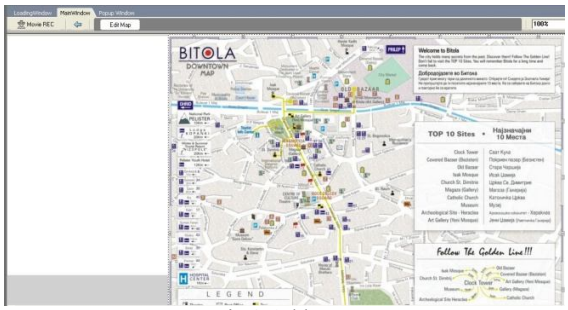


Fig.8 Add map

3. Add various hotspots, such as sound, URL link to link to another tour, link to scene, thus create vivid and colorful tours.

Fig.9 Add hotspot with Walk-Through effect

4. Adding hotspot and radar on map. Radar is actually a hotspot which contains a specified action to indicate the position and direction of the panorama in map.

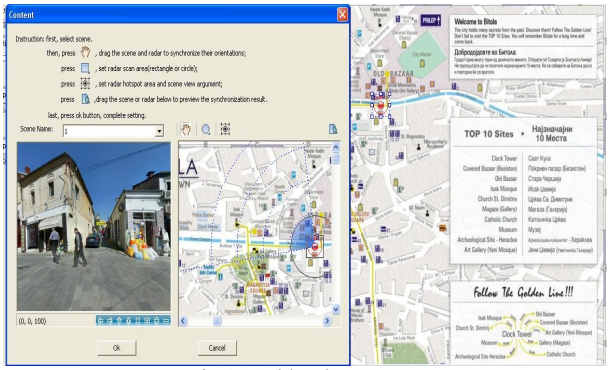


Fig.10 Add radar on map

5. Add background sound.

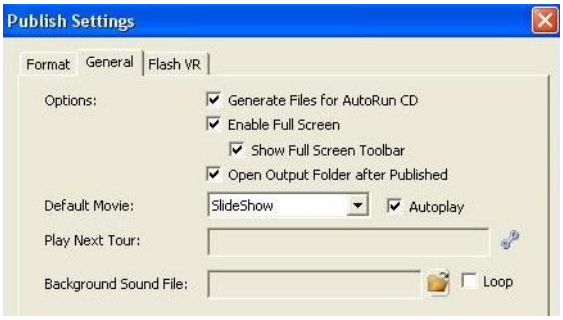


Fig. 11 Add background sound

6. Publish virtual tour.

The final step required to create a virtual tour is to prepare the virtual tour for deployment on the Internet. Virtual tours can be deployed using plug-in or Java technologies. A plug-in is a helper application which works within a Web browser, or as a standalone application. Plug-ins are usually available free of charge as downloads on the Web. The advantage of plug-ins is that they extend and improve a browser's capability by enabling the browser to play a wider and more robust selection of Web media.

Using this method we created image based virtual tour through Bitola.

Fig.12 Image based virtual tour through Bitola

Visit: http://www.visitbitola.info and enjoy walking through Bitola.

IV. CONCLUSION

Creating virtual tour is method of creating panoramas and then retouch them with appropriate virtual tour software. Depending for what purpose is virtual tour, you can use different equipment for creating panoramas (the economical method is stitch photographs taken from the same point in space but of varying pitch and yaw and arranged them in spherical projection). In this paper we demonstrated low cost technology (with use traditional digital camera) for producing continuous 360-degree spherical virtual tour on long-range path (~ 1,5 km), instead traditional use of this technology for spot virtual tours. Further research will continue toward improving procedures for efficient taking of the needed photos and avoid problems with blending of mobile objects on the scenes.

REFERENCES

[1] Anne Morgan Spalter, The Computer in the Visual Arts.Adision – Weesley, 1999.

[2] James D.Foley, Andries van Dam, Steven K.Feiner, John F. Huges, Richard L. Phillips, Introduction to Computer Graphics. Adison – Weesley, 1993.

[3] Malcom Kesson, An Introduction to 3D Computer Graphics. Malcom Kesson, 1995.

[4] Igor Nedelkovski, Computer Graphics, 3D Modeling and Animation. Faculty of Technical Sciences - Bitola 2008.

[5] http://www.autopano.net
[6] http://www.easypano.com
[7] http://www.realviz.com
[8] http://www.wikipedia.org/
[9] http://www.dpreview.com
[10] http://www.0-360.com
[11] http://www.panoramas.dk/
[12] http://www.panoguide.com
[13] http://www.realtourvision.com
[14] http://www.virtualtourusa.com
[15] http://www.thegnomonworkshop.com/
[16] http://www.cambridgeincolour.com
[17] http://www.commission5.isprs.org
[18] http://www.goodtutorials.org/
[19] http://www.photomodeler.com/
[20] http://www.canoma.com/

A Software Solution For Making Digital Systems For Data Sorting

Pavle Savković[1], Branislav Atlagić[2], Nikola Vranić [3], Vladimir Marinković[4]

Abstract - **The paper presents an approach to the design of digital data-sorting system in parameterized HDL source code, and software for its customization and generation. As a case study of presented concept, it is described an optimal solution for data sorting system that receives input data in serial way.**

Keywords - **data sorting, parallel digital systems, FPGA**

I. INTRODUCTION

An effective data sorting is an obligatory problem in every complex system, and it plays an important role in many disciplines like: automatic control in industry, robotics, artificial intelligence, genetic algorithms, medicine etc. Need for efficient sorting algorithms is evident from great number of published works on this subject. [1]. Most of those algorithms are intended to work on different processor types. Recently, due to increase in equivalent gate count, development of dedicated primitives used for image processing, as well as lowering power consumption, FPGA (*Field-Programmable Gate Array*) devices slowly starts to replace some segments or even a whole traditional image processing systems. Beside dedicated DSP (*Digital Signal Processing*) primitives, potential for parallel computing is greatest advantage of FPGA-based devices. This improvements lead to development of new algorithms for data sorting and tailoring existing ones.

One of the classifications of the sorting algorithms, realized in the digital systems, is based on the ways of data reception:

- Serial – the data arrives in the system one after another
- Completely parallel – all data arrive in the system simultaneously
- Partially parallel – the data arrives in the system partially in series and parallel way.

To make the development of digital systems for data sorting easier dedicated software environment is developed.

[1]Pavle Savković, is with Faculty of Technical Sciences at University of Novi Sad, Fruškogorska 18, Novi Sad, 21000, Serbia, E-mail: Pavle.Savkovic@krt.neobee.net

[2]Branislav Atlagić, is with Faculty of Technical Sciences at University of Novi Sad, Fruškogorska 18, Novi Sad, 21000, Serbia, E-mail: Branislav.Atlagic@krt.neobee.net

[3]Nikola Vranić, is with Faculty of Technical Sciences at University of Novi Sad, Fruškogorska 18, Novi Sad, 21000, Serbia, E-mail: Nikola.Vranic@krt.neobee.net

[4]Vladimir Marinković, is with Faculty of Technical Sciences at University of Novi Sad, Fruškogorska 18, Novi Sad, 21000, Serbia, E-mail: Vladimir.Marinkovic@krt.neobee.net

This software incorporates following modules:

- SortGen – a software tool module for calculating the parameters and synthesis the digital systems for data sorting,
- SortProbe – a software tool module for test & verification of the developed data sorting systems and
- SortLib – Verilog library of parameterized source code for definition, synthesis and verification (testbench development) of the data sorting system

Using the generated testbench, SortProbe tool facilitate testing of the sorting system by sending test data and receiving results, via serial interface to the target hardware platform. This encircles the development and testing process of the digital system for data sorting.

The simplest system for realization of all sorting systems is the system where data comes serially. That was the reason why this type of data reception was realized first. The chapter two contains description of the algorithms and digital system implementation. The chapter three contains the detailed explanation of the SortGen software tool. The forth chapter contains description of the SortProbe tool, and in the end, the chapter five contains the results/comments for few different FPGA based system realization.

II. DESCRIPTION OF DIGITAL SYSTEM FOR DATA SORTING

In this project, we consider the case of sorting in which the members of array are not known at the beginning of the sorting, but they are coming in regular time period. Besides this constraint, the system should enable that:

- data array must be sorted in any time,
- the maximal period for sorting entire data array after arrival of new data item must be less than one system cycle, and
- the source code must be organized in the way that enables parameterization.

These constrains will be meet only if the data array is sorted before the moment of new data arrival (i.e. less than period of the one system cycle). The most important parts of the system for sorting are shown in the Figure 1. Because of the large number of modules and the connections between them, only the most important parts and their connections are shown. The registers for storing inputs, from R0 to Rn, and one of the n pairs of the multiplexer-comparator, which are the basic elements of system for sorting the array of the elements length n, are shown. Output from multiplexer 0 represents minimal, and output form multiplexer n-1 represents maximal value in data array. The algorithm is based on idea to compare new

input item simultaneously to all array data, and than do the calculation of new values for multiplexer's selection.

The data items are stored in one of the registers immediately after their arrival. When system starts to work,

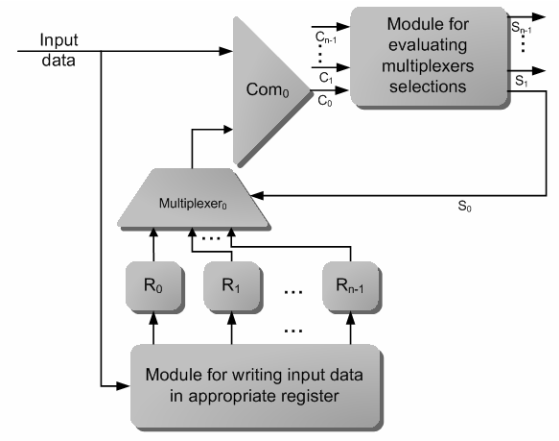


Figure 1. Block diagram of key parts of system for sorting data

the first arrived item is stored in the register R0, the next one in R1, until the nth item that is stored in the register Rn-1. The next one, n-th data item is again stored in the register R0, and so on, following the circular scheme for writing in the registers. In which one of the registers the data is stored depends on value of a dedicated counter with modulo n. Receiving data in cyclic manner is chosen to minimize number of registers. Algorithm for data sorting does not require any substitution of register's contents. The number of registers is equal to the number of members in data array that should be sorted. To avoid changing values between registers multiplexers n-to-1 are used. The number of multiplexers is equal to the number of the registers for data array. The output from multiplexer is connected to the comparator, which compares the value from the selected register and the current input item. All of the registers are brought at the multiplexers inputs. The selections come from the module for calculating the new selection values.

The fact that the new data is inserted in previously sorted array makes the calculation of the new selection values easier[2]. The outputs from comparators follow certain pattern. This will be explained in the next example, where the array of 9 elements is to be sorted. The output from comparators is shown in the Figure 2. The comparators 0 to 5 show that the new value is higher than the value that is brought on those comparators, while the comparators from 6 to 8 show that the new value is smaller. Because of that the

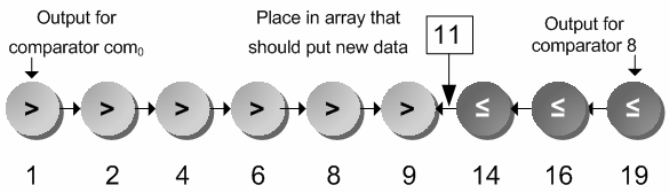


Figure 2. Example of comparators output

conclusion is that the new value should be stored at the 6th place, counting from 0. This example proves that the logic for the updating multiplexer selections is simple and it could be parameterized easily.

The module for the evaluating multiplexer selection contains logic for determining the new values for multiplexer selections, which are to be valid during the next cycle period, and the dedicated registers for their storing until this cycle begins. This solution fulfills all demanded requirements, i.e. sorting is done within one system cycle, and data array remains sorted all the time.

III. Description of SortGen Software

Possibility to parameterize the digital system leads to development of SortGen software. Software, according to user's requirements, calculates system parameters and generates custom source code using library of parameterized source code. The first version of the SortGen tool was console application and it was used for initial testing. The second version introduces a graphic user interface (GUI). Figure 3 shows the main window SortGen tools.

To generate source code, it is necessary to enter array length, the data width, sorting order (increasing or decreasing order) and finally which data will be from the sorting system will be forwarded to its output. To output from the system for

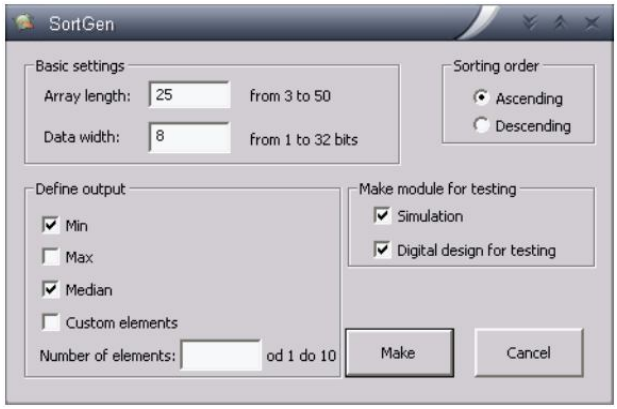


Figure 3. Main window SortGen tools

data sorting can be forwarded, individually or in any combination, minimum, maximum and median value of sorted array. Also user can define any other position in array to be forwarded to output. This option is added in order to completey customize the system to the user's needs.

When the user enters all the necessary information with confirmation on button "Make", tool will create 15 files with the source code in the same directory where the tool is settled. All the source code files are fully accessible to user.

A. The Description Of The Digital System For Data Sorting

The software can optionally make testbench and complete digital system for the testing the system for sorting on FPGA. The Figure 4 presents the testbench organization.

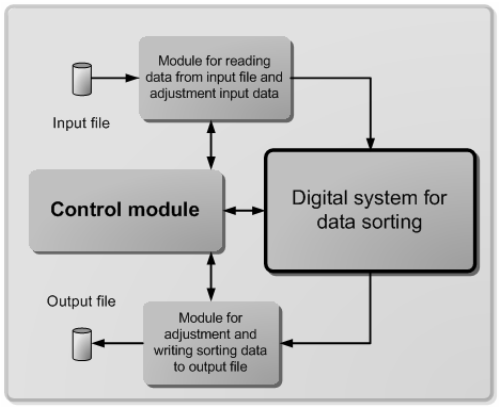


Figure 4. Testbench organization

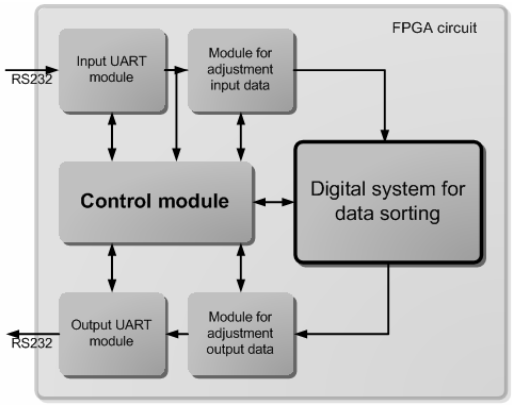


Figure 5. The organization of the digital system for testing digital system of sorting the facts

The testbench is composed of four modules. The task of the module for reading and adapting the values from input file is to read one by one data from the input file and to send it to the control module and digital system under test. The input file should be formatted in such a way to contain in every line just one value, which will be forwarded to the system under test and control module, without any other symbols, expect decimal numbers (hexadecimal and binary numbers are not supported). The duration of simulation is determined by the number of values in the input file. If the value, which goes out of the tested range, is brought in the file, or if it is brought into any sign, which is not the number, there will be indicated a mistake and the simulation will be stopped. The control module is responsible for the control of the simulation's progress and checking of the output values during the simulation. In order to speed up the testing and to shorten the time of simulation, the control module models digital system for sorting, using the constructions of Verilog language, which could not be synthesized. It compares the outputs of the digital system with results, which are got in the model of simulation's system and if any differences are detected, it will issue the warning to the user, who can in that moment stop the simulation, without need to wait its ending. Although the possibility of simulation's check during its operation is incorporated, the results of the sorting are written in the output file. The special module is responsible for formatting the data and their writing in the output file. The format of the output file is the same as the format of the input file.

B. The Description Of The System For Testing Digital Sorting System On FPGA

Beside the testbench, the SortGen software is able to generate complete digital system for testing digital system on FPGA[3]. The organization of the testing system is given in Figure 5.

The SortProbe software (it will be explained in details in the next chapter) sends via serial interface commands and data to the digital system realized in FPGA device. When system starts the operation, SortProbe first establishes the connection with system under test and then it sends the commands for setting system parameters. All of that is followed by the sending data and commands and reconvening results. The receiving UART module (Universal Asynchronous Receiver / Transmitter) is responsible for the serial receiving data and commands from SortProbe software.

The width of the word, is sent or received by UART interface, is predefined to 8 bits. The most significant bit in each word sent over serial interface indicates what is sent, data or command. The controls relate to adapting the parameters in the system and in the communication between the digital system and SortProbe software. The task of the receiving UART module is to interpret received message, to send commands to the command module and the data to the system for sorting and to the control module. The module for adapting input data takes over from the receiving UART module certain number of words and forms them in the data, which are intended to the digital system for sorting. For example, if the width of the data, which are sorted by the system, is 24 bits, it is needed to take over 4 words from the receiving UART module and to make of them one 24-bits data word and to send it to the system for sorting. It is needed four 8-bits words for sending 24-bits of the data, because every package contains one bit of the indication, which shows if the sent package, is data or command.

The results of the digital system for sorting are in the same way formed for sending. The forming of the data performs the module for the adapting the results, before it sends them to the Delivery UART module, which sends them to SortProbe software via UART interface. The control module controls the system and supervises the progress of the receiving and sending the data. If the control module finds out that there is a possibility of making the mistake in the data flow, it could stop every individual module and/or send message to SortProbe software via UART interface. The stopping of the digital system is performed synchronously in order not to exert influence on its function and not to imperil the integrity of the data flow in the system.

IV. THE DESCRIPTION OF THE PROGRAM SORTPROBE SOFTWARE

The main task of the SortProbe software is to enable sending the data and commands to the digital system for testing the system for data sorting and to enable the

supervision of results. The Figure 6 presents the main window of SortProbe software.

Before the start of the operation it is necessary to enter the following information: a array length, which should be sorted, the width of the data in bits, if the system is tested by the data from the user's file or random array of numbers should be generated, then the length of the tested sequence. After the entering all of the necessary data, by confirmation on the button "Run test", begins the testing of the digital system for data sorting.

