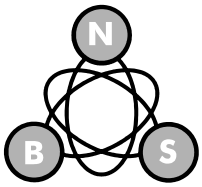
XLVI INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES



ICEST 2011

Proceedings of Papers

Volume 1

Serbia, Niš, June 29 - July 1, 2011

ICEST 2011 - XLVI INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES, Serbia, Niš, June 29 - July 1, 2011

Proceedings of Papers - Volume 1 of 3 volumes

Editor: Prof. Dr. Bratislav D. Milovanović

Technical Editor: Dr. Zoran Ž. Stanković

Technical Co-Editor: Dr. Biljana P. Stošić

Published by: Faculty of Electronic Engineering, University of Niš, Serbia
Printed by: UNIGRAF, Niš, Serbia

Number of copies printed: 80

Printing of this edition has been financially supported by Serbian Ministry of Science

ISBN: 978-86-6125-031-6

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

621.39(082)
537.8(082)
006.91(082)
681.586(082)

INTERNATIONAL Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST (46 ; 2011 ; Niš)
Proceedings of Papers. #Vol. #1 / XLVI International Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST 2011, Niš, June 29 - July 1, 2011 ; [organized by Faculty of Electronic Engineering, Niš [and] Faculty of Telecommunications, Sofia [and] Faculty of Technical Sciences, Bitola ; editor Bratislav D. Milovanović]. - Niš : Faculty of Electronic Engineering, 2011 (Niš : Unigraf).
- XVII, 260 str. : ilustr. ; 29 cm

Tiraž 280. - Bibliografija uz svaki rad. -
Registar.

ISBN 978-86-6125-031-6

a) Телекомуникације - Зборници b) Обрада сигнала - Зборници c) Микроталасна техника - Зборници d) Метрологија - Зборници e) Мерни инструменти, сензорски - Зборници
COBISS.SR-ID 186458124

Dear Colleagues,

Welcome to the XLVI International Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST 2011. The conference is going to be held from June 29 to July 1, 2011, at the Faculty of Electronic Engineering, University of Niš, Serbia. The Conference is, for the tenth time, jointly organized by the Faculty of Electronic Engineering, Niš, Serbia; the Faculty of Telecommunications, Sofia, Bulgaria, and by the Faculty of Technical Sciences, Bitola, Macedonia. This is the forth time that this big Balkan event takes place in Niš.

As to the earlier ICEST Conferences, many authors from institutions all over the Europe submitted their papers. This year, 252 papers have been presented as oral (136 papers) or poster (116 papers) presentations.

After the Conference opening one plenary invited paper “Spatio-Temporal Analysis of Multi-Source Data in Remote Sensing” will be given by Prof. Dr. Zoran Obradović from Temple University, Philadelphia, USA. Furthermore, a workshop “Interference and Noise in Electromagnetics and Electromagnetic Compatibility” will be given by Prof. Dr. Peter Russer and Dr. Johannes Russer from Technical University Munich, Germany. The Conference will also include a round table “Digital Literacy and Information Society in Serbia”.

I hope that all participants will take opportunities not only to exchange their knowledge, experiences and ideas but also to make contacts and establish further collaboration. A social program, rich in events, will provide more relaxing atmosphere for meeting the colleagues.

On the behalf of the Technical Program Committee, I wish you successful presentations and pleasant stay in Niš!

On the behalf of the Technical Program Committee,

Prof. Dr. Bratislav Milovanović,
Conference Chairman

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

ICEST 2011

organized by

Faculty of Electronic Engineering,
Niš, Serbia

Faculty of Telecommunications,
Sofia, Bulgaria

Faculty of Technical Sciences,
Bitola, Macedonia

under auspices of

- **Serbian Ministry of Education and Science**

in cooperation with

- **Academy of Engineering Sciences of Serbia**
- **IEEE Serbia and Montenegro Section**

TECHNICAL PROGRAM COMMITTEE

Chairman:

Chairman:

B. Milovanović University of Niš, Serbia

Vice-chairmen:

R. Arnaudov Technical University of Sofia, Bulgaria

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

Members:

N. Acevski University "St. Kliment Ohridski", Bitola, Macedonia
V. Aćimović-Raspopović University of Belgrade,Serbia
V. Ceselkoska University "St. Kliment Ohridski", Bitola, Macedonia
D. Dimitrov Technical University of Sofia, Bulgaria
K. Dimitrov Technical University of Sofia, Bulgaria
B. Dokić University of Banja Luka, Bosnia and Herzegovina
D. Dobrev Technical University of Sofia, Bulgaria
N. Dončov University of Niš, Serbia
R. Goleva Technical University of Sofia, Bulgaria
N. Gospić University of Belgrade, Serbia
G. Iliev Technical University of Sofia, Bulgaria
I. Iliev Technical University of Sofia, Bulgaria
Z. Jakšić IHTM Institute, Belgrade, Serbia
D. Janković University of Niš, Serbia
N. Janković University of Niš, Serbia
B. Jokanović Institut of Physics, Belgrade, Serbia
I. Jolevski University "St. Kliment Ohridski" Bitola, Macedonia
M. Jeftić University of Niš, Serbia
M. Kostov University "St. Kliment Ohridski" Bitola, Macedonia
R. Krneta University of Kragujevac, Serbia
R. Kuntchev Technical University of Sofia, Bulgaria
M. Lutovac University of Novi Pazar, Serbia
V. Marković University of Niš, Serbia
A. Markovski University "St. Kliment Ohridski", Bitola, Macedonia
S. Mirtchev Technical University of Sofia, Bulgaria
P. Mitrevski University "St. Kliment Ohridski", Bitola, Macedonia
I. Nedelkovski University "St. Kliment Ohridski", Bitola, Macedonia
A. Nešić IMTEL Institute, Belgrade, Serbia
N. Nešković University of Belgrade, Serbia
B. Nikolova Technical University of Sofia, Bulgaria
Dj. Paunović University of Belgrade, Serbia
E. Pencheva Technical University of Sofia, Bulgaria
Z. Perić University of Niš, Serbia
P. Petrović IRITEL Institute, Belgrade, Serbia
S. Pleshkova-Bekiarska Technical University of Sofia, Bulgaria
V. Poulkov Technical University of Sofia, Bulgaria
P.M. Radevska University "St. Kliment Ohridski", Bitola, Macedonia
B. Reljin University of Belgrade, Serbia

I. Reljin	University of Belgrade, Serbia
P. Spalević	University of K. Mitrovica, Serbia
R. Stanković	University of Niš, Serbia
Z. Stanković	University of Niš, Serbia
M. Stefanović	University of Niš, Serbia
M. Stevanovski	University "St. Kliment Ohridski", Bitola, Macedonia
M. Stojčev	University of Niš, Serbia
L. Stoimenov	University of Niš, Serbia
G. Stoyanov	Technical University of Sofia, Bulgaria
D. Tasić	University of Niš, Serbia
M. Temerinac	University of Novi Sad, Serbia
Lj. Trpezanovski	University "St. Kliment Ohridski", Bitola, Macedonia
B. Tsankov	Technical University of Sofia, Bulgaria
I. Uzunov	Technical University of Sofia, Bulgaria
L. Zieleznik	Brookes University of Oxford, UK

CONFERENCE ORGANIZING COMMITTEE

Chairman:

B. Milovanović	University of Niš, Serbia

Members:

M. Agatonović	University of Niš, Serbia
T. Asenov	University of Niš, Serbia
T. Dimitrijević	University of Niš, Serbia
I. Dochev	Technical University of Sofia, Bulgaria
N. Dončov	University of Niš, Serbia
D. Janković	University of Niš, Serbia
V. Jović	University of Niš, Serbia
L. Lubih	Technical University of Sofia, Bulgaria
M. Milijić	University of Niš, Serbia
J. Pargovski	University "St. Kliment Ohridski", Bitola, Macedonia
Z. Perić	University of Niš, Serbia
M. Petkovski	University "St. Kliment Ohridski", Bitola, Macedonia
Z. Stanković	University of Niš, Serbia
B. Stošić	University of Niš, Serbia
D. Tasić	University of Niš, Serbia

TECHNICAL SUPPORT

A. Atanasković	University of Niš, Serbia
J. Anastasov	University of Niš, Serbia
T. Asenov	University of Niš, Serbia
Z. Djordjević	University of Niš, Serbia
J. Joković	University of Niš, Serbia
Z. Marinković	University of Niš, Serbia
M. Miletić	University of Niš, Serbia

M. Milijić	University of Niš, Serbia
J. Nikolić	University of Niš, Serbia
A. Panajotović	University of Niš, Serbia
B. Stošić	University of Niš, Serbia

CONFERENCE SECRETARIAT

Z. Stanković	Conference Technical Editor
B. Stošić	Conference Technical Co-Editor & Technical Coordinator
T. Dimitrijević	Technical Secretary
T. Asenov	Technical Secretary

Address:

ICEST 2011 Conference
University of Niš
Faculty of Electronic Engineering
Aleksandra Medvedeva 14
18000 Niš, Serbia

phone:	+381 18 529 105, 529 302, 529 303
fax:	+381 18 588 399
e-mail:	info@icestconf.org
	icest@elfak.ni.ac.rs

CONFERENCE INTERNET SITE

http://www.icestconf.org

LIST OF ICEST 2011 REVIEWERS

Prof. Dr. Nikolče Acevski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Rumen Arnaudov
Technical University – Sofia, Bulgaria

Mr. Aleksandar Atanasković
University of Niš, Serbia

Prof. Dr. Ivaylo Atanasov
Technical University – Sofia, Bulgaria

Prof. Dr. Georgi Balabanov
Technical University – Sofia, Bulgaria

Prof. Dr. Aleksandar Bekiarski
Technical University – Sofia, Bulgaria

Prof. Dr. Boncho Bonev
Technical University – Sofia, Bulgaria

Prof. Dr. Ognian Bumbarov
Technical University – Sofia, Bulgaria

Prof. Dr. Zlata Cvetković
University of Niš, Serbia

Prof. Dr. Dejan Ćirić
University of Niš, Serbia

Prof. Dr. Stojče Deskovski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Kalin Dimitrov
Technical University – Sofia, Bulgaria

Prof. Dr. Dobri Dobrev
Technical University – Sofia, Bulgaria

Prof. Dr. Goran S. Đorđević
University of Niš, Serbia

Prof. Dr. Goran T. Đorđević
University of Niš, Serbia

Prof. Dr. Ivo Dochev
Technical University – Sofia, Bulgaria

Prof. Dr. Nebojša Dončov
University of Niš, Serbia

Prof. Dr. Ivo Draganov
Technical University – Sofia, Bulgaria

Mr. Predrag Eferica
University of Niš, Serbia

Prof. Dr. Veska Georgieva
Technical University – Sofia, Bulgaria

Prof. Dr. Snežana Golubović
University of Niš, Serbia

Prof. Dr. Nataša Gospić
University of Belgrade, Serbia

Prof. Dr. Ilia Iliev
Technical University – Sofia, Bulgaria

Prof. Dr. Georgi Iliev
Technical University – Sofia, Bulgaria

Prof. Dr. Dragan Janković
University of Niš, Serbia

Prof. Dr. Nebojša Janković
University of Niš, Serbia

Prof. Dr. Borislav Jeftenić
University of Belgrade, Serbia

Prof. Dr. Milun Jevtić
University of Niš, Serbia

Dr. Jugoslav Joković
University of Niš, Serbia

Prof. Dr. Ilija Jolevski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Lidia Jordanova
Technical University – Sofia, Bulgaria

Prof. Dr. Zoran Jovanović
University of Niš, Serbia

Prof. Dr. Pavlina Koleva
Technical University – Sofia, Bulgaria

Prof. Dr. Zora Konjović
University of Novi Sad, Serbia

Prof. Dr. Lidija Korunović
University of Niš, Serbia

Prof. Dr. Mitko Kostov
Technical University – Sofia, Bulgaria

Prof. Dr. Roumen Kountchev
Technical University – Sofia, Bulgaria

Prof. Dr. Ludvig Lubih
Technical University – Sofia, Bulgaria

Prof. Dr. Nataša Maleš Ilić
University of Niš, Serbia

Prof. Dr. Dragan Mančić
University of Niš, Serbia

Dr. Zlatica Marinković
University of Niš, Serbia

Prof. Dr. Vera Marković
University of Niš, Serbia

Prof. Dr. Aleksandar Markovski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Rossen Miletiev
Technical University – Sofia, Bulgaria

Prof. Dr. Dejan Milić
University of Niš, Serbia

Dr. Nenad Milošević
University of Niš, Serbia

Prof. Dr Danijela Milović
University of Niš, Serbia

Prof. Dr. Rumen Mironov
Technical University – Sofia, Bulgaria

Prof. Dr. Seferin Mirtchev
Technical University – Sofia, Bulgaria

Prof. Dr. Pece Mitrevski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Cvetko Mitrovski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Tsvetan Mitsev
Technical University – Sofia, Bulgaria

Prof. Dr. Milica Naumović
University of Niš, Serbia

Prof. Dr. Marin Nedelchev
Technical University – Sofia, Bulgaria

Prof. Dr. Saša Nikolić
University of Niš, Serbia

Prof. Dr. Tashko Nikolov
Technical University – Sofia, Bulgaria

Prof. Dr. Boyanka Nikolova
Technical University – Sofia, Bulgaria

Prof. Dr. Zlatka Nikolova
Technical University – Sofia, Bulgaria

Prof. Dr. Dragan Pantić
University of Niš, Serbia

Prof. Dr. Evelina Pencheva
Technical University – Sofia, Bulgaria

Prof. Dr. Mile Petkovski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Vladimir Petrović
University of Belgrade, Serbia

Prof. Dr. Antoaneta Popova
Technical University – Sofia, Bulgaria

Prof. Dr. Vladimir Poulkov
Technical University – Sofia, Bulgaria

Prof. Dr. Olivera Pronić - Rančić
University of Niš, Serbia

Dr. Zoran Stanković
University of Niš, Serbia

Prof. Dr. Blagoj Stevanovski
University "St.Kliment Ohridski"- Bitola, Macedonia

Prof. Dr. Leonid Stoimenov
University of Niš, Serbia

Prof. Dr. Dragan Stojanović
University of Niš, Serbia

Prof. Dr. Dobrivoje Stojanović
University of Niš, Serbia

Prof. Dr. Gordana Stojanović
University of Niš, Serbia

Prof. Dr. Mile Stojčev
University of Niš, Serbia

Prof. Dr. Milić Stojić
University of Belgrade, Serbia

Doc. Dr. Suzana Stojković
University of Niš, Serbia

Dr. Biljana Stošić
University of Niš, Serbia

Prof. Dr. Georgi Stoyanov
Technical University – Sofia, Bulgaria

Doc. Dr. Suzana Stojković
University of Niš, Serbia

Prof. Dr. Tomislav Šekara
University of Belgrade, Serbia

TABLE OF CONTENTS

VOLUME 1

ORAL SESSIONS

SIGNAL PROCESSING I

SIGNAL PROCESSING II

TELECOMMUNICATION NETWORKS AND SERVICES I

TELECOMMUNICATION NETWORKS AND SERVICES II

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS I

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS II

METROLOGY AND REMOTE SENSING

TELECOMMUNICATION SYSTEMS AND TECHNOLOGIES

VOLUME 2

ELECTRONIC COMPONENTS, SYSTEMS AND TECHNOLOGIES I

ELECTRONIC COMPONENTS, SYSTEMS AND TECHNOLOGIES II

EDUCATION QUALITY

INTERNET TECHNOLOGIES

CONTROL SYSTEMS

COMPUTER SYSTEMS

POWER TRANSMISSION AND DISTRIBUTION SYSTEMS I

POWER TRANSMISSION AND DISTRIBUTION SYSTEMS II

STUDENT SESSIONS

STUDENT SESSION I

STUDENT SESSION II

VOLUME 3

POSTER SESSIONS

PO1 – TELECOMMUNICATION SYSTEMS AND TECHNOLOGIES

PO2 – RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS

PO3 – SIGNAL PROCESSING, METROLOGY AND REMOTE SENSING

PO4 – CONTROL SYSTEMS AND ROBOTICS

PO5 – RENEWABLE ENERGIES

PO6 – COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES

PO7 – ELECTRONIC COMPONENTS, SYSTEMS AND TECHNOLOGIES

PO8 – POWER TRANSMISSION, DISTRIBUTION SYSTEM AND ELECTRICAL MACHINES

PO9 – EDUCATION QUALITY

ORAL SESSIONS

Session SP I:

SIGNAL PROCESSING I

Algorithm for Adaptive Color KLT of Images based on Histogram Matching of the Color Components

Peter N. Ivanov[1], Roumen K. Kountchev[2] and R. Mironov[3]

Abstract – **In this paper is proposed a new algorithm for Adaptive Color Karhunen-Loeve Transform (ACKLT) of RGB components, based on pre- and post-processing, depending of the their power distributions. The pre-processing of the components is made by the histogram of the most powerful component which is used as reference image for the transformation of the other components by the histogram-matching. On the already transformed components is applied ACKLT. Through the proposed algorithm for color transformation we increase the power in the first component at the expense of the others hence we can compress more efficiently the color components. The results demonstrated by the experiments made with the modeling of the algorithm are confirming the advantages.**

Keywords – **Color systems, Adaptive color Karhunen-Loeve transform, Representation of color images components, Histogram matching, Power color components distribution, Enhanced Adaptive Color Transform (EACT).**

I. INTRODUCTION

The transformation of the primary color space *RGB* is a very important part of the image processing. In general the color components in an *RGB* color system are very much correlated. The most important feature of the new EACT (Enhanced Adaptive Color Transform) [1] is the decorrelation of the color components. The proposed In this paper algorithm gives predictability of the power distribution of the color components in the new format. Hence the applications of the method are many: lossy and lossless image compression, color image segmentation, image recognition, and others.

There are two types of color spaces - deterministic and statistical. The deterministic transforms such as *YCrCb*, *YUV*, *YIQ*, *Lab*, *CMYK* [3, 4, 7, 8] are calculated by using fixed coefficients – hence they require less computations in order to make the color transform but the disadvantages are that they are not adapted to the image specific that is being transformed. Also they are unpredictable in terms of knowing which transformed component is the most powerful or in other words which component carries most information. The proposed algorithm has excellent predictability in terms of power distribution i.e. the first component is the most powerful then comes the second then the third etc. Also the EACT is adapted to the statistical properties of every image or group of images that is being transformed hence the transform is more accurate and from that follows the better quality of the restored image in comparison with the deterministic transforms [2] or the power distribution that we obtain by using the ACT algorithm without the pre-processing techniques [3, 7]. The disadvantage is that the algorithm requires a large number of computations for the pre- and post-processing of the color components.

The purpose of this paper is to improve the already created algorithm Adaptive Color Transform (ACT), by using the method for histogram matching as pre-processing technique on selected color components. The goal is to achieve better power distribution of the components which is with very good predictability, something that can be further improved.

This paper is organized in the following manner: in part two we give the algorithm description, in part three are the experimental results in the forms of analysis, graphics and tables. The final forth part gives the conclusions.

II. ALGORITHM DESCRIPTION

The proposed algorithm designed as an improvement of the already created algorithm ACT [1, 2]. It is also important to mention that this algorithm is a complete analytical solution to the problem of the color KLT, presented in [2]. The algorithm is simplified to reduce the necessary computations of the color transform.

Transforming an image into the new color format is made by the following steps, which are the forward part of the EACT.

Step 1: Histogram analysis and selection of the base component.

In order to select the base component which we will use in the following step as a reference image for the histogram matching we must perform histogram analysis of the power contained by all the components. For that purpose we use the following formulas:

$$Pow(R)=\sum_{i=0}^{S}R_i^2, Pow(G)=\sum_{i=0}^{S}G_i^2, Pow(B)=\sum_{i=0}^{S}B_i^2$$

Where R_i, G_i, B_i are the respective color components in the *RGB* image, "*i*" is the current pixel index and *S* is the total number of pixels in the image. For base component we select the component *R*, *G* or *B* for which the power function has greatest value. For that component we don't perform histogram matching, because it is used as a reference image in the next step.

Step 2: Histogram matching and output array expansion.

Let us consider the two histograms $h_{in}(k)$ and $h_{out}(r)$ where the first one is the component that have to be transformed and the second is the histogram of the reference

[1]Peter Ivanov is PhD student in the Faculty of Telecommunications, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: peter.n.ivanov@gmail.com

[2,3]Roumen Kountchev and R. Mironov are professors in the Faculty of Telecommunications, Technical University, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: rkountch@tu-sofia.bg, rpm@tu-sofia.bg.

component we selected. Respectively $l = r = 0,1,...,L-1$ where they are the discrete levels in the two histograms.

In accordance with the method "histogram matching" of a histogram in an image based on the intensity of every pixel with a discrete level "*k*" in first is applied the following non-linear transform:

$$g_1(k) = (L-1)\sum_{l=0}^{k} h_{in}(l) \quad \text{for} \quad k = 0,1,..,L\text{-}1,$$

where L is the number of discrete levels in the image. The function $g_1(k)$ is the first non-linear transform where $h_{in}(k) = n(k)/S$ is the intensity histogram, $n(k)$ is the number of pixels in the image with the same value (intensity) "*k*", and "*S*" is the total number of pixels in the image. This action is known "histogram equalization" or "histogram normalization". On the function $g_1(k)$ we must apply another non-linear transform in order to reshape the original histogram. The relation between the level "*r*" of the transformed intensity and the original level "*k*" is defined by the expression:

$$r = g_2^{-1}[g_1(k)] = g_2^{-1}[(L-1)\sum_{l=0}^{k} h_{in}(l)] \text{ for } k = 0,1,..,L\text{-}1,$$

The function $g_2(r)$ is calculated in the same manner as $g_1(k)$.

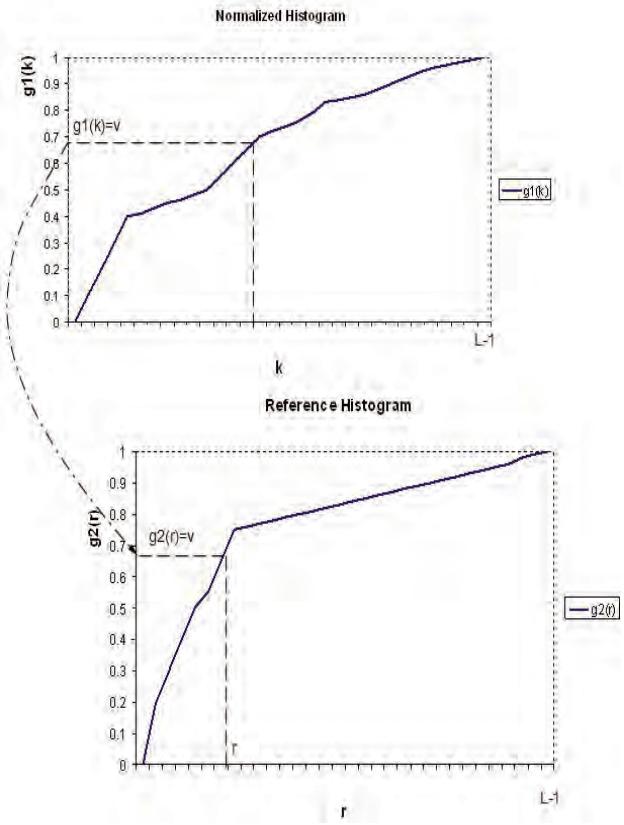


Fig 1. Histogram Matching

We must add that the two functions $g_1(k)$ and $g_2(k)$ must comply with the condition: $0 \le g_1(k), g_2(r) \le 1$ or that the two functions must be monotonically increasing and their values are always between 0 and 1 (because they are normalized).

Before actually applying the method described above, we must add a little customization. In the general case an image in the format RGB is with 24 bpp (bits per pixel) size or 8 bpp per component, so the discrete levels in every component are 256. The use of 256 discrete values in the histogram matching will generate a lot of errors in the restoration of the image. To avoid that, we must increase the number of discrete values in the transformed components. It is a good choice to have 10 bits or 1024 different values for the components that are to be transformed. So, to comply with that condition, we must expand the histograms by linear interpolation of the elements. To do that we must look at the equation $256 + 3 \times 256 = 1024$, hence to every two components in the original histogram we must add another three components. The best choice is to add them in between the known two. Let us consider two neighbor elements in the already normalized histogram $g_1(k)$ or $g_2(k)$, let $g_1(i) = a$ and $g_1(i+1) = e$ where "*i*" is a random number in the interval $0 \le i \le L-1$ and $a \le e$ (which is a given because the function is monotonically increasing). So, let the expanded histogram be denoted as $g_{1\exp}(j)$ where $0 \le j \le 4L-1$ and then let us consider a small section of the histogram, for example the section of elements a,b,c,d,e where "*a*" and "*e*" are already know so the others must be calculated by using the equations: $c = \frac{a+e}{2}, b = \frac{c+a}{2}, d = \frac{c+e}{2}$ and the result is in Fig 2.

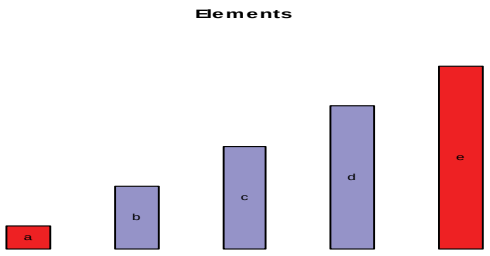


Fig 2. Histogram Expansion

In the end we have a function $g_{1\exp}(j)$ with the same shape as $g_1(r)$ but with more values. The same operation must be applied to the base component histogram in order to have two arrays with the same size for the process of histogram matching.

After completing the histogram matching we will have all the three components of the *RGB* image in the same shape. It is important to note that the base component which is selected as the most powerful component is not transformed by the method. It is passed to the EACT coder as it is. Only the other two components are transformed and expanded. In addition to that we must ensure that the reverse functions can be applied – so we must write into the header the original histograms of the two transformed components in order to be able to construct the inverse functions [3, 7].

Step 3: Calculation of the primary color vectors $\vec{C}_s = [R_s, G_s, B_s]^t$ for each pixel $s = 1,2,..,S$ from the already

pre-processed image, where $S = M \times N$ pixels and M and N are respectively the height and the width of the image [1, 2].

Step 4: Calculation of the image covariance matrix $[K_C] = \frac{1}{S}\sum_{s=1}^{S} \vec{C}_s \vec{C}_s{}^t - \vec{\mu}\vec{\mu}^t$, where $\vec{\mu}_s = \frac{1}{S}\sum_{s=1}^{S} \vec{C}_s$. It is important to mention that the matrix is symmetric across the main diagonal hence the eigenvalues are always real numbers [1, 2].

Step 5: Calculating the coefficients of the characteristic equation of the covariance matrix $[K_C]$ [1, 2].

Step 6: Calculation of the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the characteristic equation defined in the previous step. Given that the matrix $[K_C]$ is a symmetric matrix the eigenvalues can be defined by the "Cardano" relations or the so called trigonometric solution, where we have the condition [1, 2]:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$$

From where we have the predictability of the power distribution. The first component has more power or carries more information than the second etc.

Step 7: Calculation of the eigenvectors $\vec{\Phi}_1, \vec{\Phi}_2, \vec{\Phi}_3$ of the covariance matrix $[K_C]$ and from them forming the transformation matrix $[\Phi]$ [1]:

$$[\Phi] = \begin{bmatrix} \vec{\Phi}_1^t \\ \vec{\Phi}_2^t \\ \vec{\Phi}_3^t \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{bmatrix}$$

Step 8: Performing the color transform $\vec{L}_s = [\Phi](\vec{C}_s - \vec{\mu})$ using the already generated transformation matrix $[\Phi]$ to obtain the transformed color vectors $\vec{L}_s = [L_{1s}, L_{2s}, L_{3s}]^t$ where s is the current pixel in the image [1, 2].

The restoration to the primary format *RGB* is made by two steps. The first one is by applying the inverse ACT the method of which is described in [2]. The second step is applying inverse histogram matching in order to restore the original color values. For this we use the already stored in the file header original histograms of the transformed components and we use them as reference images in the same way that we used the base component in the forward transformation part of the algorithm. It is important to mention that we don't actually need to store the whole two components. We need to store at least one of them and the other can be represented as a difference array. Also, we can apply any form of lossless compression on them in order to reduce their size.

The improvement of the algorithm presented in this paper shows great results in terms of power distribution in the color components, quality of the restored images and excellent predictability of the power distribution in the transformed matrices. More detailed results are shown in the next section which is the experimental results.

III. Experimental Results

As experimental dataset was used three different sets of images of different size – the "Kodak" image set plus the image "Lena" and "Barbara" 26 images in total, of size 512×768 or 768×512, Lena - 512×512 and Barbara - 640×512 pixels. The "cgraph" image set which comprises ten computer generated images of size 1024×768. The "natural" image set which is comprises 10 images of HD size, 1920×1024 pixels. All the images are in the format .bmp (primary format for the RGB color system) with 24 bits describing each pixel (bpp).

For quality measurement was used the Peak Signal to Noise Ratio (PSNR) - equation below. The PSNR gives the objective representation of the restored image quality [3].

$$MSE = \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(X_{i,j} - \hat{X}_{i,j})^2$$

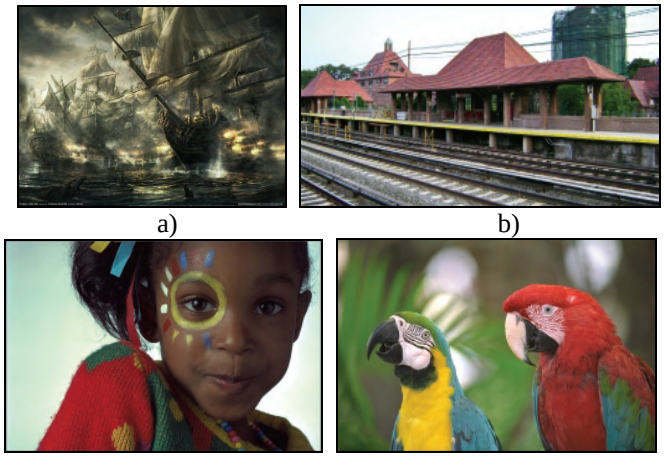
Where the MSE is the Mean Square Error of one component of the *RGB* image. The value $X_{i,j}$ is the original value from the original image and $\hat{X}_{i,j}$ is the restored value in the point *(i, j)* of the image for the given component *R*, *G* or *B*. Therefore there are three different *MSE*s which represent the error in the image. So to find the total error we must add them $(MSE_R + MSE_G + MSE_B)$ [3,7]:

$$PSNR = 10\log_{10}\frac{255^2 \times 3}{MSE_R + MSE_G + MSE_B} \text{ [dB].}$$

For measurement of the power distribution we use the following formula:

$$Pow(L_k) = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(L_{i,j})^2 \text{ for } k = 1,2,3$$ and where *M* and *N* are respectively the width and the height of the image and $L_{i,j}$ is the respective value from the *k*-th matrix.

In Fig. 3 are shown some images from the image toolboxes.



a) b) c) d)

Fig. 3 Images from the image toolboxes: a) cgraph1, b) natural7, c) kodim15, d) kodim23.

TABLE 1.

EACT MEAN PSNR VALUES FOR ALL THE TOOLBOXES

Toolbox	EACT Mean PSNR, dB
Kodak + Lena +Barbara	50.79
Cgraph	51.24
Natural	50.17

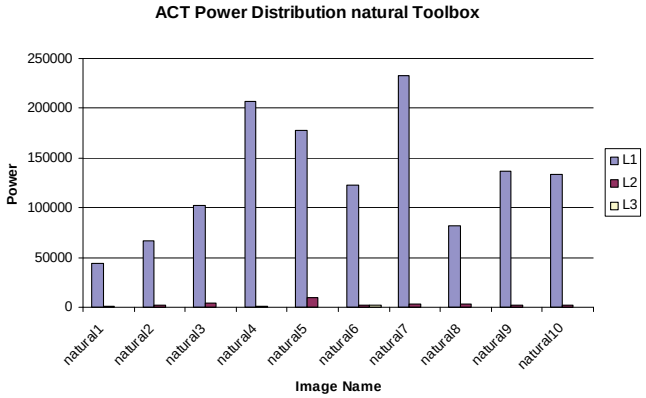


Fig 4. Power Distribution "Natural Toolbox" EACT

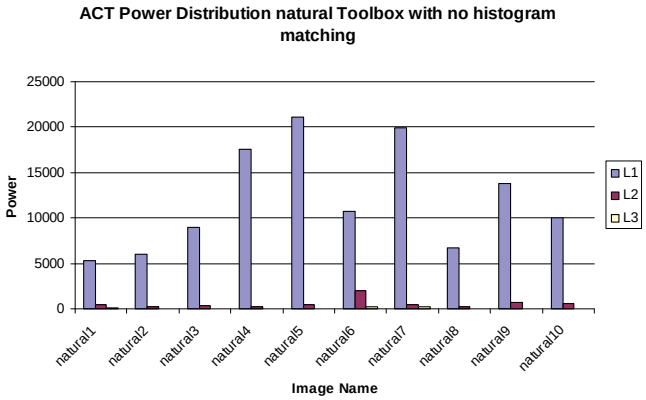


Fig 5. Power Distribution "Natural Toolbox" ACT without histogram matching

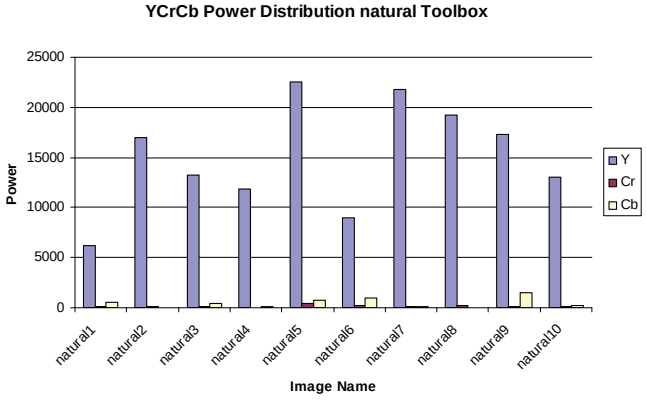


Fig 6. Power Distribution "Natural Toolbox" YCrCb

From the chart in Fig. 4 we can see that power distribution in the transformed components strongly favors the $[L_1]$ matrix, and then comes the $[L_2]$ and the $[L_3]$ matrix is not visible in the charts. From here we can tell that the power distribution that we obtain from the improved algorithm presented in this paper is with excellent predictability in terms of the power distribution as we mentioned above. We can also make a comparison with the power distribution of the ACT Algorithm presented in [1, 2] and the results prove the effectiveness of the new algorithm, Fig 5 and the YCrCb power distribution mentioned above in section two, Fig 6. Also the power ratio between the transformed components is very high. We must add that the quality of the restored images is very good as we see from Table 1.

IV. CONCLUSION

In the proposed algorithm the transformation is adapted for each image that is being transformed therefore the transform matrix generated by the algorithm is adapted to the statistical information of the image hence the transform is more accurate. Also we must add that the demonstrated in the previous part power distribution in the matrices allows the algorithm to be used very effectively for image compression and/or processing. The results of the algorithm in terms of image quality and processing time can be further improved by finding an efficient way to decide when the histogram-matching will be an advantage. Something that can be described in future papers.

ACKNOWLEDGEMENT

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science, Contract DNTS 02/19.

REFERENCES

[1] R. Kountchev, R. Kountcheva. "New Method for Adaptive Karhunen-Loeve Color Transform". Proc. of the 9th Intern. Conf. on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS'09) IEEE, Nis, Serbia, Oct. 07-09, 2009, pp. 209-216.

[2] P. Ivanov, R. Kountchev, "Comparative analysis of Adaptive Color KLT and YCrCb for representation of color images", XLV Intern. Scientific Conf. on Information, Communication and Energy Systems and Technologies (ICEST'10), Ohrid, Macedonia, 23-26 June 2010, SP I, pp. 29-32.

[3] D. Carevic, T. Caelli, "Region based coding of color images using K-L transform, Graphical Models and Image Processing"- Vol. 59, No.1, 1997, pp. 27-38.

[4] W. Pratt, "Digital Image Processing", Wiley Interscience, New York, 2007.

[5] M. Fleury, A. C. Downton and A. F. Clark, "Karhunen - Loeve Transform - Image Processing", University of Essex, Wivenhoe Park, UK 1997.