The instantaneous status of the testing is marked by three indicators: red one, yellow one and green one. If there is not any mistake or warning during the testing – the green indicator is active. If there is a warning – the yellow indicator

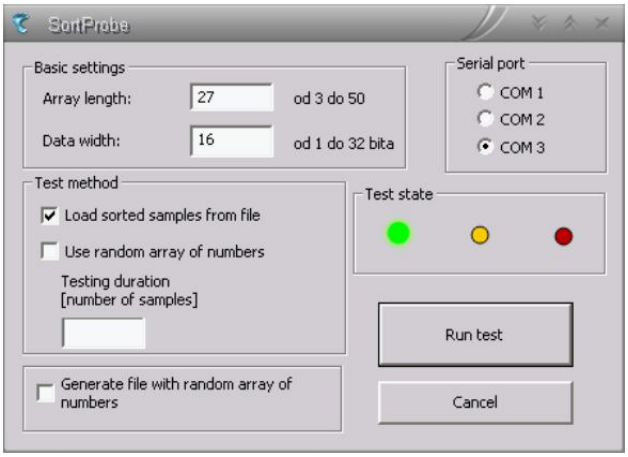


Figure: 6. The main window of SortProbe software

is active (for example, the input buffer overflow), and if there is a mistake – the red indicator is active. The receiving results and the messages are put in two separated files, which can be later reviewed by user.

V. The Results Of The Realization Of The Digital Sorting System

The testing is performed on digital sorting system for sorting a array of 25 bytes. The digital system and testbench for it was generated using SortGen software. The results of the simulation were correct. Next step was making the digital system for the testing on the FPGA. The made system is tested on the platform for fast development of the digital systems, which is based on Xilinx's FPGA Spartan 3 device. The testing is done by the help of the SortProbe software. For the testing the special test files are made. The results of the testing are checked using SortProbe software and the SciLab software package for verification. Both software gave the same results.

In order to get information about the quantity of the resources, which the digital sorting system takes, it is made a synthesis of the systems for few of the most used FPGA devices of Xilinx production, whose software for synthesis is free. The results are given in the Table 1.

TABLE I
THE OCCUPATION OF THE RESOURCE IN THE DIFFERENT FPGA CIRCUITS

			Array length		
			9	25	35
FPGA circuit	SPARTAN 3 XSC200 FG900 -5	LUTs	704 (1%)	5284 (12%)	8578 (20%)
		ROM	6	22	30
		MHz	85	65.9	62.1
	VIRTEX 2P VP50 FF1152 -6	LUTs	745 (1%)	5284 (10%)	8641 (17%)
		ROM	4	22	26
		MHz	102	72.3	68.3
	VIRTEX 4 VLX80 FF1148 -11	LUTs	713 (<1%)	5555 (7%)	8564 (11%)
		ROM	6	22	31
		MHz	132	97.8	92.1
	VIRTEX 5 LX220 FF1760 -2	LUTs	510 (<1%)	3414 (2%)	5648 (4%)
		ROM	4	18	27
		MHz	141	108.4	98.8

The Table 1 shows how many resources has the system for the date sorting, whose length is 9, 25 and 35 members, when the synthesis for different FPGA devices is done. The number of ROM primitives (Read-only memory) and logical program cells describes, how much the digital system is demanded about the resource questions. The frequency expressed in MHz presents the maximal speed, on which the system can operate without any mistake.

VI. Conclusion

In this work is described an environment for the design digital systems for data sorting and algorithm for data sorting for the system, in which data arrive serially. All of the goals, which are placed in front of the algorithm before the projecting, are satisfied. The example of the realization of the digital system for data sorting of the 25 members shows that the system is absolutely functional and that it make easier and speeds up design of the complex digital systems. The development of extension, for described system, for parallel reception and parallel sorting based on sorting networks is in progress.

References

[1] Leighton, Frank Thomson, "Introduction to Parallel Algorithms and Architectures-array, tress, hypercubes", San Mateo Morgan Kaufmann Publishers, Inc. 1992

[2] Donald E. Knuth, "The art of computer programming, volume 3 : sorting and searching" Addison Wesley Longman Publishing Co. ,Inc., Redwood city, Ca,1998

[3] Hahanov,V Kaminska,M. "Testability Analysis of Digital Design Verification", Baltic Electronic Conference, 2006

Session

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES – 2

Specialized Geo-Information System for Calculation of Distribution Network Losses

Miloš Bogdanović[1], Aleksandar Stanimirović[2], Saša Tošić[3] and Leonid Stoimenov[4]

Abstract –**This paper presents out researches in the area of geodata integration and Geo-Information Systems (GIS). As a result of this research we have developed GinisED geo-information system for public electric power supply company PD Jugoistok Niš. Main purpose of GinisED is documentation, maintenance and analysis of electric power supply networks. This paper contains a description of solutions for the calculation of losses in the distribution network. Also, we present a prototype application for the calculation of losses in the distribution network. This application was developed using GinisED system as a base.**

Keywords – **GIS, integration, distribution network losses**

I. INTRODUCTION

In recent years integration of information from heterogeneous sources is one of the most popular research directions in the field of information systems [1]. Information integration involves the usage of data from two or more databases (e.g. information sources) for the development of a new database. New database can be physically implemented, or developed as a virtual database. This database enables unified querying of integrated information. Integrated information sources may be conventional databases, Web pages, text files, mail etc. The need for such integration is present in the process of calculation of low voltage distribution network losses.

Energy crisis and current problems of efficient energy utilization demand introduction of effective mechanisms for the optimal use of available energy. From the aspect of electric power supply companies, energy efficiency is closely related to losses in distribution of electric power. Therefore, expert debates and analysis devotes a great attention to the issue of energy losses and their reduction, especially when losses are not at a satisfactory level [2].

Energy losses are one of the key elements that indicate the degree of quality of performing the business activities in electric power supply companies. Therefore, the reduction of energy losses in the distribution network has become one of the priority business goals [3] in companies engaged in distribution of electricity, such as PD Jugoistok Niš company. A significant part of losses in the distribution of electricity refers to the low voltage (LV) network losses.

The first step in the process of reducing the LV network losses is the calculation of losses based on available information concerning network infrastructure, customers and their consumption. Procedures used for calculation of LV losses are known and defined, but the problem of availability of data, required for the calculation, still remains. The data needed for the calculation of LV losses are often located in isolated information systems. Therefore, for successful calculation of LV losses, it is necessary to ensure the integration of information and access to all necessary information sources from single access point.

This rest of the paper is organized in the following way: next chapter presents the problem of LV losses calculation. In Chapter 3 GinisED system is presented along with information integration platform it uses. Chapter 4 gives an overview of the application for calculation of LV network losses. At the end of the paper, a conclusion and review of literature are given.

II. CALCULATION OF ELECTRIC ENERGY LOSSES

Losses of electric power in electric power supply network correspond to difference between amount of transmitted energy and the amount of energy sold to customers. Analysis of electricity losses is very complex due to a large number of parameters that affect them. In order to perform this kind of analysis, it is necessary for the data used to estimate losses to be as specific as possible and to reflect the real state of electric power supply network [2].

The basic structure of electricity transfer/distribution system consists of a large number of different plants, electric power lines and devices. For the purposes of managing this complex network, several IT systems that cover different areas of network functioning are used. Electric power supply systems often use: Geo-Information System (GIS), Supervisory Control and Data Acquisition (SCADA), Distribution Management System (DMS), Automatic Meter Reading (AMR), Computerized Maintenance Management System (CMMS), Consumer Information System (CIS) and other [4, 5, 6].

The basic purpose of all the above mentioned systems is to provide electric power supply companies with support for business processes such as recording, maintaining and planning of electric power supply network. From the standpoint of reducing the losses in LV network, these systems contain information necessary for the calculation of losses. GIS allows recording and georeferencing of all electric

[1]Miloš Bogdanović is with the Faculty of Electronic Engineering at University of Niš, 14 Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: milos.bogdanovic@elfak.ni.ac.rs

[2]Aleksandar Stanimirović is with the Faculty of Electronic Engineering at University of Niš, 14 Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: aleksandar.stanimirovic@elfak.ni.ac.rs

[3]Saša Tošić is with PD Jugoistok, 4 Zetska, Niš 18000, Serbia, E-mail: sasa.tosic@jugoistok.com

[4]Leonid Stoimenov is with the Faculty of Electronic Engineering at University of Niš, 14 Aleksandra Medvedeva, Niš 18000, Serbia, E-mail: leonid.stoimenov@elfak.ni.ac.rs

power supply network elements. Spatial database contains details about the length and diameter of network electric power lines. This data can be successfully used for the calculation of a part of LV network losses. On the other hand, measured data from SCADA, electric power supply network state from DMS, current consumption data acquired by AMR and charged electricity consumption data from CIS can significantly contribute to increase the accuracy of calculation of losses and possibly locate a potentially critical location [2].

Calculation of losses in the LV network is quite complex due to existence of a large number of elements in LV distribution network. In this case, if we know input power and electric current that flows through each section of LV network in advance, we can apply methodology based on the calculation of radial network [2]. Electricity losses calculation requires information that can be found in different information systems that operate within the electric power supply company. These systems are usually completely isolated which makes data they generate available only to users of a single system. Therefore, there is a problem of ensuring that different users in different locations have access to all necessary information. If the data processing is done manually, the probability of harming or losing data consistency is increased. In order to support the exchange of information, IT systems integration techniques should be applied.

III. GINISED SYSTEM AND INFORMATION INTEGRATION PROCESS

The functioning of companies engaged in the transmission and distribution of electricity depends on the existence of appropriate electric power supply network geodata [7, 8]. It is estimated that more than 80% of data used in a variety of processes (network design process, data input and update, maintenance and various analysis) posses geographic (spatial) component. Therefore, almost any electric power supply company has a need for the existence of specialized geo-information system that should provide mechanisms for collecting, storing and manipulating spatial data.

GIS applications enable connecting various types of information in the spatial context and generating new information and conclusions on the basis of these connections. GIS allows the integration of information in a manner that is impossible to achieve using any other type of tool. This significantly increases the value of GIS in everyday usage and maintenance of electric power supply network. In addition, integration with other systems reduces costs and simplifies the maintenance of GIS systems. There is no need for special hardware, software platforms or special trained GIS users. Because of its openness, GIS system is very easy to integrate with other IT systems within a single electric power supply company (Fig. 1).

For the needs of PD Jugoistok Niš, CG&GIS Lab, Faculty of Electronic Engineering in Niš, with the support Ministry of Science of Republic of Serbia, developed a geo-information system GinisED [7, 8]. GinisED is a specialized geo-information system which, in addition to standard alphanumeric data concerning electrical parameters of electric power supply network, allows recording, processing, analysis and graphic presentation of specialized information about the electric power supply network, such as spatial data, temporal data, image and multimedia.

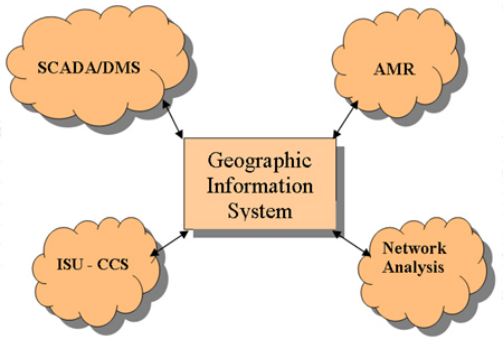


Fig. 1. GIS as a basis for IT systems integration in electric power supply company environment

GinisED system was developed using GeoNis platform for the interoperability of GIS applications. GeoNis platform provides the mechanisms and infrastructure for the exchange of information in the environment of local government [1, 9], but can be applied for integration of information on a single company level. This platform is developed for purpose of intelligent integration of information from a number of heterogeneous GIS (geographical and spatial) and non-spatial data sources. Companies, institutions or their parts that have some information of interest are considered to be data sources.

GeoNis platform is located between GinisED system, which operates as a C3 (Command Control and Communication) module, and relevant data sources (GeoInformation Community - GIC) (Fig. 2). GeoNis environment nodes can be existing applications. For each of applications, it is necessary to develop translators and domain (local) ontologies. Nodes may also be new applications developed in accordance with the OpenGIS standards and component software development methodology [10].

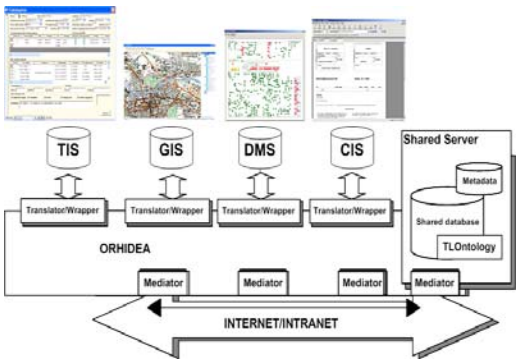


Fig. 2. GeoNis platform for the interoperability of GIS applications

Integration of information from all relevant sources provides data necessary for calculation of losses in electric power supply network. Thus, it is possible for users to receive real-time information about all parameters that are relevant to the functioning of electric power supply network.

IV. APPLICATION FOR LV NETWORK LOSSES CALCULATION

On the basis of developed GinisED system, which allows the integration of information from different IT systems in the PD Jugoistok Niš, a prototype application for calculation of losses in the LV network was developed. The architecture of this application is shown in the Fig. 3.

GIS module holds the central part in the application for the calculation of electricity losses. It is a downscaled GIS application that has retained only the minimum of required GIS functionalities. This application visualizes spatial data of electric power supply network and provides users with a simple interface to GinisED information integration system.

For the purposes of analysis and calculation of losses, data from three different information systems is currently being used. Other systems as information sources will be added with the further development and improvement of the application.

GIS system is used as a source of data related to LV network topology and technical description of LV electric power line sections (section length, section resistance, electric power line type, type of conductor, conductor diameter etc). LV network spatial data was recorded in the field and is being regularly updated.

LV network GIS data is related to information about consumers. CIS system contains consumer information [11]. Integration of GIS and CIS allow determination of consumer's exact position on LV electric power line. It also allows determination of geographical location of connection that the consumer is related to. This enables easy identification of all customers related to particular LV electric power line.

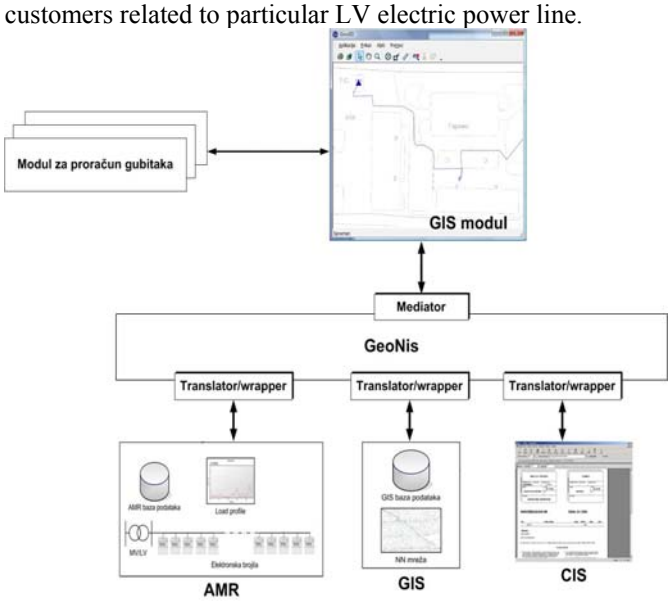


Fig. 3. LV network losses calculation application architecture

When all consumers related to a particular LV electric power line are identified, their unique consumer codes are used as input data to obtain their daily load characteristic diagrams from AMR system. AMR system uses modern electronic consumption meters. These meters allow storing of load characteristic diagrams for a period of time (load profile). Hence, load characteristic diagram is imposed as one of the basic analytical data for the calculation of energy balance and LV electric power line losses [12]. Fig. 4. shows typical load characteristic diagram.

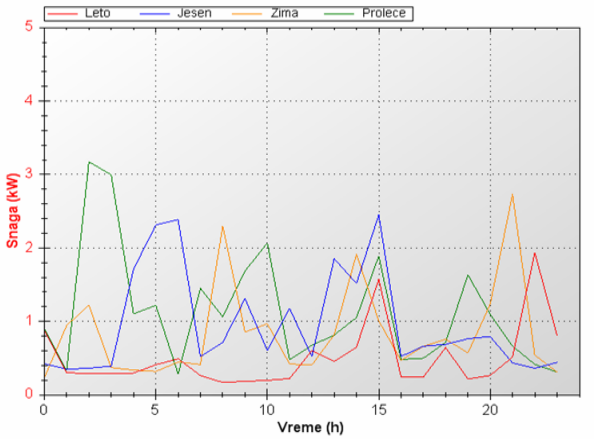


Fig. 4. Typical load characteristic diagram

Based on technical information related to LV electric power line (section length, type and diameter of the conductors) and consumer load characteristic diagram, losses calculation module determines losses on a particular LV electric power line. Losses calculation module is not based on approximative methods. Instead, it uses recursive method for calculating the electric current that flows through each LV electric power line section [2]. This module uses LV electric power line data topology as graph (from the transformer station to the end consumer). This graph consists of transmission facilities, sub facilities and consumption meters related to company clients. Based on unique customer codes, daily load characteristic diagrams are obtained from AMR system. If daily load characteristic can not be obtained, particular consumer is related to one of standard load characteristic diagrams.

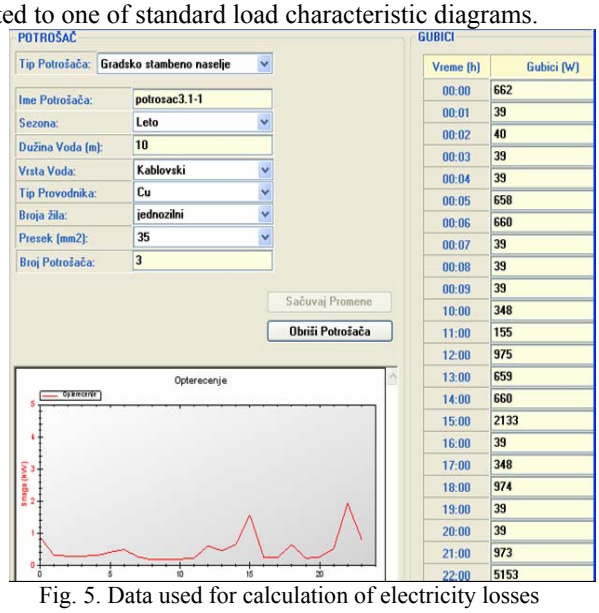


Fig. 5. Data used for calculation of electricity losses

Fig. 5. presents data related to a particular LV electric power line and consumer load characteristic diagram used for calculation of electricity losses. It is possible to alternate LV electric power line section parameters (electric power line

length, electric power line type, conductor type, number of cores and conductor diameter) and analyze how these changes affect the percentage of technical losses. It is also possible to define standard load characteristic diagrams for all four seasons.

Part of the application that is responsible for the calculation of losses has a modular architecture so it can be easily expanded with modules using a different methodology for calculation of losses. This allows recalculation of losses for LV electric power line using different methodologies and comparison and analysis of obtained results.