[6] R. Dony. "Karhunen-Loève Transform". Book Chapter in The Transform and Data Compression Handbook, Ed. K. Rao and P. Yip, Boca Raton, CRC Press, 2001.

[7] R. Gonzalez, R. Woods, "Digital Image Compression" Prentice Hall, Second Edition, 2002.

[8] S. Sangwine, R. Horne, "The colour image processing handbook", Chapman & Hall, 1998.

Enhanced Predictive Block-based Encoding for Stereo Image Compression

Alexander A. Krupev[1] Antoaneta A. Popova[2] and Ivo R. Draganov[3]

Abstract – **The block-based predictive encoding exploits both inter and intra-frame correlation to compress stereoscopic image pairs, similarly to the MPEG standard. Two variations of the method are reviewed and tested in this paper – conventional disparity estimation technique and pioneering block-based encoding. The later does not require extra overhead bits for disparity estimation - it is done in the decoder using information from previously received and decoded blocks. Reported is an enhancement of the method using a bicubic resizing block for extra compression and speed.**

Keywords – **predictive encoding, stereo image compression, DCT, disparity estimation, disparity compensation**

I. INTRODUCTION

Stereo pairs consist of two similar images intended for each eye in order to achieve depth perception. This information redundancy is not well exploited by the conventional compression standards like JPEG, GIF and others [1], [3].

A common approach is to compress one image independently (called reference image), then to compress the second image using the knowledge about the first image. For this purpose we use predictive encoding by searching for similar 8x8 blocks between the two images and transmitting their residuals along with the reference image. The DCT transformed and quantized residual blocks are usually sparse matrices (if the stereo matching is performed correctly) and the few non-zero coefficients are compacted, which is a prerequisite for efficient RLE lossless coding.

The pioneering block-based version of the algorithm, first proposed by Jiang et al. [1], uses neighboring block search instead of direct matching in order to avoid sending disparity vectors to the decoder. As we shall see, this is possible because we assume that a standard stereo setup is used (parallel camera optical axes) - the corresponding pixels/blocks lie in the same row. Thus the disparity search range is 1-dimensional. To accelerate the matching block search and achieve extra compression, we added a resize module to the scheme. Human vision is able to compensate the lack of higher frequency information in one of the images if it is present in the other [4], [5]. This allows us to downsize one of the images in the process of encoding and upsize it to its initial resolution again before stereoscopic visualization.

Our goal is to suggest a predictive algorithm for stereo image compression and simulate its impact on stereo perception. The evaluation of the reconstructed compressed images is still an open problem. We employ an objective measure (PSNR), but it doesn't take into account the peculiarities of the human visual system [5]. Subjective experiments are also performed.

[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

[3]Ivo R. Draganov is with the Faculty of Communications, Technical University - Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: idraganov@abv.bg

II. PREDICTIVE BLOCK-BASED ENCODING

The general scheme for block-based predictive encoding with a bicubic resize module, proposed by us, is given below:

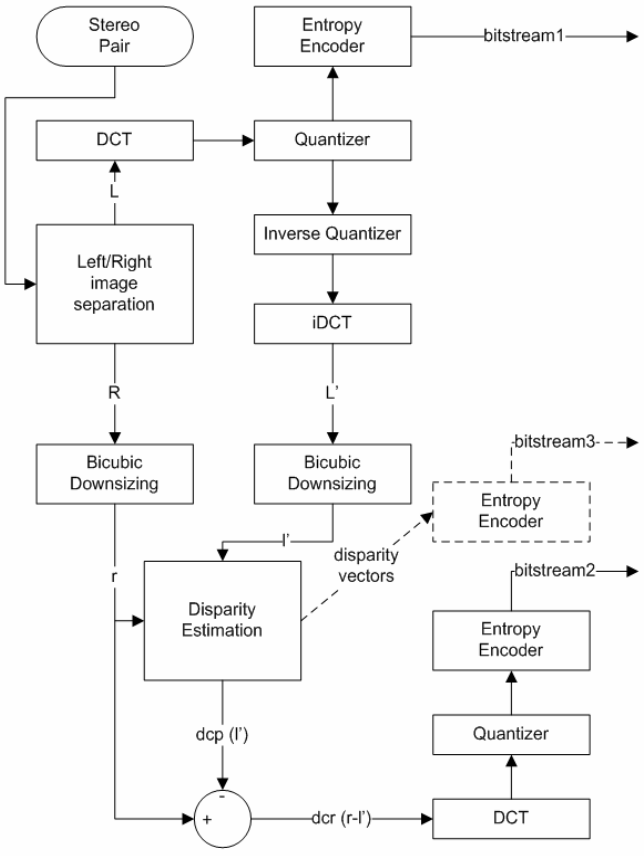


Fig. 1. Disparity compensated DCT encoder for stereo image compression

First the stereo image is separated into left (L) and right (R) images for each eye respectively. L is chosen as a reference image. It is separated into 8x8 blocks, which are DCT transformed and quantized. The intra-frame quantization matrix used as per the MPEG standard [7] is shown in Eq.1.

$$Q = \frac{1}{S}\begin{bmatrix} 8 & 16 & 19 & 22 & 26 & 27 & 29 & 34 \\ 16 & 16 & 22 & 24 & 27 & 29 & 34 & 37 \\ 19 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\ 22 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\ 22 & 26 & 27 & 29 & 32 & 35 & 40 & 48 \\ 26 & 27 & 29 & 32 & 35 & 40 & 48 & 58 \\ 26 & 27 & 29 & 34 & 38 & 46 & 56 & 69 \\ 27 & 29 & 35 & 38 & 46 & 56 & 69 & 83 \end{bmatrix}$$

$$S \in [1,\ 16] \tag{1}$$

S is the scale factor – the higher it is, the smaller the quantization steps are, which leads to higher quality and size of the output image. The fraction values are rounded to integers. The quantized values are sent to the lossless entropy encoder (zig-zag scanning, RLE and Huffman coding). The DC coefficients are DPCM coded separately before being sent to the entropy encoder as per JPG standard [6].

To minimize the prediction error, the left image blocks are dequantized and inverse DCT transformed in the encoder. Thus we will be working with the same reference image L' at the encoder and the decoder. Before disparity estimation R and L' can be downsized to increase compression. This will affect the quality of the right image only. Disparity estimation is performed on the downsized left an right images – l' and r. The right image is separated into 8x8 blocks. Let the row/column indices of a current block from r be m and n. The best match search in the left image is performed in the same row – m, on pixel basis. The minimal SSD (Sum of Square Differences) is chosen as a matching criterion (Eq.2):

$$ssd_{m,n} = \sum_{a=0}^{7}\sum_{b=0}^{7}(r_{8m+a,8n+b} - l'_{8m+a,8n+b+k})^2 \tag{2}$$
$$where \qquad k \in [-\lceil 0.5N \rceil, \lceil 0.5N \rceil]$$

The image is composed of $M \times N$ blocks, k represents the disparity vector. It is chosen to be within 1/8 of the image width interval and thus defines the search area. It's unlikely the disparity to be higher for a correctly composed stereo pair. The search area in this case belongs to the interval:

$$search\ \ area_{m,n} \in [8n - \lceil 0.5N \rceil,\ 8n + \lceil 0.5N \rceil + 7] \tag{3}$$

After the best match is found, the disparity compensated residual (dcr) $r - l'$ is formed (Eq. 4)

$$dcr_{m,n} = r_{m,n} - l'_{m,n'} \tag{4}$$

n' denotes the disparity relative to n. The residuals are DCT transformed, quantized, losslessly encoded and sent to the decoder. The quantization is uniform with default step value of 16 (inter-frame quantization, [7]), which can be scaled down also, like in Eq. 1. The disparity estimation block also sends the disparity vectors to the decoder, being DPCM encoded first. That's useful because clusters of blocks, belonging to the same objects, are displaced equally [7].

In the pioneering block-based version of the encoder, the disparity vectors (bitstream3) are not sent to the decoder. They are restored there using information from previously decoded blocks. That's why the disparity estimation process uses neighboring blocks to $r_{m,n}$, namely $r_{m-1,n}$ and $r_{m,n-1}$, because they will be already present at the decoder and the vectors could be reproduced there. The pioneering block based disparity estimation, as proposed by Jiang et al. is shown on Fig. 2.

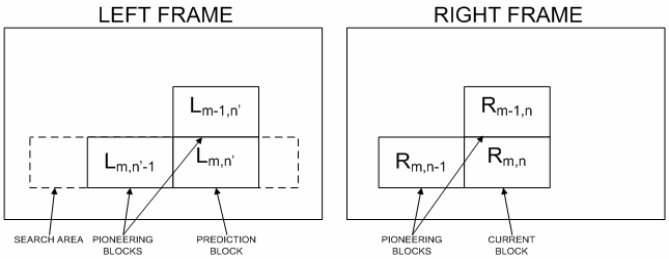


Fig. 2. Pioneering block-based disparity estimation

In our implementation we use the three neighbouring blocks to $r_{m,n}$, to improve precision (Eq. 5).

$$PR_{m,n} = \frac{r_{m-1,n} + r_{m-1,n-1} + r_{m,n-1}}{3} \tag{5}$$

This pioneering block $PR_{m,n}$ is used for disparity estimation, instead of the current block $r_{m,n}$. Analogically, a pioneering block $PL_{m,n'}$ is formed in the left image l', within the search window (Eq. 3) and SSD criterion is used again (Eq. 6) to find the best match.

$$ssd_{m,n} = \sum_{a=0}^{7}\sum_{b=0}^{7}\left(\frac{r_{8(m-1)+a,8n+b} + r_{8(m-1)+a,8(n-1)+b} + r_{8m+a,8(n-1)+b}}{3}\right. \tag{6}$$
$$\left. - \frac{l'_{8(m-1)+a,8n+b+k} + l'_{8(m-1)+a,8(n-1)+b+k} + l'_{8m+a,8(n-1)+b+k}}{3}\right)^2$$

In the decoder (Fig.3), the left reference image L' is restored after inverse quantization and inverse DCT. It's ready to be displayed, but is also used in the restoration of the right image. L' is downsized to the same size as in the coder and l' is sent to the disparity estimation block as well as the restored residuals (dcr'). The disparity is estimated the same way as in the encoder. The effect of the quantization error of the left image on the estimation process is reduced, because L' (not L) was used in the encoder for prediction. But the matching process can be affected by the quantization error in the residual image blocks (and thus indirectly by L'). There is also an error distribution effect when a disparity vector is not correctly restored. The dependence of the correctly restored vectors on the quantization step (through the scaling coefficients s for intra and z for inter-frame quantization of the residuals) is shown in the experimental results.

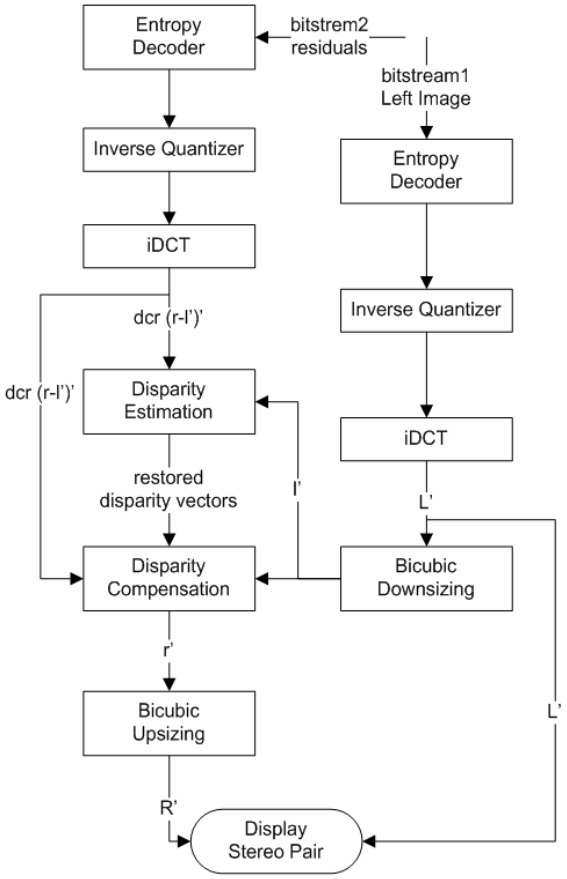


Fig. 3 Pioneering block-based disparity decoder for stereo image compression

Using the restored disparity vectors, the restored disparity compensated residuals dcr' and the downsized left image l', the disparity compensation block restores the right image r' (Eq. 7). It's upsized before being displayed as R'.

$$r'_{m,n} = l'_{m,n'} + dcr'_{m,n} = l'_{m,n'} + (r_{m,n} - l'_{m,n'})' \quad (7)$$

The conventional non-pioneering predictive decoder is a simpler version of the decoder from Fig. 3. Since the vectors are sent from the coder, the disparity estimation block is omitted. There is no risk to apply mismatched residuals in Eq.7. The compression is slightly lower, of course.

III. Experimental Results

The schemes from Figs. 2 and 3 are implemented in Matlab 7.0.3 working environment. The images are processed in RGB color space and grayscale intensity values are used for disparity estimation. The lossless entropy encoders have not been implemented since our main goal is to evaluate quality.

A PSNR evaluation module is added, comparing the processed images R' and L' to the original R and L images according to Eq. 8:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)[dB] \quad (8)$$

$$MAX_I = 255$$

Here MSE is:

$$MSE = \frac{1}{pq}\sum_{i=0}^{p-1}\sum_{j=0}^{q-1}[I(i,j) - \hat{I}(i,j)]^2 \quad (9)$$

p, q are the dimensions of the images, I and $\hat{I}$ - pixel intensity values before/after processing, ranging from 0 to 255. The images are turned to grayscale before evaluation.

The stereo pair used for the experiments is shown on Fig. 4:

Fig. 4 Mountain stereo pair, used for the experiments

For the first experiment, pioneering encoding is applied without resizing. The disparity vector tables at the encoding and the decoding ends are compared for different quantization scale coefficients (greater scale coefficient gives finer quantization - Eq.1). z affects residual quantization, s – the left/reference image quantization. The results are shown in Table 1.

TABLE I

MISMATCHED BLOCK PERCENTAGES

Inter scaling coefficient *(z)*	*z=1*	*z=8*	*z=16*
Intra scaling coefficient, *s=1*	21.2%	12.14%	3.93%
Intra scaling coefficient, *s=4*	20.35%	5.98%	0.97%
Intra scaling coefficient, *s=8*	17.39%	6.34%	2.17%
Intra scaling coefficient, *s=12*	16.79%	9.24%	4.11%
Intra scaling coefficient, *s=16*	21.27%	14.13%	7.42%

The correctly restored vectors percentage always increases with the decrease of the inter-frame quantization step (increased *z*). The effect of the inter-frame reference image quantization on the matching process is content-based/random though, since l' participates in the quantized residuals. For coarse residual quantization ($z = 1$), it is dominated by the quantization error of the residual, though. Averagely about 1/7th of the blocks are mismatched due to incorrectly restored vectors. This may manifest itself as blockish artifacts in the image (Fig. 5, Fig 7-d). It doesn't hamper stereo perception, but could be unpleasant for the viewer.

Fig. 5 Restored stereo pair for s, z=1, pioneering encoding, no bicubic resizing of the right frame

The second experiment involves measuring PSNR as a function of the quantization step, which is controlled by the coefficients s and z. The results are shown on Fig. 6. For single images optimal quality/compression is achieved for PSNR between 30-60dB [6]. This applies for the reference image (left) of the stereo pair, but the requirements for the right image are lower. That's because human vision can compensate high frequency loss from the other image.

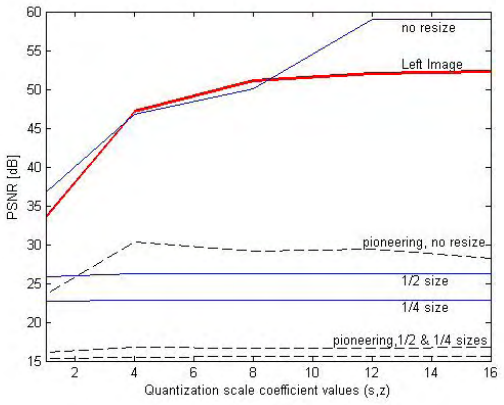


Fig. 6 The dependence of PSNR on the quantization scale coefficients s and z. Assumed is s=z.

The left image is always identical for all stereo pairs. Evidently pioneering block-based encoding does not give satisfactory results in combination with bicubic resize. For $s, z \in (4,6)$ without resizing, though, its PSNR is above 30dB. But without resizing, conventional disparity estimation technique yields 45dB in this case! The compression might be slightly lower because the vectors have to be transmitted, but it's worth the few extra bits. It even surpasses the reference image in quality, when smaller quantization steps are used. The error, induced from resizing, clearly outweighs the quantization errors – the graphs are almost flat lines for ½ and ¼ resizing. Significant compression is achieved this way.

For the purposes of subjective evaluation the processed stereo pairs are visualized using the anaglyph method, based on color separation [2]. Some of the visualized right images are shown on Fig. 7. In our opinion for ½ resize, s=4, z=1, optimal results are achieved (considering not only perceived quality, but compression). For ½ downsizing and pioneering encoding the stereo perception is preserved, but there are lots of visible annoying artifacts. Using ¼ resize yields significant granular noise perception. The stereo pair is very hard to focus on and eye strain occurs fast.

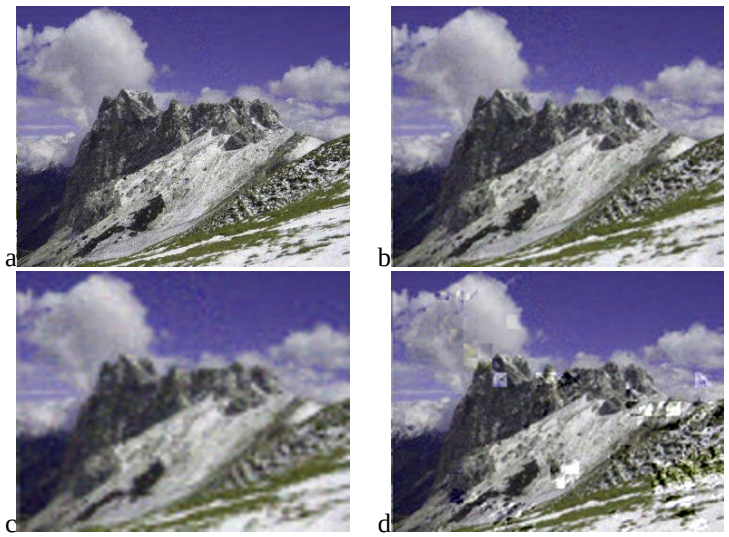


Fig. 7 Right images for s,z=1; a) no resize; b) ½ resize; c) ¼ resize; d) pioneering encoding, ½ resize

IV. Conclusion

Until further improved, the authors would rather employ the conventional predictive encoding over the pioneering block-based prediction scheme. These algorithms can be applied together with other standard compression technologies besides DCT based JPEG, like fractals, wavelets etc. to achieve extra compression for all right frames.

Acknowledgement

This paper was supported by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science (Contract – DDVU 02/13 – "Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization").

References

[1] J. Jiang, E. Edirisinghe, and H. Schroder, "Algorithm for compression of stereo image pairs" *Electronics Letters*, 33(12):1034-1035, September 1991.

[2] 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.

[3] M. Bax and A. Vitus, "Stereo Image Compression", http://www.stanford.edu/~mbax/ee392c/report.html, Stanford, Dec. 1997.

[4] I. Dinstein, M.G. Kim, A. Henik, and J. Tzelgov, "Compression of stereo images using subsampling transform coding," Optical Engineering, vol. 30, no. 9, pp. 1359-1364, Sept. 1991.

[5] L. Kaufman, *Sight and Mind: An Introduction to Visual Perception"*, Oxford University Press, 1974.

[6] W. Pennebaker and J. Mitchell, "*JPEG still image data compression standard (3rd ed.)"*, Springer. , 1993.

[7] J. Mitchell, W. Pennebaker, C. Fogg, D. LeGall ,"MPEG video compression standard", Kluwer Academic Publishers, 2002

Efficient Compression of Medical Images Based on Adaptive Histogram Modification

Roumen Kountchev[1], Roumen Mironov[1] and Roumiana Kountcheva[2]

Abstract – **In this paper is presented one new method for efficient compression of visual medical information, based on adaptive histogram modification. After that the images are losslessly compressed and saved in a new format. The restored images are then processed so that to enhance their contrast. In result of the pre-processing the compression ratio is significantly increased, retaining the visual quality of the image.**

Keywords – **Image pre-processing, Lossless image processing, Archiving of medical images.**

I. INTRODUCTION

Governments worldwide are pushing to computerize paper-based medical records in order to make them available for easy and reliable exchange. Paper-based medical record serves all the essential issues of medical care. They are used to record each patient's health status, and allow caregivers to offer the appropriate treatment to patients.

Although paper-based medical records played an important role in health care, it has several significant limitations like:

- Fragmented.
- Limited availability - only at one place at a time.
- Poorly indexed.
- Often illegible.

In many countries, the current implementations of inter-institutional transfer, share and use of patient's data, the information exchange is still performed mainly manually. This process is tedious, time consuming, error-prone and therefore expensive. In addition, the importance of linking medical specialists lies in speeding up the extraction of comprehensive patients' information where the time is critical in the medial care process. Since Electronic Medical Records (EMR) are independently developed, they often have different structure and semantics. There is a lack of communication and integration between these heterogeneous information systems. Current Health Care Information Systems (HIS) provide only local data within the medical institute. This situation creates a reality in which the medical record is scattered in various locations, different structures and semantics, and among many different caregivers. Beside, patient's can only be given care after reviewing its medical history. This would lead to a situation where patient can only get medical care in institutions that obtain a medical record, even in case of emergencies. Increasingly, healthcare organizations are considering moving to an EMR. EMR is a computer-based patient medical record that serves as the primary source of information for effective patient care. The purpose of an EMR is to gain the benefits of paper-based medical record and overcome its limitations. Thus, EMR is a single, permanent, legal document. Thus, a first step in integrating EMR is to identify and characterize their inter-schema and semantic relationships, which involves ontology mapping.

Recent scientific literature suggests that Electronic Health Record (EHR) implementation will optimize healthcare delivery to individual patients [1]. It is expected that EHRs will decrease the cost of care, increase the quality of life and allow the portability of records, while maintaining privacy of some medical information. Such systems would connect hospital patient data with clinical data as the patient moves between providers and patient care establishments. Interoperable EHRs have the potential to promote access to more detailed and accurate patient information at the time of treatment, to reduce medical errors and to improve the overall healthcare quality.

Considering the importance of the patient's medical information for the caregivers to ensure that patients receive the appropriate treatment in light of their medical history, sharing distributed medical information among caregivers is essential. On the other hand, these caregivers suffer from communication gaps and the heterogeneity of their medical records. As a solution, there has been proposed new approach for archiving and standardization of medical records. Other techniques are proposed to achieve partial automation of interoperability based on medical record schema or semantics.

The prevailing part of the information, stored in the medical databases is visual. In correspondence to the contemporary trends and requirements, the hospitals and other healthcare institutions have to maintain significant number of electronic files, comprising various kinds of information: images (ultrasound, CT, X-Ray etc.), ECGs, EEGs, scanned documents, prescriptions, and many others. Two main problems arose when all this information is archived and stored: how to archive all this information efficiently in order to create databases of size as small as possible, retaining the visual quality unchanged, and how to ensure the needed confidentiality for the patients' information.

The first problem is solved using some kind of image and data compression. The most famous standard, used for medical imagery, is DICOM, based on the JPEG standard [2,3]. As it is very well known, the JPEG-based compression is highly efficient and offers restored images with retained visual quality.

The paper presents one new approach for image pre- and post-processing aimed at the efficient compression of the medical images. The paper is arranged as follows: in Section 2 is presented the method for adaptive histogram modification

[1]Roumen Kountchev and Roumen Mironov are with the Faculty of Telecommunications, Technical University of Sofia, Bul. Kl. Ohridsky 8, Sofia 1000, Bulgaria, E-mails: rkountch@tu-sofia.bg, rpm@tu-sofia.bg

[2]Roumiana Kountcheva is with T&K Engineering, Mladost 3, POB 12, Sofia 1712, Bulgaria, E-mail: kountcheva_r@yahoo.com.

and the adaptive contrast enhancement of the restored images; in Section 3 are given some experimental results and Section 4 contains the Conclusions.

II. ADAPTIVE HISTOGRAM MODIFICATION

Typical medical images (X-ray, MRI, etc.) are of low contrast and possess significant noises. The visual quality of the medical images is not high and for the better image understanding various techniques aimed at the improvement of their visual quality are used. In this paper is offered a method for image preprocessing, which results in high compression ratio with retained visual quality. Besides, the restored image could be processed in order to enhance the image contrast also.

An Example medical ultrasound image is shown on Fig. 1. The image is of size 350×432 pixels, 24 bpp. The image contrast is relatively low. The image histogram is shown on Fig. 2.

As it is seen from Fig. 2, the brightness values, which comprise the meaning information in this test image, are in the range 37-168, i.e. about 50% of the standard 255 values.

Fig. 1. Test ultrasound image “Axial 2” of size 344× 430 pixels

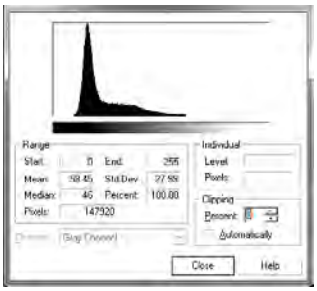


Fig. 2. The histogram of the test image from Fig. 1

The histogram modification, proposed in this work, comprises the following steps.

First, the image contrast is evaluated. As a rule, the image contrast enhancement is performed in cases, when the corresponding contrast coefficient K is low. This coefficient is calculated in accordance with the relation [4]:

$$K = \frac{k_{\max} - k_{\min}}{k_{\max} + k_{\min}} \ ,$$

where k_{max} and k_{min} are correspondingly the maximum and the minimum brightness levels in the processed image.

In case, that for some image $K<K_0$ (where K_0 is a threshold value, defining the image as one with low contrast), its brightness histogram $h(k)$ is strongly distorted and has an outlined maximum, h_{max}. The medical images (ultrasound, X-ray, MRI, etc.) are of low contrast and need corresponding processing. The processing is described below.

The image histogram is calculated in accordance with the relation:

$$h(k) = n(k)/n \ \text{ for } k = 0, 1, 2, \ldots, k_{max} \ ,$$

where $n(k)$ is the number of the pixels in the discrete brightness level k, n is the total number of the image pixels and k_{max} is the maximum number of brightness levels.

Then, (for $K<K_0$, which is the case of the low-contrast images) is performed image segmentation, using the thresholds k_1 and k_2, in result of which the brightness range is divided into three segments (A, B, C). The algorithm for the calculation of the thresholds is presented as follows:

- The maximum of the histogram is defined:

$$h_{max} = max\{h(k)\} \text{ for } k = 0, 1, 2, \ldots, k_{max},$$

- The value of the segmentation boundary is calculated as $t = \alpha h_{max}$, where $\alpha < 1$ (0.01%).
- The values of the thresholds k_1 and k_2 are defined in accordance with he relations:

$$h(k) \le t \ \text{ for } \ k = 0, 1, 2, \ldots, k_1\text{-}1,$$

$$h(k) \ge t \ \text{ for } \ k = k_2\text{+}1, k_2\text{+}2, \ldots, k_{max},$$

The histogram is then modified as follows:

The brightness values, placed in parts of the histogram, outside the "meaning" parts of the histogram are strongly reduced - for every 10 or 12 values one representative only is selected. The belonging to the closest representative is done following the rule for arithmetic rounding.

Similar approach is used for the "meaning" part of the histogram, but in this case the representatives are usually placed at a distance of 4 or 5 values.

With this, the preprocessing is finished. In result of the preprocessing the visual quality of the processed images is retained (the PSNR remains higher than 38 dB, which means, that the histogram modification is practically unnoticeable.

After the preprocessing, the image is losslessly compressed using the Adaptive Run-length Coding, developed by the authors [5].

After decoding the image is restored with retained visual quality. For better image representation is used adaptive image histogram enhancement. For this, the gray-level k for every pixel in the already defined histogram segments A, B and C is transformed in accordance with an individual table, as follows:

$$g(k) = \begin{cases} g_A(k) & if \quad 0 \le k < k_1; \\ g_B(k) & if \quad k_1 \le k \le k_2; \\ g_C(k) & if \quad k_2 < k \le k_{max}. \end{cases}$$

Here $g_A(k)$, $g_B(k)$ and $g_C(k)$ are the corresponding tables for brightness transformation in the segments *A*, *B* and *C*. In order to improve the image quality of the low-contrast image areas the thresholds k_1, k_2 of the segment *B* are widened (stretched) up to $(k_1-\delta_1) \geq 0$ and $(k_2+\delta_2) \leq 255$ shrinking the segments *A* and *C*. In this case δ_1 and δ_2 are parameters, which define the contrast enhancement for the segment *B* and correspondingly – the change of the contrast range for segments *A* and *C*. Each table for brightness transformation is defined in accordance with the condition for histogram equalization for the corresponding image segment *A*, *B* or *C* with changed (stretched or skewed) dynamic range:

$$g_A(k) = (k_1-\delta_1)\sum_{l=0}^{k} h_A(l),$$

$$g_B(k)=(k_2-k_1+\delta_1+\delta_2)\sum_{l=k_1-\delta_1}^{k} h_B(l) + (k_1-\delta_1),$$

$$g_C(k) = (k_{\max}-k_2-\delta_2)\sum_{l=k_2+\delta_2}^{k} h_C(l) + (k_2+\delta_2).$$

In particular, in case that the histogram of the corresponding segment is constant, i.e. for:

$$h_A(k)=\frac{1}{k_1} \quad for \quad k=0,1,..,k_1-1;$$

$$h_B(k)=\frac{1}{k_2-k_1} \quad for \quad k=k_1,k_1+1,..,k_2;$$

$$h_C(k)=\frac{1}{k_{\max}-k_2} \quad for\ k=k_2+1,k_2+2,..,k_{max},$$

the table for the brightness transform for each segment is linear and the relations are defined as:

$$g_A(k) = \left(\frac{k_1-\delta_1}{k_1}\right)k;$$

$$g_B(k)=\left(\frac{k_1-k_2-\delta_1-\delta_2}{k_1-k_2}\right)k+\left(\frac{\delta_2 k_1+\delta_1 k_2}{k_1-k_2}\right);$$

$$g_C(k)=\left(\frac{k_{\max}-k_2-\delta_2}{k_{\max}-k_2}\right)k+\left(\frac{\delta_2 k_{\max}}{k_{\max}-k_2}\right).$$

In this case the brightness levels in the range (k_1, k_2) are stretched in accordance with a linear relation and corresponding inverse operations are performed for the two remaining histogram segments (0, k_1-1) and (k_2+1, k_{max}).

III. Experimental results

For the experiments were used more than 300 test images (ultrasound, X-ray and MRI) of various sizes (bmp, 24 bpp). The presented method was implemented in software (*Visual C*, *Windows* environment). The processed images were archived in a specially developed format (*tk*). For the experiments were used the image database of the Technical University of Sofia and some free medical databases accessible via Internet.

On Fig. 3 is shown the test image "Axial 2" obtained after histogram modification, and on Fig. 4 – the corresponding histogram. In spite of the histogram change (Figs. 2 and 4), the visual image quality is retained. The compression ratio obtained for the processed image after lossless compression is much higher than that for JPEG2000LS (for JPEG, which is the basis of the DICOM standard, the compression efficiency of the new approach, is better also).

Fig. 3. Test ultrasound image "Axial 2" obtained in result of the histogram modification (PSNR= 42dB)

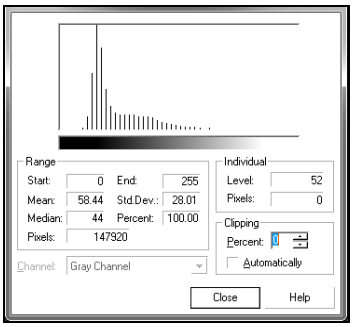


Fig. 4. The modified histogram of the test ultrasound image

The compression efficiency was compared with that of JPEG2000 LS. Some of the results obtained are shown in Table 1 and their graphic representation - on Fig. 5.

Image		Size [pixels]	$PSNR_{MH}$ [dB]	CR_{MH}	CR_{JP2K}	Size [KB]
1	Axial2	350×432	43,5	13,04	5,7	34,1
2	Axial8	344×430	42,20	12,08	5,6	36,7
3	Thorax	280×301	44,78	10,8	6,8	23,3
4	1-vh	359×283	40,44	14,1	8,81	21,57
5	1. 135	512 ×512	39,02	15,5	5,85	50,8
6	1.136	512 ×512	39,44	10,6	5,26	74
7	1.138	512 ×512	39,33	13,3	5,27	59,0
8	1.139	512×512	39,62	22,83	8,49	34,4
9	1.198	512×512	39,26	18,04	7,15	43,5
10	1.826	512 ×512	39,86	11,95	4,90	65,8

On Fig. 6 is shown the image obtained after decoding and contrast enhancement and on Fig. 7 is shown the corresponding histogram. The image histogram is widened with about 10% of the part of the histogram, corresponding to the original image. Further stretching of the histogram could produce visible distortions. The visual quality of the restored image (Fig. 6) is much better than that of the original (Fig. 3).

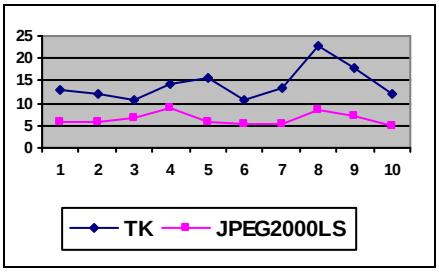


Fig. 5. Graphic representation of the results obtained after lossless compression with JPEG 2000 and the new approach (TK)

Fig. 6. Test ultrasound image (Fig. 3) obtained after decoding and contrast enhancement

Fig. 7. The histogram of the test image after contrast enhancement

For comparison, same test image was processed with Corel Photo Paint (Contrast Enhancement). It is easy to notice that the image on Fig. 8 has visible distortions, while the images with contrast enhancement in accordance with the method, presented above, have no such changes.

IV. Conclusion

In the paper is presented one new method for preprocessing of medical images aimed at their efficient archiving with retained visual quality.

The method was implemented in software (Visual C. Windows environment). A special format was developed for the coded images, which will permit the creation of efficient contemporary medical image databases. Significant number of tests was performed with various kinds of medical images, with different statistical properties. The software implementation of the method confirmed its efficiency in all cases.

The method efficiency was compared with that of JPEG2000. Some experimental results are presented in the paper.

The visual quality of the restored images is improved after their decoding using the special method for adaptive histogram enhancement, whose basic algorithm is presented in the paper. In result, the visual quality of the images was improved.

Fig. 8. The test ultrasound image "Axial2" obtained after contrast enhancement with Corel Photo Paint

Acknowledgement

This research work was supported in part by the Joint Research Project Bulgaria-Romania: "Electronic Health Records for the Next Generation Medical Decision Support in Romanian and Bulgarian National Healthcare Systems".

References

[1] Information Technology–JPEG 2000 "Image Coding System: Part 9–Interactivity tools, APIs and protocols", No. 15444-9, ISO/IEC JTC1/SC29/WG1 IS, final publication draft, rev. 3, 2003.

[2] Camapum, J. Rocha, A. Evangelista, N. Amemiya, E. "DICOM database for medical imaging". XV Congreso Argentino de Bioingieneria, 2009, pp. 1-4.

[3] ACR-NEMA "DICOM Digital Imaging and Communications in Medicine", 2003.

[4] William Pratt. "Digital Image Processing". John Wiley and Sons 2001.

[5] Vl. Todorov, R. Kountchev, M. Milanova, R. Kountcheva, C. Ford. Method and Apparatus for Contents-Based Run-Length Data Encoding, US patent 20070279261; June 2006.

3D Digital Filtering of Volumetric Images

Dimitar G. Valchev

Abstract – **This paper shows three different kinds of three-dimensional (3D) digital filters for processing of volumetric images. Ideal impulse responses are given for a separable, a non-separable and a semi-separable 3D filter.**

Keywords – **Digital filters, Digital image processing.**

I. INTRODUCTION

Volumetric images are characteristic to many applications in computer tomography and magnetic resonance imaging, as well as in subsurface sensing and imaging of various hidden spaces. These volumetric images are represented as three-dimensional (3D) arrays of voxel values. They give realistic representations of solid shapes as true 3D images, not just two-dimensional (2D) projections onto a planar display consisting of pixels. Being essentially 3D signals, volumetric images are subject to 3D signal processing [1]. The various filtering techniques from one-dimensional (1D) and 2D digital signal processing (DSP) [2, 3] may easily be extended to three dimensions. Filters used in 3D volumetric digital image processing share many common properties with their 1D counterparts and hence the various design approaches for 3D filters can be closely related to 1D and 2D algorithms.

On the other hand, there are some differences from the 1D and 2D cases, due to the additional dimensionality in 3D digital filtering. 3D filters are characterized by more degrees of freedom that can be used to achieve design optimality. Also, the available mathematical tools are more restrictive which further complicates the design process. Since 3D polynomials can not be factorized, the designer must consider the implementation aspect, especially separability. Again, due to the increased dimensionality and hence number of degrees of freedom, the separable filter can be fully separable in the three dimensions, or semi-separable – non-separable in two dimensions but separable in the third one.

The rest of the paper is organized as follows. Section II gives the basics of 3D digital filtering. Section III presents different kinds of separability in 3D digital filters. Section IV concludes the paper pointing at further research aspects in 3D digital filtering.

II. BASICS OF 3D DIGITAL FILTERING

Common to all 3D digital filtering techniques is the frequency response of an ideal lowpass 3D filter. Starting with it, various techniques known from 1D DSP [4] can be applied. Then, based on the derived impulse response, various highpass, bandpass and bandstop filters can be designed. The impulse responses of the 3D filters as well as the volumetric images are defined on generally rectangular sampling grids. The sampling grids can also be non-rectangular [2, 3] depending on optimality considerations. This paper considers rectangular sampling grids.

A. Ideal Lowpass Filter

The starting point for designing a digital filter is the ideal lowpass filter spectral characteristic from which different highpass, bandpass and bandstop filter configurations can be derived. Similarly to the digital filters in 2D signal and image processing [2, 3], the 3D volumetric image filters can also be separable with rectangular support and non-separable. In addition, the 3D filters can be also semi-separable with cylindrical support.

B. Three kinds of pass bands in 3D filters

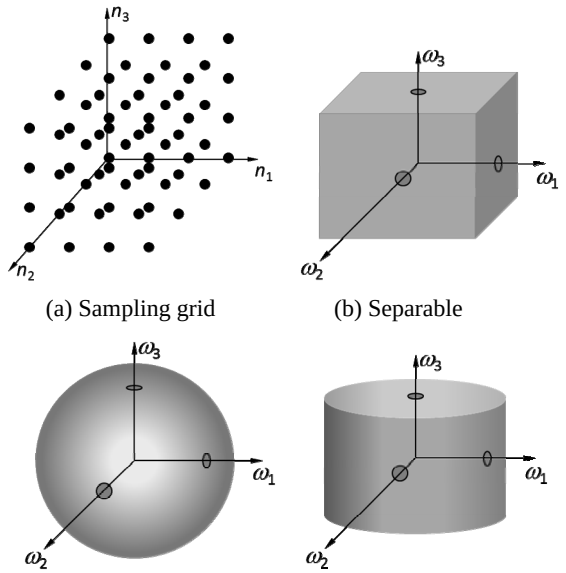
The performance difference among a separable, a semi-separable and a non-separable filter can be perceived through the spectral characteristics along different directions in 3D space. On the one hand, the separable filter is simply designed; its impulse response is equal to the product of the marginal impulse responses along the three spatial axes which determines independent spectral characteristics along the three axes. Its passband is rectangular which means that this kind of filter would pass higher frequencies along the diagonal directions of the rectangle, compared to the frequencies along each frequency axis.

On the other hand, the non-separable filter offers spectral characteristics along the different directions which are no longer independent. In particular, the non-separable filter with spherical passband offers equal spectral characteristics along any direction in 3D space. This may be very important especially in computed tomography and magnetic resonance imaging where the data details are equally important in all directions.

Between the performance of the separable and the non-separable 3D FIR filter is the semi-separable filter with cylindrical support. It has non-independent spectral characteristics in one plane and independent ones in the direction perpendicular to that plane. This can play role in subsurface imaging where a 3D image is constructed from many 2D image slices where the slicing can be performed at a much lower rate compared to the 2D sampling rate for the images within each slice.

The 3D sampling grid and the three kinds of support regions are depicted in Fig. 1.

Dimitar G. Valchev is with the Department of Radioengineering, Technical University of Varna, Varna 9010, Bulgaria, E-mail: D.Valchev@tu-varna.bg



(a) Sampling grid (b) Separable

(c) Non-separable (d) Semi-separable

Fig. 1. 3D sampling grid and support regions for separable, non-separable and semi-separable ideal 3D lowpass filters

III. THREE KINDS OF SEPARABILITY IN 3D FILTERS

The rectangular passband defines a fully separable 3D filter. The impulse response of such a filter is given by [1]

$$h_r(n_1,n_2,n_3)=\frac{\sin\omega_{c_1}n_1}{\pi n_1}\frac{\sin\omega_{c_2}n_2}{\pi n_2}\frac{\sin\omega_{c_3}n_3}{\pi n_3} \tag{1}$$

for $-\infty < n_1, n_2, n_3 < \infty$, where ω_{c_1}, ω_{c_2} and ω_{c_3} are the cut-off spatial frequencies along the corresponding three frequency axes ω_1, ω_2 and ω_3 shown in Fig. 1 (a). For $\omega_{c_1}=\omega_{c_2}=\omega_{c_3}=\omega_c$, the filter's passband turns into cubic.

The spherical passband defines a non-separable 3D filter. The impulse response of such a filter is given by [1]

$$h_s(n_1,n_2,n_3)=\frac{\omega_c j_1\left(\omega_c\sqrt{n_1^2+n_2^2+n_3^2}\right)}{4\pi\left(n_1^2+n_2^2+n_3^2\right)} \tag{2}$$

where ω_c is the cut-off spatial frequency which is the same along the three frequency axes ω_1, ω_2 and ω_3 shown in Fig. 1 (b), and $j_1(x)$ is the first-order spherical Bessel function of the first kind, also equal to $\sin(x)/x^2 - \cos(x)/x$ [5].

The cylindrical passband defines a semi-separable 3D filter. The impulse response of such a filter is given by (details in the Appendix)

$$h_c=\frac{\omega_{c_1}\sin\omega_{c_3}n_3}{2\pi^2 n_3\sqrt{n_1^2+n_2^2}}J_1\left(\omega_{c_1}\sqrt{n_1^2+n_2^2}\right) \tag{3}$$

where ω_{c_1} and ω_{c_3} are the cut-off spatial frequencies along the corresponding three frequency axes ω_1, ω_2 and ω_3 shown in Fig. 1 (c). Here, without loss of generality it is assumed that the 3D fiter is non-separable in the n_1 and n_2 axes while separable in the n_3 axis. Any other orientation of the lowpass spatial spectral cylinder is also possible for a 3D filter characterized by a semi-separability in its arguments.

IV. CONCLUSION

This paper gives ideal impulse responses for three kinds of 3D digital filters for volumetric image processing: a separable, a non-separable and a semi-separable one. The separable filter is characterized by a rectangular passband, the non-separable filter is characterized by a spherical passband, and the semi-separable filter is characterized by a cylindrical passband. Such filters are applicable in various 3D subsurface sensing and imaging applications. Future work will focus on filter design approaches for volumetric digital image processing.

APPENDIX

This Appendix outlines the derivation of the impulse response of an ideal lowpass filter with a cylindrical passband. The frequency response of such a filter is given by

$$H_c(\omega_1,\omega_2,\omega_3)=\begin{cases}1, & \sqrt{\omega_1+\omega_2}\le\omega_{c_1},\ |\omega_3|\le\omega_{c_3},\\ 0, & \text{else},\end{cases} \tag{4}$$

in the frequency cube $[-\pi,+\pi]^3$. The ideal impulse response of the filter is derived by taking the inverse 3D Fourier transform of this function in the frequency domain [2]:

$$\begin{aligned}h_c&=\frac{1}{(2\pi)^3}\int_{-\pi}^{\pi}\iint_{\sqrt{\omega_1+\omega_2}\le\omega_{c_1}}1e^{j(\omega_1n_1+\omega_2n_2+\omega_3n_3)}d\omega_1d\omega_2d\omega_3\\&=\frac{1}{(2\pi)^3}\int_{-\pi}^{\pi}e^{j\omega_3n_3}d\omega_3\iint_{\sqrt{\omega_1+\omega_2}\le\omega_{c_1}}e^{j(\omega_1n_1+\omega_2n_2)}d\omega_1d\omega_2\end{aligned} \tag{5}$$

which reduces to [2, 5]

$$h_c=\frac{\omega_{c_1}}{2\pi\sqrt{n_1^2+n_2^2}}J_1\left(\omega_{c_1}\sqrt{n_1^2+n_2^2}\right)\frac{\sin\omega_{c_3}n_3}{\pi n_3}. \tag{6}$$

ACKNOWLEDGEMENT

This work is supported by the state budget of Technical University of Varna, research project # NP4/2011.

REFERENCES

[1] D. G. Valchev, "Volumetric image processing by 3D filters", *Internat. Sci. Conf. on Communications, Electromagnetics and Medical Applications CEMA'2010, Athens, Greece*, October 7–9, 2010, pp. 62–64.

[2] J. Woods, *Multidimensional signal, image and video processing and coding*, Elsevier Academic Press, 2006.

[3] D. Dudgeon, R. Mersereau, *Multidimensional digital signal processing*, Prentice Hall, 1990.

[4] J. Proakis, D. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*, Prentice Hall, 2007.

[5] M. Abramowitz, I. Stegun, *Handbook of mathematical functions with formulas, graphs and mathematical tables*, Dover, 1972.

Precision of Some Motion Detection Methods Using Background Subtraction in Traffic Surveillance Video

Boris Nikolov[1] and Nikolay Kostov[2]

Abstract – **In these days video surveillance is used for many purposes such as security, traffic control, special measurement systems etc. In most of these systems the video stream is processed by motion detection algorithms. The goal of proposed algorithms is not only to detect motion, but to correctly subtract the foreground moving objects and to separate the foreground motion from the background motion.**

In this paper, some methods for background subtraction in video surveillance are investigated. An experiment and the results of estimating the footage of real traffic video is proposed. Conclusions about precision and computational cost for each method are given.

Keywords **– Motion detection, Background subtraction, Video surveillance.**

I. INTRODUCTION

In modern video surveillance systems motion detection is used to estimate the motion of moving objects. This is used for counting the moving objects, to recognize specific forms and to alarm if there is crossing the security border. The goal of this article is to investigate the precision of the methods used for motion detection. This is an important problem because the decision in motion detection algorithm must be correct and must avoid false detection caused by moving objects that belong to the background of the scene.

II. BACKGROUND SUBTRACTION METHODS

Background subtraction methods are used for identifying motion in video sequences using algorithm which includes creating a background model to represent the scene with no moving objects. Each pixel in the frame $I_{t(x,y)}$ is compared with the estimated background model $B_{t(x,y)}$. The pixels in the current frame that differs from the estimated background are classified as foreground. The moving objects in the scene are represented by areas of foreground pixels. To prevent incorrect classification of moving objects in the background like tree branches waving by the wind, the correctly estimation of the background model is important.

[1]Boris Nikolov is with the Technical University of Varna, Faculty of Electronics, Department of Radio Engineering, Bulgaria, Varna 9010, 1 Studentska Str.
E-mail: boris.nikolov@yahoo.com

[2]Nikolay Kostov is with the Technical University of Varna, Faculty of Electronics, Department of Radio Engineering, Bulgaria, Varna 9010, 1 Studentska Str.
E-mail: n_kostov@mail.bg

A. Temporal Averaging

Temporal averaging is a method used for background subtraction in video sequences. This method estimates background B_t by calculating the median value of the previous frames, [1]. The background is given by

$$B_{t+1} = \alpha I_t + (1-\alpha) B_t, \quad (1)$$

where α is a learning rate and is used for applying the adaptive mean.

The difference between the current frame and the background is given by

$$D_t = |I_t - B_t|. \quad (2)$$

The algorithm classifies the pixels as foreground by the rule

$$M_t(x,y) = \begin{cases} 0, & D_t(x,y) \le T \\ 1, & D_t(x,y) > T \end{cases}, \quad (3)$$

where T is a threshold preventing camera noise.

B. $\Sigma-\Delta$ background modelling

$\Sigma-\Delta$ (Sigma Delta) background modelling method is called $\Sigma-\Delta$, because the similarity to analog to digital conversion of a time varying signal using $\Sigma-\Delta$ modulation as it is interpreted in [2].

A current pixel is classified as foreground if the absolute pixel difference is greater or less than the estimated $\Sigma-\Delta$ variance. The $\Sigma-\Delta$ variance is different to the standard mathematical definition of variance. As it is defined in [2] it is a measure of the variation of the colours of each pixel over time. Depending on the difference between the current pixel value and the background the $\Sigma-\Delta$ variance are updated at each time step by incrementing or decrementing them by one.

The current background is given by

$$B_{t+1}(x,y) = \begin{cases} B_t(x,y) - 1, & D_t(x,y) < 0 \\ B_t, & D_t(x,y) = 0 \\ B_t(x,y) + 1, & D_t(x,y) > 0 \end{cases}, \quad (4)$$

where D_t is the difference, which is

$$D_t = I_t - B_t. \quad (5)$$

As it is proposed in [2], to adapt for different conditions at different areas in the image, a per pixel $\Sigma-\Delta$ variance, $V_{t(x,y)}$, is used as a threshold for each pixel. To update this variance $V_{t(x,y)}$ we implement the equation

$$V_{t+1}(x,y)=\begin{cases} V_t(x,y)-1, & C.|D_t(x,y)|<V_t(x,y) \\ V_t(x,y), & |D_t(x,y)|=0 \\ V_t(x,y)+1, & C.|D_t(x,y)|>V_t(x,y) \end{cases}, \quad (6)$$

where C is a user set parameter that determines how large the difference must be, before the variance is updated. The foreground mask is given by

$$M_t(x,y)=\begin{cases} 0, & |D_t(x,y)|\le V_t(x,y) \\ 1, & |D_t(x,y)|>V_t(x,y) \end{cases}. \quad (7)$$

High $\Sigma-\Delta$ variance is typical for pixels of repeating backgrounds, for example, leaves waving by the wind and wavy surface of water. Those pixels would not be classified as foreground because the difference is minor than the variance.

C. Mixture of Gaussians

The Mixture of Gaussians method describes each pixel in the frame by using multiple Gaussian distributions. Each pixel is represented by a distribution with its associated variance, weight and mean, [3].

The probability of observing the current pixel value *x* at time *t* at a particular pixel location is given by

$$P(x)=\sum_{i=1}^{K}\omega_{i,t}\eta(x;\mu_{i,t},\Sigma_{i,t}), \quad (8)$$

where K is the number of Gaussians distributions representing each pixel, $\omega_{i,t}$ is the weight of the i^{th} Gaussian at time t, η is the Gaussian probability density function with parameters: x is the current pixel, $\mu_{i,t}$ is the mean of the i^{th} distribution at time t, and $\Sigma_{i,t}$ is the covariance of the i^{th} distribution at time t.

The Gaussian probability density function η is given by

$$\eta(x;\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}. \quad (9)$$

A particular value *x* being observed at a pixel location has high probability if it is close to the mean of Gaussian distribution with high weight with a low variance. So, this is the Gaussian distributions that best describe each pixel. To update each distribution we use an adaptive learning rate described in [8]. This learning rate depends on the strength of the match between the current pixel value and the i^{th} distribution, q_i.

The weight of the i^{th} Gaussian at time *t* is given by

$$\omega_{i,t}=(1-\alpha)\omega_{i,t-1}+\alpha q_i. \quad (10)$$

In case there is no match it is created a new Gaussian distribution with a mean equal to the current pixel value, a low weight and a high variance. The new distribution replaces the distribution with the lowest weight and highest variance. It is assumed that the background is represented by Gaussian distributions with the highest weight and lowest variance. To estimate the background, the distributions are first sorted in order of decreasing ω/σ. The pixels that belong to the background are the first C distributions and C is given by

$$C=\arg\min_c\left(\sum_{k=1}^{c}\omega_k>T\right). \quad (11)$$

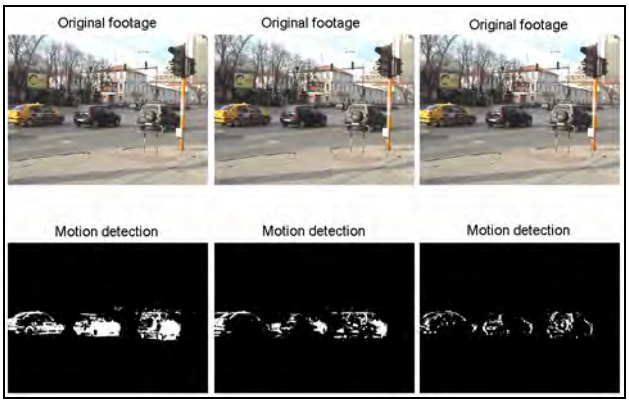
III. Experiment and Results

The proposed three methods for motion detection using background subtraction are implemented in Matlab and are shown. The processed video is captured at a city crossway where are many moving objects like cars and people. The shooting camera is stationary. The lighting of the scene is equal for all experiments. The methods, Temporal Averaging and Sigma-Delta background subtraction are ran for the original frame rate 25fps and resolution of 720x576 pixels. The Mixture of Gaussians method requires a lot of computational power and because of that it is ran for the same footage at 15fps and QVGA resolution of 320x240 pixels.

A. Temporal Averaging

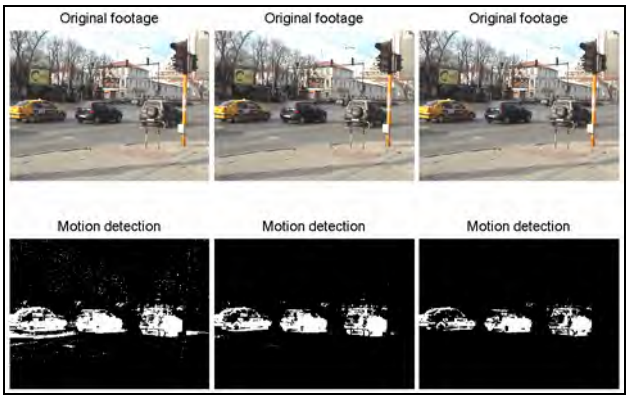
The main feature of this method is calculating historical background of the scene, Eq. (1). The learning rate α determines how fast the background is updated, so it should be between 0 to 1. As the learning rate is close to 1, the background is updating faster. This is not suitable for estimating relatively slow moving objects which are classified as a background. To estimate correctly slow moving objects α should be close to 0.1. If α is to much close to zero the natural changes in the background should not be updated in time and there will be incorrect foreground classification. In Fig. 1 are shown the original footage and the motion detection image for same sequence estimated for three different values of α. The threshold from equation (3) is equal for the three images. In Fig. 1a, there is a maximum similarity in motion detection, $\alpha=0.01$. In the next case, Fig. 1b, $\alpha=0.1$ there is an obvious trace after the moving objects, which is caused by incorrect background classification. The moving objects are assigned to the background too fast and when the object passes away the real background is estimated like a foreground. When α is to close to 1 like Fig. 1c $\alpha=0.8$, only the edges of the moving object will be detected.

The threshold T in Eq. (3) determines the resistance of noise in the image and flickering backgrounds. In Fig. 2, three screenshots of three amounts T are shown. The dimension of T is the number of the *255-th* levels of the *8 bit* depth luminance of the image in the algorithm. As the threshold is rising, the noise is more reduced, but small moving objects are hard to detect.



(a) *(b)* *(c)*

Fig. 1.Temporal Averaging method executed for three amounts of α, $\alpha = 0.01$ (a); $\alpha = 0.1$ (b); $\alpha = 0.8$ (c)

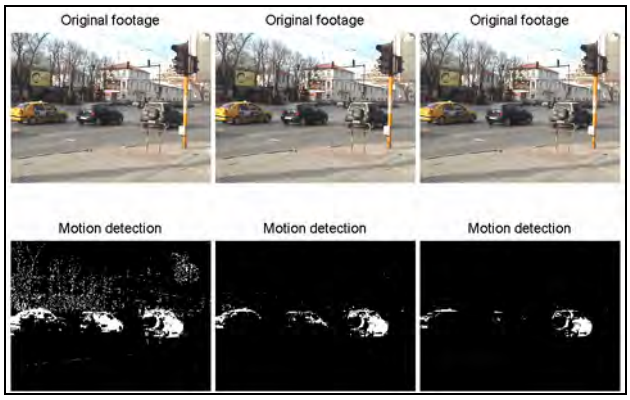


(a) *(b)* *(c)*

Fig. 2. Temporal Averaging method executed for three amounts of T, $T = 25$ (a); $T = 40$ (b); $T = 55$ (c)

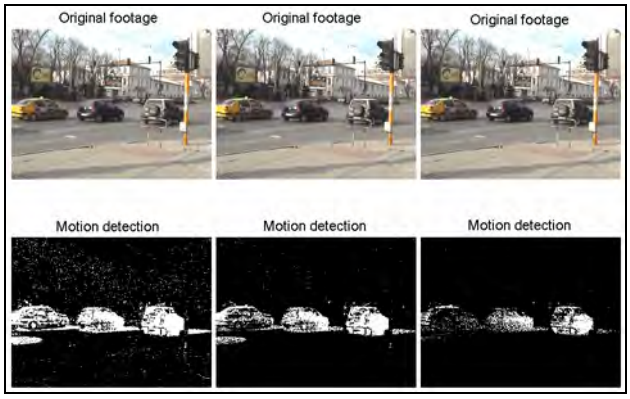
B. $\Sigma - \Delta$ background modelling

In Fig. 3 are shown the results of running the $\Sigma - \Delta$ background modeling method for the same footage as the previous method. The algorithm is executed for three amounts of the parameter C, Eq. (6). This parameter determines the speed of increasing the $\Sigma - \Delta$ variation and as consequence of that the resistance against noise and flickering background. In the original footage in Fig. 3 there are three cars moving from left to right. The first car is moving in area with low $\Sigma - \Delta$ variance and it is relatively equal represented in the three cases of motion detection, respectively Fig. 3a $C = 1$; Fig. 3b $C = 5$; Fig. 3c $C = 20$. As C goes up the second and third car are getting less visible in the motion detection image, Fig. 3b and Fig. 3c. This is determined that they are moving in the trace of the first car where $\Sigma - \Delta$ variance is high. The positive in this case is the high resistance against noise and flickering backgrounds. So, $\Sigma - \Delta$ background modeling method is not suitable for heavy traffic video estimating. But there is high precision and noise resistance when the moving objects passes in relatively long interval.



(a) *(b)* *(c)*

Fig. 3. $\Sigma - \Delta$ background modeling executed for three amounts of C, $C = 1$ (a); $C = 5$ (b); $C = 20$ (c)



(a) *(b)* *(c)*

Fig. 4. Mixture of Gaussians background subtraction method executed for three amounts of T, $T = 0.25$ (a); $T = 0.35$ (b); $T = 0.5$ (c)

C. Mixture of Gaussians

Mixture of Gaussian Background subtraction method is very precise, even the shadows of the moving objects are detected. This method is very complex to implement and configure and has relatively long running time. At every time step new distribution is created when the new observed pixel does not match previous Gaussian distribution. If this pixel belongs to a moving object which then goes away, the mean and variance of the previous distribution, which represented the true background, would not be changed. The algorithm is executed for three values of the threshold T, Eq. (11) and the results are shown in Fig. 4. High levels of the threshold determine high resistance against noise without serious corrupting of the moving objects estimation in the output motion detection image. In Fig. 4b and Fig.4c the noise is quite attenuated but the shapes of the foreground objects are still correctly represented.

IV. CONCLUSION

Three methods for motion detection using background subtraction were presented. The results of estimating real traffic surveillance video footage were proposed.

The conclusions about the precision of each method is analyzed according the purpose of implementation in practice.

The Temporal Averaging method is simple and do not require high computational power for maintains. Despite the simple algorithm, this method is surprisingly correct and resistant to noisy background, only if the learning rate α is correctly chosen. High levels of α are appropriate for scenes of fast moving objects. But in traffic surveillance sequences where most of the foreground objects are moving relatively slow, the amount of α should be near to 0.1.

The $\Sigma-\Delta$ background modeling method is a little bit complicated than the Temporal Averaging method, but it is not flexible enough when it is used for estimating heavy traffic video. Increasing the amount of the parameter C reduces the noise in the background but also creates a trace after the first moving object. This trace has high $\Sigma-\Delta$ variance, which determines low sensibility for motion in those pixels for the next consecutive frames. As a result there is a serious problem in estimating heavy traffic where most of the cars will be invisible.

Mixture of Gaussians is complex method for background subtraction. In our experiment this method shows highest precision. Even the shadows of the moving objects are detected. To achieve an optimal result using the mixture of Gaussians method there are many parameters to be adjusted and this is a disadvantage of this method. Another disadvantage is that it is computationally expensive and requires relatively large amount of memory to store, update and sort multiple Gaussian distributions.

The choice of the appropriate motion detection method depends on available computational power, specific surveillance conditions and requirements for processing time. Temporal Averaging and $\Sigma-\Delta$ background modeling are useful in real time systems, where is no need of high precision motion detection. For best precise results Mixture of Gaussians is a suitable decision but requires huge computational power.

ACKNOWLEDGEMENT

This paper is funded by National Ministry of Education and Science of Bulgaria under project "Investigation of Algorithms for Motion Detection in Motion Pictures", 2011, Technical University of Varna.

REFERENCES

[1] Janne Heikkilä and Olli Silvén. "A real-time system for monitoring of cyclists and pedestrians" Image and Vision Computing, Vol. 22, Issue 7, Visual Surveillance 1 July 2004 Pages: 563–570.

[2] Antoine Manzanera and Julien C. Richefeu. "A new motion detection algorithm based on sigma-delta background estimation". Pattern Recognition Letters In Advances in Visual information Processing, Vol.28 No.3 Pages: 320–328, 1 Feb 2007.

[3] C. Stauffer and W. E. L. Grimson. "Adaptive background mixture models for real-time tracking". In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, page 252, 1999.

[4] Vesna Zeljkovic, Dragoljub Pokrajac. "Improved Illumination Independent Moving Object Detection Algorithm in Infrared Video Sequences" In Proceedings of ICEST Conference 2005, Nis, Serbia and Montenegro, pages 342-345, 2005.

[5] Cheung Sen-Ching and Chandrika Kamath. "Robust techniques for background subtraction in urban traffic video". Proceedings of Electronic Imaging: Visual Communications and Image Processing 2004 (Part One), January 20-22 2004, San Jose, California. Bellingham WA:SPIE, 5308, Pages: 881–892.

[6] Andrea Prati, Ivana Mikic, Costantino Grana, and Mohan M. Trivedi. "Shadow detection algorithms for traffic flow analysis: a comparative study". In Proceedings of IEEE Intelligent Transportation Systems Conference , pages 340–345, Aug 2001.

[7] Ahmed Elgammal, Ramani Duraiswami, David Harwood, Larry S. Davis, R. Duraiswami, and D. Harwood. "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance" In Proceedings of the IEEE, VOL. 90, No. 7, July 2002 Pages 1151–1163, 2002.

[8] Dar S. Lee. "Effective Gaussian mixture learning for video background subtraction", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27 Issue 5, Pages: 827–832, 2005.

[9] M. Vargas, S. L. Toral, F. Barrero, J.M. Milla. "An Enhanced Background Estimation Algorithm for Vehicle Detection in Urban Traffic Video", Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15, 2008, Pages 784 – 790.

[10] Nan Lu, Jihong Wang, Q.H. Wu and Li Yang. "An Improved Motion Detection Method for Real-Time Surveillance", Proceedings of International Journal of Computer Science, IAENG, 35:1, IJCS_35_1_16, January 2008.

Spectrum Optimization of Truncated Complex Hadamard Transform

Rumen P. Mironov[1], Roumen K. Kountchev[2]

Abstract – **An algorithm for spectrum optimization of truncated Complex Hadamard Transform on the base of minimization of mean-squared error of reconstructed transform coefficients is presented. The developed algorithm is simulated on Matlab 6.5 environment and the obtained results showed increasing of signal to noise ratio with about 0.5 dB.**

Keywords – **Digital Signal Processing, Complex Hadamard Transform, Data Compression, Orthogonal Transforms.**

I. INTRODUCTION

Data compression, the art and science of reducing the amount of data required to represents 1D and 2D information, is one of the most useful and commercially successful technologies in the field of digital signal processing [1]. Dimensionality reduction in computation is a major signal processing application. One of the common compression approaches for reducing of spatial and temporal redundancy of information is based on block transform coding (BTC) [1]-[4], in which, a reversible, linear transform is used to map each block into a set of transform coefficients. Discrete orthogonal (unitary) transforms [3], [4], used in BTC, have found applications in many areas of N-dimensional signal processing, spectral analysis, pattern recognition, digital coding, computational mathematic and etc. Stated simply, these transform coefficients that are small may be excluded from processing operations, such as filtering, without much loss in processing accuracy.

The discrete integer Walsh-Hadamard Transform (WHT) is a fairly simple orthogonal transform and is an example of a generalized class of Fourier transforms [3]. The idea of using complex, rather than integer transforms matrices for spectral processing, analysis and watermarking has been shown in [5], [6], [7] and [8]. From the Complex Hadamard Transform (CHT), several complex decisions diagrams are derived and analysis of more general CHT properties for 1D and 2D signals are investigated [9],[10].

In this paper an algorithm for optimization of reduced spectrum of Complex Hadamard Transform is developed, using minimization of mean-squared error of reconstructed one- and two-dimensional signals. The obtained results showed increasing of signal to noise ratio with about 0.3 – 0.7 dB for any unitary transform.

The developed optimization algorithm is simulated on Matlab 6.5 environment for one test image "Lena" and the results of four unitary transforms – FFT, DCT, WHT and CHT are given.

II. MATHEMATICAL DESCRIPTION

Using the basic forward and inverse one-dimensional complex Hadamard transform for N coefficients [9],[11], the input signal vector $\vec{A}=[a_0,a_1,a_2,.....,a_{N-1}]$ and the output spectral vector - $\vec{B}=[b_0,b_1,b_2,.....,b_{N-1}]$ are joined by the equations:

$$\vec{B}=\frac{1}{N}[CH]\vec{A}\ ,\quad \vec{A}=\frac{1}{N}[CH]^t\vec{B}\,. \tag{1}$$

The discrete forward and the inverse orthogonal transform can be written by the following:

$$\left|\begin{array}{l} b_j=\dfrac{1}{N}\sum_{i=0}^{N-1}a_i t_{ij}\ ,\text{for}: j=\overline{0,N-1} \\ a_i=\sum_{j=0}^{N-1}b_j t_{ij}\qquad ,\text{for}: i=\overline{0,N-1}\end{array}\right. , \tag{2}$$

where: t_{ij} are transform coefficients.

The goal of the transformation process is to decorrelate the values of each sub-block, or to pack as much information as possible into the smallest number of spectral coefficients. The quantization stage than selectively eliminates or more coarsely quantizes the coefficients that carry the least amount of information in a predefined sense. These coefficients have the smallest impact on reconstructed sub-block quality. In more cases the first *k* important coefficients are saved and the next *N-k-1* coefficients are truncated.

The inverse decomposition from equation (2) can be written by the following:

$$a_i=\sum_{j=0}^{N-1}b_j t_{ij}=\sum_{j=0}^{k}b_j t_{ij}+\sum_{j=k+1}^{N-1}b_j t_{ij}\qquad ,\text{for}: i=\overline{0,N-1}\,. \tag{3}$$

The last *N-k-1* coefficients can be substituted by the value *A* and the approximated signal can be obtained from the truncated expansion:

$$\hat{a}_i=\sum_{j=0}^{N-1}b_j t_{ij}=\sum_{j=0}^{k}b_j t_{ij}+A\sum_{j=k+1}^{N-1}t_{ij}\qquad . \tag{4}$$

The difference between the input signal and its approximation can be given with the equation (5):

[1]Rumen P. Mironov is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky 8, Sofia 1000, Bulgaria, E-mail: rmironov@tu-sofia.bg

[2]Roumen K. Kountchev is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky 8, Sofia 1000, Bulgaria. E-mail: rkountch@tu-sofia.bg

$$\Delta a_i = a_i - \hat{a}_i = \sum_{j=0}^{N-1} b_j t_{ij} - \sum_{j=0}^{k} b_j t_{ij} - A\sum_{j=k+1}^{N-1} t_{ij} = \sum_{j=k+1}^{N-1} (b_j - A) t_{ij} \quad (5)$$

The mean-square error then is:

$$\begin{aligned}\overline{\varepsilon^2} &= \frac{1}{N}\sum_{i=0}^{N-1}(a_i - \hat{a}_i)^2 = \frac{1}{N}\sum_{i=0}^{N-1}\left[\sum_{j=k+1}^{N-1}(b_j - A)t_{ij}\right]^2 = \\ &= \frac{1}{N}\sum_{i=0}^{N-1}\left[\sum_{s=k+1}^{N-1}(b_s - A)t_{is}\right]\left[\sum_{l=k+1}^{N-1}(b_l - A)t_{il}\right] = \\ &= \frac{1}{N}\sum_{i=0}^{N-1}\left[\sum_{s=k+1}^{N-1}\sum_{l=k+1}^{N-1}(b_s - A)(b_l - A)t_{is}t_{il}\right] = \\ &= \sum_{j=k+1}^{N-1}(b_j - A)^2\end{aligned} \quad (6)$$

where: $t_{is}t_{il} = \begin{cases}1, \text{for } s = l \\ 0, \text{for } s \neq l\end{cases}$ is true for each orthogonal transform.

The minimum of mean-square error is obtained by differentiation:

$$\frac{\partial \overline{\varepsilon^2}}{\partial A} = -2\sum_{j=k+1}^{N-1}(b_j - A) = 0\ , \quad (7)$$

and the optimum value for truncated coefficients is:

$$A_{opt} = \frac{1}{N-k-1}\sum_{j=k+1}^{N-1} b_j = m_b\ . \quad (8)$$

The equation (8) show that the mean-square error (MSE) decreases for any orthogonal transform via approximation of reduced spectral coefficients by average of their values.

The improvement of MSE can be calculated from equations (6) and (8):

$$\begin{aligned}\Delta\overline{\varepsilon^2} &= \overline{\varepsilon^2} - \overline{\varepsilon^2_{min}} = \sum_{j=k+1}^{N-1} b_j^2 - \sum_{j=k+1}^{N-1}(b_j - m_b)^2 = \\ &= \sum_{j=k+1}^{N-1} b_j^2 - \sum_{j=k+1}^{N-1}\left[b_j^2 - 2b_j m_b + m_b^2\right] = \\ &= \sum_{j=k+1}^{N-1}\left[2b_j m_b - m_b^2\right]\end{aligned} \quad , \quad (9)$$

where: $\overline{\varepsilon^2} = \sum_{j=k+1}^{N-1} b_j^2$ is MSE for zero approximation of truncated coefficients and $\overline{\varepsilon^2_{min}} = \sum_{j=k+1}^{N-1}(b_j - m_b)^2$ is MSE for mean approximation of truncated coefficients.

This algorithm can be summarized for 2D signals (images). The forward and the inverse discrete transform of sub-image *g(x,y)* of size *N*x*N* can be expressed as the following equations:

$$\begin{aligned}S(u,v) &= \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} g(x,y).r(x,y,u,v) \\ g(x,y) &= \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} S(u,v).t(x,y,u,v)\end{aligned} \quad , \quad (10)$$

In this equations *r(x,y,u,v)* and *t(x,y,u,v)* are called the forward and inverse transformation kernels, respectively. Because the inverse kernel *t(x,y,u,v)* in (10) depends only on the indices *(x,y,u,v)* and not on the values of *g(x,y)* and *S(u,v)*, it can be viewed as defining a set of basis functions or basis images.