V. CONCLUSION

Every company that deals with the distribution of electricity paid great attention to the problem of energy losses, especially in a situation where the losses are not at a satisfactory level. Electricity losses are one of the key factors that indicate the degree of economy and quality of business in the area of electricity distribution. Due to these reasons, the problem of reducing energy losses has become one of the priority business objectives in PD Jugoistok Niš.

Calculation of electricity losses requires usage of information from various IT systems that operate in the PD Jugoistok Niš. As a starting point for the development of losses calculation prototype application, GinisED system was used. This system is developed for the purpose of recording spatial electric power supply network information. For integration with other systems within PD Jugoistok Niš, GinisED system uses GeoNis platform for interoperability. This provides transparent access to data regardless of their original format and storage location within the organization, which is of extreme importance in the application for calculation of electricity losses.

A prototype application for the calculation of electricity losses, using a recursive method for calculating the electric current that flows through each LV electric power line section, was developed. This application will enhance the detection and reduction of electricity losses.

Analysis of the results of electricity losses calculation opened the possibility for better work in the following fields:

- planning development and reconstruction of LV network where, in addition to other parameters for the financial justification of investments, technical losses reduction parameter is added
- possibility of simulating consumption increase in order to locate the parts of the network exposed to a large degree of losses
- selection of optimal LV network topology
- planning the introduction of remote reading meters

ACKNOWLEDGEMENT

Research presented in this paper were partially funded by the Ministry of Science of the Republic of Serbia and PD Jugoistok Niš, within the project in the field of technological development "Intelligent integration of geo-, business and technical information on the company level," ev. No 13003.

REFERENCES

[1] L. Stoimenov, *"Integration of semantic and distributed GIS information sources based on ontologies and mediation "*, PhD Thesis, Faculty of Electronic Engineering, University of Niš, Serbia, 2003 (in Serbian)

[2] S. Tošić, A. Krstić, B. Nikolić, *"Aplication for calculation of low voltage losses"*, CIRED 2008, Vrnjačka Banja, Serbia, 2008 (in Serbian)

[3] A. Pavić, J. Trupinić, "*Electrical Energy Losses In The Distribution Network",* Energija, Vol. 56, No. 2, pp. 185-215, 2007

[4] S. Vatland, L. S. Gundersen, G. Sande, J. Bugge, T. Asbjørnsen, T. Lund, "*Utility Systems Integration*", Nordic Distribution and Asset Management Conference, 2004.

[5] U. D. Kale, R. Lad, "*GIS integration with SCADA, DMS & AMR in Electrical Utility*", Map India Conference, 2006.

[6] A. Stanimirović, D. Stojanović, L. Stoimenov, S. Đorđević–Kajan, M. Kostić, A. Krstić, *"Geographic Information System for Support of Control and Management of Electric Power Supply Network"*, IX Triennial International Conference on Systems, Automatic Control and Measurements SAUM, ISBN 86-85195-49-7, Niš, 2007.

[7] S. Djordjević-Kajan, M. Božić, *"GeoEE - sistem u GIS tehnologiji za evidentiranje, održavanje i analizu elektroenergetske nreže"*, JUKO CIRED, Septembar 2002, Vrnjačka Banja, Serbia, 2002.

[8] L. Stoimenov, S. Đorđević-Kajan, D. Stojanović, M. Kostić, A. Vukašinović, A. Janjić, *"Geografski Informacioni Sistem za evidenciju, održavanje i analizu elektrodistributivne mreže",* YU INFO 2006, Kopaonik, CD izdanje, 2006

[9] L. Stoimenov, A. Stanimirović, S. Đorđević-Kajan, *"Development of GIS Interoperability Infrastructure in Local Community Environment",* From Pharaohs to Geoinformatics, FIG Working Week 2005 and GSDI-8 Cairo, Egypt April 16-21, TS41.2., 2005

[10] L. Stoimenov, A. Stanimirović, S. Đorđević-Kajan, *"Realization of Component-Based GIS Application Framework"*, 7th Agile Conference on Geographic Information Science, Heraklion, Crete, 29 April-1 May, pp. 11, 2004

[11] A. Stanimirović, L. Stoimenov, S. Đorđević-Kajan, M. Kostić, A. Krstić, *" Company level geodata integration within GinisED application "*, JUINFO 2007, Kopaonik, Serbia, CD Edition, ISBN 978-86-85525-02-5, 2007

[12] J. A. Jardini, C. M. V. Tahan, M. R. Gouvea, S. U. Ahn, F. M. Figueiredo, *"Daily Load Profiles for Residential, Commercial and Industrial Low Voltage Consumers"*, IEEE Trans. on Power Delivery, Vol.15, No. 1, Jan. 2000

Implementing Complex Polylines for use in GIS

Marko Kovačević[1], Aleksandar Milosavljević[2], Dejan Rančić[3]

Abstract – **This paper presents an approach for implementing visual portrayal of geographic features with polyline geometry (complex polylines) for appliance in GIS based applications. This approach relies on developed XML styling language for defining styling rules (decorations) of complex polylines. The structure of developed XML language was the foundation for designing the appropriate class library able to interpret XML definition of decorations and draw the desired complex polylines.**

Keywords – **Polylines, Style, GIS, XML, GDI+.**

I. Introduction

A geographic information system (GIS) is special type of computer-based information system tailored to store, process, and manipulate geospatial data [1]. The ability of GIS to handle and process both location and attribute data distinguishes GIS from other information systems. It also establishes GIS as a technology important for a wide variety of applications [2].

The fundamental information unit that GIS deals with is called a geographic feature. Geographic feature is an abstraction of a real world phenomenon associated with a location relative to the Earth [3]. Every feature may have a number of properties. One or more of the feature's properties may be geometric. Geometry provides the means for quantitative description of the spatial characteristics of features, including dimension, position, size, shape, and orientation. A geometric object is a combination of a coordinate geometry and a coordinate reference system. In general, a geometric object is a set of geometric points, represented by their coordinates. Basic geometric objects are points, polylines, and polygons [4].

The importance of the visual portrayal of geographic data in GIS cannot be overemphasized. The skill that goes into portraying data is what transforms raw information into an explanatory or decision-support tool. Fine-grained control of the graphical representation of geographic features is a fundamental requirement for any professional mapping community. Allowing user to define styling rules for visual portrayal of geographic features requires the existence of a styling language that the user and GIS application can both understand [5].

In this paper we present an approach for implementation of *visual portrayal of geographic features with polyline geometry* (further referred to as *complex polylines*). We developed XML styling language that enables simple and flexible way to define *styling rules* (further referred to as *decorations*) of complex polylines. The structure of developed XML language was the foundation for design and implementation of the appropriate class library aimed to interpret XML definition of decorations and draw the desired complex polylines.

Decorations include all graphic and text elements of a complex polyline. In this paper we focus on implementation of the following decorations:

- *Start cap* – graphic symbol at the start of the polyline.
- *End cap* – graphic symbol at the end of the polyline.
- *Pattern* – element set (includes graphic symbols, lines of different types (e.g. 2 pixels solid green line), empty spaces and text elements) that repeats itself along the polyline.
- *Label* – pattern like element set placed at defined positions on the polyline.

The paper is organized as follows: Section 2 presents XML language for defining decorations of complex polylines. Section 3 discusses the architecture and implementation issues of developed class library. Section 4 summarizes the achieved results.

II. XML Definition of Complex Polylines

XML language for defining decorations of complex polylines is specified using *XML Schema Definition Language*. A valid XML document contains a definition of decorations for a specific complex polyline, which can be interpreted using developed class library. Similar approach, based on defining styling language using XML Schema, is used by Open Geospatial Consortium in developing The Styled Layer Descriptor (SLD) Profile of the WMS [5].

Top level element that is used for specifying decorations is based on XML complex type named *ComplexLineType* (see Fig. 2). Bellow we briefly describe all containing elements of this XML type.

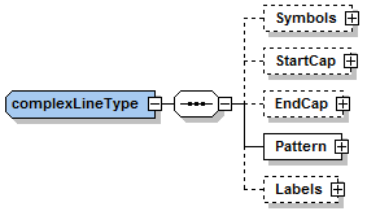


Fig. 2 - Structure of XML definition of complex polyline

Fig. 3 shows the structure of element *Symbols*. This element contains the description of every graphic symbol that is going to be used in decorations of the complex polyline. Contained element *SymbolLineWidth* is used to optionally setup the width

[1]Marko Kovačević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: markko.marce@gmail.com

[2]Aleksandar Milosavljević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: alexm@elfak.ni.ac.yu

[3]Dejan Rančić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: ranca@elfak.ni.ac.yu

of the line used for drawing graphic symbols. There is also an array of elements *Symbol*. Element *Symbol* has two attributes: *id* and *src*. Attribute *id* is a unique identificator of the graphic symbol, used for referencing in other elements. Attribute *src* is the path to the SVG file that contains the shape description of the graphic symbol. Contained elements of the element *Symbol* are used to optionally setup the appearance of graphic symbol: its line and interior color (*LineColor* and *FillColor*, respectively), width of the symbol (*Width*) and displacement from the polyline segment (*Offset*).

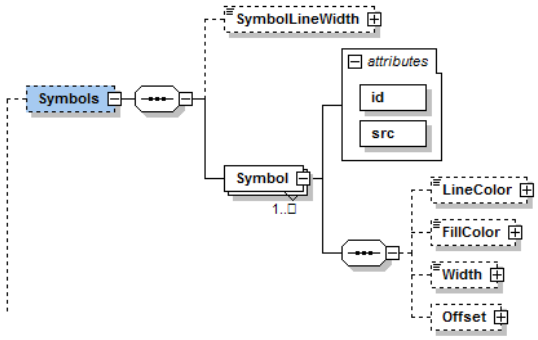


Fig. 3 - Structure of an XML definition of graphic symbols in complex polyline

Elements *StartCap* and *EndCap* are used to define graphic symbol at the start and at the end of the polyline, respectively. They only contain one attribute, *idRef*, a reference to the appropriate graphic symbol.

Fig. 4 shows the structure of an element *Pattern*, that is based on XML complex type *patternType*.

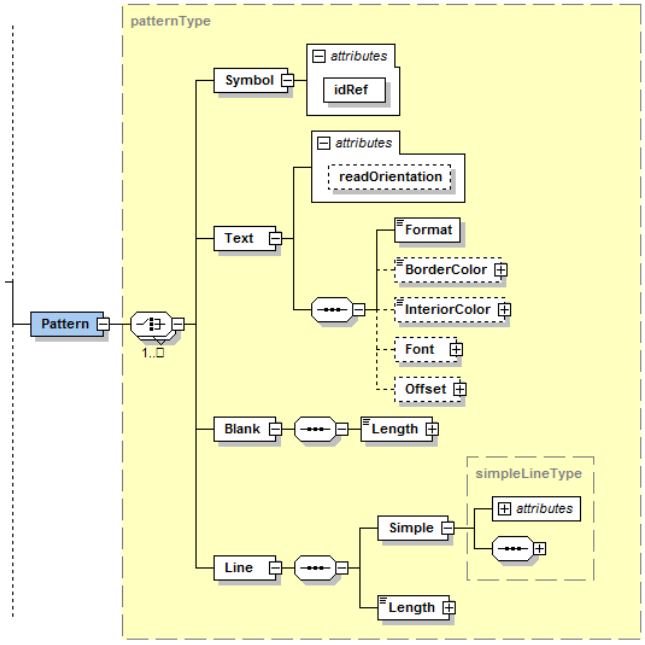


Fig. 4 – Structure of an XML definition of a pattern of complex polyline

Element *Pattern* is used to define a pattern of the complex polyline. Pattern is a set of graphic and text elements that repeats itself along the polyline. For creating the pattern, four contained elements can be used: *Symbol*, *Text*, *Blank* and *Line*. *Symbol* defines graphic symbol, referenced by the attribute *idRef*. *Text* defines text in pattern. Its attribute *readOrientation* is used to setup how the text will be oriented – in the direction of the polyline or in the way so it can be easily readable. First contained element *Format* holds the text. The rest of the contained elements of the element *Text* are used to optionally setup the appearance of text element: its line and shadow color (*InteriorColor* and *BorderColor*, respectively), font (*Font*) and displacement from the polyline segment (*Offset*). *Blank* defines the length (via element *Length*) of an empty spacing in pattern. *Line* defines length (via element *Length*) and properties (via element *Simple*) of the line in pattern.

Fig. 5 shows the structure of element *Labels*, which is used to define labels of the polyline. There are six types of labels, which correspond to the contained elements of element *Labels*. *PreStaringLabel* and *PostEndingLabel* define label before the start and label after the end of polyline, respectively. *StaringLabel* and *EndingLabel* define label at the start and label at the end of polyline, respectively. *EverySegmentLabel* defines label which will be in the middle of every segment of polyline. The array of elements *PositionLabels* defines labels at the specified positions on polyline.

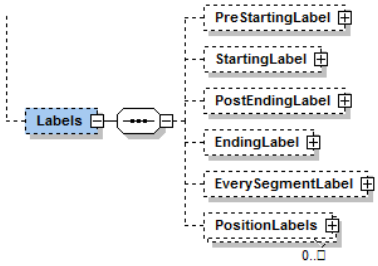


Fig. 5 – Structure of an XML definition of types of labels of complex polyline

All of these elements are based on XML complex type *labelType*, shown in fig. 6. Each label can consist of maximum 5 rows, which are defined by contained elements *RowAbove2*, *RowAbove1*, *Row0*, *RowBellow1* and *RowBellow2*. These elements are based on previously explained XML complex type *patternType*. The rest of the contained elements define the position of label (*Positions* and *Segment*), alignment of label (*Alignment*), orientation of label (*ReadOrientation*) and minimum distance between labels of the same type (*MinimumDistance*).

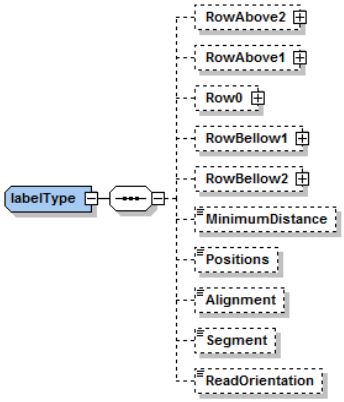


Fig. 6 – Structure of an XML definition of labels of complex polyline

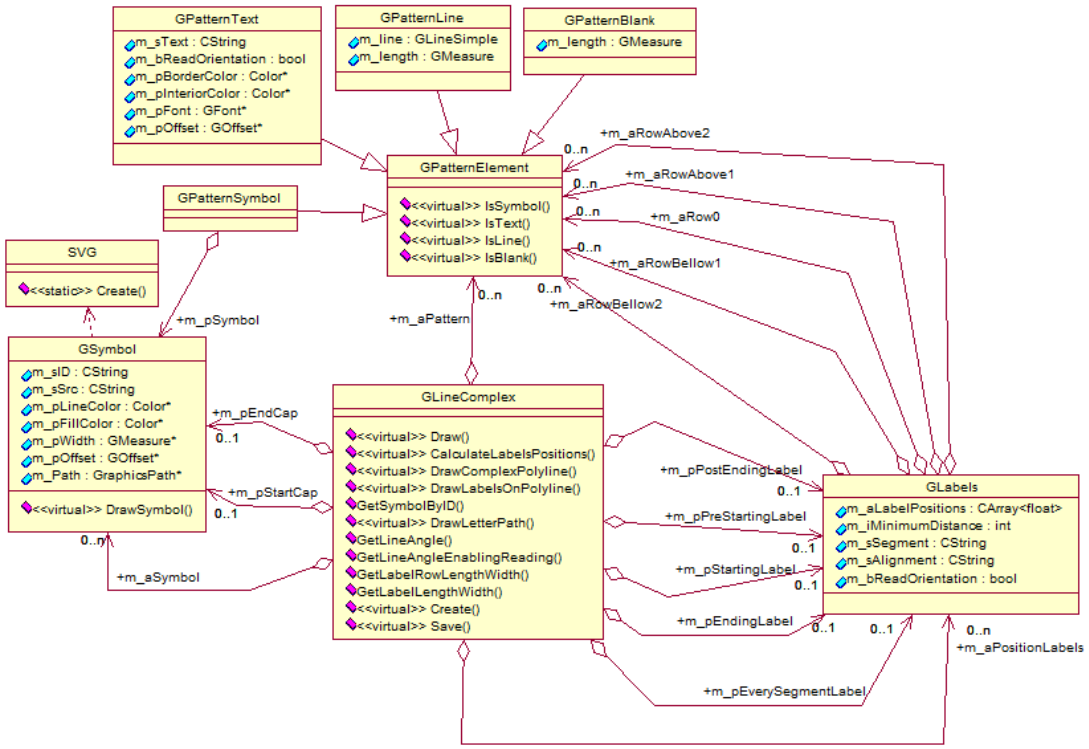


Fig. 7 – Logical model of class library for complex polylines

III. CLASS LIBRARY FOR COMPLEX POLYLINES

XML definition of decorations for complex polylines was the foundation for design and implementation of the class library aimed to interpret XML definition of decorations and draw the desired complex polylines. Similar approach, where XML definition and logical model of class library are closely related, is described in [6].

Fig. 7 shows UML logical model of the developed class library. The main tasks that this class library must perform are following:

- Reading data from a valid XML document and creating required objects.
- Calculating size, position and rotation of complex polyline decorations for submitted endpoints.
- Drawing complex polyline using GDI+ [7].

The class library for complex polylines is implemented in Visual C++. The main class is *GLineComplex*, which corresponds to XML complex type *ComplexLineType*. Attribute members of this class correspond to the contained elements of XML complex type *ComplexLineType* (Fig. 2). These members are set by calling the method *Create* that reads and interprets XML document with complex polyline definition. Method *Draw* does all necessary calculations and then draws complex polyline, which endpoints are passed to this method as one of the parameters.

Class *GSymbol* corresponds to the definition of XML element *Symbol* (Fig. 3). The graphic shape of each symbol is defined in appropriate SVG file [8]. Path to the SVG file is contained in the attribute member *m_sSrc*, which corresponds to XML attribute *src* of XML element *Symbol*. An attribute member *m_Path* contains GDI+ GraphicsPath object, obtained by calling the method *Create* of the class *SVG*. This method creates GDI+ GraphicsPath object from the submitted SVG file [7].

Abstract class *GPatternElement*, is a base class for four classes: *GPatternSymbol*, *GPatternText*, *GPatternBlank* and *GPatternLine*, which correspond to XML elements *Symbol*, *Text*, *Blank* and *Line*, respectively (Fig. 4). Class *GLabels* corresponds to XML complex type *labelType* (Fig. 5).

The main method of class *GLineComplex* is *Draw* method. Endpoints of polyline are passed to this method as one of its parameters. This method performs three essential tasks:

- Calculating *drawing parameters* (size, position and rotation) for defined decorations of complex polyline.
- Accepting decorations that satisfy certain conditions.
- Drawing complex polyline (with accepted decorations), using GDI+.