This interpretation becomes clearer if the equation is modified in matrix form:

$$\mathbf{G} = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} S(u,v)\mathbf{T}_{uv}\ , \quad (11)$$

where: **G** is *N*x*N* matrix containing the pixels of *g(x,y)*, the matrices $\mathbf{T}_{uv}$ are the basis images and S(u,v) are the spectral coefficients.

We can define a transform coefficients masking function:

$$\chi(u,v) = \begin{cases}0, \text{if } S(u,v) \text{satisfies a specified truncation criterion} \\ 1, \text{otherwise.}\end{cases}$$

An approximation of **G** can be obtained from the truncated expansion:

$$\begin{aligned}\hat{\mathbf{G}} = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\chi(u,v)S(u,v)\mathbf{T}_{uv} &= \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\chi(u,v)S(u,v)\mathbf{T}_{uv} - \\ &- \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]A\mathbf{T}_{uv}\end{aligned} \quad ,(12)$$

The mean-square error between sub-image **G** and its approximation $\hat{\mathbf{G}}$ then is:

$$\overline{\varepsilon^2} = E\left\{\left\|\mathrm{G} - \hat{\mathrm{G}}\right\|^2\right\}, \quad (13)$$

$$\overline{\varepsilon^2} = E\left\{\left\|\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} S(u,v)\mathbf{T}_{uv} - \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\chi(u,v)S(u,v)\mathbf{T}_{uv}\right\|^2\right\} =$$

$$= E\left\{\left\|\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]S(u,v)\mathbf{T}_{uv} - \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]A\mathbf{T}_{uv}\right\|^2\right\}$$

$$= E\left\{\left\|\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]\left[S(u,v)-A\right]\mathbf{T}_{uv}\right\|^2\right\}$$

$$= \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]\left[S(u,v)-A\right]^2$$

The minimum of mean-square error is obtained by the following:

$$\frac{\partial \overline{\varepsilon^2}}{\partial A} = -2\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]\left[S(u,v)-A\right] = 0\ , \quad (14)$$

and the optimum value for the truncated coefficients is:

$$A_{opt} = \frac{\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]S(u,v)}{\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left[1-\chi(u,v)\right]}\ . \quad (15)$$

III. Experimental Results

The developed optimization algorithm is simulated on Matlab 6.5 environment for four 2D unitary transforms – Discrete Fourier Transform, Discrete Cosine Transform,

Discrete Walsh-Hadamard Transform and Complex Hadamard Transform. The obtained experimental results for the test image "Lenna" (512x512, 8 bits) with sub-image kernel 8x8 are given on Table 1.

TABLE 1.

Reduction type	MSE	NMSE	SNR, dB	PSNR, dB
48 Reduced Coefficients, 16 saved				
Zero FFT	107.383	8.65×10^{-6}	50.6263	27.8554
Mean FFT	97.1544	7.83×10^{-6}	51.0610	28.2902
Zero DCT	0.5078	4.09×10^{-8}	73.8785	51.1076
Mean DCT	0.6590	5.31×10^{-8}	72.7471	49.9762
Zero WHT	19.6889	1.58×10^{-6}	57.9934	35.2226
Mean WHT	16.7618	1.35×10^{-6}	58.6924	35.9216
Zero CHT	16.0925	1.30×10^{-6}	58.8694	36.0985
Mean CHT	15.2343	1.23×10^{-6}	59.1074	36.3365
55 Reduced Coefficients, 9 saved				
Zero FFT	116.777	9.41×10^{-6}	50.2620	27.4912
Mean FFT	106.358	8.57×10^{-6}	50.6679	27.8971
Zero DCT	0.6828	5.50×10^{-8}	72.5924	49.8216
Mean DCT	0.5800	4.67×10^{-8}	73.3013	50.5304
Zero WHT	43.6631	3.52×10^{-6}	54.5345	31.7636
Mean WHT	42.5557	3.43×10^{-6}	54.6461	31.8752
Zero CHT	46.4509	3.74×10^{-6}	54.2657	31.4948
Mean CHT	42.2766	3.40×10^{-6}	54.6747	31.9038

In the first part of the table the results for 48 reduced coefficients (4x4 are saved) approximated with zero and mean values are shown, and in the second part the same experiments for 55 reduced coefficients (3x3 - saved) are shown. The calculated values for the mean-square error (MSE), normalized mean-square error (NMSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) showed increasing of each parameter with about 0.5 %. The input image is shown on Fig.1, and the output images for 48 reduced coefficients are showed on Fig.2.

Fig.1. Input test image "Lena" (512x512, 8 bits)

IV. CONCLUSION

Method for spectrum optimization of truncated orthogonal transforms is presented. The improvement of quality by the compression of one-dimensional and two-dimensional signals is theoretically proved. An algorithm for block truncation coding is developed on the base of minimization of mean-squared error of reduced spectral coefficients reconstruction. The experimental results with four transformations are given.

The main advantages of the developed algorithm are:

- increasing the signal to noise ratio by the compression with truncated discrete orthogonal transforms of one- and two-dimensional signals;
- decreasing the preserved coefficients and increasing compression ratio by the using any discrete orthogonal transforms;
- using the CHT instead most complicated Fourier transform and keep the possibilities for working with complex spectrum.

The presented spectrum optimization for discrete orthogonal transforms can be used in digital signal processing for spectral analysis, pattern recognition, digital watermarking, transformation, coding and transmission of one-dimensional and two-dimensional signals.

V. 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 DNTC 02/19-1.10.2010.

VI. REFERENCES

[1] R. C. Gonzalez, R. E. Woods. *Digital Image Processing*, Third Ed., Pearson Prentice Hall, N.J., 2008.

[2] W. K. Pratt. *Digital Image Processing*, Fourth Ed., John Wiley & Sons, N.Y., 2007.

[3] N. Ahmed, K. R. Rao. *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag Berlin, Heidelberg, 1975.

[4] A. D. Poularikas. *The Transforms and Applications Handbook*, Second Ed., CRC Press, 2000.

[5] B. Falkowski, S. Rahardja. "Complex Spectral Decision Diagrams", *Proc. of the 26th Int. Symposium on Multiple Valued Logic*, Vol. ISMVL'96, 1996.

[6] S. Rahardja, B. Falkowski. "Complex Composite Spectra of Unified Complex Hadamard Transform for Logic Functions", *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 47, No. 11, November 2000.

[7] B. Falkowski, S. Rahardja. "Complex Hadamard Transforms: Properties, Relations and Architecture", *IEICE Trans. Fundamentals*. Vol. E87-A, No.8, August 2004.

[8] R. Kountchev, R. Mironov. "Audio Watermarking in the Phase-Frequency Domain", *XL Intern. Scientific Conference on Information, Communication and Energy Systems and Technologies*, ICEST'2005, Nis, Serbia and Montenegro, 2005.

[9] R. Mironov, R. Kountchev. "Analysis of Complex Hadamard Transform Properties", *XLI International Scientific Conference on Information, Communication and Energy Systems and Technologies*, ICEST 2006, 26 June – 1st July, Sofia, Bulgaria, 2006, pp.173-176.

[10] R. Mironov, R. Kountchev. "Algorithm for Fast Complex Hadamard Transform", *XLV International Scientific Conference on Information, Communication and Energy Systems and Technologies*, ICEST 2010, 23-26 June, Ohrid, Macedonia, 2010, pp.179-182.

Fig.2. Output images after reduction with 48 coefficients for FFT, DCT, WHT and CHT

Music Genre Recognition and Classification

Milos Djuric[1] and Milena Stankovic[2]

Abstract – **This work describes a system for the automatic recognition and classification of music according to genres, based exclusively on audio content of the signal. Described system is based on sound characteristics that should have influence on human spontaneous and natural ability to classify music into genres. In order to examine the needed characteristics, every melody is divided into segments which are analyzed individually and final result for each characteristic is presented as arithmetic mean or variance of all proper segment values. A Support Vector Machine classifier is chosen for training and classification, reaching a classification accuracy of between 85 % - 98 % for the three test genres: Heavy Metal, Techno Dance and Classical music.**

Keywords – **Music genre recognition, Music genre classification, Support Vector Machine method.**

I. INTRODUCTION

The amount of multimedia now available online has created an impact for efficient tools to organize, search and manage this huge amount of data [1, 2]. At present, multimedia data is usually classified based on textual meta–information. Although such kind of information is very useful for indexing, sorting, comparing and retrieval, the problem is that it is manually generated and, as so, the process is more expensive and very likely arbitrary and subjective. Extracting the information through an automatic and systematic process might overcome most of these problems. So, a challenge for the pattern recognition community is to develop intelligent algorithms for searching and indexing what in recent years became a huge amount of data. While there has obviously been a great deal of work around speech recognition and music processing, in this work, we will focus on the music genre recognition.

Distinguishing between music genres is a trivial task for human beings. A few seconds of music is usually sufficient to allow us to do a rough classification, such as identifying a song as punk or rap, or rock or classical music. The question that this paper attempts to address is whether it is also possible for a machine to make such a classification. The present work investigates ways to automatically classify music files according to genre, based exclusively on the audio content of the files.

Machine can "experience" music only in a digital format, and for it, it is nothing but a sequence of bits whose values correspond to the sound-pressure levels in an analogue acoustic waveform. These bits are for instance easily interpreted by a machine to find out certain facts, such as the overall amplitude of the signal at a given time, which is, of course, impossible to the humans. But understanding music, like humans do it all the time without effort, is far more complex matter. The recognition of music genres undertakes these advanced tasks.

[1]Milos Djuric is with the Metropolitan University, Vozd Karadjordje 47, 18000 Nis, Serbia, E-mail: djura042@gmaill.com.

[2]Milena Stankovic is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: milena.stankovic@elfak.ni.ac.rs

II. THE APPROACH

A. Main Idea

For this discussion, a music genre can be considered as a specific class of music with a set of common properties that in the perception of the average listener distinguish music in that category from other songs.

Human listeners have remarkable music genre recognition abilities. This was shown in a study conducted by R.O. Gjerdigen and D. Perrot [3]. They used ten different genres and eight sample songs for each genre. Five excerpts with different durations were taken from each song. The subjects of the study who did not have any higher-level knowledge in musical theories, were presented with the short excerpts and asked to decide on one of the ten genres for each excerpt. The accuracy of genre prediction for the longest samples was around 70%, compared to the CD companies' original classification.

Taking into account that music genres are a fuzzy concept, and that even the music industry is sometimes contradictory in assigning genres, this percentage is, according to [3] unexpectedly high. The results of the study are especially interesting, since they show that it may be possible to accurately recognize music genres without using any higher-level abstractions. So, the basic assumption of this paper is that some form of classification is possible based on spectral and timbral characteristics alone, because music samples used in [3] are much too short for recognizing the rhythm, melody or conceptual structure of a song.

In order to increase classification success of 70 % for the longest samples, we divided each signal into windows and so we analyzed every part of the song and calculated arithmetical mean and variance of all feature values for every sound signal (song). In this way the possibility that the part of the song that is atypical for genre is chosen is smaller then it was the case in [3].

B. Features used for recognition

For characterization of music songs a feature set originally proposed by Tzanetakis [4] is used. This feature set in combination with other types of features (e.g. from cepstral analysis) is also used in many other works from related area, e.g. musical instrument recognition [5].

Those features are based on the short time Fourier transform (STFT) and are calculated for every short–time

frame of sound [6]. The basic ideas and knowledge for creating MatLab algorithms for calculating values of the following features are extracted from [7] and [8].

Spectral Centroid is the balancing point of the spectrum. It is a measure of spectral shape, and is often associated with the notion of spectral brightness, or in this case – sound brightness. The spectral centroid can be calculated as:

$$C = \frac{\sum_{n=1}^{N} M_t[n] \cdot n}{\sum_{n=1}^{N} M_t[n]} \quad (1)$$

where $M_t[n]$ is the magnitude of the Fourier transform at frame t and frequency bin n.

Spectral Flux is a measure of local spectral change, and it is defined as:

$$F_t = (N_t[n] - N_{t-1}[n])^2 \quad (2)$$

where $N_t[n]$ is the normalized magnitude of the Fourier transform at window *t*. In this particular case spectral flux could be considered as a flow rate of musical events in one song (sound variations).

Spectral Rolloff is, like the centroid, a measure of spectral shape. It is defined as the frequency R_t below which 80% of the magnitude distribution is concentrated. It is computed as

$$\sum_{n=1}^{R_t} M_t[n] = 0.8 \sum_{n=1}^{N} M_t[n] \quad (3)$$

Time Domain Zero–Crossings occurs when successive samples in a digital signal have different signs. The zero-crossing rate is a simple measure of the noisiness of a signal. It can be calculated as

$$Z_t = \sum_{n=1}^{N} | s(x[n]) - s(x[n-1]) | \quad (4)$$

where $x(n)$ is the time domain signal, and the *s* function has value 1 or 0 for positive and negative arguments respectively. Unlike spectral centroid, rolloff and flux, which are frequency-domain features, the zero-crossing rate is a time-domain feature.

There is one additional feature, called **Low Energy**. It is defined as the percentage of windows that have less energy than the average energy of all windows. Music that contains silent parts will have a larger low energy value than continuous sounds.

C. Forming Feature Vector

The features proposed before are concatenated to form a 9–dimensional feature vector. Eight features present mean values and variances of *spectral centroid*, *rolloff*, *flux* and *zerocrossing*, that are calculated from all windows, and there is also an additional *low–energy* feature. So in this case, each signal is divided in 4.5 *s* windows for the sample rate of 44.100 kHz and by accounting all of them there is a good possibility that this model will have better results than in the case of Gjerdigen–Perrot study.

Every song, wheter it has a role in training or in testing part of the process, is presented with one feature vector.

III. CLASSIFICATION METHOD

The basic problem in musical genre classification is to assign a class, i.e. a musical genre $g \in G$, that best matches to the input vector representing one music clip $X_D = (x_1\ x_2\ \ldots\ x_D)$ where *D* is the dimension of the vector. For such an aim, we use method called Support Vector Machine (SVM). SVM is a set of related supervised learning methods invented by Vladimir Vapnik that analyze data and recognize patterns, used for classification and regression analysis.

SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. So the goal of SVM modeling is to find the optimal hyperplane that in idealized example separates clusters of vectors in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the support vectors. As the Fig.1. below depicts, the optimal hyperplane is oriented in a way that maximal margin between the clusters is present.

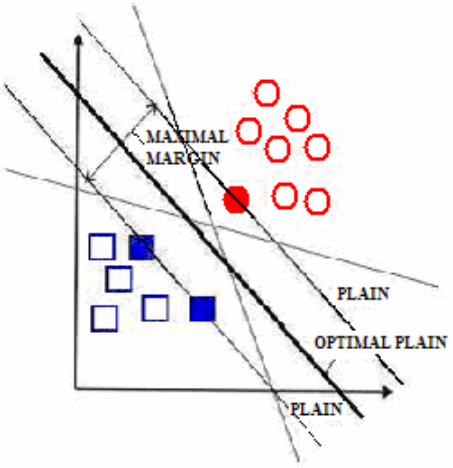


Fig. 1 Optimal plain that separates two clusters of vectors in SVM - method

In cases like the one we are analyzing there are more than two categories (three music genres). For resolving this kind of classification several approaches have been suggested, but two are the most popular. (1) "one against many" where a data point would be classified under a certain class if and only if that class's SVM accepted it and all other classes' SVMs rejected it (the classifier with the highest output function assigns the class). (2) "one against one" where the classification is done by max-wins voting strategy, in which

every classifier assigns the instance to one of the two classes when the vote for the assigned class is increased by one vote, and the class with the most votes finally determines the instance classification. While accurate for tightly clustered classes (as is not the case in problem analyzed in this paper), method (1) leaves regions of the feature space undecided where more than one class accepts or all classes reject. For this reason approach used in this case is "one against one".

For more detailed information about the Support Vector Machine method see [9].

IV. System and Experiment Overview

A. System Overview

System described in this paper consists of two parts or subsystems. The first one is the *extractor* that transforms signal on the entry if necessarily and extracts features crucial for the particular problem, and is denoted as *E* on Fig. 2. Second one is the *classifier* and it classifies data given from the output of the extractor, and is denoted as *C*. So, the two steps can be clearly separated: The output of the feature extraction step is the input for the classification step. We can substitute this subsystem into the black box introduced above, resulting in the model shown in Fig. 2. This is the basic music genre recognition system in its most simple representation.

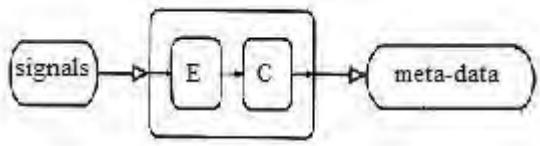


Fig. 2 Global view of the music genre recognition system

B. Experiment Procedure

After forming feature vectors the system operates into two modes: training and testing. For the implementation of SVM-method, MatLab LSSVMLab1.6 library is used, and feature vectors are imported through standard *.xls* document.

In the training mode, the feature vectors are used by a learning algorithm to train the classifier. For this purpose 80 songs from each of three genres are used.

Because of the fact that the problem of music genre recognition is fuzzy and is often matter of subjective arbitration, it is important to underline that modality of forming the training kernel is of the crucial importance in this experiment. This means that 240 training vectors should present the songs that are "very typical" for what today is called classical, metal and dance music, i.e. classical music often has relatively small spectral flux values (less local spectral changes) compared to the dance music, and relatively small spectral centroid values (less brightness) than heavy metal, or high values of zero-crossing-rate feature (simple measure of noisiness of the music) should be expected in music genres that could be described as "noisy".

At the classification mode, a music whose genre is unknown is submitted to the system. From such a music clip feature vector is extracted and for this purpose 100 songs are used, i.e. 30 for heavy metal, 40 for classical music and 30 for techno dance (this songs were chosen randomly). The results of this classification follow in next section.

V. Test Results

Special attention was paid to the necessity of variance in the selection of test music. For instance, the Metal collection contains songs from the Trash Metal, Speed Metal, Classic Metal, Hard Rock, Gothic Metal, Black Metal, Punk, Grunge, and Death Metal categories. The same is true for the Dance genre, which covers a spectrum from Eurodance to Rave. Test music data for the classical genre is somewhat more limited. This is due to the fact that classical music is based on different concepts than popular music. It requires the listener to actively pay attention in order to fully appreciate a piece. So, the limitation for the case of the classical music mentioned above, refers to the fact that both test and training classical music sets mostly consist of the compositions made by three different and, in musical terms, especial classic composers: Strauss (orchestral music and his famous waltzes were used only), Bach (only his concertos were used), and Vivaldi (also with his concertos and Four Seasons). Almost 80 % of both training and test sets consist of Strauss's, Bach's and Vivaldi's compositions, and the rest of the sets are filled with compositions that, according to author's opinion, belong to the same classical music subgenres. This does not seem to jeopardize scientific approach to the problem because, as mentioned above (I and II. *A*), the classification of the music into genres is a fuzzy concept that often generates a lot of uncertainty and requires arbitrary solutions. So, it is clear that in this case classical music set is reduced to three subsets (subgenres), and it does not fully represent current understanding of what term "classical music" really means.

The results obtained from recognition and classification process presented in this paper are given in Table 1.

TABLE 1.

EXAMINATION OF TRAINING AND CLASSIFICATION OF MUSIC GENRES.

	Classical	Dance	Metal
Training vectors	80	80	80
Test vectors	40	30	30
Correct hits	39	25	28
Misses	1	5	2
Classification accuracy	97.5%	83.3%	93.3%

As can be seen from the Table 1. (and as noted in IV.*B*), 240 training vectors were used, 80 for each of three genres. For the testing process 100 vectors were used (40 – 30 – 30, for classical, dance and metal music, respectively).

Subgenres are, in classical music, mostly based on certain composer, particular epoch and sometimes are totally arbitrary. It should be stated that without dividing the classical music set on subgenres in both training and testing sets, classification results had 10 % to 20 % lower values.

Achieved classification accuracy (Table 1.) is 97.5% for classical music, 83.3% for dance music, and 93.3% for metal music. These results are in rank with other classifiers that operate with similar problems [4].

VI. CONCLUSION

Automatic musical genre classification is a difficult pattern recognition task. In this paper we have presented an approach to musical genre classification that combines Support Vector Machine Method with basic assumption that music can be successfully classified into genres without involving complicated music theory and by proposed set of nine stated features.

The results achieved by the proposed approach are similar to some results from the literature [4]. However, it should be stressed that these studies have used different datasets and experimental conditions, which makes a direct comparison very difficult.

Future work should include other combinations of strategies that include inserting some different kind of features, e.g. rhythm features.

Also, in this case only tree very typical music genres are considered. It will be interesting to apply similar approach for classification of music songs from larger number of genres. In this case, sound recognition of some specific musical instruments could also be helpful in achieving better results.

ACKNOWLEDGEMENT

The work presented here was supported by the Serbian Ministry of Education and Science (project III44006).

REFERENCES

[1] E. Pampalk, A. Rauber, and D. Merkl. “Content–based organization and visualization of music archives”, In *ACM International Conference on Multimedia*, 2002.

[2] M. Fingerhut. The ircam multimedia library: A digital music library. In *IEEE Forum on Research and Technology Advances in Digital Libraries*, pages 19–21, 1999.

[3] D. Perrot and R. O. Gjerdigen. Scanning the dial: An exploration of factors in the identification of musical style. In Proceedings of the 1999 Society for Music Perception and Cognition, 1999. S. Haykin, *Neural Networks*, New York, IEEE Press, 1994.

[4] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals”, *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[5] J. A. Charles, D. Fitzgerald, E. Coyle, "Violin Timbre Space Features", *IEE Irish Signals and Systems Conference, Dublin, 2006.*

[6] L. Rabiner and B. H. Juang, “Fundamentals of Speech Recognition”, Prentice Hall, 1993.

[7] S. Theodoridis and K. Koutroumbas, “Pattern Recognition – fourth edition”, 2009, Elsevier Inc.

[8] Richard G. Lyons, “Understanding Digital Signal Processing”, Prentice Hall PTR, 2004.

[9] Corinna Cortes and V. Vapnik, "Support-Vector Networks", *Machine Learning, 20, 1995*.

Session SP II:

SIGNAL PROCESSING II

Application of Switched-Capacitive Filters in Anti-Aliasing Filtering

Dragisa Milovanovic[1], Sasa Nikolic[2] and Darko Ilic[3]

***Abstract* – In this paper we introduce possible applications of switched-capacitive filters for anti-aliasing filtering. In order to avoid aliasing effects in the process of sampling an analog signal, it is necessary to use an analog anti-aliasing filter to eliminate noise of outside the useful signal band. Today we have the importance of implementation of integrated solutions that can be packed in small form of chip and aspiration to minimize the integration of number of electronic blocks in one electronic circuit. In addition switched-capacitive filters bring many advantages compared to active filters, require no external capacitors, precision, accuracy and limit the frequency up to ± 0.2%, and less sensitivity to temperature changes. A very important feature of switched-capacitive filter is their flexibility, because changing the frequency of overlap changes the cut-off frequency of the filter and thus avoids the need for changing the values of capacitors in the circuit filter.**

***Keywords* – Anti-aliasing, analog filters, switched-capacitive filters.**

I. Introduction

Aliasing effect emerged as a product of side effects or false signals outside the desired frequency band is a problem in many applications that use the A/D conversion. These signals, if not properly filtered, can seriously damage performance of the system for A/D conversion.

Components of the signal from outside the useful band, which for some reason, we have on input of the A/D converter, due to the aliasing effect will appear in the output samples with frequencies that fall into the desired frequency band. This imposes the need for the necessity of filtering the input signal of the A/D converter. Standard analog filters require a larger number of passive components, which becomes a problem especially in today's highly integrated electronic circuits. For these applications switched-capacitive filter becomes more significant considering the possibility of integration of higher order filters in a small area, but also all the other benefits that this technology brings.

This work will mainly illustrate the occurrence of aliasing effects in order to underscore the importance of removing, then the basis of realization of switched-aliasing filter with the implementation of an integrated solution switched-capacitive filter and simulation results of application EE-Sim for design and simulation of the filter, from the company Maxim Integrated Products, Inc..

II. Anti-aliasing Effect

A/D converters typically operate with a constant sampling frequency when digitalize an analog signal. Using a sampling frequency (f_s), which is usually called the Nyquist rate, all components of the input signal with frequencies below $f_s/2$ will be reliably digitalized. If the input signal has components over the frequency domain $f_s/2$, after A/D conversion that components will appear in a useful range of output signal with preserved values of the amplitude. This phenomenon leads to the fact that it is impossible to discern the difference between the signals from the useful range (below $f_s/2$) and components over the useful range (above $f_s/2$).

The appearance of aliasing effect is illustrated in the frequency domain in Fig 1.

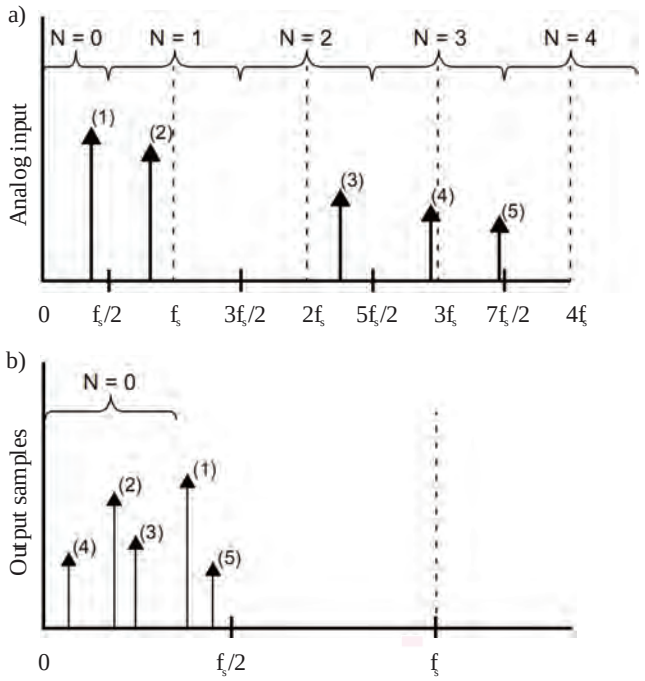


Fig 1. a) A system that sampled input signal with the frequency f_s identify signals with frequency $f_s/2$, and signals with higher frequency b) Signals with a frequency above $f_s/2$ will be represented as components of frequencies below $f_s/2$ in the digital output

[1]Dragisa Milovanovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dragisa.milovanovic@elfak.ni.ac.rs

[2]Sasa Nikolic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: sasa.nikolic@elfak.ni.ac.rs

[3]Darko Ilic is with the Radius South East Europe, Hilandarska 20, 18000 Nis, Serbia, E-mail: darko.ilic@radius-see.com

On the both graphics a) and b) on Fig 1. the x-axis shows the sampling frequency f_s. On the graphic a) can be seen five frequency segments. Segment n = 0 includes components from DC to half the sampling frequency. In this range, the sampling system will accurately display the frequency content of the analog input signal. Sampling high frequency signal components from segments N > 0 is also done by a system for the components segment N = 0, ie the frequency of fs/2. Mathematically, the high frequency components projected into a useful signal band will generate unwanted components which can be described by the following formula:

$$f_{ALIASED} = \left| f_{IN} - Nf_s \right| \quad (1)$$

For example, if we have the sampling rate (f_s) 100kHz and have a frequency content:

$$f_{IN}(1) = 41kHz$$
$$f_{IN}(2) = 82kHz$$
$$f_{IN}(3) = 219kHz$$
$$f_{IN}(4) = 294kHz$$
$$f_{IN}(5) = 353kHz$$

Digital output after sampling will contain the true values of the amplitudes of all five components of the signal, but the four of them lie outside the range of the signal DC to $f_s/2$ or in this case the DC to 50kHz. Using the formula $f_{OUT} = \left| f_{IN} - Nf_s \right|$, frequency input signals are transformed into:

$$f_{OUT}(1) = \left| 41kHz - 0x100kHz \right| = 41kHz$$
$$f_{OUT}(2) = \left| 82kHz - 1x100kHz \right| = 18kHz$$
$$f_{OUT}(3) = \left| 219kHz - 2x100kHz \right| = 19kHz$$
$$f_{OUT}(4) = \left| 294kHz - 3x100kHz \right| = 6kHz$$
$$f_{OUT}(5) = \left| 353kHz - 4x100kHz \right| = 47kHz$$

This frequency phenomenon can be eliminated or reduced by using low-pass filter in front of the A/D converter. This situation is shown in Fig 2, where we see that the filter passed only components of the signal to the fs/2, in this case only the first component f_{IN} (1), while all other components will be suppressed.

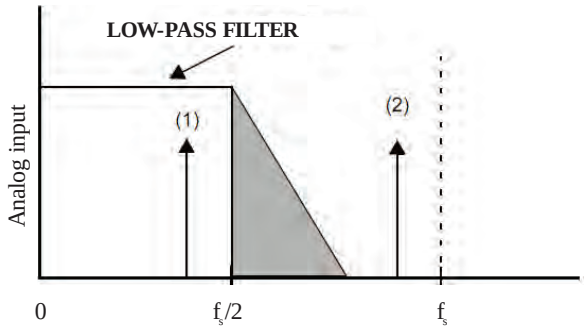


Fig 2. If the A/D system has a low pass filter ahead of the sampling mechanism, high frequency signals will be weakened and will not be sampled

For example, a 12-bit A/D conversion characteristics of the designed filter should be such as to provide at least that the SNR of the system for sampling 74dBA. For the purposes of design of anti-aliasing filters different realisations are used from an ordinary RC cell through standard filters, such as architectures Bessel, Chebyshev, Butterworth, Eliptic.

Important role here plays switched-capacitive filter for the realization of anti-aliasing filter.

III. Realization of Switched-capacitive Filter

Many designers have an aversion to analog filters because they are difficult to design, have a great tolerance and are difficult to produce, especially with the strong requirements. The rule that should be taken in the design is that the tolerances of discrete components accumulate. So with standard tolerances of resistors and capacitors, we can expect an error in relation to the estimated frequency. A good alternative to solves this problem is to use integrated solution filters.

There are two types of integrated filters, continuous time and switched-capacitive filters. Continuous time filters usually have a need for some external components to adjust the cutoff frequency, which makes them limited in the layer of flexibility. Switched-capacitive filters can, because of its architecture to be very flexible. If used properly can be an excellent alternative to discrete and continuous time integrated filters.

Switched-capacitive technology has long been known architecture that can be reliably and multiplicative implemented on silicon. How it works and some mathematical foundations are given in Fig 3.

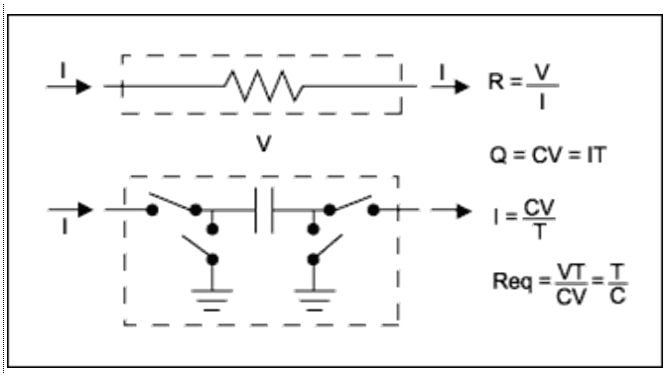


Fig 3. Switched-capacitive schematic

The operation is as follows: capacitor charged and discharged periodically depends to the switches on both sides. This causes the charge transfer, leading to the appearance of a pulsating current. The mean current can be calculated, and if the frequency of switching is large enough, this current will be equal to the current through the resistor. In basic terms, this means that the resistor is replaced by a capacitor. The value of current or indirect resistor values depend on two variables: the size of capacitors, and frequency of the switching. If the filter is created using this architecture,

frequency behavior of the filter can be changed by changing the value of the capacitor or switching frequency. In the integrated solutions, the value of capacitor is fixed, so the frequency characteristics of the filter should be changed by changing the switching frequency. The principle scheme of such a filter is shown in Fig 4.

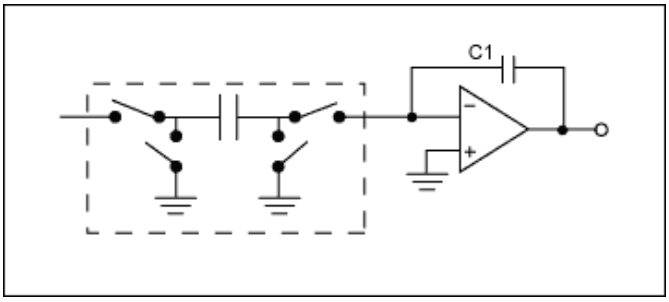


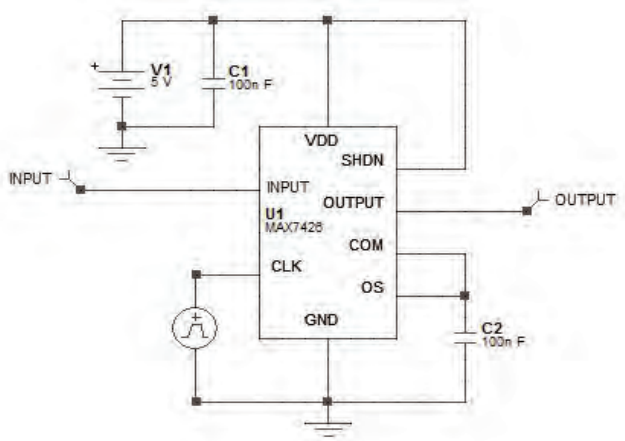
Fig 4. Simple filter implemented using Swiched-capacitive technology

In the case of a discrete filter, we can only use the unmatched components, while the integrated solutions matching these components can go in the range of 0.1%. So, we can expect good control characteristics of the filter. For example, round firm Maxim, MAX7490 has a precision of the cut-off frequency of about 0.2%, which is impossible to achieve discrete components. Also, the temperature drift is a great 10ppm /° C.

It can be said to switch capacitive filter is actually the same sampled signal. It actually converts the time-continuous signal at discrete time signal.

IV. RESULTS

Below is displayed the realization of a low-pass filter using the MAX7426 integrated circuit that can be used as anti-aliasing filter. MAX7426 integrated circuit is a switched-capacitive low-pass filter 5-th order with architecturez, making it suitable for the realization of doing anti-aliasing filter. Circuit is equipped with tuning of cut-off frequency from 1Hz to 12kHz. It is envisaged that the clock frequency 100 times greater than the desired cut-off frekvencije. Most of the switched-capacitive filters are achieved via biquadratic sections, each with two pairs of zeros and sections can be cascaded to tie in order to obtain higher order filter. The advantage of this approach is a simpler design. On the scheme of implementation, shown in Fig 5, we can see that the chip has no external passive components, except that two capacitors are used for decoupling of the power circuits.



.

Fig 5. Implementation scheme of the switched-capacitive filter with integrated solution Maxim, MAX7426 circuit

In figures below are shown the transmission characteristics of filters, as well as the launch-response functions, and a squared sinus function of the input filter.

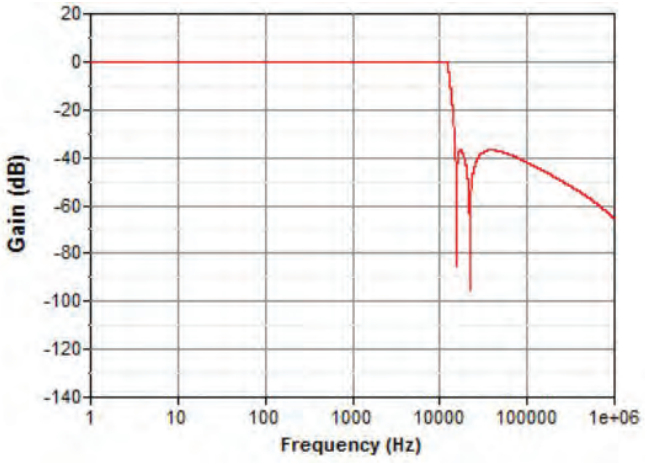


Fig 6. AC transmission characteristics of filter

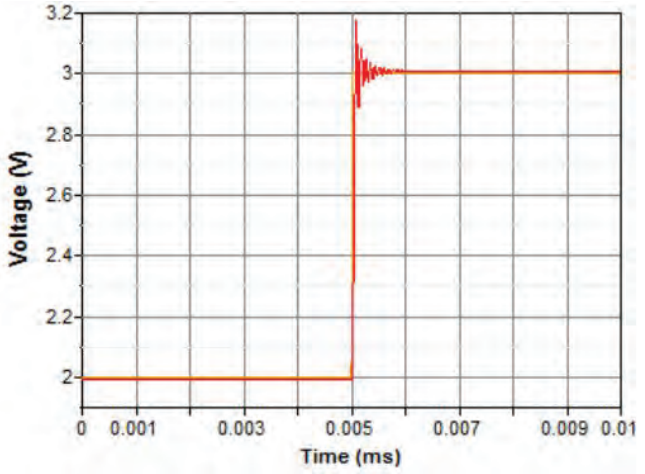


Fig 7. Response to a step function (input signal - orange, the output signal - red)

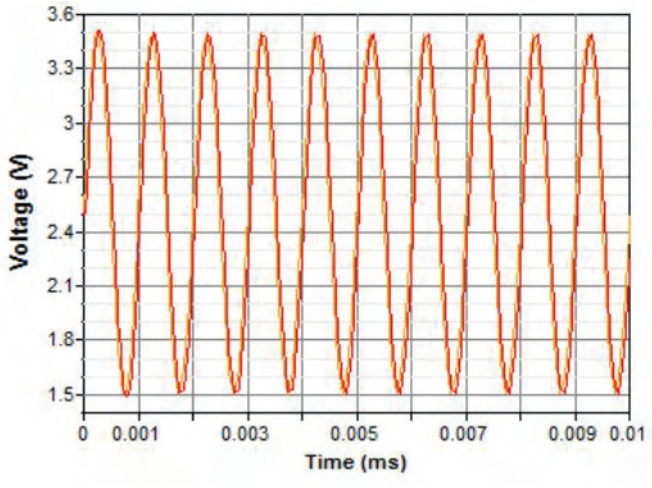


Fig 8. The response of the sinus function (input signal - orange, the output signal - red)

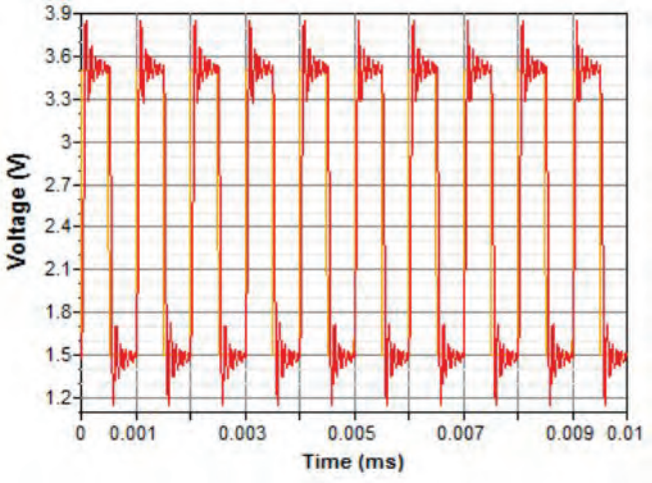


Fig 9. Response to a rectangular function (input signal - orange, the output signal - red)

IV. CONCLUSION

As a conclusion we have all the advantages of switched-capacitive technology that are mentioned. First of all we should mention the little area that offers a solution especially in the realization of higher order filters. In addition this filters allow a simple implementation, without external passive components that contribute to the instability of the filter due to tolerance, temperature drift, etc.

Of course we should be aware that this technology is like any other has its drawbacks and limitations. Possible problem is the appearance of jitter in clock time distribution in the circuit, which leads to distortion. Since in circuit where they would use such a filter we have some A/D converter, uC or DSP, which all use a clock, it is necessary to solve the problem of synchronization of all clocks. Therefore, the commonly way is to use single oscillator whose signal causes the circuits to work as a clock generator, and that generates a clock for each component in the circuit, with the lowest jitter of order 10^{-10}s.

REFERENCES

[1] Microchip Technology Inc., “Anti-Aliasing, Analog Filters for Data Acquision Systems”, *AN699 Application Note*, www.microchip.com

[2] Maxim Integrated Products, Inc., “The Basics of Anti-Aliasing: Using Switched-Capacitor Filters”, *AN3494 Application Note*, www.maxim-ic.com

[3] Maxim Integrated Products, Inc., “Filter Basics: Anti-Aliasing”, *AN928 Application Note*, www.maxim-ic.com

[4] Kyaw Soe Lwin, “Design and Implementation of Analog Filter with Anti-Aliasing Parameter”, *GMSARN International Conference on Sustainable Development: Issues and Prospects for the GMS*, 12-14 November 2008.

[5] Charles Yager and Carlos Laber, Micro Linear, “Switched-Capacitor Filters Beat Active Filters at Their Own Game”, *TechOnLine Publication*, Jun. 29, 2000.

Modified Legendre Filters with Minimization of Summed Sensitivity

Vlastimir Pavlović[1], Maja Lutovac[2] and Miroslav Lutovac[3]

Abstract – **The paper presents the design of active RC filter based on a new class of all-pole approximation. The approximating function is derived using Legendre orthogonal polynomial with the appropriate weights and usage of two Legendre multiplication factors at the origin and the passband edge frequency. Detailed analysis is done for the frequency response and tolerance analysis of active RC filters.**

Keywords – **Approximation, Tolerance analysis.**

I. INTRODUCTION

Continuous-time active RC filters are suitable for integration into the analog front end of mixed mode VLSI (Very-Large-Scale Integration) chips for communication systems. Programmable analog filters, and SC (Switched Capacitor) integrated filters, are replacing the classic analog filters (such as active opamp RC – operational amplifier + Resistor-Capacitor), but the design procedure is still based on the sensitivity and tolerance analysis [1, 2] in order to manufacture robust filters, to increase the production yield and to minimize the cost of mass production. The sensitivity and tolerance analysis allows the designer to predict variations of the filter performances and to predict the production yield (which is defined as a ratio of the number of manufactured filters satisfying the specification to the total number of manufactured filters) [1, 2].

It is well known that in the design of filters with real elements, the most influential factors on the filter performance may be the finite tolerances and temperature changes of their components. However, although the deviation between the designed and the measured attenuation characteristics, especially in the pass-band, is very important in the design of filter functions, which satisfies the specified characteristics when implemented with real components, it does not seem to be fully covered in the literature available.

It should be stressed that realizations using ideal components may disregard influence of changes due to the temperature variations. We are trying to overcome this problem using appropriate transfer function in such a way to minimize the summed sensitivity and thus to minimize temperature changes using component with the same temperature coefficient. In some papers, the problem was solved by designing filters with low Q factor of critical pair of poles [3–6].

[1]Vlastimir Pavlović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vlastimir.pavlovic@elfak.ni.ac.rs.

[2]Maja Lutovac is with Lola Institute, Belgrade, Serbia, E-mail: majalutovac@yahoo.com.

[3]Miroslav Lutovac is with SUNP, Novi Pazar: lutovac@etf.rs

II. APPROXIMATING METHOD

The most general form of lowpass prototype filter all-pole transfer function is

$$H_n(s) = \frac{K\prod_{r=1}^{n}(-s_r)}{\prod_{r=1}^{n}(s-s_r)} \tag{1}$$

The filter order is n, K is a constant to specify the attenuation at s=0 (for example 0 dB), and the poles of the transfer function are

$$s_r = \sigma_r + j\,\omega_r, \qquad r(1, 2, 3, \ldots n)$$

The squared magnitude response is

$$H_n(\omega)\,H_n(\omega) = \frac{1}{1+\varepsilon^2\,A_n(\omega^2)} = \frac{1}{1+\varepsilon^2\sum_{i}^{n} a_{2i}\omega^{2i}} \tag{2}$$

The denominator of the squared magnitude response is a polynomial in ω, with rael contstants a_{2i}. The parameter $\varepsilon^2 = \rho^2/(1-\rho^2)$ determines the attenuation at the pass-band edge, while ρ is the pass-band reflection factor.

The characteristic function is normalized to1 at the pass-band edge frequency

$$A_n(\omega_p) = 1 \tag{3}$$

The squared magnitude response of new class becomes

$$H_n(\omega)\,H_n(\omega) = \frac{1}{1+\varepsilon^2\left(\sum_{r=0}^{n} b_{2r}\,P_{2n}(\omega)\right)} \tag{4}$$

$P_r(\omega)$ is the rth-order Legendre orthogonal polynomial.

The minimum of the ratio of the reflected power and the transmitted power is obtained by minimizing the following integral

$$I_{\min}(\omega) = \int_0^1 p(\omega)\,A_n(\omega^2)\,d(\omega) \tag{5}$$

The weight $p(\omega)$ is 1. Firstly, we define a function using conditions of the proper prototype low-pass approximation for odd order

$$\phi(b_0, b_2, b_4, b_6, \ldots b_{2n}, \lambda_0, \lambda_1) = \int_0^1 [\sum_{r=0}^{r=n} b_{2r} P_{2r}(\omega)]^2 d(\omega) - \lambda_0[\sum_{r=0}^{r=n} b_{2r} P_{2r}(0)] - \lambda_1[\sum_{r=0}^{r=n} b_{2r} P_{2r}(1) - 1] \quad (6)$$

Next, we derive partial derivatives and set a system of equations that should be solved in terms of b_0, b_2, $b_4, b_6, \ldots, b_{2n}$

$$\frac{\partial}{\partial b_0}\phi(b_0, b_2, b_4, b_6, \ldots b_{2n}, \lambda_0, \lambda_1) = 2b_0 - \lambda_0 P_0(0) - \lambda_1 P_0(1) = 0 \quad (7)$$

$$\frac{\partial}{\partial b_{2r}}\phi(b_0, b_2, b_4, b_6, \ldots b_{2n}, \lambda_0, \lambda_1) = \frac{2}{1+4r} b_{2r} - \lambda_0 P_{2r}(0) - \lambda_1 P_{2r}(1) = 0 \quad (8)$$

$$r = 1,2,3,\ldots,n$$

$$\frac{\partial}{\partial \lambda_0}\phi(b_0, b_2, b_4, b_6, \ldots b_{2n}, \lambda_0, \lambda_1) = \sum_{r=0}^{r=n} b_{2r} P_{2r}(0) = 0 \quad (9)$$

$$\frac{\partial}{\partial \lambda_1}\phi(b_0, b_2, b_4, b_6, \ldots b_{2n}, \lambda_0, \lambda_1) = \sum_{r=0}^{r=n} b_{2r} P_{2r}(1) - 1 = 0 \quad (10)$$

The properties of the new class of functions will be demonstrated by example.

III. Example Design

Let us design a filter of the order $n = 10$ with $\rho = 0.25$. The poles of the transfer function are computed using equations (7)-(10). Magnitude response in dB is presented in Figure 1 for the 10th-order filter. It is important to notice that attenuation oscillates around 0dB. This type of approximation cannot be implemented using lossless LC filter because it has negative attenuation (gain) in the passband. The poles of the designed transfer function are

$$\begin{pmatrix} -0.3638978001828931 - 0.1566353531179695\,i \\ -0.3638978001828931 + 0.1566353531179695\,i \\ -0.3231936893533610 - 0.4592472353776506\,i \\ -0.3231936893533610 + 0.4592472353776506\,i \\ -0.2515205288535666 - 0.7206516018040339\,i \\ -0.2515205288535666 + 0.7206516018040339\,i \\ -0.1571211264738653 - 0.9145512196545400\,i \\ -0.1571211264738653 + 0.9145512196545400\,i \\ -0.0524125445027391 - 1.0204240278599806\,i \\ -0.0524125445027391 + 1.0204240278599806\,i \end{pmatrix}$$

The poles of the 11th-order transfer function are

$$\begin{pmatrix} -0.3423280721141051 \\ -0.3251529281079403 - 0.2819569693507116\,i \\ -0.3251529281079403 + 0.2819569693507116\,i \\ -0.2806469111451467 - 0.5454420792584142\,i \\ -0.2806469111451467 + 0.5454420792584142\,i \\ -0.2140524158334217 - 0.7671968743259037\,i \\ -0.2140524158334217 + 0.7671968743259037\,i \\ -0.1319460394353278 - 0.9293038103800189\,i \\ -0.1319460394353278 + 0.9293038103800189\,i \\ -0.0437164371089922 - 1.0171395361471176\,i \\ -0.0437164371089922 + 1.0171395361471176\,i \end{pmatrix}$$

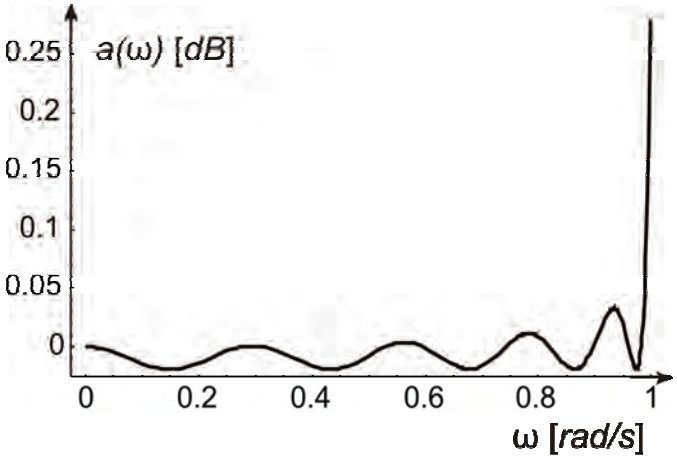


Fig. 1. Attenuation in dB of the 10th order filter.

Figure 2 shows the summed sensitivity in the passband expressed in dB. Very small sensitivity in the passband implies that the variation of the magnitude response due to temperature changes will be small for the case of the same temperature coefficient of all elements of the same type.

More details about summed sensitivity can be found in [1] and [2].

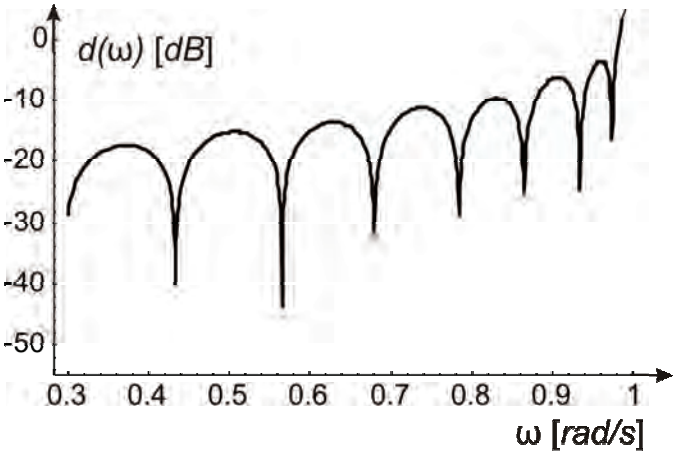


Fig. 2. Summed sensitivity of the 10th order filter.

Figure 3 shows the summed sensitivity in the passband expressed in dB for 11th-order filter. Again, the summed sensitivity is very low in the passband.

The passband variation in dB is illustrated in Figure 4. This type of approximation provides very low passband variation

in the most of the passband, and attenuation increases at frequencies close to the edge frequency. Since the larges sensitivity and larges deviation of the attenuation due to element values changes are at the passband edge frequency, the filter can be more robust after moving the edge frequency into the transition region, as it is done in [7].

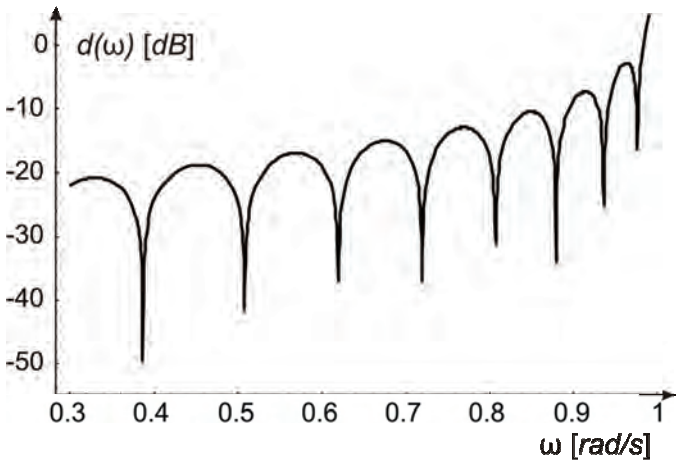


Fig. 3. Summed sensitivity of the 11th order filter.

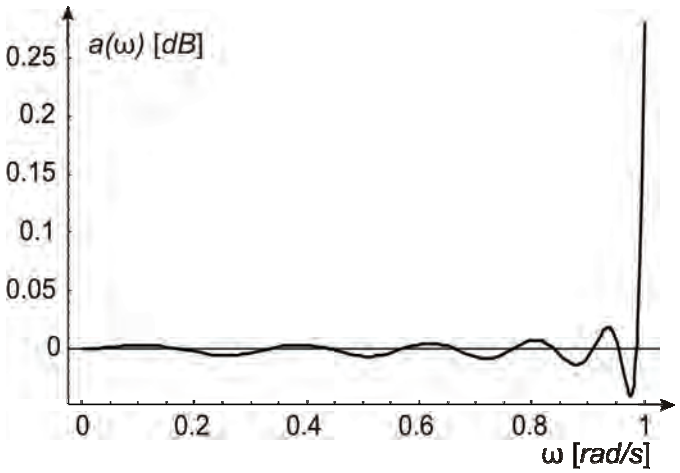


Fig. 4. Attenuation in dB of the 11th order filter.

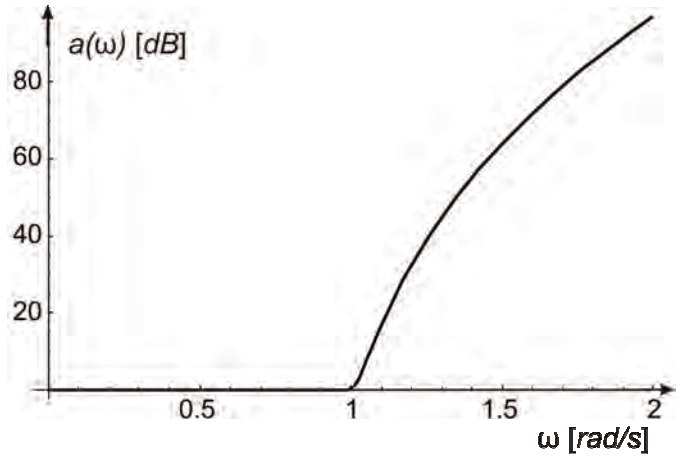


Fig. 5. Attenuation in dB of the 11th order filter.

Figure 4 shows that the maximum deviation can be controlled in a similar way as in the case of reducing the effect of imperfection by reducing Q factor of second order sections.

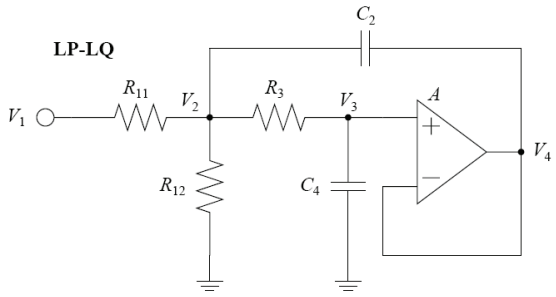


Fig. 6. Low-pass low-Q factor biquad.

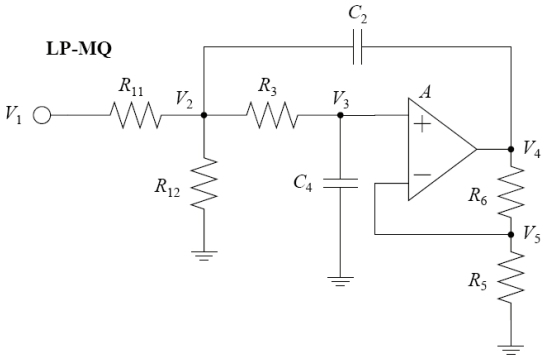


Fig. 7. Low-pass medium-Q factor biquad.

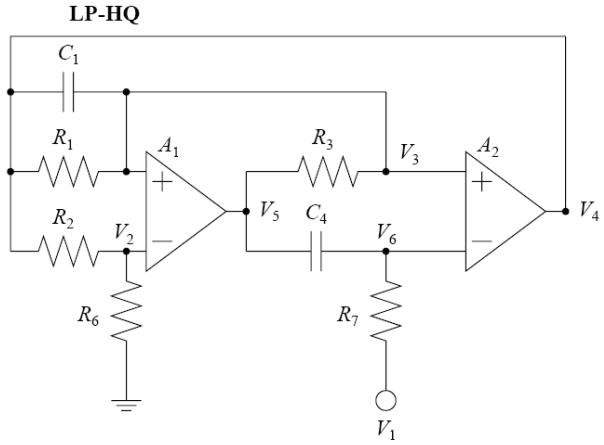


Fig. 8. Low-pass high-Q factor biquad.

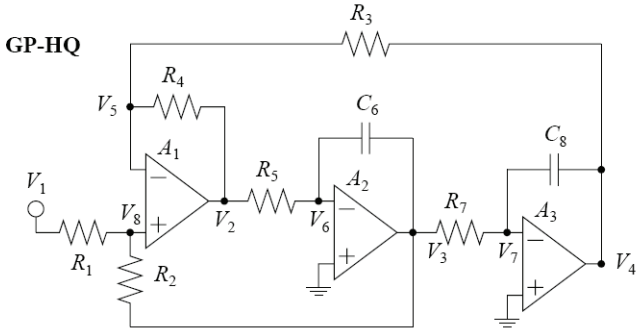


Fig. 9. Low-pass high-Q factor filter - general purpose biquad.

Figure 6 illustrates that the attenuation increases monotonically in the stopband, because this type of approximation belongs to all-pole approximations.

The design procedure is based on the second order filter sections described in [2]. Some of typical biquads are swown in Figures 6, 7, 8, and 9.

The general purpose second-order section can be implemented using programmable analog integrated circuits, such as biquad shown in Figure 10.

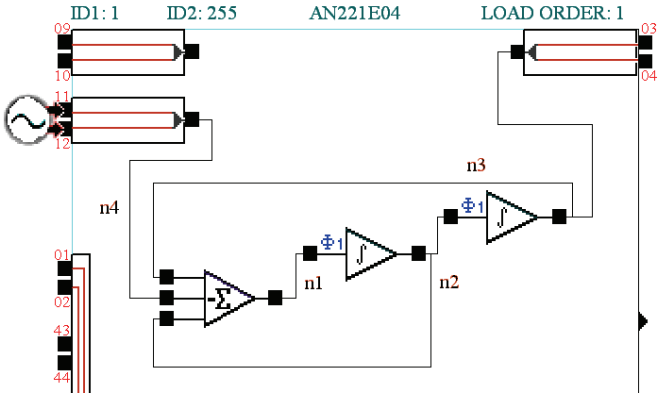


Fig. 10. General purpose opamp biquad implemented using programmable analog integrated circuits.

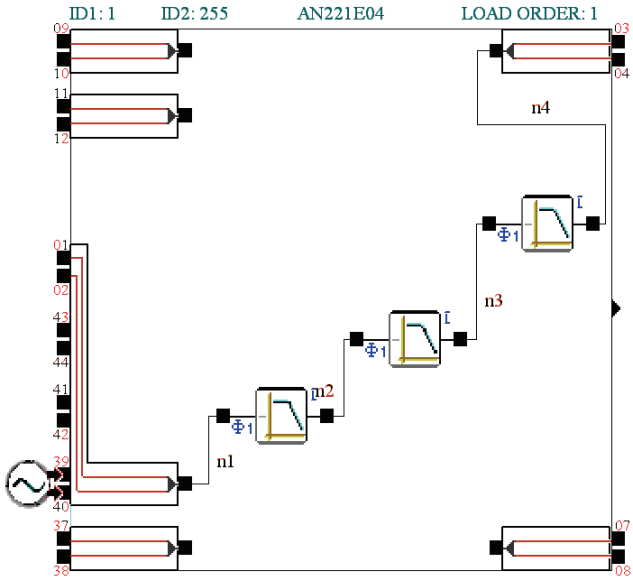


Fig. 11. Sixth-order filter implemented using programmable analog integrated circuits and 3 biguads.

The designed second-order sections as a building block can be used for implementing highest order filters, see Figure 11. Q factor can be extremely high with respect to classic opamp implementations, see Figure 12.

The design procedure is programmed in Mathematica [8]. The whole analysis is evaluated for arbitrary precision, and automated using equations (7) – (10).

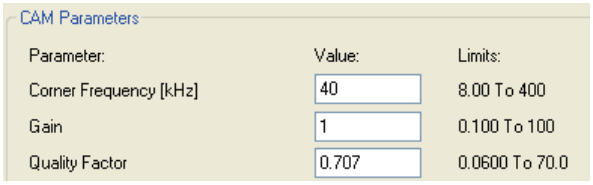


Fig. 12. Programmable parameters of biguads.

IV. Conclusion

Detailed analysis of the frequency response and summed sensitivity in the passband for active RC filters are presented for a class of modified Legendre filters. The filter parameters can be computed for optimization of some of the filter property, such as to minimize the magnitude response deviation in the passband.

In practice, it is more comfortable to have an application as executive programme, using C++ or Java, or web application for computing element values, filter parameters and filter characteristics. The future work is to develop software that will be available for users on different platforms.

Acknowledgement

This work was partially supported by the Ministry of Science of Serbia under Grant TR-32023.

References

[1] V. D. Pavlović, “An explicit form of all-pole filter function with decreasing envelope of the summed sensitivity function” *Int. J. Circ. Theor. Appl.*, vol. 39, no. 5, pp. 515–531, 2011.

[2] M. D. Lutovac, D. V. Tošić, and B. L. Evans, *Filter Design for Signal Processing Using MATLAB and Mathematica*, New Jersey, Prentice Hall, 2000.

[3] B. Djuradj, G. Goussetis, “Design of asymmetrical RF and microwave bandpass filters by computer optimization,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 4, pp. 1174–1178, 2003.

[4] S. Huang, Y. Lee, “Tapered dual-plane compact electromagnetic bandgao microstrip filter structures,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 9, pp. 2656–2664, 2005.

[5] Z. Hao et al., “Compact super-wide band-pass substrate integrated waveguide (SIW) filters,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 9, pp. 2968–2977, 2005.

[6] G. Mathaei, “Narrow-band, fixed-tuned and tunable band-pass filters with zig-zag hairpin–comb resonators,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 4, pp. 1214–1219, 2003.

[7] M. Lutovac, *Design of Analog and Digital Filters Using Computer Algebra Systems*, in Serbian, Belgrade, Academic Mind, 2011.

[8] S. Wolfram, *The Mathematica Book*, 5th ed., Champaign: Wolfram Media, 2003.

FIR Filter Design using Compressed Cosine Polynomial Approximation

Peter Apostolov[1]

Abstract – **This paper considers a new method for FIR filters design. The method uses an L_∞ optimality norm. To achieve a better approximating effect, a new modulating function which compresses the oscillations of the cosine is proposed. A parameter sets the gradient of the modulating function, with respect to the oscillations' compression. The approximating polynomial is carried out using Remez' exchange algorithm. An optimal polynomial with lowest possible (four) degree, that approximates an ideal filter's response with high precision is proposed. With the proposed method an FIR filter with arbitrary specifications can be designed. Design example of low pass FIR filter with a minimization of calculation is performed. The obtained filter's response is close to the ideal low pass filter response.**

Keywords – **FIR digital filters, Frequency response, Polynomial approximation.**

I. INTRODUCTION

The filters' design is a mathematical problem for an approximation of the ideal filter's response. This is an ideal function, comprising rectangular shape with two areas: pass band and stop band. The analytical expression of the ideal low pass filter response is.

$$D(\omega)=\begin{cases}1, \omega\in[0,\omega_c]\\ 0, \omega\in(\omega_c,\pi]\end{cases}, \quad (1)$$

where ω_c is the cut-off frequency. In the FIR filter design, the approximation function is a cosine polynomial

$$A(\omega)=\frac{1}{\sqrt{2}}a_0+\sum_{n=1}^{m}a_n\cos(n\omega)\triangleq\sum_{n=0}^{m}a_n\cos(n\omega). \quad (2)$$

The coefficients of the filter's impulse response are determined by the coefficients of the polynomial. The difference between the ideal function and the approximating polynomial defines the error function.

$$E(\omega)=D(\omega)-A(\omega),\ \omega\in[0,\pi]. \quad (3)$$

When the approximation is optimal, the graph of the approximating polynomial oscillates with equal amplitude close to the ideal function's graph. The amplitude of the oscillations determines the approximation's error. The three most popular norms used in FIR design are as follows:

-- weighted L_1 norm. The approximation's error is

$$\|E(\omega)\|_1=\int_0^{\pi}W(\omega)|E(\omega)|d\omega; \quad (4)$$

-- L_2 error - weighted integral least-squares norm

$$\|E(\omega)\|_2^2=\int_0^{\pi}W(\omega)|E(\omega)|^2 d\omega; \quad (5)$$

-- L_∞ error - weighted Chebyshev's norm

$$\|E(\omega)\|_\infty=\max_{\omega\in[o,\pi]}[W(\omega)|E(\omega)|]. \quad (6)$$

In the all above cases $W(\omega)$ is a positive weighting function, used in order to weight certain frequencies. When $W(\omega)=1$, the maximal error in the pass band and stop band is equal. In this case the approximation using L_∞ norm is equiripple.

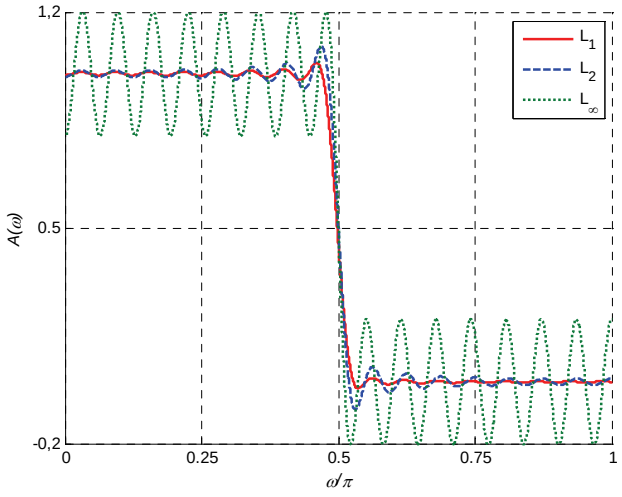


Fig. 1. Approximations with 32-degree polynomials using L_1, L_2 and L_∞ norms

Figure 1 shows a comparison of 32-degree approximating polynomials using L_1, L_2 and L_∞ norms. It is seen that in the FIR filters design a suitable trade-off between flatness and the transition bandwidth must be made. In all the criteria, the functions have the oscillations in the pass band and stop band. These oscillations are undesirable. The goal in the design is to

[1]Peter Stoyanov Apostolov is from the Institute for Special Technical Equipment - MI, Bulgaria, Sofia, 1799, POB 83; e-mail: p_apostolov@abv.bg.

obtain a rectangular shape of the ideal filter's response, that is maximally flat pass band and stop band, and narrowest possible transition band.

In L_1 and L_2 cases the oscillations increase near to the function's transition band. This is due to the Gibbs' phenomenon [1]. The filter design using L_2 criterion is studied in details in many publications. A various methods for reduction of the Gibbs' phenomenon have been proposed: with window functions, "don't care" transition band, optimal change the transition band etc.

Methods for L_2 FIR filters design, where the function's overshot near the transition band is constrained, are presented in [2]. The method success is in reducing the Gibbs' phenomenon. The design is known as "Constrained Least Square" filters. The synthesis is performed with iterative algorithm, having set the maximum value of the overshot. As a result an alignment of the oscillations' amplitude near the transition band is obtained. This leads to the transition's band expansion. The transition band is either not set as an input parameter in the synthesis, or set only at one of the frequencies that define it. The transition bandwidth is the result of the approximation, and is called "induced" transition band. In [3] computationally efficient method for L_2 FIR filter design is proposed.

With approximation using L_1 norm [4] filters with flatter response, but with wider transition band are obtained (Fig.1).

The L_∞ norm is the most suitable for the filter design problem. The design is performed with the well known Parks-McClellan's method [5]. This is minimax approximation using the Chebyshev's norm. With this method, the transition band can be accurately defined and arbitrarily narrow. However this increases the approximation's error and the amplitude of oscillations (Fig.1). It was found that with the same specification (pass band ripple, stop band attenuation and transition band width) with Parks-McClellan's method, an approximation with polynomial of lowest degree has been done. That means that the L_∞ filters will be implemented with the least number of coefficients.

A new polynomial approximation using L_∞ norm producing FIR filters with response close to the ideal will be proposed in the article.

II. Background

As noted in all criteria, the approximating polynomial is a sum of cosines. Approximation using L_∞ norm can be done if the polynomial is a linear combination of Chebyshev's polynomials (Fig.2)

$$A(\omega)=\sum_{n=0}^{m} a_n \cos\left(n \arccos\frac{\omega}{\pi}\right),\ \omega\in[0,\pi]. \qquad (7)$$

The approximation is optimal, but inefficient, because the function's graph compresses its oscillations on both ends of the definition domain, while in the transition band they are most sparse. Thus, a high slope of the function in the transition band can not be obtained.

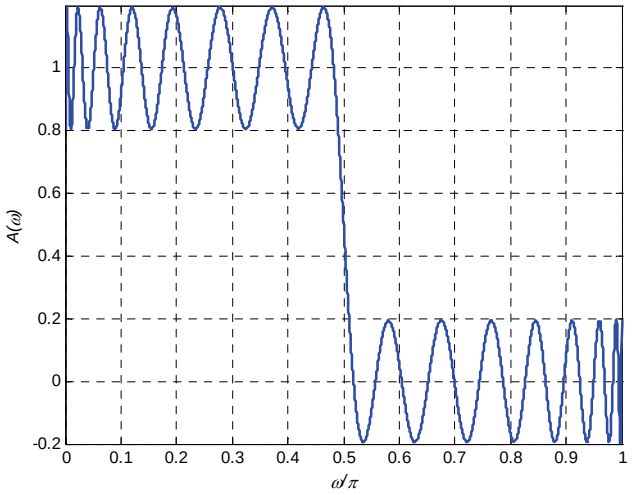


Fig. 2. Approximation with linear combination of Chebyshev's polynomials

From (7) it is obvious that the oscillation's compression is owing to the $\arccos(.)$ function, which modulates the $\cos(.)$ function's argument, as is shown in figure 3.