Fig. 8 shows pseudo code of *Draw* method. As it can be noticed, pattern of the polyline is drawn first. Drawing parameters for other decorations are calculated prior to drawing the pattern, but these decorations are drawn after drawing a pattern. In this way, pattern is drawn only where is visible (not where it is covered by other decorations), which in most cases gives better visual representation of polyline pattern.

Implemented approach also tackles the frequent problem of empty spaces in complex polylines. If the space left for drawing one of pattern's elements is slightly smaller than needed, it is better to draw this element than to leave this space empty. An implemented approach allows this element to be drawn without overlapping other, more important, decorations.

```
Set of decorations SI = {start cap, end cap, labels}.
Set of accepted decorations SA = {}.

foreach decoration D in SI
begin
   if (D exists)
   begin
      Calculate drawing parameters (params) of D.

      if ((Calculated polyline segment is big enough for D to be drawn on it) and
      (D can be drawn without overlapping with decorations from SA) and
      (if D is a label, minimum distance condition is satisfied))
      begin
         Put D in SA and memorize previously calculated drawing params of D.
      end if
   end if
end foreach

Calculate and draw pattern on parts of the complex polyline not reserved for
decorations from SA.
Draw decorations from SA, using previously calculated and memorized params.
```

Fig. 8 – Pseudo code for *Draw* method

We applied the class library presented in this paper to the GIS based application for graphical representation of military situation maps (see Fig. 9). All polylines (including polygons) shown in Fig. 10 are essentially complex polylines, drawn using the developed class library. Defining new types of polylines and modifying decorations of drawn polylines and polygons can easily be achieved due to the developed XML language. Part of the XML document that defines position and appearance of label "XX", of the polyline used to draw polygon in the top right corner of Fig. 9, is shown in Fig. 10.

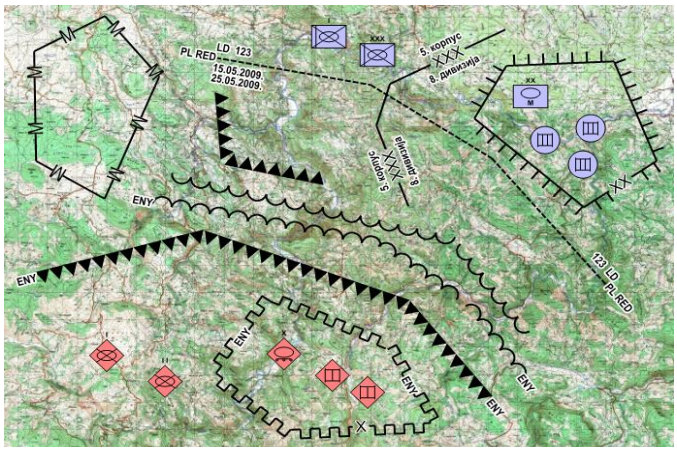


Fig. 9 – Using complex polylines for military situation maps

```
<PositionLabels>
   <Row0>
      <Text>
         <Format>XX</Format>
         <BorderColor>#FFFFFF</BorderColor>
         <InteriorColor>#000000</InteriorColor>
      </Text>
   </Row0>
   <Alignment>Center</Alignment>
   <Segment>South</Segment>
   <ReadOrientation>true</ReadOrientation>
</PositionLabels>
```

Fig. 10 –Defining label of complex polyline via XML

IV. Conclusion

Great importance of the visual portrayal of geographic features in GIS is undisputable. Implementing user-defined visual portrayal of geographic features requires the existence of a styling language that the user and GIS application can both understand.

In this paper we presented an approach for implementing visual portrayal of geographic features with polyline geometry (complex polylines). We developed XML styling language that enables simple and flexible way to define styling rules (decorations) of complex polylines. The structure of the this XML language was the foundation for designing the class library able to interpret XML definition of decorations and draw the desired complex polylines using GDI+.

In this paper we discussed implementation of the following decorations:

- *Start cap* – graphic symbol at the start of the polyline.
- *End cap* – graphic symbol at the end of the polyline.
- *Pattern* – element set repeated along the polyline.
- *Label* – element set located in defined locations on the polyline.

Element set specified for pattern and labels can contain one or more of the following elements:

- Graphic symbol (which shape is loaded from SVG file).
- Text with the specified properties.
- Line with the specified properties and length.
- Empty space with the specified length.

The core of the developed class library is an algorithm that calculates drawing parameters for specified decorations, eliminates some of the decorations in order to make the display distinctive and readable, and draws complex polylines.

References

[1] Worboys, M., and Duckham, M., *GIS: A Computing Perspective, Second Edition*, CRC Press, Boca Raton, FL, 2004.

[2] Chang, K., *Introduction to Geographic Information Systems, Third Edition*, McGraw-Hill, New York, NY, 2005.

[3] *The OpenGIS Abstract Specification, Topic 5: Features* (Version 5.0), document 08-126, Open Geospatial Consortium Inc., January 2009, http://www.opengeospatial.org/standards/as

[4] *OGC Reference Model* (Version 2.0), document 08-062r4, Open Geospatial Consortium Inc., November 2008, http://www.opengeospatial.org/standards/orm

[5] *Styled Layer Descriptor profile of the Web Map Service Implementation Specification* (Version 1.1.0), document 05-078r4, Open Geospatial Consortium Inc., June 2007, http://www.opengeospatial.org/standards/sld

[6] Milosavljević, A., Đorđević-Kajan, S., Stoimenov, L., *An Application Framework for Rapid Development of Web based GIS: GinisWeb*, Chapter 3 in Geospatial Services and Applications for the Internet (eds. J. T. Sample, K. Shaw, S. Tu, M. Abdelguerfi), Springer, 2008, pp. 49-72, ISBN: 978-0-387-74673-9.

[7] *Microsoft Windows GDI+*, Microsoft Corporation, http://msdn.microsoft.com/en-us/library/ms533798(VS.85).aspx

[8] *Scalable Vector Graphics (SVG) 1.1 Specification*, World Wide Web Consortium, January 2003, http://www.w3.org/TR/SVG11

WEB Model for Remote Data Acquisition Based on Google Maps GIS

Siniša Ranđić[1], Uroš Pešović[2], Željko Jovanović[3], Vlade Maksimović[4]

Abstract – **In this paper, we presented the Web model which can be used for storage and representation of data acquired from remote objects. These objects, which can be either mobile or stationary, send acquired data to Web server using GSM/GPRS networks. Data is stored in database, from where it's displayed to user in various forms. Standard table view is used to presents data history, while live data with position of remote station is shown on geographic map which is based on Google Maps API, global Geographic Informational System (GIS).**

Keywords – **Remote object, data acquisition, Google Map API, Web application.**

I. INTRODUCTION

This article is striving to show how is possible to remotely monitor measured values for different locations. Remote stations are used for acquisition of various environmental values, such as metrological, hydrological, air and water pollution and many other applications. For exact representation of some environmental property on observed area, wide network of these stations is necessary. Usually, only acceptable mean of connecting these remote stations into network is by some kind of wireless network, for example cellular GSM network.

Data acquisition system presented in this paper is based on GSM/GPRS modules, which can be used both for localized and distributed data acquisition systems. These kind of modules use cellular GSM networks which are widespread in every country around the world and cheap and easy to use.

Web application of our remote data acquisition system is based on Java platform. For geographic maps, we use Google Maps API. This global geographic informational system Geographical Information System (GIS) offers precise worldwide coverage which is continuously updated.

[1]Siniša Ranđić is with Technical Faculty in Čačak at University of Kragujevac, Sv. Save 65, Čačak 32000, Serbia, E-mail: rasin@tfc.kg.ac.rs

[2]Uroš Pešović is with Technical Faculty in Čačak at University of Kragujevac, Sv. Save 65, Čačak 32000, Serbia, E-mail: pesovic@yahoo.com

[3]Željko Jovanović is with Technical Faculty in Čačak at University of Kragujevac, Sv. Save 65, Čačak 32000, Serbia, E-mail: zjovanovici@gmail.com

[4]Vlade Maksimović is with Technical Faculty in Čačak at University of Kragujevac, Sv. Save 65, Čačak 32000, Serbia, E-mail: vlade.maksimovic@gmail.com

II. DATA ACQUISITION

Data acquisition system based on Telit GM862-GPS module which has GSM/GPRS capabilities. This device besides its GSM/GPRS functions, has built in GPS receiver with high accuracy of acquired position. Module also has one analog input and thirteen digital multipurpose I/O pins. Digital I/O pins beside its primary functions can be use as interface for digital sensors, which are connect to device with SPI or I^2C bus. Telit module itself can be use for localized data acquisition where sensors are not too far away from module, as shown in Figure 1.

Figure 1. Localized data acquisition system

In case for dislocated sensors, acquisition system must be base on remote I/O modules, which are connect together with some kind of industrial network, such as RS485 or Ethernet. These modules are capable to measure various analog signals such as voltage, current, temperature, etc., acquire, and update values of discrete signals. This configuration of measuring system has many advantages over centralized measuring systems. Analog signals could not travel long distances because voltage-drops on cables and noise which can be easily picked-up in noisy industrial environment. Also for every sensor, separate cable should be used which could be quite costly if the cables are long. Drawback of this configuration is low speed, which is limited by speed of the network which is used for communication between remote I/O.

Module used in our application is ADAM-4017, which is 8-channel A/D converter, with various input ranges for voltage and current signals. It is capable to communicate over RS485 network bus, using ASCII character protocol. Telit module is connected to this RS485 network by its serial port using RS232/RS485 converter module ADAM 4520. Layout of distributed data acquisition system used in our application is shown on Figure 2.

Figure 2. Distributed data acquisition system

Telit module can work in two operating modes, as a microcontroller and as modem. When working as microcontroller is control by program written in, Python script language, which enables full autonomy of the module. If module works as modem, it needs to be control by external microcontroller or computer using AT commands.

Basic principle of both data acquisition systems is based on continuous polling of analog, digital or remote I/Os and sending of acquired data to web server using GPRS protocol. Data transfer between module and server is carried by HTTP protocol, using GET request. One cycle of data acquisition and communication with server is shown in following code listing.

```
Reading ADAM module channels
Response: >+7.2111+7.2567+7.3125+7.1000
+7.4712+7.2555+7.1234+7.5678

Send data to: 91.187.132.16:8080
Socket: CONNECT
GET
CSLGPS/GISRazvojnoOkruzenje/NewPoint?Telit_id=3
&lon=02021.1222E&lat=4352.9047N&parm1=+7.2111&p
arm2=+7.2567&parm3=+7.3125&parm4=+7.1000&Status
=01 HTTP/1.0

Response: HTTP/1.1 200 OK
Server: Apache-Coyote/1.1
Content-Length: 0
Date: Fri, 23 Jan 2009 19:11:19 GMT
Connection: close

Response: NO CARRIER
```

III. WEB APPLICATION

As regards of the practical part of this project, its main goal is to represents the data in a way, which was explain, in the previous chapter. Normally, it is possible to pass parameters and values in some other way only if the specified URL http request made. The entire software solution has been realized using the open source J2EE [2] technology. For the realization of this solution, it is possible to use any programming language like PHP, but because of the large number of functions that provides JDK (Java Development Kit) the Java programming language be selected. One of the reasons why the J2EE technology was select is the possibility of extension to mention the system used on mobile phones in a J2ME applications. Completed software solution can be dividing roughly into two parts:

- Part that exchange data with hardware
- Part that presents results to the client

Part of the software used to exchange data with hardware, is the interface with the hardware support realized in described way. This part of the software has no visual interpretation and is execute only when the http request is pass to the Servlet by a device that forwards data. It has been develop as a Java Servlet, which call by hardware is using the HTTP GET request. To the appropriate data Servlet that performs the same processing and saves them in the MySQL database. The server confirm a successful reception of data and it is able to send the parameters which correct the way that device work. Since the Servlet can be accessed with the using HTTP requests from anywhere in the world, this way of communication gives this project one global level. There is ability of usage in different cities, countries and even continents. The preference listed in the realization of communication in addition to the great advantages is also a defect that openness a project to the attacks and simulation of some other persons who would like to emulate values of passed parameters. Therefore, it is very important to pay extra attention to the protection and checking the validity of the device that sends data. One of the first ways to protect validity of data is to assign to the device a unique identification number to be verifying on each received data. In this field, there are more possibilities for the improvement of security using certain cryptographic methods, which would preserve “the privacy” of data.

The J2EE platform is not able to handle all necessary requirements so implementation of the JavaScript language used in combination with AJAX [3] technology was required. The reason for this is that the dynamics of reading separate data from the database is required. Data in the database changes after a certain time, i.e. after each sending data from the hardware. His is the only way that allows tracking the desired settings for the desired station without any user action. There was no need to use some additional framework like hibernate, struts-and others. From the platform of the work, Servlets are use to process requests and JSP (Java Server Pages) for presentation of processed request. Some JSP pages are implementing using the Expression Language, which provides very easy access to the data stored in JavaBeans components.

As the container Apache Tomcat version, 6.0.16 was use. The reason for using Apache Tomcat Applications Server before the other variations of JBoss, Jrun, Glashfish Applications Servers, etc. is a very simple usage with the Eclipse development tool that was use during construction. As a geographical support in the project, the Google Maps API [6] was use. It provides great opportunities thanks to the entire globe coverage with satellite and aero-photo shots of high resolution. In addition, a very well developed mapping system that is in Western Europe and America developed to such an extent that charted the streets with their names in almost every inhabited place. Serbia is now cover only with the main motorways and roads network, but in

the near future it is expected a complete GIS coverage. The principle of working with Google Maps API is that a complete GIS system is on Google's server. The user passes the coordinates and parameters for the display, to the corresponding server that replies sending the required graphic content.

The main condition that must be fulfill to be able to work with Google Maps API is that in each Web page (in our case, the JSP page) that should be displayed must be imported script tag in the HTML head tag of the (JSP) page. The key is supplied free of charge by the Google and it is required to have a Gmail account (mail). In order to get key it is needed to enter the URL of the Web server on which Google Maps will be used. Google will generate a unique key and send it to the Google mail account. It is necessary to set the key to the previously described position.

```
<script
       src=http://maps.google.com/maps?
       file=api&v=2.x&key=

       ABQIAAAA6HHnyYiT2EVEKFG_RERksh
       QqcwwaKzD9OpoV4WAgC4lSGJ_ZoxSrievm
       QkroThPpsog3YNQBW33JsQ

       type="text/javascript">
</script>
```

After that, the image with the desired coordinates could be display. In the JavaScript that will start when the page is loading, it is necessary to define the element in the HTML of the page to display the image in it:

```
if (GBrowserIsCompatible ()) (
       var map=new GMap2
       document.getElementById (
       "map_canvas"),
       (size: new GSize (570,380)));
```

Google Maps has three types of the maps:

- Map - a base geographic-information map with drawed roads and street names
- Satellite - a satellite or aero-photo image record of the desired location
- Hybrid - a combination of previous two types

During the presentation of desired location, it is necessary to forward a map type and coordinates of the locations that want to be show to the server. Shown coordinates correspond to the central position of the image.

```
       map.setMapType (G_SATELLITE_MAP);
       map.setCenter (new GLatLng (latitude,
       longitude), 13];
```

In addition, user has the possibility of post-election of the map type that would be display. Optional it is possible to add zoom control scale that is on the left side of the map. Map clarity depends on the quality of recordings. Some of the more important locations have high-quality aero-photo shots with the possibility of zoom increasing up to 17 times. If necessary, map can be zooming without additional control with double-clicking the left mouse button, and un-zoom with double-clicking the right mouse button.

```
       map.addControl(newGMapTypeControl());
       map.addControl(newGSmallMapControl ());
```

The project was implementing as a Web application [7], which is located on the server in Computer Science Laboratory of Technical Faculty Cacak. The application consists of two parts. One provides currently monitoring of more stations and displays their view in real time with information about latitude and longitude as well as the values of parameters related to each of the station, which is show on the Figure 3. Since this system has a role in monitoring, the value of parameters for easier viewing, marker, which is a station on the map, may be green, yellow or red depending on the values of parameters, i.e. whether they allowed or not allowed critical range, respectively. In addition to transferring data, hardware has a role that does the processing of data, and as a parameter sends its status. Processing is not on the server from the simple reason that for n stations which monitoring m parameters would representing an m * n processing every few seconds. In addition to status of station by the colors of the marker, it is possible to see the value of all parameters by clicking on the marker.

In addition, there is a option to see the value for a specific station during the time interval as shown in the Figure 4. When reviewing the values shown in the table there is a visual assistance in the form of colors that represents status of certain station at a time.

It is important to note that Google Maps allows the execution of up to 500 000 requests during the day in a free license. If a higher number of requests are need for cases to monitor a large number of devices, Google Maps need to contact with the request to increase that number.

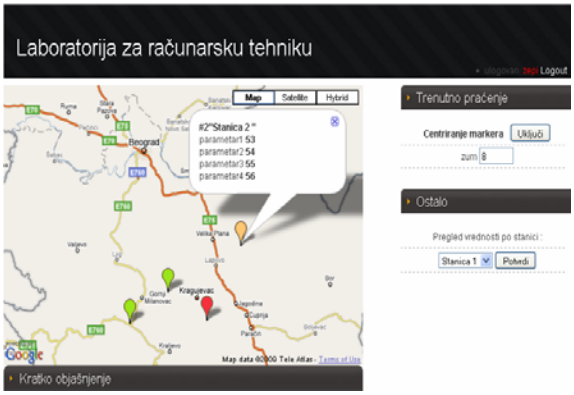


Figure 3. View the current follow-up station

Complete realization of solutions in the form of a web service provides one more level in the field of globalization. Is possible to develop users applications on any Web platform and the necessary information for presentation are receive from the web services, and they can be use in various ways. Building the web service is also available on the J2EE platform [4] required the use of Apache AXIS framework that is responsible for the

creation of WSDL and UDDI file that describes the structure of the web service and its location on which it would be find.

Figure 3. View values of parameters for the selected channel

When the structure of the service are known client application communicates with the web service using SOAP protocol which is independent of platform and represents the HTTP protocol which send the XML files.

IV. Conclusion

Since the project in the stage of development, we are facing with a lot more work on the selection of appropriate hardware, and the software, primarily to increase the reliability and security of data presentation. The plan is to develop specialized protocols for the exchange of data between devices and servers, in order to increase the rate and reduce the amount of data that is exchanged, which in the case of the HTTP protocol is charge with large overhead. It has enough space for development and implementation of Web service and possibly his connection with some of the companies which goal is to give and use a value of some parameters, of the measuring stations around the world.

References

[1] *GM862-GPS Hardware User Guide*, Telit Communications S.p.A., 1vv0300728 Rev. 8 -2007
[2] *ADAM 4000 series user manual,* Advantech Co., Ltd., Edition 10.7, May 2008
[3] Paul Perrone Venkata S.R. Tom Schwenk, J2EE Developer's Handbook**,** 2003
[4] Kris Hadlock, Ajax for Web Application Developers (Developer's Library) , 2006
[5] Steve Graham, Doug Davis, Building Web Services with Java, Sums publishing, second edition
[6] http://code.google.com/apis/maps
[7] http://csl.tfc.kg.ac.yu:8080/GISRazvojnoOkruzenje/

CDMA versus TDMA Transfer over Shared Bus

Tatjana R. Nikolić[1] and Mile K. Stojčev[1]

***Abstract* – CDMA interconnect is a new interconnect mechanism for SoC design. This paper presents how a multiprocessor system can benefit from the use of concurrent data transfers: CDMA interconnect has been adopted to implement the shared bus of a multiprocessor system. Unlike the conventional TDMA bus-based multiprocessor system that shows degradation in performance as the number of processing cores increases, the proposed CDMA bus-based multiprocessor system shows higher performance up to 188 percent for four processors. The obtained results show that increased data transfer latencies involved by CDMA data transfer are compensated by simultaneous master-slave transfers.**

***Keywords* – TDMA, CDMA, Bus, Wrapper, Latency.**

I. INTRODUCTION

Current VLSI design trends are shifting toward the System-on-Chip (SoC) incorporating several homogeneous or heterogeneous processing elements. As systems become more complex, the interconnection between subcomponents is becoming increasingly difficult and costly. Therefore, communication architecture has become the bottleneck for improving system-level performance, and it requires special design attention. To design an efficient interconnection network, the system's bandwidth and latency sensitivities should be taken into account together [1].

In general, on-chip communication architectures can be categorized into three main classes: point-to-point interconnects, bus, and network-on-chip [2]. However, there is no standard solution how to establish fast, flexible, efficient, and easy-to-design communication network to connect a large number of IP cores that have heterogeneous requirements.

Until now, most of interconnect networks in the SoC rely on parallel system bus, which apply time division multiple access, TDMA. The popularity and wide acceptance of the shared bus architecture is perhaps due to the fact that it is easy to adopt, well known among the computer industry, and also relatively inexpensive to implement [3]. In those systems, the bus masters perform read and write operations with slave memory or I/O modules [4]. However, bus-based architectures cannot scale up with an increasing number of components.

In this paper, the Code Division Multiple Access, CDMA, is used to implement the memory/peripheral shared bus of multicore systems consisting of multiple processing cores. The motivation of this paper is to explore how the standard memory/peripheral bus could be modified for higher system performance without increasing cost and complexity.

In the proposal, only bus lines that carry address and data signals are CDMA coded. We implemented a pair of master-slave wrapper as a component of CDMA, described in VHDL, using a Xilinx Spartan and Virtex. We also present results which show that benefits of involving CDMA coding relates both to decreasing a number of bus lines, and accomplishing simultaneous multiple master-slave connections. Convenient range indices R_W and R_R to determine data transfer rate for `Write` and `Read` operations in multiprocessor bus systems that use TDMA and CDMA data transfer techniques.

[1]Tatjana R. Nikolić and Mile K. Stojčev are with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: {tatjanas,stojcev}@elfak.ni.ac.yu

II. IP CORES INTERCONNECTING

Among the various conventional interconnection networks, multipoint shared bus is the most simple and cost-effective solution for small-scale shared-memory multiprocessor systems. A number of different standard busses are already being used in custom SoC designs [5]. Different IP blocks can be used to build a SoC as long as they are designed according to the standard bus specifications. Bus interfaces usually have groups of wires for command, address, and read or write data.

As we have already mentioned, a bus is a resource shared by multiple cores. Therefore, before using bus, cores must go through an arbitration phase. If there is a single arbitration for a request-response pair, the bus is called non-split. Typical shared bus memory/peripheral access scenario in a multiprocessors system is given in Fig. 1. In this case, the bus remains allocated to the master of the transaction until the response is delivered, even when this takes a long time. Alternatively, a split bus is released after the request to allow transactions from other masters to be initiated [1], [5].

The communication between the masters and slaves of bus-based systems is traditionally based on a TDMA. This interconnect is simple but not efficient in dealing with multiple data transactions simultaneously. The TDMA-based buses may have long request latency for concurrent requests due to the bus contention and arbitration delays [1].

The bus allocation in single master-slave connections is determined by an arbitration protocol. However, by using CDMA data transfer technique, which is based on a concept that each master-slave set can use its unique code subset, it is possible, over shared bus, to realize multiple master-slave data transfers, simultaneously. In order to implement this approach, we need to: i) modify the arbiter's hardware (by implementation of logic for solving the contention problem); and ii) involve a wrapper that is used as interface logic between the shared bus and IP connecting to it.

The traditional bandwidth-focused approach is not enough to improve overall system performance. To compare the `Read` and `Write` operation latencies, the bidirectional shared bus implementations based on a conventional TDMA and a proposed CDMA are discussed. Both busses have the same number of data and address lines with the same aggregate bandwidth per wire line. In a conventional TDMA bus, if two read requests arrive at arbiter at the same time, the second

request must stall until the first request finishes using the shared interconnect [6]. As the channel becomes narrower, the address and data latency of TDMA approach increases more rapidly. The goal of this paper is to illustrate that a TDMA based bus has a longer latency than the CDMA based bus.

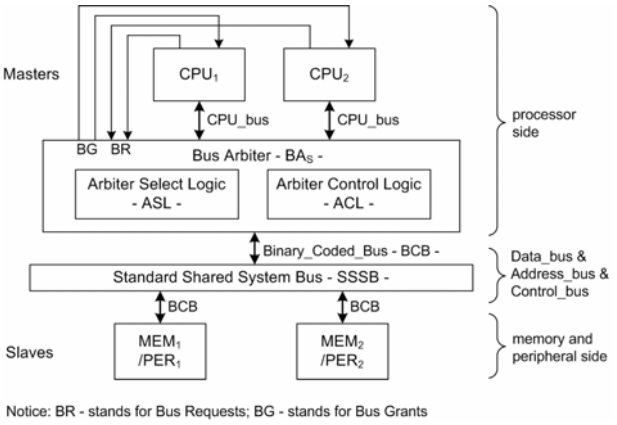


Fig. 1. A standard bus topology

III. Multiprocessor System Based on CDMA Shared System Bus

In order to eliminate variance of data transfer latency and complexity, an bus based interconnect which applies a CDMA technique is introduced in this paper. As one of the spread-spectrum techniques, the CDMA technique has been widely used in communication systems because it has great bandwidth efficiency and multiple access capability. The CDMA technique applies a set of orthogonal codes to encode the data from different users before transmission in a shared communication media. Therefore, it permits multiple users to use the communication media concurrently by separating data from different users in the code domain [7].

When a CDMA technique is implemented on standard multiprocessor system presented in Fig. 1, we obtain a scheme given in Fig. 2. In order to make the discussion clear, we will assume that the system consists of two local computers and two shared memory or peripheral modules. By comparing the structures sketched in Fig. 1 and 2, we catch sight of the following three basic differences:

1. The Standard Shared System Bus, `SSSB` (see Fig. 1) is substituted with a CDMA Shared System Bus, `CSSB` (Fig. 2).

2. The Bus_Arbiter, `BA`, given in Fig. 2 is realized using the following two building blocks: Arbiter_Switching_Logic, `ASL`, and Arbiter_Control_Logic, `ACL`. According to the implemented algorithm for bus priority assignment the ACL's output `Switch`, defines which CPU bus will drive a corresponding BCB_1 or BCB_2, respectively.

3. Two types of bus wrapper are involved. The master bus wrapper, `BW_CPU`, converts data and address signals present at BCB_1 and BCB_2 into CDMA coded bus signals and conversely. The slave wrapper, `BW_MEM/PER`, connects memory or peripheral IP blocks to the `CSSB`. In all standard solutions based on AMBA bus [8], CoreConnect, STBus, etc., the data transfer protocol over system bus is mainly defined by the timing (signaling) of the `Control_bus`. In order to improve IP core reusability and be protocol compliant with any standard on-chip bus we involve a CDMA coding for data and address signal lines, while signal lines that belong to the `Control_bus` remain unchanged.

In general, the hardware structures and principles of operation of the `BW_CPU` and `BW_MEM/PER` are similar. The following six modules, presented in [9], are parts of the bus wrapper logic: 1) Bus Wrapper Control Unit, `BWCU`; 2) Control Protocol Transfer Block, `CPTB`; 3) CDMA Data Encoder/Decoder, `DED`; 4) CDMA Address Encoder, `AE`; 5) Configuration Register, `CR`; and 6) Clock Generator, `CG`. The specifity of the wrapper structure, sketched in Fig. 2, relates to the implementation of CDMA encoder and decoder blocks.

IV. CDMA Coded Bus Transfer Operation

The operation of a CDMA coded wrapper-based bus we will explain on execution of CPU `Read` and `Write` cycles. Figure 3 shows a case when the `CPU` IP core initiates a `Read` cycle. The `BA` logic acts as a transparent block, i.e. it maps standard CPU bus signals into binary coded bus signals.

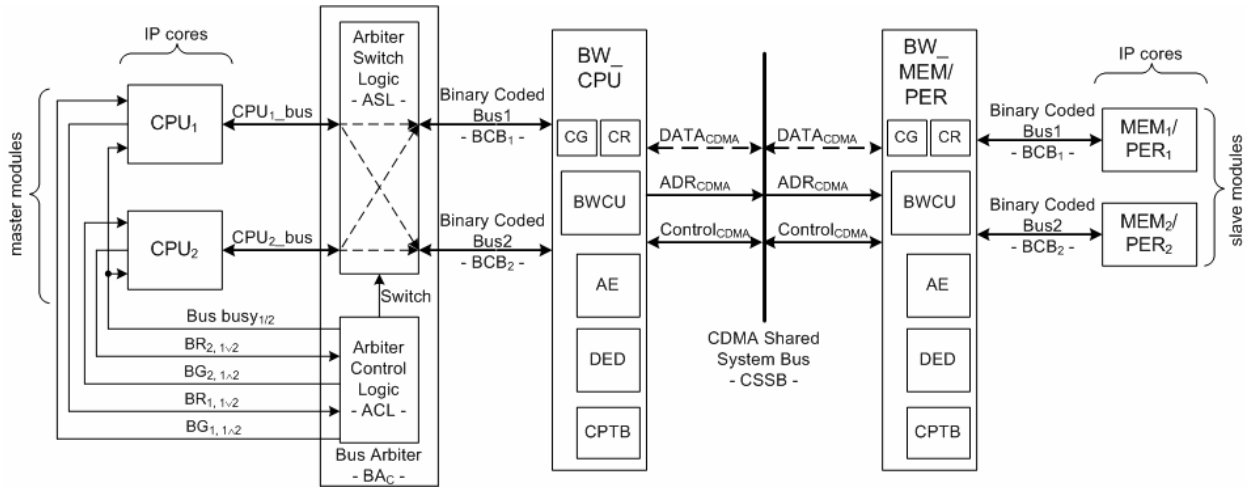


Fig. 2. Multiprocessor system based on CDMA transmission technique

The Read cycle begins at instant t_0. At t_1 the CPU IP core sets its address and status lines at valid states and, after passing through BA logic, the BW_CPU accepts them and asserts a signal RDY_activate. It signals to the CPU IP core to insert wait states. In addition, the BW_CPU converts address from binary to CDMA form and sends the address via CSSB. The BW_MEM/PER wrapper decodes CDMA coded address and at instant t_2 drives the MEM/PER IP core with a binary coded address via BCB_ADR bus. After the access time, t_{ACC}, has expired, at instant t_3, a MEM/PER IP core sets its data lines at valid states. The BW_MEM/PER wrapper encodes a binary coded data into a CDMA coded and forwards them via a CSSB back to BW_CPU. The BW_CPU decodes them and at t_4 passes CPU_DATA signals to the CPU IP core. At instant t_4 the BW_MEM/PER generates a signal RDY_deactivate and signals to the CPU IP core to deassert the wait state period, t_w. At t_5 the CPU IP core terminates its Read cycle.

The timing of a Write cycle is simpler in respect to the Read cycle. The main difference is that, during a Write cycle, at instant t_1 the CPU IP core generates an address, and after that, it generates a valid data.

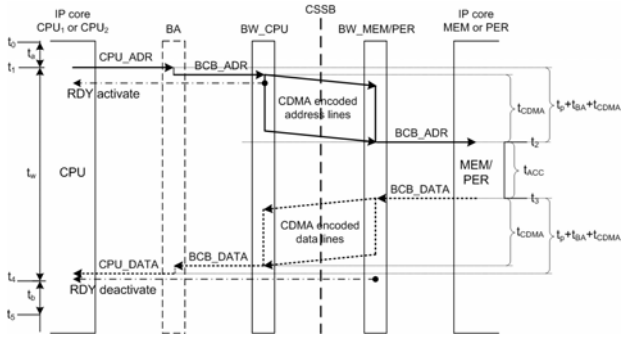


Fig. 3. Signaling scenario of a Read processor cycle

V. Performance Metrics

In order to evaluate latencies of a system based on standard Binary Coded Bus, BCB (Fig. 1), in respect to CDMA Shared System Bus, CSSB (Fig. 2), we will assume the following:

a) Multiprocessor systems given in Fig. 1 and 2 are composed of k processors (masters), and master-slave data transfer rates, for both kinds of systems, will be considered.

b) Binary Coded Bus (BCB_1 or BCB_2) consists of 32 bit address bus, 32 bit data bus, and Control bus (see Fig. 2).

c) The velocity of signal propagation over bus wires is $2*10^8$ m/s, and the master-slave distance is, in average, 30 cm. Accordingly, the signal propagation delay is t_p = 1,5 ns.

d) Time delays involved by both types of arbiter logic, given in Fig. 1 and 2, are identical, $t_{BAS} = t_{BAC} = t_{BA}$ = 10 ns;

e) Access times to all slave modules are identical, $t_{ACCMEM} = t_{ACCI/O} = t_{ACC}$ = 30 ns;

f) An address decoder is installed in each slave module of Fig. 1, only, and involves time delay t_D. In our case t_D = 3 ns.

g) The total time delay involved by a CDMA coding and decoding process (t_{CDMA} - see Fig. 3), for different spreading code sizes is generated by a CAD tool.

h) According to a bus allocation policy, bus requests in a multiprocessor system given in Fig. 1 will be served in a sequential manner. Contrary, for a multiprocessor system pictured in Fig. 2 all requests will be served simultaneously.

In order to evaluate data transfer latency for both multiprocessor systems (Fig. 1 and 2) we will consider first a timing related to Write, and Read operation. We will assume that $t_{BA} \cong 7 \cdot t_P$, $t_D = 2 \cdot t_P$, and $t_{ACC} = 20 \cdot t_P$ is valid.

A. Standard system (Fig. 1): When there is single issue for master-slave data transfer, Write and Read operation will be served for

$$T_{TS1W} = t_P + t_{BA} + t_D + t_{ACC} = 30 \cdot t_P \quad (1)$$

$$T_{TS1R} = 2*t_P + 2*t_{BA} + t_D + t_{ACC} = 38 \cdot t_P \quad (2)$$

When all k master modules issue requests for data transfer, the access times for Write and Read operation will be

$$T_{TSW} = k * T_{TS1W} = k * 30 \cdot t_P \quad (3)$$

$$T_{TSR} = k * T_{TS1R} = k * 38 \cdot t_P \quad (4)$$

B. CDMA based system (Fig. 2): In this case all k master-slave connections, during Write and Read operation, are performed simultaneously for

$$T_{TCW} = t_P + t_{BA} + t_{CDMA} + t_{ACC} = 28 \cdot t_P + t_{CDMA} \quad (5)$$

$$T_{TCR} = 2*t_P + 2*t_{BA} + 2*t_{CDMA} + t_{ACC} = 36 \cdot t_P + 2*t_{CDMA} \quad (6)$$

where t_{CDMA} is a time interval generated by a CAD tool.

Form now a performance metric, Q, as a product of the following three parameters: a) a number of bus lines, L; b) total access time, T; and c) communication bandwidth, B:

$$Q = L \cdot T \cdot B \quad (7)$$

For the standard system (Fig. 1) and CDMA bus based system (Fig. 2), respectively, we have:

$$Q_{1X} = L_1 \cdot T_{TSX} \cdot B_1 \quad (8)$$

$$Q_{2X} = L_2 \cdot T_{TCX} \cdot B_2 \quad (9)$$

where a subscript X is W(R) for Write(Read) operation.

Let define now a data transfer ratio for Write and Read operation, R_W, and R_R, respectively, if $B_1 = B_2$ (i.e. an equal amount of information is transferred by both systems), as

$$R_W = \frac{Q_{1W}}{Q_{2W}} = \frac{L_1 \cdot T_{TSW}}{L_2 \cdot T_{TCW}} \quad (10)$$

$$R_R = \frac{Q_{1R}}{Q_{2R}} = \frac{L_1 \cdot T_{TSR}}{L_2 \cdot T_{TCR}} \quad (11)$$

We use R_W and R_R as convenient range indices by which we will evaluate the latency of the simultaneous CDMA bus transfer in respect to a sequential TDMA bus transfer.

VI. Results

An on-chip interconnect scheme based on CDMA technique is proposed here and its performance related to data transfer latency are evaluated. The wrapper logic was described at RTL using VHDL. For synthesis, routing, and mapping a Xilinx development CAD tool ISE 9.1i was used. The wrapper was implemented on FPGA devices from Spartan2, Virtex4, and Virtex5 series. The results generated by a CAD tool relate to a signal propagation time which corresponds to the total latency of a communication channel (the time interval t_{CDMA} in Fig. 3).

TABLE I
DATA TRANSFER RATIO FOR WRITE AND READ OPERATIONS

Target device from series	Number of masters	TDMA based system			CDMA based system			Data transfer ratio, R_W	Data transfer ratio, R_R
		Number of lines	Access time, T_{TSW}	Access time, T_{TSR}	Number of lines	Total access time, T_{TCW}	Total access time, T_{TCR}		
Spartan3E	2	64	60 t_p	76 t_p	16	169 t_p	318 t_p	1.42	0.95
	4	64	120 t_p	152 t_p	18	311 t_p	602 t_p	1.37	0.90
	8	64	240 t_p	304 t_p	20	593 t_p	1166 t_p	1.29	0.83
	16	64	480 t_p	608 t_p	22	1159 t_p	2298 t_p	1.20	0.77
VirtexE	2	64	60 t_p	76 t_p	16	164 t_p	308 t_p	1.46	0.99
	4	64	120 t_p	152 t_p	18	300 t_p	580 t_p	1.42	0.93
	8	64	240 t_p	304 t_p	20	572 t_p	1124 t_p	1.34	0.86
	16	64	480 t_p	608 t_p	22	1116 t_p	2212 t_p	1.25	0.80
Virtex5	2	64	60 t_p	76 t_p	16	88 t_p	156 t_p	2.72	1.94
	4	64	120 t_p	152 t_p	18	148 t_p	276 t_p	2.88	1.96
	8	64	240 t_p	304 t_p	20	268 t_p	516 t_p	2.86	1.88
	16	64	480 t_p	608 t_p	22	508 t_p	996 t_p	2.74	1.78

Notice: t_p = 1.5 ns; Column referred as Number of lines includes address and data bus lines.