It is logical to assume if another modulating function with inverse slope to $\arccos(.)$, and higher gradient of the graph is applied, a polynomial with higher oscillations' compression in the transition band, and greater slope will be obtained. A new function called "Third basis function" in [6] is proposed.

$$y(x)=\cos\left\{\frac{m\pi}{2}\left[\tanh\left(\beta x-\frac{\beta}{2}\right)+1\right]\right\};\ x\in[0,1]. \qquad (8)$$

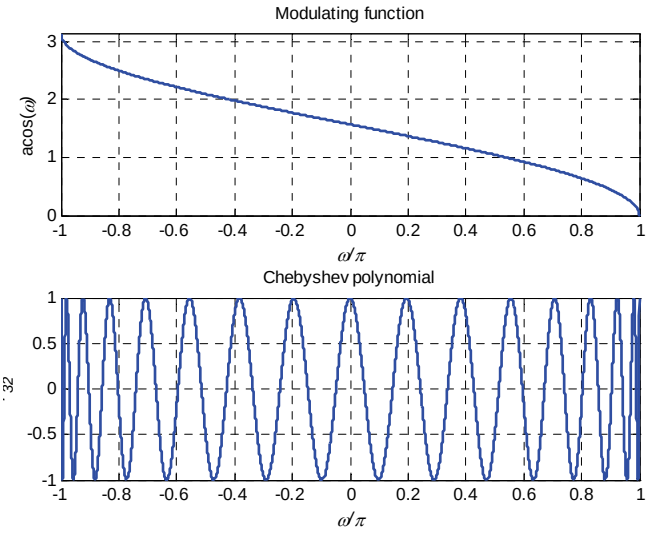


Fig. 3. Modulating function $\arccos(.)$ and Chebyshev's polynomial

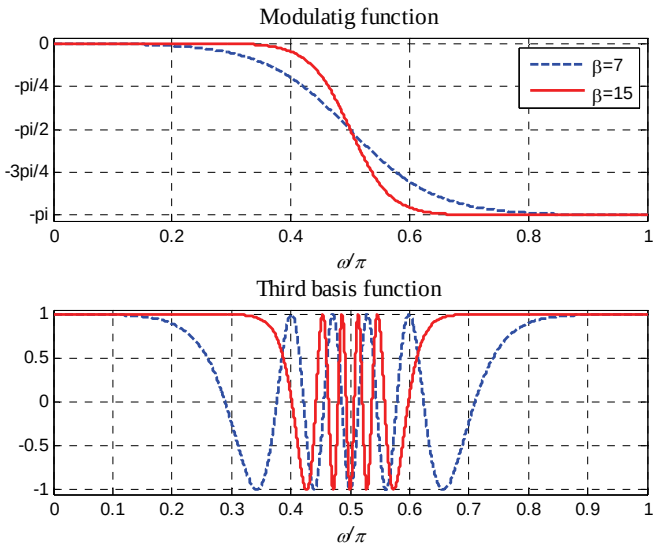


Fig. 4. Modulating function $\tanh(.)$ and Third basis function, $m = 10$

The modulating function is $\tanh(.)$, which comprises parameter $\beta > 4$. Changing the parameter sets the slope of the modulating function, and the oscillations' compression, as is shown in figure 4. Similar to the Parks-McClellan's method, the approximating polynomial with Remez' exchange algorithm [7] is performed. Concerning the specific requirements of the algorithm, the degree of the polynomial *m* is an even number. For the purposes of the FIR filters design the polynomial has the form.

$$A_m(\omega) = \sum_{k=1}^{m+1} a_k \cos\left\{(k-1)\frac{\pi}{2}\left[\tanh\frac{2\beta(\omega-\omega_c)}{\omega_s}+1\right]\right\}, \quad (9)$$

where $\omega \in [0, \omega_s/2]$, and ω_s is the sampling frequency. The goal in the proposed method is that the amplitude of the function's overshot decreases when the parameter β increases, *without changing the width of the transition band, or increasing the degree of the polynomial.*

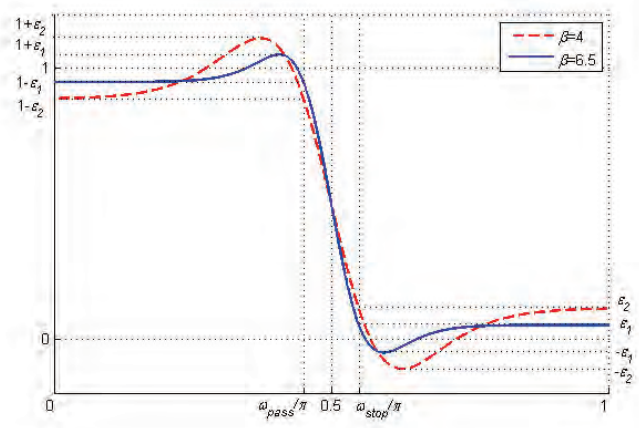


Fig.5. Approximation's error ε depending of the parameter β by fixed transition band

In the other methods it is not so. Figure 5 shows the graph of polynomial with the lowest possible (four) degree for two values of the parameter β. The pass band ripple *DP* and stop band attenuation *DS* are determined by the approximation's error ε as follows:

$$DP = 20\lg(1-|\varepsilon|)\,\text{dB};\ DS = 10\lg|\varepsilon|\,\text{dB}. \quad (10)$$

III. Design Example

The properties of the filters are illustrated through low pass FIR filter example design.

Filter's specification: Cut-off frequency $\omega_c = 0.2\pi$ rad/s; transition band width $\Delta\omega_c = 0.005\pi$ rad/s, sampling frequency $\omega_s = 2\pi$ rad/s; stop band attenuation $DS \geq 60$ dB, power of the polynomial $m = 4$; weighting function $W = 1$.

Parameter's value β can be determined approximately

$$\beta = \frac{\omega_s DS}{32.9\Delta\omega_c} \approx 730. \quad (11)$$

The polynomial's coefficients are determined with the Remez' algorithm: $a_1 = 0.5$; $a_2 = 0.5628$; $a_3 = a_5 = 0$; $a_4 = -0.0628$; and approximation's error $\varepsilon = 7.82\text{e-}7$. The exact values of the pass band ripple and stop band attenuation are calculated by (10); $DP = -6.8\text{e-}6$ dB; $DS = -61.07$ dB.

Equation (9) is the filter's magnitude response. It should be noted that the coefficients of the impulse response are not obtained directly from polynomial's coefficients as by the other methods, as the argument of the function $\cos(.)$ contains other, non-linear function. The impulse response can be determined with IFFT of the 2^N samples of the magnitude response. The filter's design is done using a method, known as "frequency sampling filter" [8]. It is based on FFT in 2^N samples, where *N* is an integer positive number. For the purposes of the design a window function with 2^{N-1} values of the magnitude response (9) are calculated. The signal filtration is performed with convolution between FFT of the input signal and the window function. The filter's structure is shown in figure 6. It should be noted that a larger number of samples should be determined to realize a narrow transition band. In this particular case the appropriate value is $N = 13$. Therefore the window function consists of 4096 values, which determine the filter's length. Figure 7 shows the filter's magnitude response.

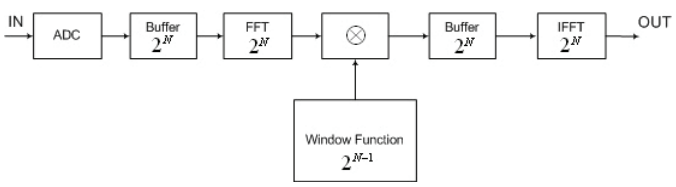


Fig.6. Filter's structure

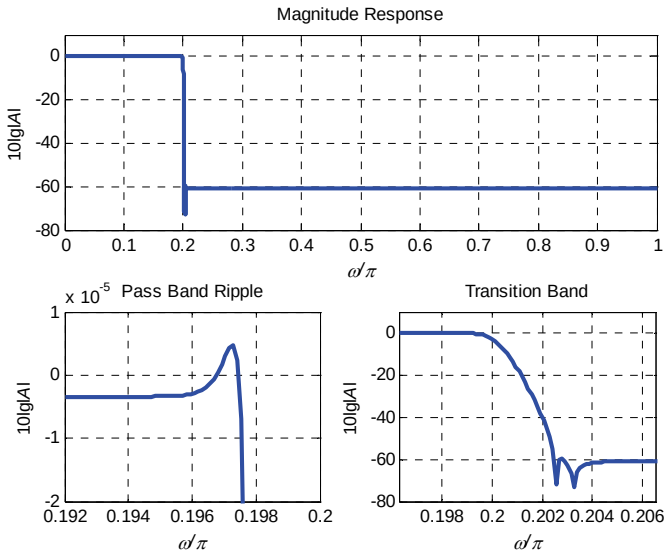


Fig.7. Low pass FIR filter magnitude response

It is obvious that the characteristic is very close to the rectangular shape of the ideal low pass response. The magnitude response is maximally flat. It is seen that large parts of pass band and stop band are constant: $1-\varepsilon$ and ε. Since ε is very small number (approximately 1e-6), it can be assumed that most of the values of the magnitude response in the pass band are equal to one, and in the stop band they are zeros. This circumstance leads to significant reduction of the calculation process of the filter. For the performance of the signal's filtration, it is necessary to execute convolution only in the frequency band slightly wider than the transition band. Of course, it must include the values of the actual transition band. The rest of the pass band frequencies are transferred directly to output buffer (since it is not necessary to multiply by one), and those of the stop band are equal to zero. Figure 8 shows the magnitude response of the same filter with described calculations' minimization. Thus the filter of the considered example is realized with only 32 multiplications.

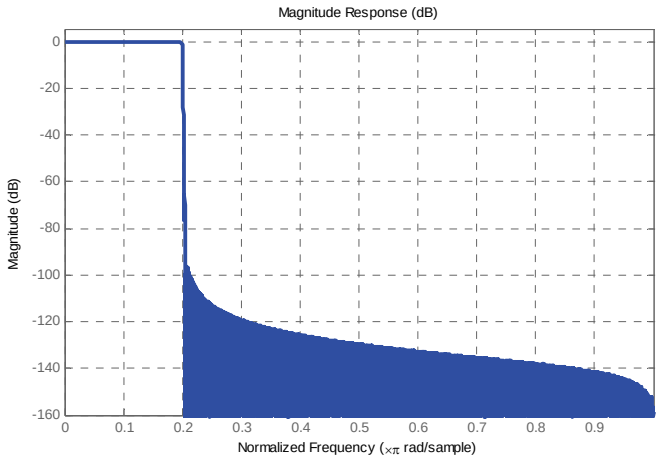


Fig.8. Magnitude response with minimization of the calculations

IV. Conclusion

FIR filters with responses close to ideal filter's response can be designed with the proposed method. A basis function that compresses the oscillations of the function in the transition band is proposed. The method gets its name from this compression. The compression ratio is determined by the parameter β. The approximation is carried out using L_∞ norm that most closely approximates the transition band. The approximating polynomial is obtained by Remez' algorithm, which has fast convergence and low computational complexity. In this case, the low degree of the polynomial (four) involves iterative solving of a system of six linear equations. In other methods they are much more. The method has very low computational cost and design complexity, which is an important advantage. With 4th degree polynomial FIR filter with arbitrary specifications in a *fixed transition band* can be designed. The approximation error ε depends on the parameter β, and defines the pass band ripple and the stop band attenuation. No other polynomial of 4th or lower degree with better approximating properties is known till date. When the width of the transition band is equal to zero and parameter $\beta = \infty$, the polynomial's graph coincides with the ideal response's rectangular shape. Of course, this is impossible in reality. The theoretical design possibilities are limited by the computer calculation accuracy. The implementation of the proposed method makes sense in the design of filters with extreme, close to ideal response. The repeated reduction of the computational operations (Fig.8) is effective in approximations with values of ε close to zero and a very narrow transition band. The proposed method may be a good alternative in several applications in the FIR filters' design.

References

[1] B. Porat, A Course in Digital Signal Processing. New York: Wiley, 1997.

[2] I. W. Selesnick, M. Lang and C. S. Burrus "Constrained Least Square Design of FIR Filters without Specified Transition Bands," IEEE Trans. on Signal Processing, vol. SP- 44, pp. 1879-1892, Aug. 1996.

[3] P. P. Vaidyanathan and T. Q. Nguyen, "Eigenfilters: A new approach to least squares FIR filter design and applications including Nyquist filters," IEEE Trans. Circuits Syst., vol. CAS-34, no. 1, pp. 11–23, Jan. 1987.

[4] Liron D. Grossmann and Yonina C. Eldar, "An L_1-Method for the Design of Linear-Phase FIR Digital Filters, "Signal Process., vol. 55, no. 11, pp. 5253–5265, 2007.

[5] T. W. Parks, J. H. McClellan, "A Program for the Design of Linear Phase FIR Digital Filters", IEEE Trans. on Audio and Electroacoustics, Vol. AU – 20, pp. 196-199, August 1972.

[6] P. S. Apostolov, "Mathematical approximations in the technical communications," DSc thesis, Sofia, 2010.

[7] E. Ya. Remez, "Fundamentals of numerical methods for Chebyshev approximations, " Naukova Dumka, Kiev, 1969.

[8] Lynn P. A. (1975) Frequency sampling filters with integer multipliers. Introduction to Digital Filtering, Bogner R. E. and Constantenides A.G. (eds), New York, Wiley.

Attacks on Digital Image Watermarks in the Discrete Wavelet Transform Domain

Andreja Samčović

***Abstract* - In the last few years, a large number of schemes have been proposed for hiding copyright marks and other information in digital images. Watermaking is a potential method for protection of ownership rights on digital images. This paper presents a number of attacks that enable the information hidden by them to be removed or otherwise rendered unusable.**

I. INTRODUCTION

Digital information is now readily available due to advances in the compression, storage and communication technologies. The amount of digital information that can be found in the Internet and the popularity of the Internet corroborate this observation. Unfortunately, the protection of this information, especially in circumstancies where the owners hope to generate revenue through controlled dissemination, is yet to be standardized. There are at least two consequences of the status quo, First, digital information that is already available is being illegally re-distributed and thereby robbing the legal owners of deserved revenue. Second, this situation discourages content generators from sharing their work with the wider community.

Digital media could be copied easy without loss of quality. Digital watermarking is an appropriate tool for protection of ownership. Digital image watermarking is a communication method in which additional information called watermark is embedded directly and imperceptibly into a digital image, also called original data or host data, to form watermarked data. Loosely analogous to watermarks in article documents, the embedded information is bound to the watermarked data wherever if goes. The embedded information should still be decodable from the watermarked data, even if the watermarked data is processed, copied, or redistributed. Potential applications of digital watermarking include copyright protection, distribution tracing, authentication and authorized access control [1].

Some types of image processing methods can be applied with the explicit goal of hindering watermark reception. In watermarking technology, an attack is any processing that may impair detection of the watermark. Attacks on digital watermarking schemes have two effects: either they reduce the effective channel capacity of fully disable the detection of the embedded watermark. Because it is not possible to enumerate all possible attacks, it is very difficult or even impossible to assess if given system is robust in the general case. What we should realize here is the fact that robustness requirements are application dependent. As a logical consequence, for a given application we first need to define the desired level of robustness and security and then to test against corresponding types of attacks. The danger is, of course, that some attacks may be forgotten or new attacks will emerge in the future. There is no way to avoid this kind of threat and it exists for all types of security related applications. What we should remember from the above comments is that attacks and their efficiencies are application dependent and that is only possible to guarantee robustness to attacks that are known at the time of application development [2].

Better understanding of the mechanisms of possible attacks will lead to the development of more efficient and robust watermarking techniques [3].

Andreja Samčović is with University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11040 Belgrade, Serbia, E-mail: andrej@sf.bg.ac.rs

II. CLASSIFICATION OF WATERMARKING ATTACKS

Five different groups of attacks can be identified: removal attacks, geometrical attacks, cryptographic attacks, protocol attacks and other attacks, which is illustrated in Fig. 1.

2.1 Removal Attacks

Removal attacks aim at the complete removal of the watermark information from the watermarked data without cracking the security of the watermarking algorithm. This category of attacks includes denoising, quantization, remodulation, and averaging. Not all of these methods always come close to complete watermark removal, but they may damage the watermark information significantly.

The main idea of removal attacks consists of assuming that the watermark is additive noise relative to the original image. The removal attacks are those which further add noise to the watermarked image. This noise may have any of a number of different statistical distributions such as Gaussian or Laplacian. The removal attacks exploit the linear additive model in order to derive optimal estimators used for denoising and consequently removing of the watermark. In other cases both the removal attacks and the interference attacks can be combined such as in the denoising with perceptual remodulation attacks [4].

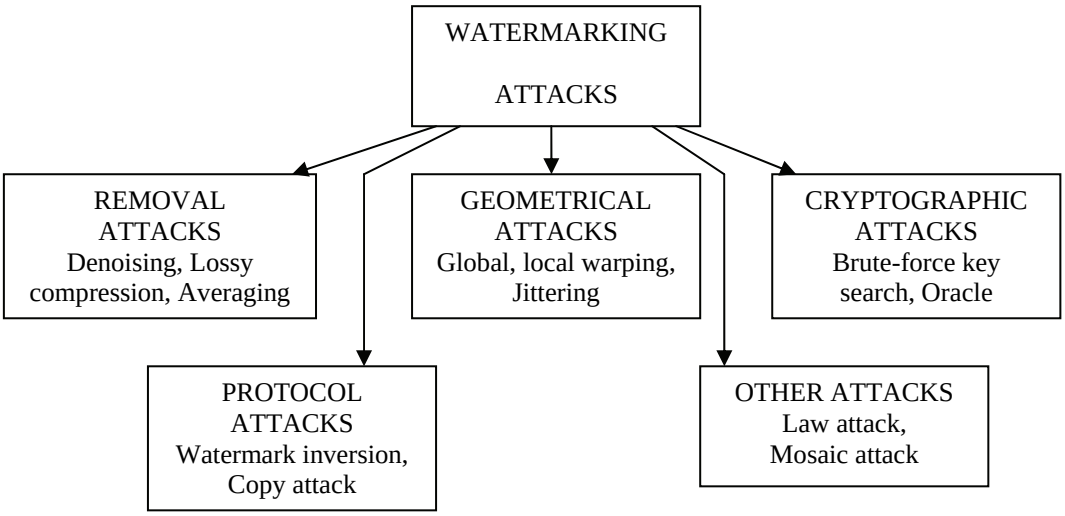


Fig.1. Classification of watermarking attacks

2.2 Geometrical Attacks

In contrast to the removal attacks, geometrical attacks intend to remove the embedded watermark itself, but to distort it through spatial alterations of the stego-data. The attacks are usually such that the watermark detector loses synchronization with the embedded information.

The most well known integrated software versions of these attacks are *Unzign* and *StirMark.* Unzign introduces local pixel jittering and is very efficient in attacking spatial domain watermarking schemes. StirMark introduces both global geometrical and local distortions. The global distortions are rotation, scaling, change of aspect ratio, translation and shearing that belong to the class of general affined transformations. The line / column removal and cropping / translation are also integrated in StirMark. Most recent watermarking methods survive after these attacks due to the usage of special synchronization technique. If robustness to global affined transformations is a solved problem, the local random alterations integrated in StirMark still remain an open problem almost for all techniques.

The so-called random bending attack exploits the fact that the Human Visual System (HVS) is not sensitive against shifts and local affined modifications. Therefore, pixels are locally shifted, scaled and rotated without significant visual distortions. The synchronization removal attacks belong also to this category. The synchronization consists of inserting peaks in the Discrete Fourier Domain (DFT). This method is called „template“.

2.3 Cryptographic Attacks

Cryptographic attacks aim at cracking the security methods in watermarking schemes and thus finding a way to remove the embedded watermark information or to embed misleading watermarks. One such technique is brute-force search for the embedded secret information. Practically, application of these attacks is restricted due to their high computational complexity.

Cryptographic attacks cover, for example, direct attacks to find the secret key or attacks called collusion attacks. Cryptographic attacks are very similar to the attacks used in cryptography. There are the brute-force attacks, which aim at finding secret information through an exhaustive search. Since many watermarking schemes use a secret key, it is very important to use keys with a secure length.

Another attack in this category is so-called Oracle attack which can be used to create a non-watermarked image when a watermark detector device is available.

2.4 Protocol Attacks

Protocol attacks neither aim at destroying the embedded information nor at disabling the detection of the embedded information (deactivation of the watermark). Rather, they take advantage of semantic deficits of the watermark's implementation. The protocol attacks aim at attracting the concept of the watermarking application. The first protocol attack was proposed by Craver et al. [5]. They introduced the framework of invertible watermark and showed that for copyright protection applications watermarks need to be non-invertible. The idea of inversion consists of the fact that an attacker who has a copy of the stego-data can claim that the data contains also the attacker's watermark by subtracting his own watermark. This can create a situation of ambiguity with respect to the real ownership of the data. The requirement of non-invertability on the watermarking technology implies that it should not be possible to extract a watermark from non-watermarked image. As a solution to this problem, the authors proposed to make watermarks signal-dependent by using a one-way function.

Consequently, a watermark must not be invertible or to be copied. A copy attack, for example, would aim at copying a watermark from one image into another without knowledge of the secret key used for the watermark

embedding to create ambiguity with respect to the real ownership of data. It also belongs to the group of the protocol attacks. In this case, the goal is not to destroy the watermark or impair its detection, but to estimate a watermark from watermarked data and copy it to some other data, called target data.

If the watermarking system or protocol makes not only the watermarked image, but also additional devices publicly available, the presence of such devices can be exploited. When exploiting the presence of a watermark detector, a test-image should be created near the detection boundary and then successively change single pixels until the detector response indicates that a particular pixel value has significant influence on the watermark. This way, a set of influential pixels can be determined which has the largest influence on the detector while introducing low disturbance into the image when manipulated. This process has linear complexity. With the presence of a watermark inserter, the difference image between the watermarked and the original image can be easily computed and analyzed. A public watermark inserter is provided by the *Digital Versatile Disc* (DVD) system for copy generation management.

2.5 Other Attacks

The Mosaic attack consists of chopping an image up into a number of smaller subimages, which are embedded in a suitable sequence in a web page. Common web browsers render juxtaposed subimages stuck together, so they appear identical to the original image. This attack appears to be quite general. All marking schemes require the marked image to have some minimal size. Thus, by splitting an image into sufficiently small pieces, the mark detector will be confused.

III. Discrete Wavelet Transform in Watermarking

The wavelet transform (WT) has been extensively studied in last decade. Many applications of the wavelet transform, such as compression, signal analysis and signal processing have been found. There are many good tutorial books and papers on this topic. Here, we just introduce the necessary concepts of the Discrete Wavelet Transform (DWT) for the purpose of this paper.

The basic idea of the DWT for a one-dimensional signal is the following. A signal is split into two parts, usually high and low frequencies. The edge components of the signal are largely confined in the high frequency part. The low frequency part is split again into two parts of high and low frequency. This process is continued until the signal has been entirely decomposed or stopped before by the application at hand. For compression and watermarking application, generally no more than five decomposition steps are computed. Furthermore, from the DWT coefficients, the original signal can be reconstructed.

The wavelet transform decomposes an image into three spatial directions, i.e. the horizontal **HL**, the vertical **LH** and the diagonal **HH**. At each level of decomposition, the magnitude of the DWT coefficients is larger in the lowest subbands ("approximation" **LL** subband), and smaller for other subbands ("detail" subbands: **HL**, **LH** and **HH**). The most significant coefficients in a subband are those with large magnitudes. The high resolution subbands help in locating the edge and texture patterns for an arbitrary image.

IV. Simulation Results

For the purpose of robustness testing the following set of ten standard test-images with the size of 512 x 512 pixels are used: *Barbara, Boat, Cameraman, Couple, Einstein, Elaine, F16, Goldhill, House* and *Lena*. The watermark is firstly converted into ASCII code and than encoded with the error correction code (ECC) in order to improve the robustness. Here, the robustness of the algorithm will be tested for the watermark sequence encoded with three different ECCs and for the watermark sequence that is directly embedded without using ECC. The following ECCs are used in order to determine which ECC performs the best from the robustness point of view:

(15,7) Bose-Chaudhuri-Hocquenghem (BCH) code,
(7,4) Hamming code, and
(15,7) Reed-Solomon (RS) code

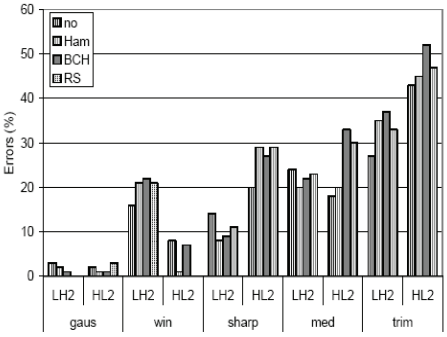
The same watermark is embedded in all detail subbands of the two-level DWT according to the embedding procedure. In order to fit our sequence to the codeword of the ECC for Hamming code, the 8-bit representation of the particular character will be used. For other ECC as well as for the directly embedded watermark sequence, the 7-bit representation will be used. The characteristic of the embedded watermark will be given in the Table I:

TABLE I
CHARACTERISTICS OF THE EMBEDDED WATERMARK

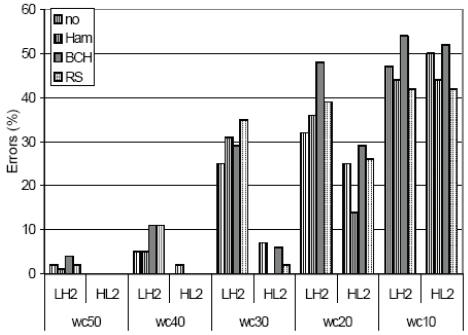
	MESSAGE LENGTH (bits)	ENCODED MESSAGE LENGTH (bits)	ADDITIONAL INFORMATION (bits per character)
NO ECC	147	147	7
BCH	147	315	7
HAMMING	168	294	8
RS	147	360	7

The Table I shows that with the Reed-Solomon coding more then twice of bits have to be embedded into the DWT subband compared to the approach without ECC. This fact must be taken into account when designing the watermark scheme due to the possible problem with the capacity of the cover image.

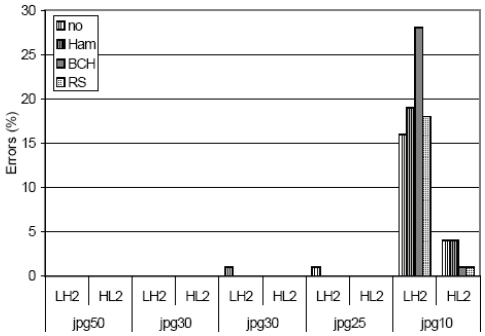
In the testing, several non-geometrical processing operations are applied watermarked test-images: median filtering with 3 x 3 window size (med), Gaussian filtering with 5 x 5 window (gaus), Wiener filtering with 5 x 5 window (wien), trimmed mean filtering with 7 x 7 window (trim), sharpening with 3 x 3 high-pass filter (sh), JPEG compression with different quality factors from 50 to 10 (jpg50, jpg40, jpg30, jpg25, jpg15, jpg10), as well as JPEG compression with different bit rates from 0,5 to 0,1 bits per pixel (bpp) (wc50, wc40, wc30, wc20 and wc10).



(a)



(b)



(c)

Fig. 2. Simulation results for: (a) different filtering attacks, (b) JPEG 2000, (c) JPEG compression attacks

The watermark is extracted separately from every subband in order to compare the robustness of the watermark embedded in that subband. The results for the *Lena* image are given in Fig. 2. The similar results are obtained for other test-images. All graphs in Fig. 2 present different attacks on the x-axes. The results are calculated as the total number of not correctly extracted watermark bits (errors) divided by the total number of watermark bits, expressed in percentage and presented on the y-axes of all three graphs. The best results are obtained for the watermark embedded in the subbands $\mathbf{HL_2}$ and $\mathbf{LH_2}$ and only results for these subbands are presented. The results for other tested subbands were not good and they were not being further considered. This was expected due to the fact that the common signal processing operations like filtering and compression will be most effective in the high frequencies (level 1 of the DWT decomposition).

From Fig. 2 it can be concluded that for the most attacks Reed-Solomon code gives less errors than other ECCs. It can also be concluded that the results strongly depend on the subband in which the watermark sequence was embedded. In some cases like trimmed mean filtering better results are obtained without using ECC.

V. Conclusion

Although the above classification makes it possible to have a clear separation between the different classes of attacks, it is necessary to note that very often a malicious attacker applies not only a single attack at the moment, but rather a combination of two or more attacks. The better understanding of possible attacks will lead to the development of more efficient and robust watermarking techniques.

References

[1] G.Voyatzis, I.Pitas, "The use of watermarks in the protection of digital multimedia products", *Proceedings of the IEEE*, Vol.87, No.6, pp 1197-1207, July 1999.

[2] B.Macq, J.Dittmann, E.Delp, "Benchmarking of image watermarking algorithms for digital rights management", *Proceedings of the IEEE*, Vol.92, No.6, pp 971-984, June 2004.

[3] Z.Bojković, J.Turán, A.Samčović, L.Ovsenik, "Coding, streaming and watermarking – some principles in multimedia signal processing", *Acta Electrotechnica et Informatica*, Vol.4, No.3, pp 13-20, 2004.

[4] F.Petitcolas, R.Anderson, M.Kuhn, "Attacks on copyright marking systems", 2nd Workshop on Information hiding, in *Vol. 1525 of Lecture Notes in Computer Science*, Portland, Oregon, USA, pp 218-238, 14-17. April 1998.

[5] S. Craver, S. Katzenbeisser, „Copyright Protection Protocols Based on Asymmetric Watermarking: The Ticket Concept", in *Communications and Multimedia Security Issues of the New Century*, Kluwer Academic Publishers, pp 159-170, 2001.

System for Acquisition and Analysis of Transesophageal ECG

Yuliyan Velchev[1], Boycho Boychev[2], Elena Boycheva[3] and Kalin Dimitrov[4]

Abstract – **In this paper, a system for acquisition and processing of transesophageal electrocardiogram (TEECG) is described. The system is intended to support the expert analysis in terms of diagnosis of abnormal electrical activity in the heart. The proposed low power acquisition subsystem assures excellent signal-to-noise ratio (SNR) and wideband coverage. The expert analysis is supported by automatic segmentation of the signal as well as Self-Organized Map (SOM) based clustering procedure.**

Keywords – **Esophageal electrocardiography, ECG, clustering, SOM.**

I. Introduction

The acute conduction abnormalities in the heart rhythm are some of the most common causes for hospitalization in the coronary care units. In some cases the differential diagnosis is troubled because of poor visualization of the P wave seen using the conventional 12-lead electrocardiography (ECG).

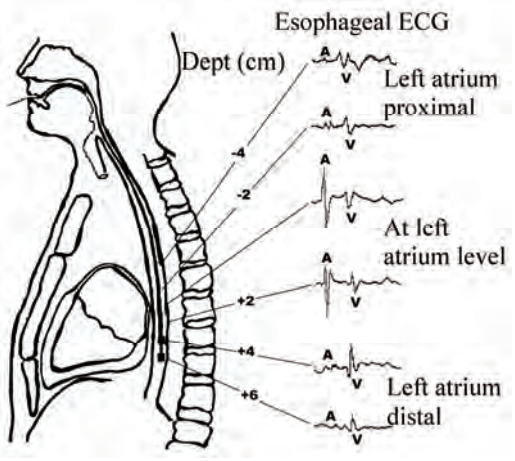


Fig. 1. Transesophageal electrocardiography and P wave morphology depending on probe position [1]

[1]Yuliyan Velchev is with the Faculty of Telecommunications, Technical University of Sofia, 8 "Kliment Ohridski" boulevard, 1000 Sofia, Bulgaria, E-mail: julian_velchev@abv.bg

[2]Boycho Boychev is with the MHAT "Dr. Hristo Stambolski", 16 "Starozagorska" Str., 6100 Kazanlak, Bulgaria, E-mail: dr.boichev@abv.bg

[3]Elena Boycheva is with the ARSENAL JSCo., 100 "Rozova Dolina" Str., 6100 Kazanlak, Bulgaria, E-mail: elena.boicheva@gmail.com

[4]Kalin Dimitrov is with the Faculty of Telecommunications, Technical University of Sofia, 8 "Kliment Ohridski" boulevard, 1000 Sofia, Bulgaria, E-mail: kld@tu-sofia.bg

The TEECG [1] is a diagnostic tool in which the electrodes are formed as an esophageal probe (Fig. 1). This probe is passed into the human esophagus usually under anesthesia. The anatomic closeness of the probe to the human heart as well as the lower electrical impedance give the possibility to achieve signal relatively free of noise and artifacts, thus the visualization of some weak ECG components is considerably improved (Fig. 2).

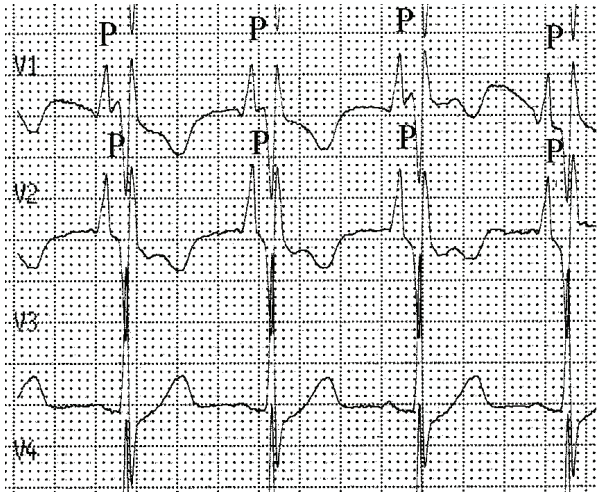


Fig. 2. Transesophageal ECG recordings (V1 and V2) compared with conventional surface ECG (V3) [1]

The method was used for first time in 1906 when an esophageal probe was inserted into the esophagus of a professional sword swallower [2]. Since 70's this technique has been used mainly to analyze atrial arrhythmias [3]. In recent years the interest concerning this method has been renewed [4].

The main challenge in such signal acquisition is the noise level, because the frequency band has to be much wider than conventional ECG registration devices. The frequency band has to cover the range from DC level up to 0.5-1kHz.

This paper introduces a system for acquisition and analysis of TEECG signals. The remainder of the paper is organized as follows: in section II is described the subsystem for signal acquisition; section III continues the work with TEECG signal processing and analysis; in section IV some experimental results and brief discussion are given; section V concludes the work and some aspects for future investigation are mentioned.

II. TEECG Signal Acquisition

The signal acquisition subsystem is based on integrated analog front-end for biopotential measurements type ADS1298 made by Texas Instruments [5]. The concept is to

use Analog-to-Digital Converters (ADC) with very high resolution (24 bit) and amplifiers with relatively low gain (1-12), thus the analog processing of the signals is minimized. A simplified block diagram of the described subsystem is shown in Fig. 3.

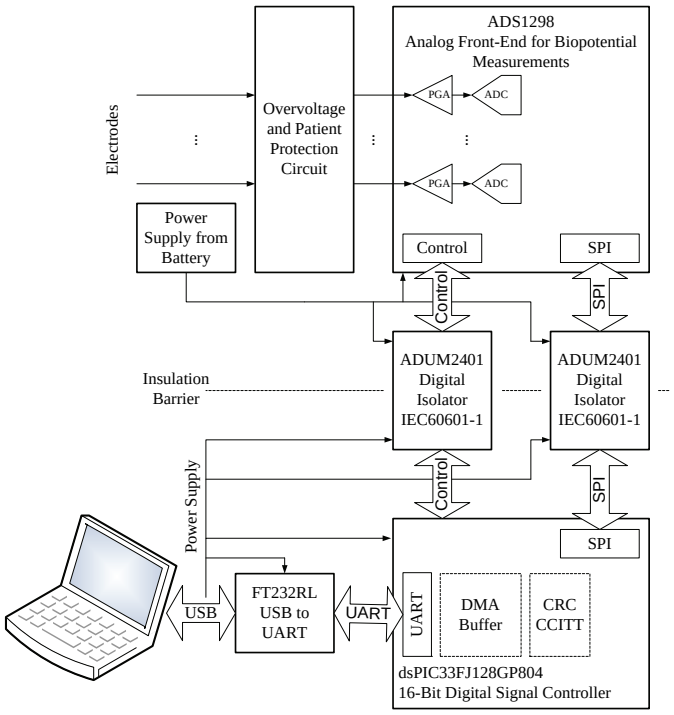


Fig. 3. Simplified block diagram of the subsystem for TEECG signals acquisition

The electrical safety of the patient is assured using digital isolators, which meet the standard IEC60601-1 [6]. The insulation barrier is placed between the analog front-end and the signal controller in order to minimize the battery consumption as well as to decouple the low noise signal acquisition part from digital transients. A Direct Memory Access (DMA) buffer is formed into the RAM memory of the signal controller for fast data transfer between Serial Peripheral Interface (SPI) and Universal Synchronous Receiver Transmitter (UART). The UART baud rate is 921 600 bps. The data is transmitted to the personal computer via UART to USB converter using 999 bytes long packets. The integrity of each packet is assured with 16-bit Cyclic Redundancy Check (CRC) with polynomial according to CRC-CCITT. The sampling rate can be either 1kHz or 2kHz.

III. TEECG Signal Analysis

The signal analysis subsystem is intended only to support expert decision. A simplified block diagram of the proposed sequence is shown in Fig. 4. As can be seen the referent surface ECG taken from external leads is used to determine the positions of R-peaks and the approximate positions of the middles of the baseline between T and next P wave. The R-peaks are found using the algorithm described in [7]. These positions are used to segment the TEECG into subsets which consist of PQRST complexes.