Table I reports the results which relate to data transfer latencies for standard TDMA and CDMA bus based systems. Multiprocessor systems composed of 2, 4, 8, and 16 master and slave modules were considered. We assumed that all master modules simultaneously access to different slave modules, i.e. a contention problem was omitted.

By analyzing the results given in Table I we can conclude the following:

1) Spartan3E and VirtexE series: transfer rates for `Write` operations are faster for CDMA based bus architecture in respect to TDMA bus architecture. Namely, for Spartan3E series `Write` operations are from 42% (for 2 processors) to 20% (for 16 processors) faster, while for VirtexE series are from 46% to 25% faster. Contrary, transfer rates for `Read` operations TDMA bus architecture is faster from 5.2% to 30% for Spartan3E series, and from 1% to 25% for VirtexE series.

2) Virtex5 series: for both `Write` and `Read` operations performance related to data transfer rates for CDMA based bus in respect to standard bus architecture are superior. Namely, `Write` operations are from 188% to 172% and `Read` operations are from 96% to 78% faster.

Let note that the results given in Table I are illustrative only. Other factors such as bus allocation policy, physical bus wiring limitations, propagation delay involved by the bus arbiter, etc. have to be considered in real applications.

In general, the results given in Table I show that: a) involvement of CDMA bus based system is a trade-off problem between decreased number of bus lines and communication time and it may appeal to applications where bus size (wiring) reduction is imperative; and b) increased data transfer latencies involved by CDMA data transfer are compensated by simultaneous master-slave transfers.

VII. CONCLUSION

With aim to combine the positive attributes of smaller address and data buses as well as to achieve control bus compliance with existing bus conversion protocols (such as AMBA to BVCI) we have proposed a CDMA encoding technique both for address and data buses, but not for a control bus. At behavioral level a wrapper structure was described using VHDL code. For synthesis, routing, and technology mapping a Xilinx development CAD tool ISE 9.1 was used. The wrapper was implemented on FPGA devices from Spartan and Virtex series. We involved convenient range indices called R_W and R_R to determine data transfer rate for `Write` and `Read` operations in multiprocessor bus systems that use TDMA and CDMA data transfer techniques. In addition, let note that increased data transfer latencies involved by CDMA data transfer are compensated by simultaneous master-slave transfers.

REFERENCES

[1] J. Kim, et. all, "A Cost-Effective Latency-Aware Memory Bus for Symmetric Multiprocessor Systems" *IEEE Transactions on Computers*, vol. 57, no. 12, pp. 1714-1719, May 2008

[2] T.S.T. Mak, et. all, "On-FPGA Communication Architectures and Design Factors", *In: Proc. Internat. Conference on Field Programmable Logic and Applications*, Madrid; 2006, pp. 1-8

[3] S. Dey, et. all, "Design of Communication Architectures for High-Performance and Energy-Efficient Systems-on-Chips", *In: Jerraya AA, Wolf W, editors. Multiprocessor Systems-on-Chips*, San Francisco: Morgan Kaufman; 2005, pp. 187-222

[4] T. D. Richardson, et. all, "A Hybrid SoC Interconnect with Dynamic TDMA-Based Transaction-Less Buses and On-Chip Networks", *19th International Conference on VLSI Design*, January 2006, pp. 657-664

[5] L. Benini, G. De Micheli, "Networks on chips: A new paradigm for component based MPSoC design", In: Jerraya AA, Wolf W, editors. Multiprocessor Systems-on-Chips, San Francisco: Morgan Kaufmann, 2004, p. 187-222

[6] J. Kim, et. all, "Design of an Interconnect Architecture and Signaling Technology for Parallelism in Communication", *IEEE Trans on VLSI Systems* 2007, 15(8), pp. 881-94

[7] X. Wang, et. all, "Applying CDMA Technique to Network-on-Chip", *IEEE Transactions on VLSI Systems*, Vol. 15, No. 10, October 2007, pp. 1091-1100

[8] ARM AMBA Specification and multi-layer AHB specification (rev 2.0), [Online], available at http://www.arm.com, July 2008

[9] T. Nikolić, M. Stojčev, "CDMA Coded Wrapper-Based SoC Interconnect", *ICEST 2008*, Vol. 2, No. 1, June 2008, Nis, pp. 399-402

Visualization and Structure Learning of Gene Regulatory Networks using Bayesian Networks

Blagoj Ristevski[1], Suzana Loskovska[2]

Abstract - **The cell functions and development are regulated by complex networks of genes, proteins and other components by means of their mutual interactions. These networks are called gene regulatory networks (GRNs). The gene regulatory networks are used to reveal the fundamental gene regulatory mechanisms, to determine the reasons for many diseases and interactions between drugs and their targets, to produce a clear and comprehensible notion for cell regulation,. The introduction of experimental technologies such as microarrays and chromatin immunoprecipitation ChIP-chip, has provided a large number of available datasets related to gene expression and transcription factors (TFs). These datasets are basis for further analysis to reveal the gene regulation mechanisms. We implemented and visualized the dynamic Bayesian network which is able to cope with missing data and can include a prior knowledge about transcription factors. Also, we describe the obtained results and survey the common structure learning algorithms for learning of GRN's structure.**

Keywords - **Gene regulatory networks, Bayesian network, Bioinformatics.**

I. INTRODUCTION

The living cells during their life span carry out many different tasks controlled by the cell genome which is encoded in the DeoxyriboNucleic Acid (DNA) molecule. The genes are transcribed into messenger RiboNucleic Acid (mRNA), and then translated in proteins.

The necessity to generate, analyze and integrate the large scale expression data led to the development of microarray technology [2]. Genes and their products – proteins work coordinately in complex networks. The cell functions and development are regulated by complex networks of genes, proteins and other components by means of their mutual interactions. These networks are called gene regulatory networks (GRNs). The proteins which activate or inhibit the transcription of the other genes are called transcription factors (TFs). Transcription factors are important components in gene regulatory networks. The GRNs are commonly used to study influences between cell components because they provide a clear and understandable notion for cell regulation as well as reveal the fundamental gene regulatory mechanisms and find out the reasons for many diseases.

[1]Blagoj Ristevski is with Faculty of Administration and Information Systems Management, Department of Information Systems Management, St. Kliment Ohridski University – Bitola, Partizanska bb (Kompleks kasarni), 7000 Bitola, Macedonia. e-mail: blagoj.ristevski@uklo.edu.mk

[2]Suzana Loskovska is with Faculty of Electrical Engineering and Information Technologies, Skopje, Department of Computer Science and Computer Engineering, Ss. Cyril and Methodius University – Skopje, Macedonia, e-mail: suze@feit.ukim.edu.mk

Many destructive diseases such as cancer are related to different genetic disorders. Modeling of the GRNs represents one of the most powerful techniques to describe the fundamental cellular mechanisms and associated intracellular and intercellular processes. The goal of many researches which include the experimental and simulating methods is by studying the GRNs to reveal therapeutic and prognostic relevant knowledge about many diseases.

Besides microarray technology, the introduction of other experimental technologies such as chromatin immunoprecipitation ChIP-chip, provides a lot of available datasets related to gene expression and transcription factors. ChIP-chip provides an insight into interaction between transcription factors and promoter region of the gene when it is combined with microarray analysis [5]. These data types are basis for further analysis and they are means of revealing the gene regulation mechanisms and essential knowledge about cell processes on genomic and molecular level.

To represent GRNs, more models are used, such as state space model, Bayesian networks, dynamic Bayesian networks, Boolean networks, linear and nonlinear differential and difference equations model, fuzzy logic model, information theory model, and others models.

Finding out the most reliable and accurate structure of GRNs from high dimensional microarray data is a machine learning problem known as structure learning of graphical models. A subset of the data is used for model fitting and the residual data for the model validation [3]. Cross-validation methods are very useful for validation and training of regulatory networks. But, the obtained networks which fit the best to the training set are overtrained [4]. Such overfitted networks lose their ability to generalize the suitable networks for the data outer the training set.

The remainder of this paper is organized as follows. In the second section we present the models based on Bayesian networks and their advantages and disadvantages. The consequent section describes the dynamic Bayesian networks. A survey of structure learning algorithms of reconstructed gene regulatory networks is given in the Section 4. The following section is devoted to the reconstructed GRN using Bayes Net Toolbox (BNT). The concluding remarks are given in the last section.

II. BAYESIAN NETWORKS

The Bayesian networks are a special case of graph models consisted of two components and based on statistical principles [6]. The first part is a directed acyclic graph *G=(V,E)* where $V = \{1, \cdots, n\}$ is a set of nodes and *E* is a set of edges. Each node $i \in V$ refers to random variable $x_i \in x$ that represents the gene expression - in the case of GRNs. The set

of edges corresponds to the conditional dependence among nodes. The second part of network is a set of conditional probability distributions that describe the conditional probability of each variable (the gene expression).

If $x = \{x_1, \cdots, x_n\}$ denotes a set of random variables, G - structure of graph, θ - set of parameters the joint probability distribution is given by Eq. 1.

$$p(x) = \prod_{i=1}^{n} p\left(x_i \middle| x_{\{1,\cdots,i-1\}}, \theta, G\right) \quad (1)$$

If the pa_i denotes the parent nodes of the node x_i that means the state of each variable x_i depends on the states of its parent pa_i (Eq. 2):

$$p(x) = \prod_{i=1}^{n} p\left(x_i \middle| pa_i, \theta, G\right) \quad (2)$$

Bayesian networks are suitable to show regulatory mechanisms between network components quantitatively as well as qualitatively. The qualitative description of regulatory mechanisms refers to presence/absence of an edge between network nodes whereas quantitative representation is made by a set of conditional probability distributions.

The modeling of the gene regulatory networks is made by structural and parameter learning. The goal of structural learning is to determine topology of network. For a given network structure, the parameter learning includes parameter estimation of unknown model for each gene. This is performed by determination of conditional dependencies between network components.

Bayesian networks can deal with noisy and stochastic nature of gene expression data and with incomplete knowledge about system. The small number of data points (samples) and the big number of genes are common problems for learning Bayesian networks. Another disadvantage is that this model cannot model feedback connections, which exist in the gene regulatory networks.

III. DYNAMIC BAYESIAN NETWORKS

To overcome the problems of Bayesian networks, **dynamic Bayesian networks** are used to model gene regulations. The dynamic Bayesian networks are capable to deal with stochastic variables, time series gene expression data, to include prior knowledge, feedback loops and to handle missing values and hidden variables. The hidden nodes (variables) can capture effects that cannot be directly measured in a microarray experiment.

The joint probability distribution is given by Eq. 3, where x_t^i is the i-th node at time t.

$$p\left(x_t \middle| x_{t-1}\right) = \prod_{i=1}^{n} p\left(x_t^i \middle| pa(x_t^i), \theta, G\right) \quad (3)$$

Dynamic Bayesian networks can apply and learn for real biological data, and there is a relationship to Hidden Markov Model, Boolean networks, stochastic Boolean networks, dynamic Bayesian networks with continuous state and other probabilistic models. The hidden nodes provide a way of linking similar data types and analysis of other network parameters. When some dependency exists between variables in the network, the hidden node can model that dependency.

Relatively low prediction accuracy and excessive computational time are two problems which reduce the performance of the dynamic Bayesian network model. To overcome these problems it is suggested a **dynamic Bayesian network approach** which limits the potential regulators to those genes with either earlier or simultaneous expression changes (up- or down-regulation) in regard to their target genes. Then, the genes with either earlier or simultaneous expression changes are assigned as possible regulators of those genes with a later expression change.

IV. STRUCTURE LEARNING

To choose the most appropriate network for a given dataset it is necessary to carry out a validation of the modeled networks. The precision and the reliability of models predictions are commonly examined in respect with input experimental data during the process of model validation. Structure learning of Bayesian network consists of finding a directed acyclic graph (DAG) that best fits the dataset. The structure learning performs by means of scoring function that evaluates how well the DAG explains the data and then to search for the best DAG that optimized the scoring function. There are two approaches for structure learning: constraint-based and search-based approach. In the constraint-based approach, the algorithm starts with a fully connected graph and removes edges.

The number of directed acyclic graphs as a function of the number of nodes G(n) is super-exponential dependent (Eq. 4)

$$G(n) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} G(n-k) \quad (4)$$

After modeling of GRNs, the model which provides a good fit to the data should be selected. The criteria such as Bayesian score [8], maximal likelihood, Bayesian information criterion, minimum description length are used for model validation of gene regulatory networks. Model is validated by the *n*-cross validation too, where the input set is divided to *n* parts. The *n*-1 parts are used as a training set, and the remaining as a test set.

Let M denotes the structure of a dynamic Bayesian network, D is the data set, $P(M)$ is the prior probability of the network structure and $P(D|M)$ is its marginal likelihood. θ is a parameter vector of the conditional probability distributions. The marginal likelihood is an average of the likelihood $P(D|M,\theta)$ over all possible parameters associated to the network. The Bayesian score is based on the marginal likelihood of the data is defined as follows:

$$P(D \mid M) = \int P(D \mid M, \theta) P(\theta \mid M) d\theta \quad (5)$$

and provides a matching between model complexity and the data size.

The goal of the Minimum Descriptive Length criterion (MDL) is to provide an optimal matching between the precision of the data fitting and the complexity of network

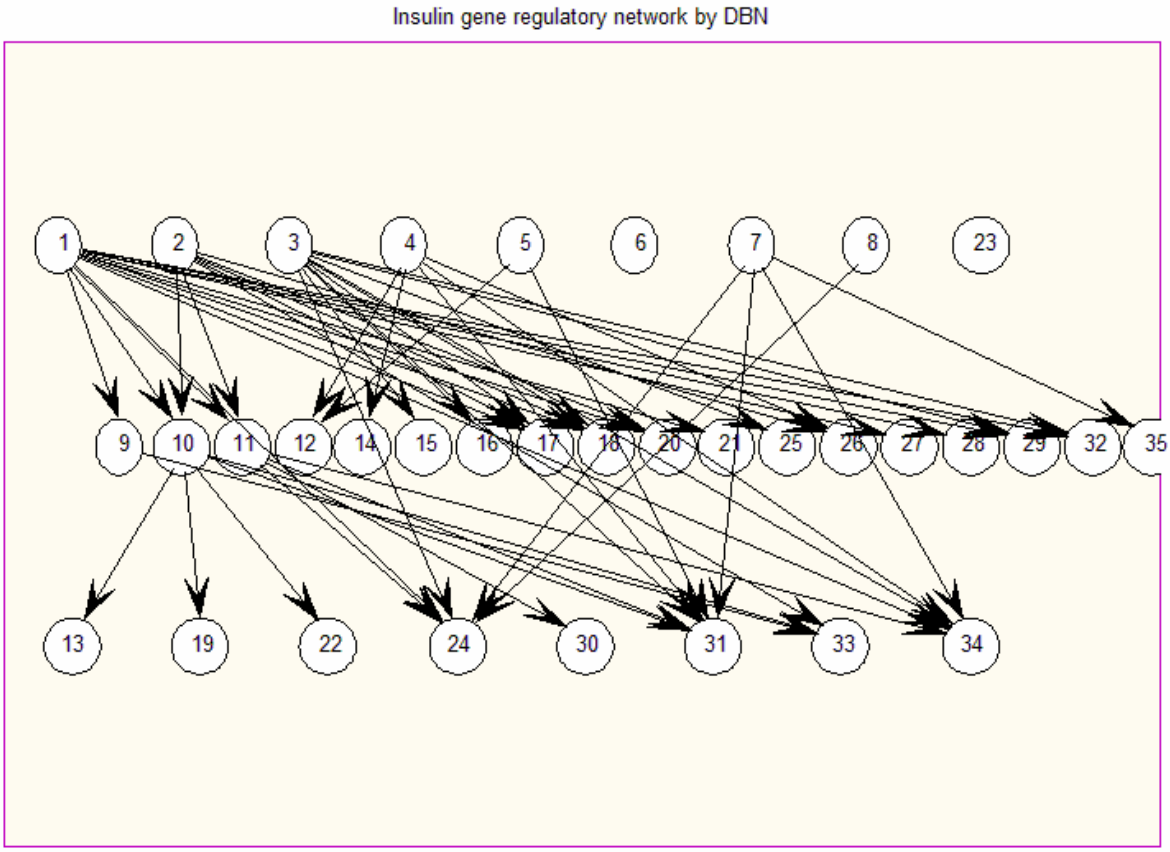


Figure 1. Insulin gene regulatory network with 35 genes.

model. The MDL score consists of model and data set encoding, hence the MDL criterion L is a sum of the network coding length L_M and data set coding lengths given by the following equation:

$$L = L_M + \sum_{j=1}^{m-1} H(x_{j+1}|x_j) = L_M - \sum_{j=1}^{m-1} \log(p(x_{j+1}|x_j)) \qquad (6)$$

where $H(x_{j+1}|x_j)$ is the state transition conditional entropy, $p(x_{j+1}|x_j)$ is the transitional probability and m – number of sample points [9].

At the Bayesian Information Criterion (BIC) n denotes the sample size, k – numbers of parameters and RSS is the residual sum of squares from estimated model. If it is assumed that the distribution of the models errors is normal, then the aim is to minimize BIC, given by Eq. 7.

$$BIC = n \log\left(\frac{RSS}{n}\right) + k \log n \qquad (7)$$

The Maximum Likelihood (ML) criterion expresses the likelihood $L(\theta)$ as a function of the unknown vector parameter θ and aims to find the parameters of all possible values which maximizes $L(\theta)$. The solution can be a function of one or many parameters and often this problem is nonlinear optimization problem.

To obtain a good balance between the accuracy of data fitting and the complexity of gene network models, some of above mentioned criteria are applied. Their application infers GRNs with significant biological performances of inferred interactions between networks elements.

Because of super-exponential dependency between number of directed acyclic graphs and number of nodes, the local or global search algorithms are used (K2, Hill-climbing, MCMC, Structured EM).

In addition to the search procedure, the scoring function should be specified.

The K2 algorithm is a greedy search algorithm that works as follows. In the beginning each node has no parents, it then adds incrementally that parent whose addition most increase the score of the resulting structure. When the adding of no single parent can increase the score, the adding parents to the node stops. It attempts to select the network structure that maximizes the posterior probability of network given the experimental data [1].

Hill-climbing starts at a specific point in space, considers all nearest neighbors, and moves to the neighbor that has the highest score. If no neighbors have higher score than the current point (a local maximum is reached), the algorithm stops.

To evaluate the obtained results, the inferred network structure should be compared with the reference network. The Receiver Operator Characteristics (ROC) curves are used to evaluate inferred network structure quantitatively [7]. The ROC curve is a chart of the ratio between sensitivity and (1-specificity), where sensitivity corresponds to proportion of

actual positives edges which are correctly identified and specificity is proportion of negatives edges which are correctly identified. The ROC curves can be summarized by computing the AUC (Area Under the ROC Curve).

V. RESULTS

To implement dynamic Bayesian gene regulatory network we use insulin gene expression data, Bayes Net Toolbox [10] and Bayesian Network Structure Learning – software package [11]. The obtained gene regulatory network is shown on Fig. 1. From insulin data, we specified the interaction between genes (we take 35 genes) which is in regard to interactions transcription factors - target genes.

VI. CONCLUSION

Besides the amount of microarray data sets, the reconstructing of gene regulatory networks is still a hard and challenging problem. Bayesian networks especially dynamic BNs are powerful tools which provides elucidation of interaction among genes. The main shortcoming of utilized Bayes Net Toolbox and Bayesian Network Structure Learning is their limitation to cope with networks with small number of nodes, especially for structure learning. For large gene networks with hundreds and thousands of genes, these tools are not advisible.

Future tools should be able to deal with large gene networks.

REFERENCES

[1] X.-wen Chen, G. Anantha and X. Wang, An effective structure learning method for constructing gene networks, *Bioinformatics*, Vol. 22 no. 11, 2006, pp 1367-1374.

[2] S. Qing Ye, *Bioinformatics A Practical Approach*. Chapman & Hall/CRC, Taylor &Francis Group, Boca Raton, 2008.

[3] P. Baldi and S. Brunak, *Bioinformatics:The Machine Learning Approach*.The MIT Press, Cambridge, 2001.

[4] H. Lahdesmaki et *al*., On Learning Gene Regulatory Networks Under the Boolean Network Model, *Machine Learning*, 52, 147-167, 2003.

[5] T. I. Lee et *al*., Chromatin immunoprecipitation and microarray-based analysis of protein location, *Nature protocols*, Vol.1 No.2, 2006.

[6] N. Friedman and M. Goldszmidt, Learning Bayesian Networks with Local Structure, *Proceedings of the NATO Advanced Study Institute on Learning in graphical models*, 1998.

[7] L. Kaderali and N. Radde, Inferring Gene Regulatory Networks from Expression Data, *Computational Intelligence in Bioinformatics 2008*, 33-74, 2008

[8] G. F. Cooper and E. Herskovits, A Bayesian Method for the Induction of Probabilistic Networks from Data, *Machine Learning*, 9, 309-347, 1992.

[9] W. Zhao et *al*. Inferring gene regulatory networks from time series data using the minimum description length principle, *Bioinformatics*, Vol. 22 no. 17 2006, pp. 2219-2135.

[10] K. Murphy, Bayes Net Toolbox for Matlab, www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html

[11] D. Eaton and K. Murhy, Bayesian Network Structure Learning - A Software Package for Matlab, www.cs.ubc.ca/~deaton/struct/bnsl.html

[12] B. Ristevski and S. Loskovska, Modeling of Gene Regulatory Networks by Boolean Networks, Proceedings of the 11th International Multiconference, Information Society – IS 2008, 97-100, pp. 2008.

A Service Platform for Context-Aware Mobile Transport-Related Information Services

Dragan Stojanovic[1], Bratislav Predic[2] and Slobodanka Djordjevic-Kajan[3]

Abstract – **This paper describes an architecture and implementation of the service platform for development of mobile information services to support fast, efficient and secure travel of people and transport of goods. These services are able to integrate, store and manage geo-referenced information related to traffic and transport and to deliver valuable functionality and content according to mobile user requirements, past, current and predicted movement and context.**

Keywords – **location-based, context-aware, mobile services, transport information systems, traffic management**

I. INTRODUCTION

With the recent advance and spreading of 2.5G and 3G mobile communication networks and development of new methods and technologies for mobile positioning, a huge interest in specialised mobile information services for support of various business, tourist and recreational activities emerges. Increasing demand for mobility of people and goods and dependence of modern economies on secure and efficient traffic and transport raises the importance of mobile information services that support safer and more efficient traffic and transport. Such services provide support for movement of users and transport of people and goods with aim to increase efficiency according to time, minimize resources needed (energy, financial, human) while, at the same time, decrease the risks of road accidents and minimize energy consumption and air pollution. Thus current research in intelligent transportation systems, such as EU ICT FP7 Challenge 6, addresses ICT for Intelligent Vehicles and Mobility Services [1]. Such services are designed for mobile users as support of their mobility and everyday activities. Their features, functionality, content and interface/interaction are adapted to users' past, current and predicted location and context.

II. MOBILE DATA AND INFORMATION SERVICES IN TRAFFIC AND TRANSPORT

Advances in mobile and ubiquitous computing, wireless communications, mobile positioning and sensor technologies enable collection of huge amount of data about mobile objects. Such data represent movement of users and objects of their interest as a sequence of locations at particular time instants. They also represent the context in which such movement occurs and define user situation while accessing services. The context, besides location and time, includes mobile computing and communication features, time of day/week/year, weather conditions, light and noise levels, nearby people, objects and devices, user activities and goals, user profile and preferences, etc. If a user drives a vehicle, the context includes the telemetry vehicle data, such as speed, direction, fuel level, engine status, and other data acquired from integrated sensors. Such a huge amount of continuously changed and acquired location and context data requires adequate methods, tools and techniques for data management, [2, 3, 4]. Data management includes efficient representation, storage, retrieval, processing, analysis, exchange and visualization of data about mobile object and is integrated into a platform for mobile information service development [2, 5].

A service platform should provide efficient representation of contextual description of users/mobile objects movement, as well as methods, index structures and algorithms for query processing and retrieval of mobile data [2, 5]. The processing and analysis of movement data enable generation of dynamic travel times along the road segments and, based on them, detection of traffic jams and stops. Analysis of history of mobile object movement provides knowledge discovery from these data, detection of movement patterns and generation of movement profiles according to context in which such movement occurs [6, 7]. Acquisition of huge amount of useful movement and context data related to transport of mobile users/vehicles have given a rise to a new class of mobile and ubiquitous applications and services, that are aware of the location and context of user/object of interest. Such services are capable to adapt their behavior and functionality according to the context with minimal distraction of the user and are called location-based (LBS) and context-aware services. LBS and context-aware include applications like automatic vehicle location, fleet management, tourist services, transport management, traffic control and digital battlefield. Market trend reports forecast that in 2013 up to 200 million mobile users regularly use some kind of LBS and context-aware services with generated revenue to about 13.3 billion of dollars [8]. Modern location-based services for navigation and transport purposes available on the market can be two fold:

- Navigation services (e.g. TomTom[1], Garmin/SCG Route[2]).
- Fleet management (e.g. Euman A/S[3], FleetOnline™[4])

[1]Dragan Stojanovic is with the Faculty of Electronic Engineering, University of Nis, A. Medvedeva 14, 18000 Nis, Serbia, E-mail: dragan.stojanovic@elfak.ni.ac.rs

[2]Bratislav Predic is with the Faculty of Electronic Engineering, University of Nis, A. Medvedeva 14, 18000 Nis, Serbia, E-mail: bratislav.predic@elfak.ni.ac.rs

[3]Slobodanka Djordjevic-Kajan is with the Faculty of Electronic Engineering, University of Nis, A. Medvedeva 14, 18000 Nis, Serbia, E-mail: slobodanka.djordjevic.kajan@elfak.ni.ac.rs

[1] http://www.tomtom.com
[2] http://www.garmin.co.yu/
[3] http://www.euman.com/
[4] http://www.fleetonline.net/

The functionality of commercial services is mostly based on current location of user/tracked object and on static data about road network, without or with minimal use of dynamic, real-time data like dynamic travel times, traffic accidents, jams, road conditions, etc. Such services do not take into account the whole history of user/vehicle movement and the context which define such movement which define the situation of the users. Obviously there is a need for service developers and providers to provide appropriate integration and management of such data and delivery of mobile information services completely customized and adapted to past, present and predicted movement and context of users/vehicles [9, 10].

III. MOWIS SERVICE PLATFORM

MOWIS service platform represents foundation for development of mobile, location-based and context-aware information services. It includes support for data management about movement and context of mobile objects:

- Methods for representation and storage of mobile objects data representing user movement and context.
- Methods, techniques and algorithms for querying and retrieval of movement and context data.
- Methods and techniques for user/mobile object movement analysis and pattern recognition and prediction of movement and context.

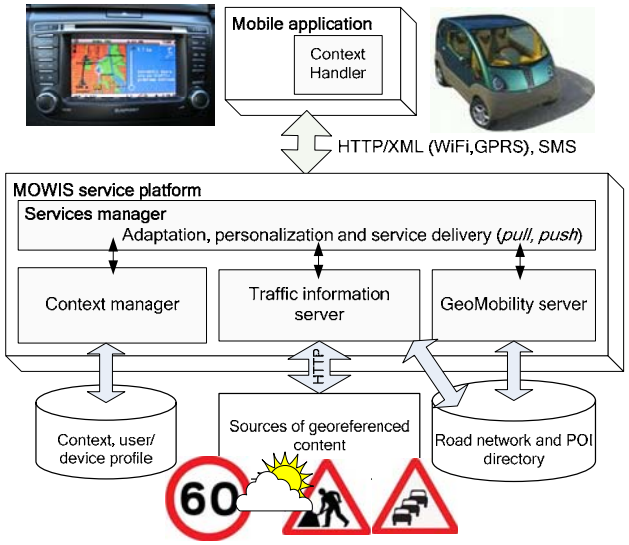


Fig 1. MOWIS service platform general architecture

Architecture of MOWIS platform includes following main components and is shown in figure 1:

- *GeoMobility server* represents implementation of OpenLS services for geocoding, reverse geocoding, routing, POI directory, presentation of information on mobile devices and gateway to location servers [11].
- *Traffic information server* integrates dynamic multimedia data (roads condition, weather forecast, traffic accidents, road segments under construction, traffic congestions and similar) from different sources with digital geo-referenced data representing road network. Access is enabled using standard interfaces and protocols like those defined by Open Geospatial Consortium (WMS, WFS, WCS) [12] and DATEX [13].
- *Context manager* provides acquisition, storage, transformation and interpretation of context data acquired from various sensors which finally results in complete description of user's situation. This server delivers context information to *Service manager* for the purpose of service adaptation and delivery either on user's request or on notification by *Services manager*.
- *Services manager* represents unified access point for MOWIS services and provides services adaptation and personalization in accordance with user's situation and context.

Service platform provides support for development of two major categories of mobile/Web information services supporting safe and efficient transport of people and goods:

- Mobile information services oriented toward mobile users, based on their past trajectory, current location and predicted movement, as well as users' context and situation.
- Mobile/Web information systems and services for tracking and monitoring of vehicles included in transport of people and goods with adequate selection of routes and means of transport.

IV. MOBILE INFORMATION SERVICES BASED ON MOWIS PLATFORM

The main objective of MOWIS service platform is to support development of mobile information services based on location and movement of user/mobile object of interest, their trajectories and their continually changing context. Based on MOWIS platform the users such services should be able to:

- see his current position, or position of every other vehicle being tracked over digital map on mobile device,
- show users' status represented by physical and computing context acquired from sensors attached to client mobile device, as well as vehicles' status obtained from vehicles' sensors,
- view vehicles' movement in real-time while maintaining minimal load on wireless network, server memory and processing resources of mobile devices and the server,
- obtain information and notifications about occurrence of unpredictable events concerning vehicle (user/vehicle has left defined route/area, remaining fuel low, speed exceeds maximum defined, user/vehicle is in vicinity of another object etc.),
- obtain information concerning road condition, occurrence of traffic jams and stops, accidents and roadwork, and based on them obtain alternative routes toward destination,
- analyze previously recorded and stored trajectories to support decision making in the process of planning transport of people and goods that are safer and more efficient.

Implementation of service platform started with the realization of *Traffic information server* and *GeoMobility server*. Java Servlet technology and Apache Tomcat Java application server was used for implementation and deployment. PostgreSQL database with spatial extension PostGIS was used to store geospatial data (roads network, addresses, points of interest – POI), as well as multimedia data referenced to road network. All services offered by *Traffic information server* and *GeoMobility server* were made available through XML protocol defined by OGC [11, 12]. The service chaining is enabled when it is necessary to invoke multiple services to process single, complex user request. For example, user requests routing service by specifying origin and destination as addresses. Before invoking routing service the system transforms origin and destinations addresses to points specified by coordinates by invoking geocoding service twice, for start and end point. Only after that system can invoke routing service in order to acquire optimal route based on current road network state.

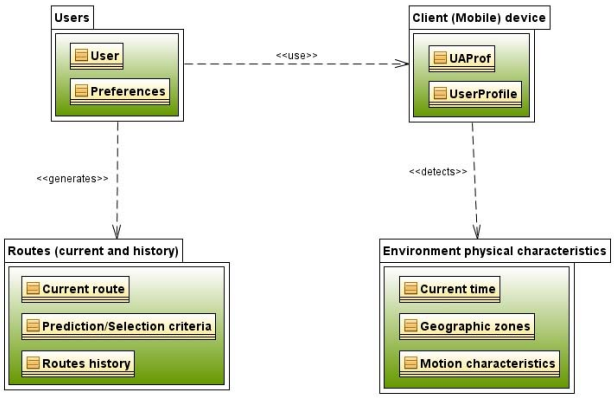


Fig 2. Context model used by MOWIS platform

Context manager implements components for acquisition, storage, processing and interpretation of user's context. Contemporary mobile applications almost exclusively interpret only technical characteristics of mobile device as context which represent only a subset of information that can be interpreted as context. Today, users of mobile applications move along the road network and have various sensors (location, lighting, noise, orientation, acceleratometers etc.) integrated with their mobile devices. So the context of such users can be described much more accurately. Apart from mobile device hardware characteristics, user's profile/preferences and geographic environment, MOWIS platform includes routes as context information that characterize situation the user is in and the goal of his movement. (figure 2.)

An important component implemented as part of *Context manager* provides user's mobile device profile management – *DeviceProxy*. This component accepts initial request sent by the user. Considering the fact that typical users of the presented system employ various types of mobile devices the first step is to detect relevant technical characteristics of the employed device that affect service operation and deployment. Initial HTTP request to log onto the system contains some form of identification of the mobile device. Currently, there are a number mechanisms being used in the industry for this purpose, such as UAProf (User Agent Profile), WALL (Wireless Abstraction Library) or WURFL. UAProf, which is used in the MOWIS implementation, uses separate header in the HTTP request to provide link to XML document containing detailed description of hardware and software features of the device sending the request. These XML descriptions are maintained by mobile device manufacturers and, based on this information, the server decides whether to reply to the client with link to XHTML Web application, link to J2ME application to be installed OTA or link to download .NET Compact Framework Windows Mobile application.

After login, each registered user becomes a continuous source of positional updates containing also movement characteristics (speed, heading etc.). *Context manager* component which performs spatial analysis bases its functionality on known road network. This component generates a current user's route by matching received stream of positional updates to underlying road network. This component also generates and stores the histories of routes users take. The key issue in interpreting route as wider contextual information is matching current user's route with routes recorded in the history. If currently generated route is not matched to any route in the history, the system adds it to the history allowing the user to classify newly created route. Route classification ("going to work" and similar) is a basis for further reasoning about travel purpose and user's intent. If there is matched route in route history, the system uses matched route to predict user's future trajectory. It is important to emphasize that route matching takes into account not only spatial component of the routes being compared, but also time component, matching the time (time of day, day of the week etc.) when route in history was recorded and current time. Also, movement profiles, such as average speed per road segment and number of stops, represent additional weight factors for matching process and allow fine classification of routes in history. After the spatial analysis is completed, generated data (route ID, users' intent and similar) are included into a rule engine. Current MOWIS implementation uses JESS (Java Expert Shell) rule engine. The rule engine applies defined rules to the constantly changing set of contextual information contained in memory to generate contextual information at a higher logical level (user's state or intention). It also includes generated information in the rule engine to trigger appropriate *Service manager* action therefore adapting offered service to user's changing environment. The usage of general type rule engine like JESS makes *Context manager* flexible and enables dynamic registration and deletion of large number of rules and tracking requests. Thus is can appropriately react to context changes of high number of users. Apart from performance improvement this approach doesn't require frequent system restarts and initializations when additional rules are introduced. In the environment where the system tracks potentially large number of users procedurally based implementation would not be possible.

Services manager performs adaptation of services which are delivered to mobile users in accordance with their current

location, past and predicted trajectory and the context. This component offers two service delivery models:

- Upon request, when client initiates interaction with *Service Manager* and requests services in accordance with his current/predicted movement and context.
- Notification by service, when *Service Manager* proactively initiates interaction with the user and delivery of appropriate service in response to occurrence of certain event, or fulfillment of defined rules over movement and context data.

a) b)

Fig 3. Mobile applications based on MOWIS platform a) navigation service b) public bus transport service

Based on presented functionalities two mobile information services intended for proof of the concepts and demonstration purposes have been implemented:

- Navigation information service enabling user to receive dynamic data concerning traffic conditions along defined route. These dynamic data include road status, meteorological data, and information about congestions, accidents and road work. User can subscribe to this service and either requests or be notified about occurrence of important events in accordance with his movement and context. Based on this information user can request alternative route which can lead him to the destination more quickly and efficiently (Figure 3a).
- Information service for tracking buses in public transport which provides prediction of bus arrival times (figure 3b) [9]. Users are able to receive this information either on request or by notification in accordance with his location, movement, context and interests.

V. CONCLUSION

MOWIS service platform and mobile information services based on it are aimed to support ICT solutions of problems caused by ever-increasing mobility of people and goods. These problems include traffic jams, congestions, fuel and energy consumption, pollution emission and above all increasing number of traffic accidents causing significant loss of life and material damage. Having in mind important role that transport of people and goods plays in modern society and economy it is reasonable to expect appropriate ICT support for safe and efficient traffic and transport. Further, service providers are able to offer users, individual or corporate, information services accessible by mobile computing devices connected to Internet/Web to improve their safety and efficiency in road transport. Individual users or corporations whose operations include transport of people and goods, can subscribe to such services and achieve significant savings in human, material and financial resources making their operations safer and more efficient. State institutions that manage and plan urban traffic will have opportunity to manage and plan future infrastructure development and maintenance for faster and more efficient traffic.