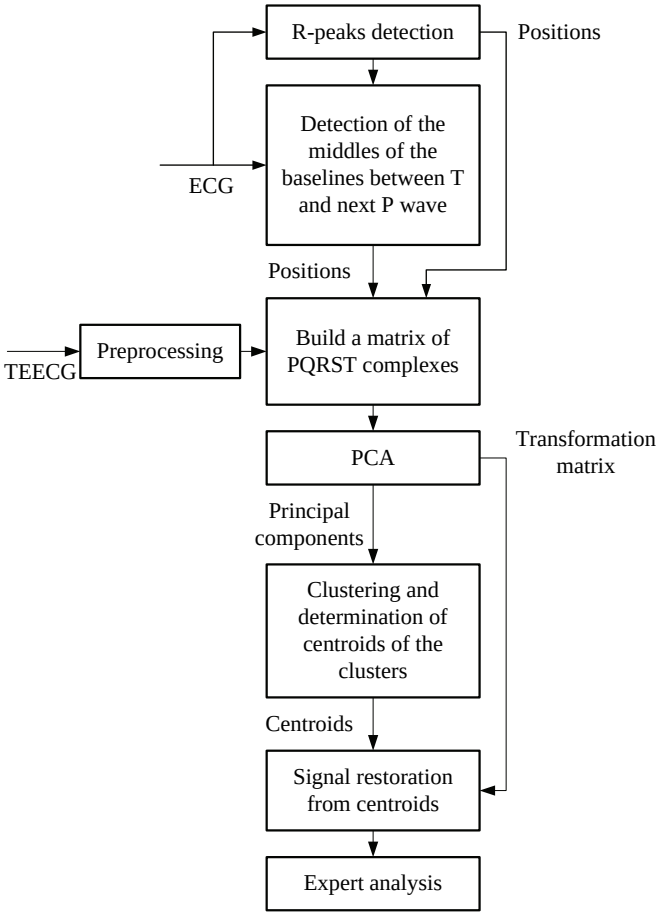


Fig. 4. Sequence for TEECG signals analysis

The rows of the input matrix $\mathbf{X} = \left[x_{ij} \right]_{N \times M}$ are arranged form the PQRST complexes in such way R-peaks are found in the middle column. The next procedure is to perform Principal Component Analysis (PCA) on the matrix $\mathbf{X}$ in order to reduce the very high dimensionality of the input data. PCA is a well known statistical technique for dimensionality reduction based on eigendecomposition, [8]. Let the input data be normalized in terms of unit standard deviation and zero mean according to:

$$\mathbf{Z} = \left[z_{ij} \right]_{N \times M} = \frac{1}{\mathbf{u}\boldsymbol{\sigma}} . \left(\mathbf{X} - \mathbf{u}\boldsymbol{\mu} \right), \tag{1}$$

where $\boldsymbol{\sigma} = \left[\sigma_1, \sigma_2, ..., \sigma_M \right]$ is the vector of standard deviation of $\mathbf{X}$, $\boldsymbol{\mu} = \left[\mu_1, \mu_2, ..., \mu_M \right]$ is the mean vector of $\mathbf{X}$, $\mathbf{u}$ is a column vector of units with number of elements N .and the operator "." denotes Hadamard product. The key procedure is to solve the eigendecomposition of the covariance matrix of the normalized input matrix $\boldsymbol{\Sigma}_Z = \left[\sigma_{Z,ij} \right]_{M \times M} = \frac{1}{N-1} \mathbf{Z}\mathbf{Z}^T$:

$$\mathbf{\Sigma}_Z \mathbf{V} = \mathbf{V}\mathbf{\Lambda}\,, \tag{2}$$

where $\mathbf{V} = \left[v_{ij} \right]_{M \times M}$ is a matrix which columns are the eigenvectors of $\mathbf{\Sigma}_Z$ and $\mathbf{\Lambda} = \left[\lambda_{ij} \right]_{M \times M}$ is a diagonal matrix of corresponding eigenvalues. The vectors in $\mathbf{V}$ and eigenvalues in $\mathbf{\Lambda}$ are rearranged in descending order of the eigenvalues. This leads to resulting matrices $\mathbf{V}' = \left[v'_{ij} \right]_{M \times M}$ and $\mathbf{\Lambda}' = \left[\lambda'_{ij} \right]_{M \times M}$ respectively. The matrix of principal components $\mathbf{Y} = \left[y_{ij} \right]_{N \times M}$ is:

$$\mathbf{Y} = \mathbf{V}'\mathbf{Z}\,. \tag{3}$$

As a result the covariance matrix of the principal components $\Sigma_Y = \left[\sigma_{Y,ij} \right]_{M \times M} = \frac{1}{N-1}\mathbf{Y}\mathbf{Y}^T$ is diagonalized. The elements of the vector of Mean Squared Error (MSE) between the original data $\mathbf{X}$ and restored data $\hat{\mathbf{X}}$ when using a reduced number of principal components $\mathbf{mse}\left(\hat{\mathbf{X}}\right) = \left[mse\left(\hat{\mathbf{X}}\right)_1, mse\left(\hat{\mathbf{X}}\right)_2, ..., mse\left(\hat{\mathbf{X}}\right)_M \right]$ are [9]:

$$mse\left(\hat{\mathbf{X}}\right)_i = \sum_{j=1}^{M} \lambda'_{jj} - \sum_{k=1}^{i} \lambda'_{kk}, i = 1, ..., M\,. \tag{4}$$

The new data dimensionality $L \le M$ is found according to:

$$L = \min_i \left(i = 1, ..., M \right) \left| mse\left(\hat{\mathbf{X}}\right)_i \le mse_{thr}\,, \right. \tag{5}$$

where mse_{thr} is a given threshold value. The matrix of principal components with reduced dimensionality is $\hat{\mathbf{Y}} = \left[\hat{y}_{ij} \right]_{N \times L}$ and its elements are : $\hat{y}_{ij} = y_{ij}, i = 1, ..., N; j = 1, ..., L$.

The clustering procedure is performed on $\hat{\mathbf{Y}}$ and centroids are formed from the resulting clusters as a mean vector or the codebook vectors of SOM can be used. The restored PQRST complex $\hat{\mathbf{x}}_c$ from each centroid $\hat{\mathbf{y}}_c$ is calculated according to [10]:

$$\hat{\mathbf{x}}_c = \boldsymbol{\sigma}.\left(\mathbf{W}\hat{\mathbf{y}}_c^T \right)^T + \boldsymbol{\mu}\,, \tag{6}$$

where $\mathbf{W} = \left[w_{ij} \right]_{M \times L}$ is the transformation matrix whose elements are: $w_{ij} = v'_{ij}, i = 1, ..., M; j = 1, ..., L$.

The clustering procedure is based on a variant of SOM called Growing Hierarchically Self-Organized Maps (GHSOM) in which the SOM topology doesn't have to be set a priory [11]. As usual the Euclidean distance serves as similarity measure.

IV. Experimental Results

The hardware part of the system was tested with the built-in square wave test generator with amplitude of 1mV. Next the noise level was investigated by shorting the input terminals. The achieved value is 5.1μV_{RMS} at 1kHz sampling rate and gain set to 6.

Because the hardware subsystem hasn't passed clinical tests yet, a synthetic signal was used to validate the algorithm for analysis. It is suitable to use the signal proposed in [12]. The signal morphology was changed in order to resemble the typical TEECG waveform. Because the positions and morphology of P waves are in a special interest, a synthetic TEECG signals were prepared with 5 different amplitudes and 5 different positions of P waves. The total number of used signals is 10. The length of each signal is about 1min. An additive uniform noise was also added in the signal. The total number of the PQRST complexes is 640 (Fig. 5)

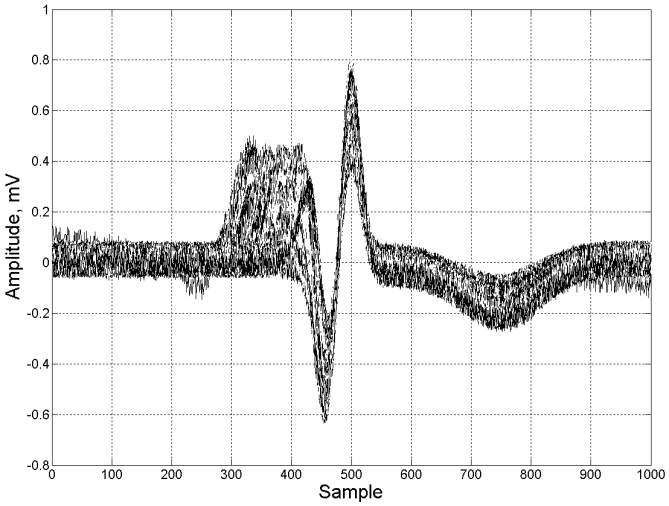


Fig. 5. Common plot of all synthetic PQRST complexes for clustering

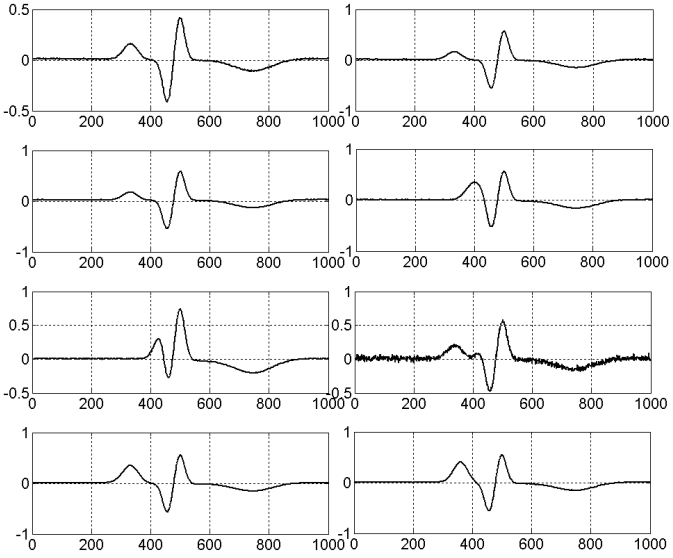


Fig. 6. Restored PQRST complexes from the centroids of the clusters

The restored signals from the centroids after GHSOM clustering are shown in Fig. 6. The clustering was performed on data with dimensionality reduced from 1000 down to 200. As can be seen the back restoration from reduced number of principal components also performs denoising.

V. Conclusion

In this paper, a system for transesophageal electrocardiogram acquisition and analysis was presented. The achieved SNR from acquisition subsystem is high considering the wide frequency range. The analysis algorithm performs correct also. In future it is planned to investigate the usefulness of PCA applied on multichannel transesophageal ECG. The obtained principal components can reveal information, which could be associated with particular cardiac disorder.

References

[1] Б. Бойчев, *Основи на трансезофагеалната електрокардиостимулация*, Казанлък, Янита ЯС, 2010.

[2] M. Cremer, "Über die direkte Ableitung der Aktionsströme des menschlichen Herzens vom Ösophagus und über das EKG des Föten", Münch Med Wschr, vol. 53, pp. 87-93, 1906.

[3] J.M. Jenkins, D. Wu, R.C Arzbaecher, "Computer diagnosis of supraventricular and ventricular arrhythmias. A new esophageal technique", Circulation, vol. 60, no. 5, pp. 977-987, 1979.

[4] B. Boychev, S. Rahimi, M. Mihalev, V. Kushev, "Transesophageal electrocardiography - methods and importance", Bulgarian Cardiology, vol. 4, pp. 31-50, 2002.

[5] Texas Instruments, "Low-Power, 8-Channel, 24-Bit Analog Front-End for Biopotential Measurements", http://focus.ti.com/lit/ds/symlink/ads1298.pdf.

[6] *International Standard IEC 60601-1*, Third edition, 2005, http://www.jbhealthcare.com/download/info_iec60601-1%7Bed3.0%7Den_d.pdf

[7] Y. Velchev, "ECG QRS Identification Using Morphological Filtering and Histogram Analysis", Computer & Communications Engineering, ISSN 1314-2291, vol.4, no.1, pp. 51-55, 2010.

[8] F. Castells, P. Laguna, L. Sornmo, A. Bollmann, J. Millet-Roig, "Principal component analysis in ECG signal processing", EURASIP Journal on Applied Signal Processing, vol. 2007, no. 1, pp. 98-119, 2007.

[9] S. Sanei, J. A. Chambers, *EEG Signal Processing*, Wiley-Interscience, 2007.

[10] V. Franc, V. Hlavac, "The statistical pattern recognition toolbox", 2004, http://cmp.felk.cvut.cz/cmp/software/stprtool/index html.

[11] M. Dittenbach, D. Merkl, A. Rauber, "The Growing Hierarchical Self-Organized Map", Proceedings of the Int'l Joint Conference on Neural Networks (IJCNN' 2000), pp. VI-15-VI-19, 2000.

[12] P. McSharry, G. Clifford, L. Tarassenko, L. Smith, "A dynamical model for generating synthetic electrocardiogram signals," IEEE Transactions on Biomedical Engineering, vol. 50, no. 3, pp. 289-294, 2003.

Programmable Jitter Generator

Goran Jovanovic[1], Mile Stojcev[1] and Tatjana Nikolic[1]

Abstract – **As CMOS technology has scaled, supply voltage have dropped, chip power consumption has increased, and clock frequency/data rates increase effects of jitter become critical and jitter budget get tighter. Knowing how to inject/isolate jitter components with time-convolution/correlation will enhance designer ability to determine and locate the root causes so that he/she can then proceed to 'beat down' individual error components one at a time in order to improve reliable system performance. Jitter can be decomposed into several subcomponents, each having specific sets of characteristics and root causes. This paper begins with a short review of jitter fundamentals including a discussion of the various random and deterministic jitter components, and injection method of jitter subcomponents into computer clock signal and/or communication data stream. The jitter injection technique gives test engineers an insight into how jitter components interact. In the rest of the paper a global hardware structure of a jitter generator, which uses digital techniques, based on a voltage controlled delay line is described. A Xilinx xc3s500e-5fg320 FPGA chip is used to validate this design. The programmable jitter generator can be used in jitter tolerance test for computer system and jitter transfer function measurement in communication systems.**

Keywords – **Jitter, Jitter generator, Jitter classification.**

I. INTRODUCTION

With continuous scaling of CMOS IC technology and increase of clock frequency, it is becoming increasingly difficult to guarantee the availability of correct clock signal throughout the chip and system due to: i) the increasing likelihood of manufacturing defects in VLSI ICs (process variation, power supply noise, etc.); and ii) influence of external environment (injected noise, crosstalk, etc.). Therefore, understanding what timing jitter is, and how to characterize it, is the first step during the process of designing high-performance, high-speed, and reliable digital systems [1, 2].

The variations of the timing signal's rising and falling edges as compared to the perfect reference are defined as jitter, and the corresponding time-variation measurements are specified in time units such as picoseconds [2, 3].

In order to evaluate system performance with respect to jitter tolerance, a jitter generator is indispensable. The jitter generator should generate and injects different kinds of jitter into the input data stream in a controllable fashion. Many works have been reported on jitter measurement and analysis [4, 5, 6, 7]. It is relatively simple to measure each component but is challenging to generate, measure, and analyze them if multiple jitter components are simultaneously injected into a data stream or in distributed clock signals.

In this paper, we propose a programmable jitter generator, which targets all the aforementioned challenges. Our intent is to determine how jitter's components can be modeled and combined, and how the total jitter can be changed according to different injection sequences.

This paper is structured as follows: Section II deals with jitter classification. Different types of jitter are discussed and their root causes are identified. Section III explains the meaning of terms unit period and eye diagram. Section IV concentrates on decomposition and jitter modeling. In Section V a global structure of jitter generator is involved and the role of its main building blocks is described. Section VI concentrates on combining method for jitter generation. In Section VII implementation details of jitter generator are given. Section VIII summarizes the conclusions.

II. JITTER CLASSIFICATION

Jitter can be categorized in many different ways [8]. We will use categorization based on the phenomenological properties of the jitter itself. This classification is between jitter that is random and jitter that is not. Jitter that is not random is bounded, that is, its magnitude is finite. In contrast, random jitter is unbounded and, within physical limits, can theoretically reach any magnitude. Fig. 1 shows diagram of jitter classifications, moving from total jitter at the left to the detailed jitter mechanisms at the right of the diagram. Total jitter, J^T, results from combination of random jitter, J^R, and deterministic jitter, J^D (see Fig. 1).

Random jitter is the jitter generation from the accumulation of random processes including thermal noise and shot noise. J^R cannot be predicted, because it has no discernable pattern. In regards to J^R, the following are examples of sources that can cause random jitter: a) thermal noise - due to electron flow in conductors; and b) shot noise - due to electron and hole flow in semiconductors. By its nature J^R is theoretically unbounded and Gaussian in distribution which is characterized by a mean, μ, and a width, σ, as shown in Fig. 2.

Deterministic jitter is the jitter generation from a variety of systematic effects. It arises from the interaction of different system components. The major causes of J^D include electromagnetic interference, crosstalk, signal reflections, driver slew rate, skin effects, and dielectric loss. J^D can be divided further into the following two subclasses: j) jitter that is correlated to data sequence or pattern; and jj) jitter occurring independent of the data.

Jitter that is correlated to the data pattern can be broken into two categories: data dependent jitter J^D_{DD} and periodic jitter J^D_P. Duty cycle distortion jitter, J^D_{DCD}, and inter-symbol interference jitter, J^D_{ISI}, are pattern dependent, and they are part of a jitter class called data dependent jitter, J^D_{DD}. There are two common causes of J^D_{DCD}: i) The slew rate of the rising

[1]Goran Jovanovic, Mile Stojcev and Tatjana Nikolic are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: goran.jovanovic{mile.stojcev, tatjana.nikolic}@elfak.ni.ac.rs.

edges differs from that of the falling edges; ii) The decision threshold for a waveform is higher or lower than it should be.

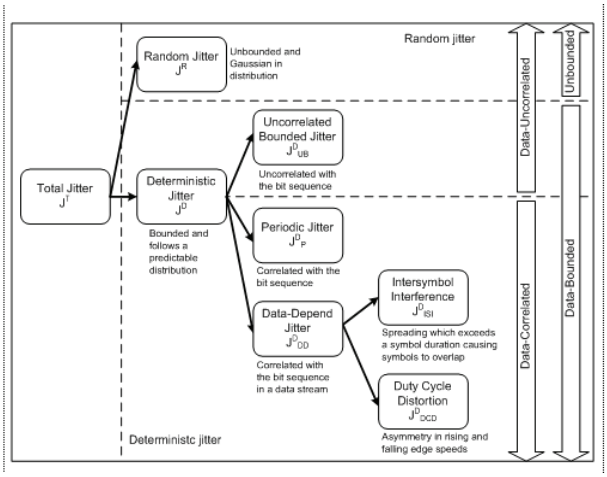
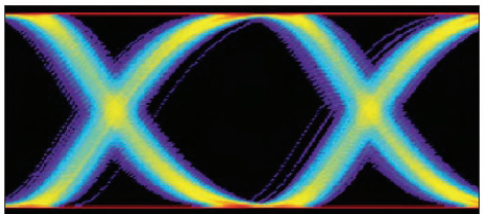
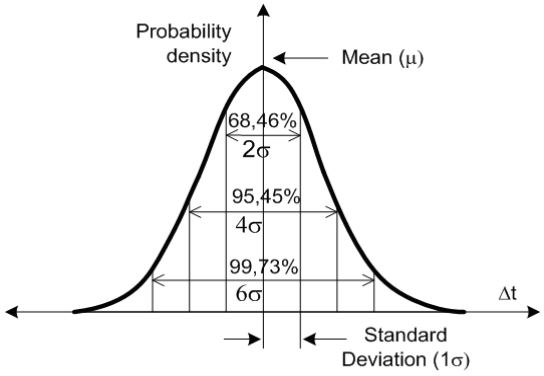


Fig. 1. Total jitter subcomponents



a)



b)

Fig. 2. a) An eye diagram; and b) Histogram for Gaussian distribution
Notice: A Gaussian power density function (PDF) with corresponding probability area of 2σ, 4σ i 6σ widths covering 68,46%, 95,45% and 99,77% of the underneath area

Fig. 3 shows waveform typical for duty-cycle distortion caused by non-symmetrical rise/fall times and incorrect detection threshold.

Fig. 3 (upper part) demonstrates the first case. Here, the decision voltage is at 50% amplitude point but the slow rise time of the waveform causes the rising edges to cross the threshold latter than the falling edges. Fig. 3 (bottom part) presents the second case, in which the waveform has balanced rise and fall times but the decision threshold is not set at the 50% amplitude point.

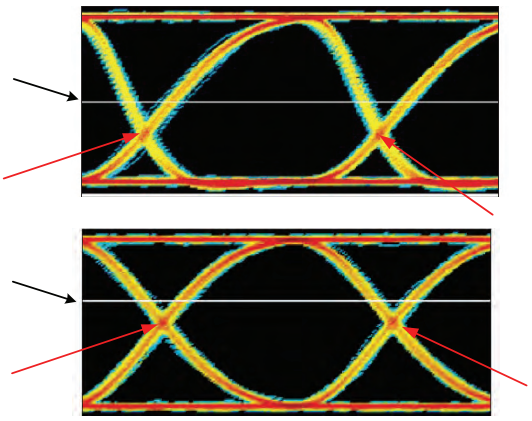


Fig. 3. Waveforms showing duty cycle distortion

J^D_{ISI}: jitter is usually the result of a bandwidth limitation problem in either the transmitter or physical media. Fig. 4 presents how bandwidth limitations produce inter-symbol interference timing.

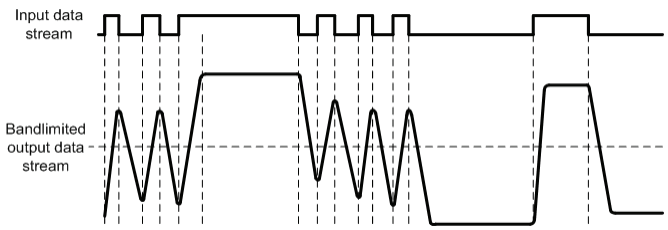


Fig. 4. Inter-symbol interference due to bandwidth limited problem

J^D_{ISI} has three main causes. a) Bandwidth limitation problem: limited bandwidth produces limited edge speeds, and limited edges speeds will result in varying pulse amplitudes at high-speed data rates. Varying pulse amplitudes will then result in transition timing errors; b) Signal reflections: arise due to improper terminations or impedance anomalies within the physical media. Signal reflections usually produce distortions in the amplitude of the data signal; and c).Non-linear phase response of the transmission media: these limitations cause frequency-dependent group delay. For example, the skin effect is proportional to the square root of the frequency, while the dielectric loss is proportional to the frequency [9]. Therefore, the skin effect dominates data loss at a lower frequency, whereas the dielectric loss dominates at a higher frequency. The nonlinear response causes edge shifts that depend on the transition density within the data stream.

Periodic jitter deals with jitter that is periodic and correlated to the data, and whose frequency is an integer sub-rate of the data rate. The source of J^D_P is usually interference from signals related to data pattern, ground bounce or power supply variations.

Bounded uncorrelated jitter, J^D_{UB}, can be due to bounded but non periodic sources. This kind of jitter appears as a consequence of coupling from adjacent data-carrying links or on-chip random logic switching [9].

III. Referent Period and Eye Diagram

Jitter is typically measured and specified in picoseconds from some reference edge and over some number of specified cycles. It's possible to express jitter in absolute time normalized to a unit interval, UI. A UI is the ideal or average time duration of a single bit or the reciprocal of the average data rate. Another method is to express the jitter as a percentage of the reference period. In many applications the reference period is call a unit interval, UI, and the jitter is specified as a ratio of percentage of the UI. For example, for clock frequency of 200 MHz with ± 50 ps of jitter or a total jitter of 100 ps, we have 1 UI= 5000 ps. Jitter as ratio of UI is 100ps/5000ps=0.02 UI, or 2% UI.

Traditionally, an eye diagram, like that shown in Fig. 5a), has served to specify signal integrity limits, including jitter. An eye diagram is a composite of all bit periods of the captured bits superimposed on each other relative to a bit clock (recovered or available from the source). We call the area within the eye the eye opening. In Fig. 5b) arrows are used to show the vertical and horizontal events in the eye opening. As the noise in a signal increases the eye becomes less open, either horizontally or vertically or both. The eye is said to be closed when non open area remains in the center of the diagram.

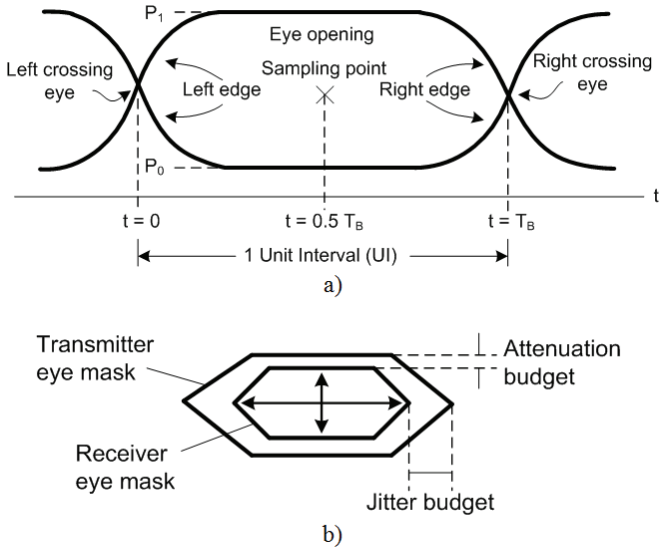


Fig. 5. Eye diagram; a) Eye diagram with definition of terms; b) Overlapping the masks gives an indication of attenuation and jitter budget

IV. Jitter Decomposition and Modelling

To understand how real systems behave, it is often useful to use a mathematical model of the system. The behavior of such a model can be tuned by adjusting the parameters of its individual components. If the parameters of the model are chosen based on observations of the real system, then the model can be used to predict the behavior of the system in other situations. Thus one of the motivations for jitter decomposition (also called jitter separation) is to extrapolate system performance to cases that would be difficult or time-consuming to measure directly.

Another motivation for modeling the system this way has to do with analysis. If each of the model components is associated with one or more underlying physical effects, an understanding of the model can provide insight into the precise cause or causes of excessive jitter.

Researchers and engineers commonly model J^R by the Gaussian distribution function [2, 10]

$$J^R(\Delta t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\Delta t-\mu)^2}{2\sigma^2}} \quad (1)$$

where $J^R(\Delta t)$ denotes J^R probability density function, PDF, σ is a standard deviation of the Gaussian distribution, and (Δt–μ) is time displacement relative to the ideal time position. This function is characterized as un-bounced because its PDF is not zero unless the jitter Δt approaches infinity. No matter, how large value Δt we peak, the probability is never zero. From mathematical point of view the mean of Gaussian form equals μ, and its standard deviation equals σ.

J^D_{DD} can be modeled through linear time invariant, LTI, system in which an ideal data pattern is the input of the LTI. J^D_{DD} is calculated from the output waveform via its deviation of the edges transition times from the corresponding ideal edges transition times [10].

The J^D_{DCD} can be best modeled by the dual-Dirac delta function [2, 10]. The sum of two δ functions that represent J^D_{DCD} PDF is:

$$J^D{}_{DCD}(\Delta t) = \frac{\delta\left(\Delta t - \frac{w}{2}\right) + \delta\left(\Delta t + \frac{w}{2}\right)}{2} \quad (2)$$

where w is peak-to-peak duty-cycle-distortion magnitude, and Δt is the time displacement relative to the ideal time position.

J^D_{ISI} is caused by timing spread of various pulses with different run lengths within the transmitted pattern. To calculate total J^D_{ISI} it is necessary to know probability of occurrence of each edge pattern and the corresponding jitter magnitude [10]. Let pi denotes the probability that given bit pattern i will occur, and Δt_i corresponds to the magnitude of the bit pattern. If the jitter magnitude of each distinct edge pattern remains constant over time, than weighted some of δ functions can be used to represent the PDF of each edge, with the weights corresponding to the edge pattern probability. Accordingly J^D_{ISI} PDF can be expressed as [10].

$$J^D_{ISI}(\Delta t) = \sum_{i=1}^{N} p_i * \sigma(\Delta t - \Delta t_i) \quad (3)$$

where: N is the number of distinct edge patterns, pi is the probability of occurrence of edge pattern i, Δt_i is the jitter magnitude of the i-th edge pattern, Δt is the time displacement relative to the ideal time position.

J^D_P is a repeating jitter signal at a certain period or frequency, and is viewed as bounded and uncorrelated

narrow-band jitter. If we assume that the J^D_P is sinusoidal it can be described mathematically by the following

$$\Delta t = A\cos\left(\omega t + \phi_0\right) \tag{4}$$

where ω is angular frequency and ϕ_0 is the initial phase.

The conclusions established can apply well to other periodic jitters with different profiles, such as rectangular, triangular, saw-tooth, or trapezoid.

In general, the model of the total periodic jitter J^D_P is summation of cosine functions with phase deviation, modulation frequency, and peak amplitude [2].

$$J_P^D\left(\Delta t\right) = \sum_{i=1}^{N} A_i \cos\left(\omega_i t + \phi_i\right) \tag{5}$$

where *N* is the number of cosine components (tones).

PDF for sine J^D_P (single tone)

$$J_{P\_SIN}^D\left(\Delta t\right) = \begin{cases} \dfrac{1}{\pi\sqrt{\dfrac{m}{2} - \left(\sqrt{\dfrac{2}{m}}\Delta t\right)^2}} & for\ \left|\Delta t\right| < \dfrac{m}{2} \\ 0 & otherwise \end{cases} \tag{6}$$

Uncorrelated bounded jitter, J^D_{UB}, can be modeled by many independent periodic jitters if it is composed by many independent root sources. Its time-domain PDF is truncated Gaussian, and is defined as [10]:

$$J_{UB}^D\left(\Delta t\right) = \begin{cases} \dfrac{p_{UB}^J}{\sigma_{UBJ}\sqrt{2\pi}} e^{-\frac{\Delta t^2}{2\sigma_{UBJ}^2}} & for\ \left|\Delta t\right| \le A_{UB} \\ 0 & for\ \left|\Delta t\right| > A_{UB} \end{cases} \tag{7}$$

where A_{UB} is the peak value, σ_{UBJ} is the sigma value, and P^J_{UB} is the normalized probability for the J^D_{UB} PDF.

Assuming that each jitter component is already known, the total jitter probability density function is given by the convolution of the PDF's of each component [8, 10].

$$J^T\left(t\right) = J^R\left(t\right) * J_{DCD}^D\left(t\right) * J_{ISI}^D\left(t\right) * J_P^D\left(t\right) * J_{UB}^D\left(t\right) \tag{11}$$

where "*" is a convolution operation.

V. Structure of Jitter Generator

The structure of a jitter generator, JG, is given in Fig. 6. Main building blocks of JG are:

Voltage controlled delay line, VCDL: composed of 32 controlled buffer-cells (D_0, D_1, ..., D_{31}) of the same structure and delay time.

Histogram logic, HL: consists of two constituent. The first is called pseudorandom generator and is implemented as a linear feedback shift register, LFSR, with five D flip-flops and five feedback taps. Having in mind the speed of operation, Galois method was used for realization of LFSR. The second constituent is realized as memory block in which different PDF patterns are stored.

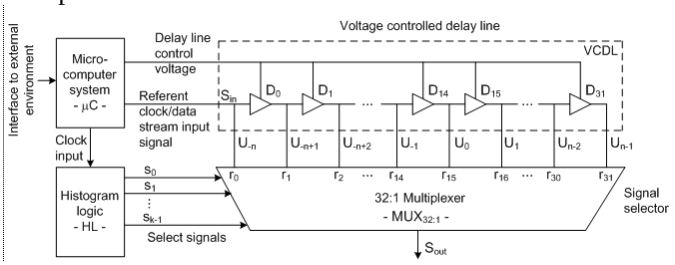


Fig. 6. Structure of jitter generator

Signal selector, MUX32:1: digital multiplexer which has one output pin S_{out} and 32 input pins named as r_0, r_1, ..., r_{31}, respectively. The signal selection lines s_0, s_1, ..., s_{31}, select a delay buffer-cell and connect it to the output pin S_{out}.

Microcomputer system, μC: used as a control block for JG. From one side it is coupled with the external environmental (keyboard, display, etc.) and from the other side it generates control signals for driving VCDL (delay line control voltage, and referent clock/data stream input signal), and clock signal for driving the HL.

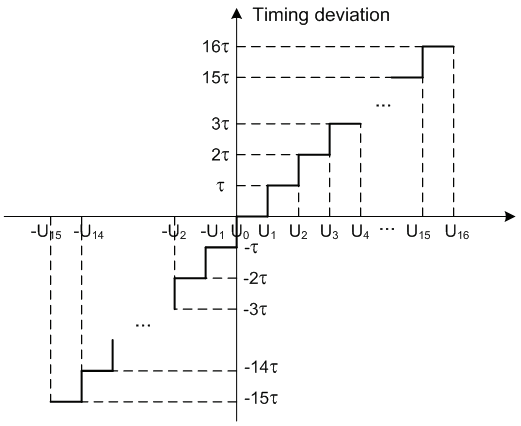


Fig. 7. Time deviations at output of JG

In computer system the signal Sin is input jitter-free clock signal, while in communication systems Sin corresponds to a data stream. In our case, for both applications, Sin is used as a signal into which jitter is injected. All buffer-cells involve identical delay τ, so that the total timing deviation $\Delta T = 32*\tau < T/2$, where T corresponds to the period of a clock signal in computer systems, or data–bit–interval in communication systems. If we take that the central buffer-cell's output represents a referent signal, its rising edge is referred to as phase zero (see Fig. 7). This means that the delay chain is able to generate 17 positive and 15 negative phase deviations.

VI. Combining Methods for Jitter Generation

In real signals multiple jitter components are simultaneously present. Bearing this in mind, our objective now was to determine how different jitter components, in a

controlled way, can be combined, and how the composition of a total jitter can be changed according to different injection sequences. Fig. 8 illustrates the combining method. The sequence of each jitter model can be interchanged to study the impact of different injection sequences. The jitter combining scheme was simulated using Matlab. In order to predict the overall system jitter the separate components were first analyzed individually, after that developed and characterized, and then combined using Matlab. The Matlab simulation shows us how jitter components combine and how the total jitter depends on the jitter injection sequence.

During the combining process some of the jitter components can be switched on or off. Our goal during jitter analysis was to understand how different jitter components combine to form the jitter distribution. In that sense, like illustration, some examples of individually injected jitter components are shown in Fig. 9. The left upper part of Fig. 9 shows random jitter with Gaussian distribution, while its right upper part corresponds to mixture of triangular periodic jitter and uncorrelated bounded jitter. The left bottom part of Fig. 9 presents convolution of sinusoidal periodic jitter and data correlated intersymbol interference jitter. Finally, the right bottom part of Fig. 9 gives combination of Gaussian, sinusoidal periodic and uncorrelated bounded jitter.

From jitter-measurement point of view, there are several benefits to separating jitter and injecting components in a clock/data stream [11]:

Once the jitter has been broken down in this way it is possible to extrapolate the performance of a system under test at very low probability levels without measuring trillions of events, offering the promise of greatly reduced testing times.

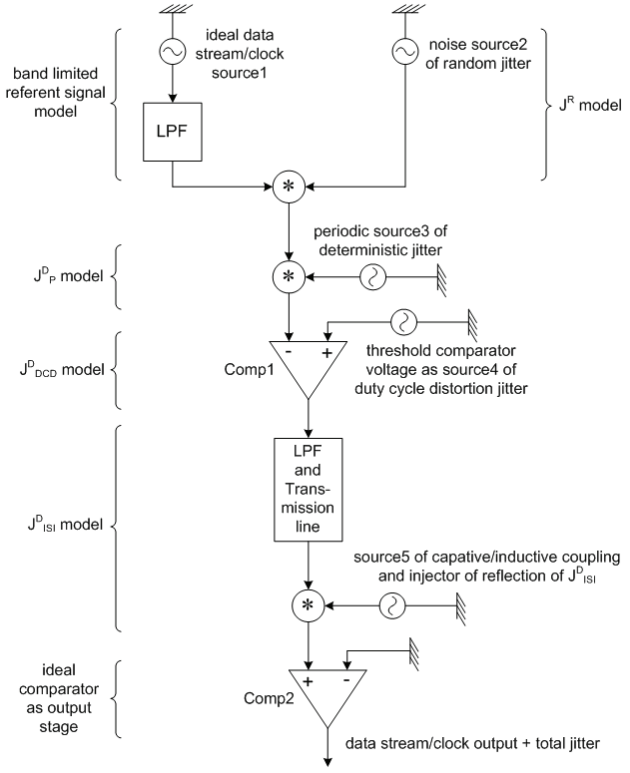


Fig. 8. Combining method for jitter

Each jitter component has one or more known causes and well-understood effect on the probability of bit errors. Thus, an efficient design strategy for reducing some jitter components can be developed.