ACKNOWLEDGEMENT

Research presented in this paper is partially funded by Ministry of Science and Technological Development, Republic of Serbia as part of project “Mobile/Web information services for support of secure and efficient transport of people and goods”, Nr. TR-13008.

REFERENCES

[1] CORDIS FP7 ICT Programme Challenge 6 - ICT for Intelligent Vehicles and Mobility Services, http://cordis.europa.eu/fp7/ict/intelligent-vehicles/home_en.html

[2] D. Stojanovic, A. N. Papadopoulos, B. Predic, S. Djordjevic-Kajan, A. Nanopoulos, “Continuous Range Monitoring of Mobile Objects in Road Networks”, Data & Knowledge Engineering, Elsevier, Vol 64 (1), January 2008, pp 77-100.

[3] Civilis A., C. S. Jensen, and S. Pakalnis. Techniques for Efficient Road-Network-Based Tracking of Moving Objects. IEEE TKDE, 17(5): 698–712 (2005)

[4] Z. Ding and R. H. Guting. Managing Moving Objects on Dynamic Transportation Networks. In Proc. International Conference on Scientific and Statistical Database Management, pp. 287–296 (2004)

[5] S. Djordjevic-Kajan, D. Stojanovic, B. Predic, "Context Aware Mobile Geographic Information Systems", Encyclopedia of Mobile Computing and Commerce, Volume 1, IDEA Group Publishing, 2007, pp. 129-137.

[6] E. Tiakas, A.N. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic, S. Djordjevic-Kajan: "Trajectory Similarity Search in Spatial Networks", Proceedings of the 10th International Database Engineering & Applications Symposium (IDEAS 2006) , New Delhi, India, 2006, pp. 185-192 .

[7] S. Brakatsoulas, D. Pfoser, N. Tryfona. Modeling, Storing and Mining Moving Object Databases, Proceedings of the 8th IDEAS, 2004 , pp: 68 – 77.

[8] ABI Research, Mobile Location-Based Services http://www.abiresearch.com/products/market_research/Location_Based_Services , 2008

[9] B. Predic, D. Stojanovic, S. Djordjevic-Kajan, A. Milosavljevic, D. Rancic, “Prediction of Bus Motion and Continuous Query Processing for Traveler Information Services”, 11th ADBIS 2007, pp. 234-249.

[10] COOPERS - Integrated Project - FP6 IST, 2006 – 2010. http://www.coopers-ip.eu/

[11] OGC Location Service (OpenLS) Implementation Standards, http://www.opengeospatial.org/standards/ols

[12] Open Geospatial Consortium, http://www.opengeospatial.org

[13] DATEX - EC Directorate General for Transport and Energy, http://www.datex2.eu Networks", TELSIKS'99, Conference Proceedings, pp.214-217, Nis, Yugoslavia, 1999.

Computer Aided Design of Knitting Pattern Structures

Elena Zaharieva-Stoyanova[1]

Abstract – **This paper suggests an approach to CAD system development based on its creation as a set of programming modules. Some of these modules can be used in next generation of CAD systems or they can be embedded in another CAD software. The main reason to create such project is that it can be used as software application from fashion designers to make boutique collections of knitting products. The project develops a graphic editor for design of knitting pattern structures. It gives designers possibility to create different types of knitting structures, such as: jacquard, intarsia, gusset, lace and others Suggested CAD module would be used as an application for making boutique collections by fashion designers or as a part of another system.**

Keywords – **CAD/CAM systems, knitting industry, data structures, knitting patterns.**

I. INTRODUCTION

The development of CAD/CAM systems for knitting industry automation is conditioned by technological improvement of knitting machines and by the advance in computer science, as well. Software rapidly changes along with hardware technologies development. New Operation Systems are continuously being introduced to the market plus a number of platforms and programming environments.

Normally, the companies manufacturing knitting machines also put on the market CAD/CAM/CAI software packages delivered with whole hardware equipment. Non-standard equipment was usually in use before due to the than existing a trademark protection policy. However this increases CAD/CAM systems' prices several times [6], [7], [8], [9], [10].

It's important to mention that Computer-aided design is not only used in the industrial production. There exist other users, who are interested in such kind of software. Many people perceive hand knitting as hobby and they are searching new knitting models continuously. Usually, they take such samples from magazines. There are a lot of magazines for hand crafts – knitting, embroidery, gobelins. Some of them are published by famous companies, which offer also materials and accessories for hand crafts [11], [12], [13].

Creating knitting patterns editors of such kind of magazines use a multy-functional software packages as word processors and spreadsheet programs. To use a professional CAD/CAM system in this case is non-efficient. First of all, the functionality of each CAD/CAM system in knitting industry is bound by knitting machines' features. It means that only parts of CAD module might be used, so, the most of whole CAD/CAM system would be useless.

The second reason to do not use CAD/CAM system is that software delivered with knitting machines is custom developed. The user of such kind of a system must be well familiar with knitting machine features and knitting technologies process as a whole. For a fashion designer, who develops knitting patterns only, such software is hard to handle.

CAD/CAM systems are expensive software and their usage must be profitable. On the other side, a wide range of customers are interested of CAD modules as software for knitting design.

The goal of this paper is to offer a project to a programming module of specific CAD system development for knitting pattern structures. It can be used in next generation of CAD systems or they can be embedded in CAD software. [2] The users of this software application would be: designers of new knitting boutique collections; editors of knitting hobby magazines; people, who are interested in knitting as hobby; ethnographers, who make investigations in the area of old traditional costumes and hand-made knitting. The project aims at more user-friendly software.

This CAD module is a part from another CAD software package. The whole project covers all aspects in knitting design process. The base feature of this system is that it is developed as set of programming modules [1], [7],[8],[9]. The main reason to develop such kind of software is that software development demands a lot of efforts and human resources it is better to use relevant modules from previous CAD/CAM systems versions. Moreover, knitting manufacturing technologies are not so frequently modified, as is the case with software development. Accordingly, it is important to develop CAD/CAM software in modules, which can be alternatively used in other systems.

II. DEMANDS TO CAD SOFTWARE FOR KNITTING PRODUCTS DESIGN

Generally, application of computer graphics in knitting industry CAD/CAM systems has two aspects: knitting structures design and design of knitting pattern shapes. The second one is related to knitting machine's capacity to make products by fully-fashion (FF) method. It means that the machine knits the cuts of products, or the whole products. This method allows to avoid cutting the materials, reduces the number of the operations and it leads to reducing the waste products to a minimum. FF knitting method is based on the technology of hand knitting.

Because there is a big difference between the images presented knitting structures and knitting pattern shapes as graphic objects, the CAD systems must contains two graphics editors – the first one to create a knitting structure, the second

[1]Elena Zaharieva-Stoyanova is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 H. Dimitar str, Gabrovo 5300, Bulgaria, E-mail: zaharieva@tugab.bg

one to create a model of the product made by FF knitting method.

According to the conventional design, the model of a knit product (jacket, sweater, cardigan, etc.) is created based on some standard constructions. The term "construction" determines the whole style of the product; the number of its parts and their shape [6], [10].

Each model can be manufactured in several sizes determined by the Standards of knit products. They define the sizes of knitting products' cuts. To create original and authentic models designers make use of different kinds of knitting techniques such as jacquard, intarsia, gusset, lace and other knitting structures. The techniques employed are involved as a part of the model. It means that creation of knitting structures is a part of a whole process of knitting design [6],[7].

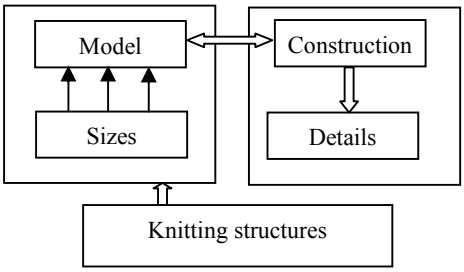


Fig. 1. Aspects of a knitting design process.

All aspects of a knitting design process and their relations to each other is shown on fig.1. As a process, knitting products design is to cover all these aspects, so that the demands to CAD/CAM systems for knitting industry in general can be described as follows:

- Capacity to create and edit a construction of model including all its parts (cuts) of the product.
- To support a set of standard constructions of knitting models.
- To create a knitting model including all sizes variation.
- To create different types of knitting structures according to machine abilities to manufacture different knitting techniques (jacquards, intarsia, gusset, lace, and so on).
- Automatic generation of knitting control programs implementing the models.

The last one is a CAM module function; the rest is demands to CAD software. As a software application, a CAD system can be determined as a graphics editor creating and describing knitting structures and cuts of knit products [7], [8].

Nevertheless, the last feature refers to industrial production; it has its application in CAD system for hand knitting, too. Introducing a model designer supports it with instructions how to realize the product. For example, fig. 2 is a list of instructions for knitting a *cable panel*. So, the last demand can be converted to "Automatic generation of knitting instruction set"

CABLE PANEL (make 3):
With pair of needles, cast on 49 sts.
Row 1 (wrong side): K3, p6, k11, p3, k3, p3, k11, p6, k3.
Row 2: P3, k6, p11, Cr9B, p11, k6, p3.
Row 3: As Row 1.
Row 4: P3, C6B, p11, k3, p3, k3, p11, C6B, p3.
Row 5: As Row 1.
Row 6: P3, k6, p11, k3, p3, k3, p11, k6, p3.
Row 7: As Row 1.
Row 8: P3, k6, p8, k3, Cr9B, k3, p8, k6, p3.
Row 9: K3, p6, k8, p6, k3, p6, k8, p6, k3.
Row 10: P3, C6B, p8, k6, p3, k6, p8, C6B, p3.
Row 11: As Row 9.
Row 12: P3, k6, p8, k6, p3, k6, p8, k6, p3.
Row 13: As Row 9.
Row 14: P3, k6, p5, k3, Cr6B, k3, Cr6F, k3, p5, k6, p3.
Row 15: K3, p6, k5, p6, k3, p3, k3, p6, k5, p6, k3.
Row 16: P3, C6B, p5, k6, p3, k3, p3, k6, p5, C6B, p3.
Row 17: As Row 15.
Row 18: P3, k6, p5, k6, p3, k3, p3, k6, p5, k6, p3.
Row 19: As Row 15.

Fig. 2. Example of a knitting model's description.

III. Features of Knitting Structures Images

There exist diverse knitting techniques helping the designers to create a big variety of knitting patterns. The features of knitting structures images are determined by the used knitting technique. This paper treats the problems of computer-aided design of knitting structures so, their features will be discussed from computer graphics' point of view.

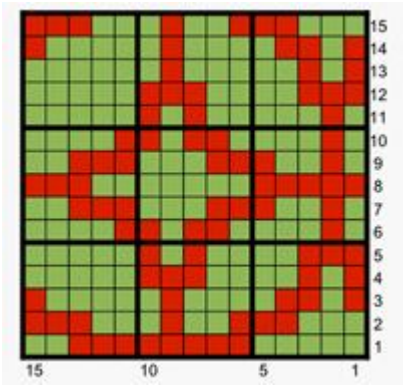


Fig. 3. Knitting pattern structure "jacquard".

Generally speaking, a pattern of knitting structure can be drawn as a raster graphics image, where the smallest item of the grid corresponds to a knitting stitch. To describe a knitting pattern structure, the grid's items are filled with different colors (fig. 3) or they are consisted of icons (small bit-map images) as it shown on fig. 4.

According the image view, the knitting patters, describing knitting structures can be divided in two groups:

- color images;
- images, consisted of icons – small bit-map images.

The first group presents knitting structures, realized by using of the following knitting techniques: *jacquard*, *intarsia*, and *stripe*. The second group includes images presenting *laces*, *arans*, *cables*, for example. The common feature of all these technique is that they use a different kind of stitch together. It allows to obtain interesting and attractive knitting structures. The different icons present different stitches such as: *knit stitch*, *purl stitch*, *yarn over*, *cross* and so on.

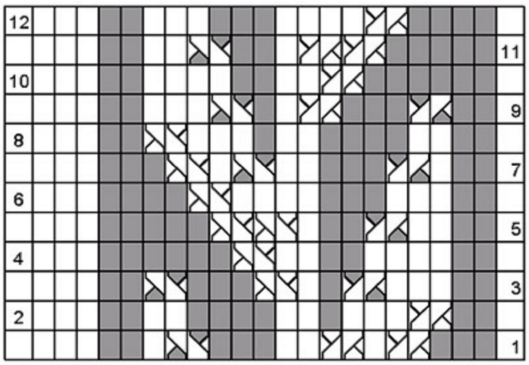


Fig. 4. Knitting pattern structure “aran”.

It is important to mention that not all knitting technique cannot be used together. For example, knitting on the right side and purling on the other side realize the colored knitting structures obtained by usage of intarsia or jacquard technique. Knitting stripes is an exception. It can be realized knitting structures by using different stitches and transverse stripes in different colors together.

IV. Project of CAD Module for Knitting Structures

A. General overview

The CAD module for knitting structures is a part of whole CAD system for knitting industry automation. It can be realized as an editor of raster graphics images put into rectangular grid. The size grid elements must be conformable to the size of the icon corresponding to a stitch – *knit stitch*; *purl stitch*; *yarn over*; *knit 2 stitches together*; *purl 2 stitches together* and so on.

The editor will support two types of images: colored images and images consisted of icons (small bit-maps). They would be chosen alternatively according to the selected knitting technique.

To create a colored image, the designer need color palettes. Practically, the number of the yarns for one knitting pattern is limited because of knitting technique’s features. Usually, two, three or four different yarn colors are used for one pattern. The situation when more than four yarn colors are involved is very rare happens. Nevertheless, the modern computer technique has possibilities to handle with thousands, so there is no reason to restrict the designer when he or she selects the workbench colors.

To create an image like to the one presenting on fig. 3, the designer needs a group of icons corresponding to the different types of stitches. The problem is that there is not unification for these symbols. So, one kind of stitch, purl for example, could be present by two or more different symbol. The want of unification could be get over by using of some sets of icons (symbols) describing the stitch types. The designer will select the set of symbols, which is like him.

The general overview of the CAD module structure is given on fig. 4.

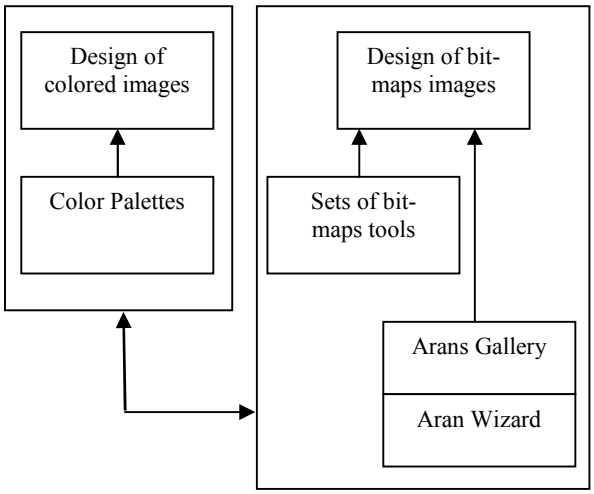


Fig. 4. Structure of the CAD module for knitting structures.

B. Arans Gallery

The feature of the *arans* is that they consist of more than one stitch. It means that they need more than one raster’s element – two, three or more stitches are joined to each other and they are taken as one whole element in. Because there is a big variety of arans and the designer could be make kinds of arans by himself, the decision of the problem is to create a Gallery of arans. The Galley will be supported by the Wizard program, which allows creating arans. The Wizard needs following information:

- number of crossing stitches;
- the kind of each stitch (knit or purl stitch);
- the cross point;
- the crossing direction (to the right ot to the left).

C. Image Editing

Unless color palettes, sets of stitch icons and aran Gallery, the editor can be supplied with tools to draw, fill, or erase the part of the images. Existing of menu items Cut, Copy, and Paste is recommendable.

Saving image to standard graphic formats is convenient and reasonable because of the unification but by reason of further image processing it is better to duplicate information presenting an image. Because each item of the raster corresponds to the knitting stitch, a matrix form of an image data representation is suitable.

Usually, graphic editors use some tools to select areas (rectangular selection, free-hands selection), to erase selected areas, to cut or copy them, to fill them, or to put drops in different colors.

Flipping and rotation is used, too. To flip or to a colored image is not a problem but to do this operation on an image presenting lace or aran means that its structure have to be changed. For example, flipping an item *knit 2 st. together* will be changed to *cross and knit 2 st. together*. This feature must be recognized when flip and rotate operations are realized.

V. CONCLUSION

This paper suggests an approach to CAD module for knitting structures. The project develops a graphic editor for design of knitting pattern structures. It gives designers possibility to create different types of knitting structures, such as: jacquard, intarsia, gusset, lace and others.

The project is a part of CAD system development based on its creation as a set of programming modules. Some of these modules can be used in next generation of CAD systems or they can be embedded in another system - CAD software. The main reason to create such project is that it can be used as software application from fashion designers to make boutique collections of knitting products.

The CAD system would be used as an application for making boutique collections by fashion designers or as a part of another system. The users of this software application would be also editors of knitting hobby magazines; people, who are interested in knitting as hobby; ethnographers, who make investigations in the area of old traditional costumes and hand-made knitting. The project can be used as an example for CAD system development.

REFERENCES

[1] Y. Angelova, E. Zaharieva-Stoyanova, "Precision of the shape of garment pieces at Fully Fashion Knitting, *Textile & Cloths*, 7/2006.

[2] E. Zaharieva, "Computer Aided Design of Knitting Products Using FF Method", *International Scientific Conference UNITECH'2007,* 23-24 November, Gabrovo, Bulgaria.

[3] E. Zaharieva-Stoyanova, "Methods of Graphic Representation of Curves in CAD systems in Knitting Industry", *XLII International Scientific Conference on Information, Comunications and Energy Systems and Technologies ICEST 2007*, 24-27 June, 2007, Ohrid, Macedonia.

[4] E. Zaharieva-Stoyanova, "Application of Bezier Curves in Knitting Industry CAD/CAM systems", *XL International Scientific Conference on Information, Communication and Energy Systems and Technologies – ICEST 2005*, June 29 – July 1, Nis, Serbvia and Montenegro, 2005.

[5] E. Zaharieva-Stoyanova, "Algorithm for Computer Aided Design Curve Shape Form Generation of Knitting Patterns", *International Conference on Automation, Quality and Testing, Robots AQTR 2006(THETA 15),* May 25-28 2006, Cluj-Napoca, Romania.

[6] Trend Collection 2004/2005, Stoll, Department Fashion & Technology

[7] www.stoll.de

[8] Sirix 110/210, Instructions, Part I, II, Reutlingen, 2001

[9] www.knitwares.ca

[10] www.universal.de

[11] http://www.verenaknitting.com/

[12] http://www.burdafashion.com/en/index.html

[13] http://www.milenastyle.com/catalog/catalog/index.php

[14] http://www.allfiberarts.com/cs/software.htm

AUTHOR INDEX