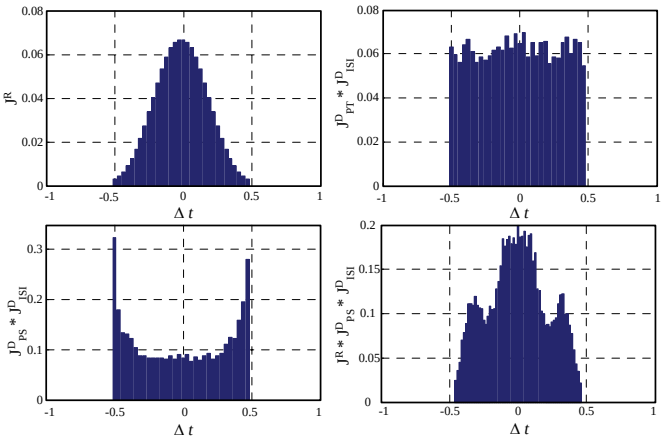


Fig. 9. PDF of jitter using the convolution method

There are two major approaches to separating individual jitter components at the receiver side. One method is based on the jitter PDF or cumulative distribution function, CDF, measurement. PDF is the normalized histogram of the signal edge times (i.e., how often the signal transitions at a particular time point). CDF, on the other hand, sorts the sampled edge time data in an ascending order to show the distribution profile. Another method is based on jitter time record. Jitter PDF can be measured by instrument such as sampling oscilloscope or time interval analyzer. Jitter CDF (also called BER CDF) can be measured by bit-error-rate tester. The jitter time record can be measured by time interval analyzer or real time oscilloscope [11, 12].

VII. Implementation

In order to estimate the performance, the jitter generator structure was described at RTL level using VHDL and has been implemented using Xilinx FPGA technology. Target device xc3s500e-5fg320 from Spartan3E series is selected. For synthesis, routing and mapping a Xilinx development CAD tool Xilinx ISE Design Suite 12.3_1 was used.

In the current design, the voltage controlled delay line was realized with 32 fixed buffer delay cells since in the FPGA chip there is no voltage-controlled buffer. In the histogram logic the pseudorandom generator was implemented as five taps LFSR, while the RAM block as an array of size 1024*5 bits. Fig. 10 depicts the simulation results that correspond to phase deviation generation. The waveforms correspond to the outputs of a delay line. At seventeen VCDL's outputs we have positive, while at fifteen outputs negative phase deviation.

Details which relate to implementation of a jitter generator on a FPGA chip xc3s500e-5fg320 from Spartan 3E series are given in Table I.

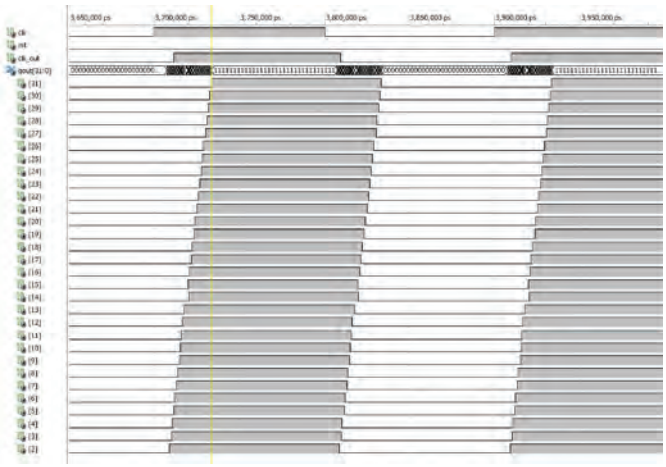


Fig. 10 Simulation results

TABLE I
DEVICE UTILIZATION SUMMARY

Logic Utilization	Used	Available	Utilization
Number of Slice Flip Flops	10	9,312	1%
Number of 4 input LUTs	49	9,312	1%
Number of occupied Slices	29	4,656	1%
Number of Slices containing only related logic	29	29	100%
Number of Slices containing unrelated logic	0	29	0%
Total Number of 4 input LUTs	49	9,312	1%
Number of bonded IOBs	35	232	15%
Number of RAMB16s	1	20	5%
Number of BUFGMUXs	1	24	4%
Average Fanout of Non-Clock Nets	3.24		

According to the results given in Table I we can conclude that the hardware overhead due to implementation of a jitter generator on FPGA chip is really minor. Utilization of flip/flops, four-input LUTs, and number of occupied slice is always 1%, and number of RAM bits is about 5%.

VIII. CONCLUSION

As speed as well as the desire for improved performance increases, the amount of jitter in system or product will become more of concern than it is today. However, by having a fundamental understanding of these effects, better choice will be made in design and characterization of a product, which in turn, will lead to improved system reliability.

A good jitter generator must be capable of controlling the jitter frequency and the magnitude of jitter amplitude. Here, the jitter amplitude represents the amount of the phase shift of a data stream signal or the computer clock signal, while the jitter frequency is the times of generation a phase shift. Currently in the market, there are several jitter generator instruments that may satisfy these demands, however they are expensive and not suitable for being utilized in mass production. In this paper we present relatively simple and low-cost programmable jitter generator which can be used for jitter tolerance test and jitter transfer function measurement. The jitter generator is implemented using digital circuits based on chip from FPGA family, and software. In near future we plan to build this jitter generator into a CMOS VLSI IC, such as dual-processor fault-tolerant system for automotive.

ACKNOWLEDGEMENTS

This work was supported by the Serbian Ministry of Science and Technological Development, Project No. TR-32009 – "Low-Power Reconfigurable Fault-Tolerant Platforms".

REFERENCES

[1] C. Metra, M. Omana, TM. Mak, A. Rabman, S. Tam, Novel On-Chip Clock Jitter Measurement Scheme for High Performance Microprocessors, IEEE International Symposium on Defect on Fault Tolerance of VLSI Systems, 2008, pp. 465-473

[2] N. Ou, T. Farahmand, A. Kuo, S. Tabatabaei, A. Ivanov, Jitter Models for the Design and Test of Gbps-Speed Serial Interconnects, IEEE Design & Test of Computers, Vol. 21, No. 4, July-August 2004, pp. 302-313

[3] J. Solorzano, Jitter basic, part I, Wavecrest Corp., Application note 2004, http://www.micrel.com/hbw_news/Jitter_Article_Part%201.pdf

[4] S. Sunter, A. Roy, On-Chip Digital Jitter Measurement from Megahertz to Gigahertz, IEEE Design & Test, Vol. 21, No. 4, July-August 2004, pp. 314 – 321

[5] "Jitter Analysis Techniques for High Data Rates, Agilent Application Note 1432, http://cp.literature.agilent.com/litweb/pdf/5988-8425EN.pdf

[6] T. Xia, P. Song, K. A. Jenkins, J.C.Lo, "Delay Chain Based Programmable Jitter Generator", Proc. of the 9-th IEEE European Test Symposiums (ETS), May 2004

[7] M. Fleischer-Reumann, "Jitter Tolerance Testing", Chapter 10, pp. 533-561 in Digital Communications Test and Measurement, by Dennis Derickson, Marcus Muller, Editors, Prentice Hall, Upper Saddle River, 2008.

[8] Marcus Muller, Ransom Stephens, Jitter Basics, Chapter 2, pp. 29-60, in Digital Communications Test and Measurement, by Dennis Derickson, Marcus Muller, Editors, Prentice Hall, Upper Saddle River, 2008.

[9] Horward W. Johnson, Martin Graham, High-Speed Digital Design: A Handbook of Black Magic, PTR Prentice Hall, New Jersey, 1993.

[10] M.P. Li, Jitter, Noise, and Signal Integrity at High Speed, Prentice Hall, Upper Saddle River, NJ, 2008

[11] Greg D. Le Cheminant, Characterizing High Speed Digital Communications Signals and Systems with the Equivalent – Time Sampling Oscilloscopes, Chapter 7, pp. 329-419, in Digital Communications Test and Measurement, by Dennis Derickson, Marcus Muller, Editors, Prentice Hall, Upper Saddle River, 2008.

[12] "Understanding and Characterizing Timing Jitter" Tektronix application note 55W-161146-0, Sept. 2007.

Displacement Signal Error Approximation for Uncorrelated Noise of Laser Illuminated Object

Žarko P. Barbarić[1], Miroslav D. Lutovac[1] and Ivan D. Đokić[1]

Abstract – **The displacement signal error for uncorrelated noise of laser illuminated object (target) is analyzed. A mathematical relation for probability density function of displacement signal for a laser illuminated object is given. The displacement signal error of laser illuminated object versus signal-to-noise ratio (SNR) and the mean value displacement signal ($\bar{\varepsilon}$) is derived. The displacement signal error is derived in closed form for given value of the spot position probability 0.5, using a system for symbolic computing Mathematica.**

Keywords – **Quadrant photodiode, Displacement signal, Probability density function, Displacement signal error.**

I. Introduction

In industry and army applications in which accurate positioning of the objects are very important [1]. The main characteristic of those applications is accuracy of positioning measurement of an object. Laser positioning systems have a special treatment, because their resolution is better than radar and other positioning systems. A number of lasers positioning applications include tracking of laser illuminated target and measurement of its angular position [2], [3], estimation of satellite vibration effect [4], and measurement of multidimensional displacement in space [5]. These and other applications can use position sensitive detectors with quadrant photodiode (QPD) [6] or lateral effect photodiode (LEP) [7] to measure lateral displacement in two perpendicular planes. Theoretical displacement signal analysis had been discussed in [1], [2], [3], [8], and [9]. Those papers have presented that the main parameter is signal-to-noise ratio, which limits positioning accuracy in laser systems with quadrant photodiode. Also, they have found that the position error is changed with mean value of displacement signal [8]. The expression for probability density function of displacement signal for Gaussian noise distribution is presented in [9].

In this paper a new approach to estimate the position error of laser illuminated object by quadrant photodiode is given. This is statistically based approach, where the displacement signal error is obtained from probability density function of displacement signal.

II. Probability Density Function of Displacement Signal

Consider a light spot, of radius *r*, falling on QPD with radius *a* as shown in Figure 1. The position of the photodiode center is at (0, 0) and the center of the spot at (*x*, *y*), where *x*, and *y* are displacement positions of the spot center along the X and Y axes, respectively.

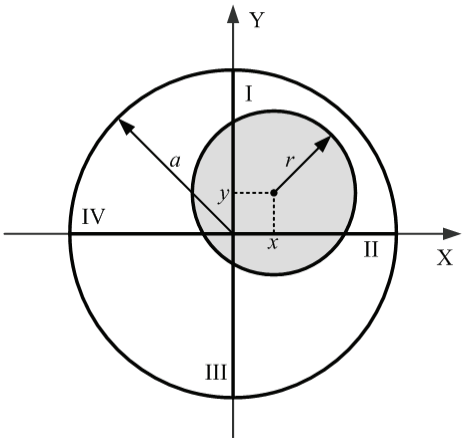


Fig. 1. Geometry of the light spot position on the QPD surface

Displacement signals depend on the spot positions in both directions. These signals can be obtained by processing the output currents of four photodiode quadrants. The normalized displacement signal ε in the X or Y direction is the function of currents of the quadrants proportional to the area of the spot falling on the quadrant [10].

We assume that each quadrant generate noise with Gaussian distribution with zero mean value. Then the probability density function (*pdf*) of displacement signal is obtained, based on known theory for jointly normal random variables, probability density functions and some transformations [11].

Useful way to analyze *pdf* of displacement signal *f(ε)* is substituting the mean value pair with mean value of the displacement signal ($\bar{\varepsilon}$) and the total signal-to-noise ratio in the channel is *SNR*. For uncorrelated noise, *f(ε)* is obtained in [10],

$$f(\varepsilon)=\frac{1}{\pi(1+\varepsilon^2)}\cdot\exp\left(-\frac{SNR}{2}(1+\bar{\varepsilon}^2)\right)\times \left[1+\sqrt{\frac{\pi}{2}}\cdot B\left(erf\left(\frac{B}{\sqrt{2}}\right)\right)\exp\left(\frac{B^2}{2}\right)\right] \quad (1)$$

where the parameter *B* is

$$B=\frac{1+\varepsilon\bar{\varepsilon}}{\sqrt{1+\varepsilon^2}}\sqrt{SNR} \quad (2)$$

Probability density function *f(ε)* is the function of the mean value $\bar{\varepsilon}$, the displacement signal ε and signal-to-noise ratio

[1]Ž. P. Barbarić, M. D. Lutovac, and I. D. Đokić are with State University of Novi Pazar, Vuka Karadžića bb, 36000 Novi Pazar, E-mail: barbaric@etf.rs.

SNR, as shown (1) and (2). Diagrams of *pdf* from (1) for SNR=10 dB, and three values of $\bar{\varepsilon}$ are given in Figure 2 as a function of ε.

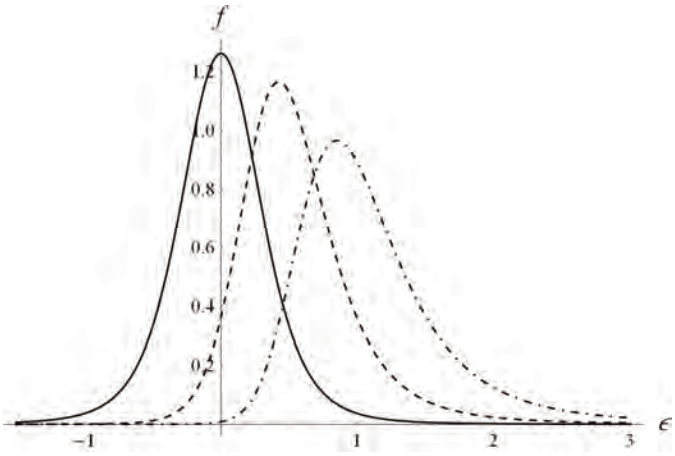


Fig. 2. Probability density function of displacement signal for *SNR*=10 dB and $\bar{\varepsilon}$ =0 (solid line), $\bar{\varepsilon}$ =0.5 (dashed line), $\bar{\varepsilon}$ =1 (dot-dashed line).

Figure 2 shows that the *pdf* maximum decreases and the *pdf* width increases by increasing $\bar{\varepsilon}$. Maximum width and minimum amplitude of *pdf* was obtained for $\bar{\varepsilon}$ =1. From Figure 2, it can be seen that the probability density function of the displacement signal $f(\varepsilon)$ for uncorrelated noise does not have the maximum value at the mean value of the displacement signal (see dot-dashed line for $\bar{\varepsilon}$ =1).

III. The Spot Position Probability

The displacement signal error presents a small range $\Delta\varepsilon$ around the mean value $\bar{\varepsilon}$ in which the spot position probability is calculated. We define the spot position probability *Pp*

$$P_p = \int_{\bar{\varepsilon}-\Delta\varepsilon}^{\bar{\varepsilon}+\Delta\varepsilon} f(\varepsilon)d\varepsilon \tag{3}$$

where $\Delta\varepsilon$ is the displacement signal error.

From (1), and (2), the spot position probability for the displacement signal ε in range $\Delta\varepsilon$ around $\bar{\varepsilon}$ becomes

$$P_p = \frac{Pp1 + Pp2}{2}$$
$$Pp1 = erf\left(\frac{\sqrt{SNR}}{\sqrt{2}}\frac{\Delta\varepsilon}{\sqrt{1+(\bar{\varepsilon}+\Delta\varepsilon)^2}}\right) \tag{4}$$
$$Pp2 = erf\left(\frac{\sqrt{SNR}}{\sqrt{2}}\frac{\Delta\varepsilon}{\sqrt{1+(\bar{\varepsilon}-\Delta\varepsilon)^2}}\right)$$

Diagrams of *Pp* (4) are given in Figure 3 as a function of $\Delta\varepsilon$.

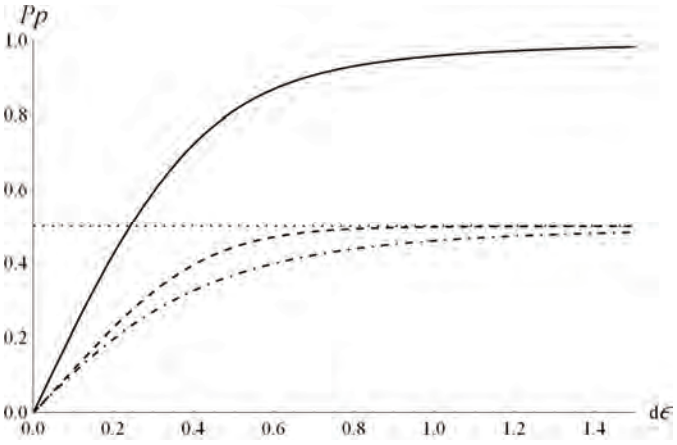


Fig. 3. Spot position probability for *SNR*=10 dB and $\bar{\varepsilon}$ =0.5 (solid line), *Pp*1 (dashed line), and *Pp*2 (dot-dashed line).

The spot position probability is the sum of Pp_1 and Pp_2 and it is not possible to find exact value of $\Delta\varepsilon$ in closed form for known *Pp*. An approximation of *Pp* in closed form is derived in [10]. In order to find another good approximation of $\Delta\varepsilon$ for specified value of *Pp*, we will analyze several examples. For the analysis purpose we are using a system for symbolic computing Mathematica [12].

IV. Displacement Signal Error

In order to determine displacement signal error for specified value of *Pp*, we enter the knowledge into Mathematica

$$\texttt{f[}\epsilon\texttt{\_, m\_ , SNR\_] :=}\ \frac{e^{-\frac{SNR\,(-m+\epsilon)^2}{2\,(1+\epsilon^2)}}\sqrt{SNR}\ (1+m\,\epsilon)}{\sqrt{2\,\pi}\ \left(1+\epsilon^2\right)^{3/2}}$$

$$\texttt{Pp[d}\epsilon\texttt{\_, m\_, SNR\_] :=}\ \frac{1}{2}\,\texttt{Erf}\left[\frac{d\epsilon\sqrt{SNR}}{\sqrt{2}\ \sqrt{1+(-d\epsilon+m)^2}}\right] +$$
$$+\frac{1}{2}\,\texttt{Erf}\left[\frac{d\epsilon\sqrt{SNR}}{\sqrt{2}\ \sqrt{1+(d\epsilon+m)^2}}\right]$$

where **m** is used for $\bar{\varepsilon}$, and **dε** for the error displacement $\Delta\varepsilon$. Next, we compute the value of $\Delta\varepsilon$ for specified value of *Pp*, say **c1**=0.5, using symbol **d** for $\Delta\varepsilon$. The squared error of the displacement signal is a linear function of the squared mean value of displacement, as it is shown in Figure 4.

```
SNR1 = 10;
m1 = 0;
c1 = 0.5;
t2 = {{m1^2, d^2} /. FindRoot[Pp[d, m1, SNR1] - c1 == 0, {d, 0.3}]}
Do[
 t1 = d^2 /. FindRoot[Pp[d, m1, SNR1] - c1 == 0, {d, 0.3}];
 t2 = Append[t2, {m1^2, t1}]
 , {m1, 0 + 0.1, 1, 0.1}]
```

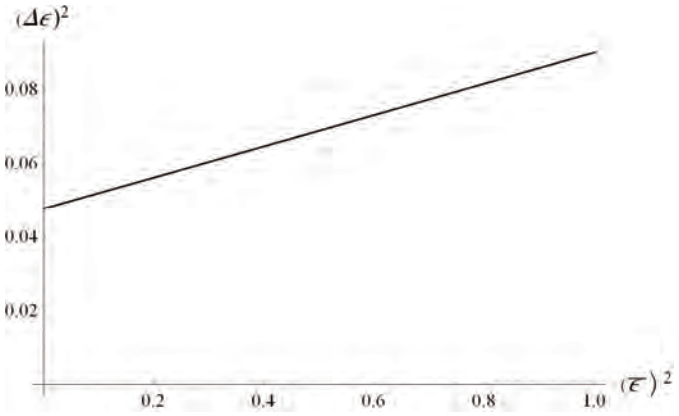


Fig. 4. The squared error of the displacement signal for *SNR*=10 dB is liner function of $\bar{\varepsilon}$.

The similar diagram is computed for different values of SNR

```
s[SNR_] := 10^(SNR/10)
SNR10 = 10.;
m1 = 0.5;
c1 = 0.5;
t2 = {{1/s[SNR10], d^2} /. FindRoot[Pp[d, m1, s[SNR10]] - c1 == 0, {d, 0.3}]}
Do[
 t1 = d^2 /. FindRoot[Pp[d, m1, s[SNR1]] - c1 == 0, {d, 0.3}];
 t2 = Append[t2, {1/s[SNR1], t1}]
 , {SNR1, SNR10 + 1., 30, 1}]
```

Figure 5 illustrates that the squared error of the displacement signal is a linear function of 1/*SNR*. Those properties of the displacement signal error, for specified value of *Pp*, gave us an idea to find a good approximation of $\Delta\varepsilon$ in terms of $\bar{\varepsilon}$ and *SNR*.

We define a new function **d2** as a linear function of $(\bar{\varepsilon})^2$ and **n**=1/*SNR*

```
d2 = (k0 + k1 m^2) n
```

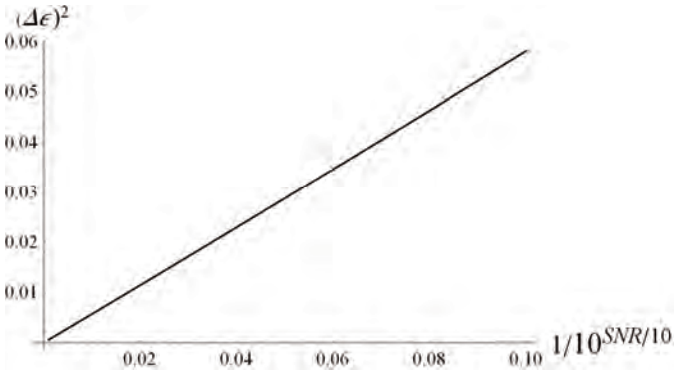


Fig. 5. The squared error of the displacement signal for $\bar{\varepsilon}$ =0.5 is liner function of $1/10^{SNR/10}$, *SNR* in dB.

Next, we compute the values for two characteristic cases $\bar{\varepsilon}$ =0 and $\bar{\varepsilon}$ =1, and for **n**=(1/*SNR*)=1/10.

```
d2m0 = d2 /. {m → 0, n → 1 / 10}
d2m1 = d2 /. {m → 1, n → 1 / 10}
```

$$\frac{k0}{10}$$

$$\frac{k0 + k1}{10}$$

Solving the equations, we determine two coefficients

```
d0 = d^2 /. FindRoot[Pp[d, 0, s[10]] - 0.5 == 0, {d, 0.3}]
```

0.047662

```
d1 = d^2 /. FindRoot[Pp[d, 1, s[10]] - 0.5 == 0, {d, 0.3}]
```

0.0900052

```
sol1 = Solve[{d2m0 == d0, d2m1 == d1}, {k0, k1}]
```

{{k0 → 0.47662, k1 → 0.423432}}

Finally, we define a new function as a good approximation of the displacement signal error

```
d2a[m_, SNR_] = Sqrt[(k0 + k1 m^2)/SNR] /. sol1
```

$$\left\{\sqrt{\frac{0.47662 + 0.423432\, m^2}{SNR}}\right\}$$

The approximation of the displacement signal error is

$$\Delta\varepsilon = \sqrt{\frac{0.4766196 + 0.42343225\,\bar{\varepsilon}^2}{SNR}} \qquad (5)$$

Figure 6 shows the displacement signal error for *SNR*=10 dB using (5). The maximal error of the approximated value with respect to the exact is 0.0000668, which is sufficiently low because the error of the displacement signal is from the range 0.2 – 0.3.

Figure 7 shows the displacement signal error for $\bar{\varepsilon}$ =0.5 using (5). The maximal error of the approximated value with respect to the exact is 0.000599, which is sufficiently low because the error of the displacement signal is in the range 0.04 – 0.3.

The displacement signal error can be computed using function **Table**

```
SNR1 = 10;
m1 = 0;
t2a = Table[{m1, d2a[m1, SNR1]}, {m1, 0, 1, 0.1}]
```

Figures 6 and 7 are obtained using functions **Table** and **Plot** command

```
ListLinePlot[{t2, t2a}, AxesLabel → {"$\bar{\varepsilon}$", "$\Delta\epsilon$"},
 PlotStyle → {{Black}, {Dashed, Black}, {Dotted, Black}
 PlotRange → {0.21, 0.31}]
```

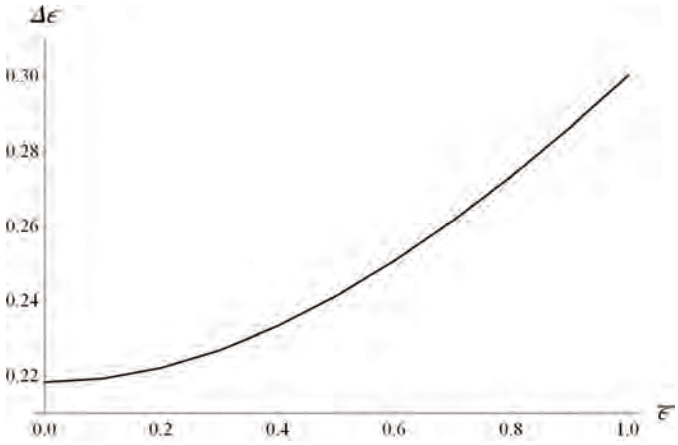


Fig. 6. The error of the displacement signal for *SNR*=10 dB.

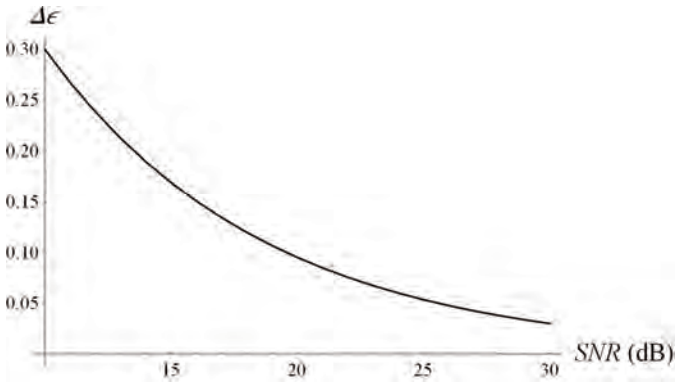


Fig. 7. The error of the displacement signal for $\bar{\varepsilon}$ =0.5.

The differences of numeric calculation using integration and closed form approximation cannot be identified visually in Figures 6 and 7.

V. Conclusion

The displacement signal error of laser illuminated object versus signal-to-noise ratio (SNR) and the mean value of displacement signal ($\bar{\varepsilon}$) is analyzed. The approximation of the displacement signal error is derived in closed form for uncorrelated noise. The displacement signal error is inverse proportional to the root square signal-to-noise ratio, and increases with increase of the mean value of displacement signal. The displacement signal error slightly increases with growth of the mean value of displacement signal.

Acknowledgement

This work was partially supported by the Ministry of Science of Serbia under Grant TR-32023.

References

[1] XA. Mäkynen, "Position-sensitive devices and sensor systems for optical tracking and displacement sensing applications," Ph.D. Thesis, Department of electrical engineering, Oulu, 2000.

[2] H. Andersson, "Position Sensitive Detectors - Device Technology and Applications in Spectroscopy," Ph.D Thesis, Sundsvall, Sweden, 2008.

[3] M. Alkeryd**,** "Evaluation of Position Sensing Techniques for an Unmanned Aerial Vehicle," Ph.D. Thesis, Department of Electrical Engineering, Linköping University, Linköping-Sweden, 2006.

[4] A. S., Kopeika, "Laser Satellite Communication Network-Vibration Effect and Possible Solutions," Proceedings of the IEEE, Vol. 85, No. 10, 1997., pp.

[5] L. Neville, Y. Cai, A. Joneja, "High-resolution multidimensional displacement monitoring system," Optical Engineering, Vol.36, No. 8, 1997., pp.2287-2293.

[6] www.pacific-sensor.co, dl100-7pcba3, "Dual Axis Psd Sum And Difference Amplifier Specification," 5700 Corsa Avenue, 105 Westlake Village, CA 91362

[7] www.pacific-sensor.co, qp50-6sd2, "Dual Axis Psd Sum And Difference Amplifier Specification," 5700 Corsa Avenue, #105 Westlake Village, CA 91362

[8] W. L. Wolfe, G. J. Zissis, *The Infrared Handbook, Chapter 22*, Ann Arbor, Michigan: Environmental Research Institute of Michigan, 1978.

[9] R.G. Kazovsky,"Theory of tracking accuracy of laser systems, Optical Engineering," Vol.22. No.3, 1983. pp.339-347.

[10] Ž. Barbarić, "Position Error of a Laser Illuminated Object," Scientific Technical Review, Vol. LII, No. 5-6, 2002. pp. 25-30.

[11] A. Papoulis, *Probability Random Variables, and Stochastic Processes*, McGrow Hill, 1965.

[12] S. Wolfram, *The Mathematica Book*, 5th ed., Champaign: Wolfram Media, Inc./Cambridge University Press, 2003.

Session TN I:

TELECOMMUNICATION NETWORKS AND SERVICES I

Role Game Theory Approach for LTE Uplink Power Control

Vladimir Poulkov[1], Pavlina Koleva[2], Oleg Asenov[3]

Abstract – **In this paper the application of Role Game Theory to Uplink Power Control (UPC) for Long Term Evolution (LTE) is considered. The authors propose an approach based on defining different roles of the subscribers within a cell of the LTE network. The main idea is to apply different methods and algorithms for UPC depending on the classified by role subscriber. Such an approach will give the possibility for flexible application and utilization of the different UPC methods to obtain better performance.**

Keywords – **LTE, Uplink Power Control, Game Theory.**

I. INTRODUCTION

In mobile communication systems it is necessary to balance between the need for sufficient transmitted energy per bit to achieve the required service quality (in respect of higher throughput and lower bit error rate performance), against the needs to minimize interference to other users of the system and to maximize the battery life of the mobile terminal. To achieve this, Uplink Power Control (UPC) methods are applied in order to adapt the transmission to the characteristics of the radio propagation channel, including path loss, shadowing and fast fading, as well as overcoming interference from other users - both within the same cell and in neighboring cells [1].

Currently, two major approaches for the realization of Power Control (PC) for the Long Term Evolution (LTE) uplink are under consideration. They are based on open and closed-loop schemes to control energy per resource element applied for an uplink transmission. With the *Open Loop Power Control* (OLPC) the devices set their power depending on their own measures and the ones obtained from the Evolved NodeB (eNB). For the *Closed Loop Power Control* (CLPC) User Equipments (UEs) also send feedback to the eNB, so they can receive more accurate information to set their transmission power [2].

LTE PC algorithms work in terms of Power Spectral Density (PSD) rather than total power. In traditional systems as 3G the used spectrum was fixed so total power was the usual term. Contrarily, LTE uses Orthogonal Frequency Division Multiple Access (OFDMA) in which the available spectrum depends on the Physical Resource Blocks (PRBs) assigned, so the PC is only capable to fix the power in each resource [3, 4].

Another important feature of PC, which must be taken under consideration when choosing and PC method for practical implementation, is the speed of the control algorithm. *Slow* PC aims to correct for shadow fading or path loss changes and *fast* PC is meant for fast channel variations like fast fading.

The implementation of an effective LTE Uplink Power Control method is of key importance for achieving steady service quality of User Equipment, regardless of their location in relation to eNB and other UE in the neighboring service area. The application of game-theoretic tools for finding an effective solution to this problem dates back to more than 10 years ago and in most cases they are based on the rule game principles. There the mobile users play non-cooperative or cooperative games, aiming at maximizing their utility functions, which from PC point is increasing their signal-to-interference-and-noise-ratio (SINR) and decreasing in the power level.

Most of these approaches are based on setting management rules depending on the distance between the UE and the cell; defining areas where service quality oriented algorithms for power management are applied; analysis and formation of user coalitions for which specific functions of power management could be applied. All this proves the lasting implementation of the game theory in this subject area.

In this paper the proposed approach is not rule based, it applies a role game theoretical methodology. The mobile users are classified by roles, where each role corresponds to predefined parameters, such as distance, traffic type, service quality, user satisfaction, etc. In this case to each role a specific UPC method could be assigned, to achieve such utility functions which are closest to the required uplink transmission quality of the mobile users.

The paper is organized as follows. In section II the basic LTE Uplink Power Control methods are reviewed. A short comparison of the rule and role game theoretic approaches to power control is done in section III. A role game method for LTE uplink power control is proposed in section IV.

II. BASIC LTE UPLINK POWER CONTROL METHODS

A. Fractional Power Control (FPC)

In this method each User Equipment (UE) sets its transmission power as follows [5]:

[1]Vladimir Poulkov is with the Faculty of Telecommunications, Technical University of Sofia, Sofia 1756, 8 Kl. Ohridski Blvd., Bulgaria, e-mail: vkp@tu-sofia.bg

[2]Pavlina Koleva is with the Faculty of Telecommunications, Technical University of Sofia, Sofia 1756, 8 Kl. Ohridski Blvd., Bulgaria, e-mail: p_koleva@tu-sofia.bg.

[3]Oleg Asenov is with the Faculty of Mathematics and Informatics, St.Kiril and St.Metodius University of Veliko Turnovo, 5003 Veliko Turnovo, 2 T. Tarnovski str., Bulgaria, e-mail: olegasenov@abv.bg

$$P_{tx} = 10\log_{10}(M) + P_{0,PUSCH} + \alpha PL\,, \quad (1)$$

where M is the number of scheduled Physical Resource Blocks (PRBs), which are 180 kHz wide. Usually the assumption is that only one PRB is scheduled to a particular UE. $P_{0,\,PUSCH}$ is a parameter broadcasted by the eNB (BS). The term PL is the result of the long-term Path Loss (PL) measurement by the UE, and $0 \leq \alpha \leq 1$ is a compensation factor broadcasted by the BS. It is obvious that if the calculated value of P_{tx} is greater than the maximum power of the UE it will transmit with its maximum power. Using Eq. 1 it is seen that UE can set its transmission power only by using the control parameters broadcasted by BS. This method is used to set the transmission power on the Physical Uplink Shared Channel (PUSCH). Using Eq. 1 each UE can calculate its transmission power (P_{tx}) without having to give any feedback to the serving BS. This is why the method is called Open Loop Power Control (OLPC). This method requires signaling between the eNB over the X2 interface.

The role of α is to act as a compensating factor for the attenuation experienced by each UE. The parameter α can take values between 0 to 1, and in particular α = [0; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1]. Values between 0.1 and 0.3 are not used in practice. Assigning zero to α means that the power control mechanism is turned off. Changing the value of α determines how much an UE will be compensated for its poor channel conditions. Higher values of α ensure more compensation. Full compensation is achieved when α = 1. Analyzing the performance of this method, it could be seen that when α increases from zero to one the overall throughput of the cell decreases while the outage throughput increases considerably. Thus, implementing this type of power control gives advantage of the UEs in the outage area of the cells on the expense of the overall cell throughput [5].

B. Interference Based Power Control (IBPC)

In the case of FPC, UEs in the outage area which are suffering from weak channel conditions significantly increase their maximum data rate when power control is applied. On the other hand UEs with low path gain, close to the serving eNB, generate most of the interference as they transmit with higher power. This is as a result of the similar values of the path gain to the serving and to the neighboring eNB used in the calculations. Let PG_s be the Path Gain (PG) to the serving eNB and PG_I the sum of the path gains to all non serving eNB [6]. The total generated interference by the UE is:

$$I_i = p_{txi} \cdot \sum_{j \neq i} PG_{ij}\,, \quad (2)$$

where p_{txi} is transmitted power from the *i-th* UE and ΣPG_{ji} is equal to PG_I.

In order to eliminate such high interference levels a method called Interference Based Power Control (IBPC) is proposed. In this method an interference limit is determined, as a result of which no eNB in the system can generate more than a given amount of interference.

It is known from Shannon's formula that the capacity doesn't increase linearly with the increase of the Signal to Noise Ratio (SNR). When the SNR is below a given value, a small increase of the SNR could result in a large increase of the capacity of the system, but for higher values of the SNR a relatively high increase won't result in such a significant increase in capacity. This philosophy lies in the core of IBPC method. If an UE increases the transmission power, the increase of the SNR will lead to an increase of its capacity, but on the other hand it will generate more interference. If ΔI is the increase of the generated interference in the cell after an UE increases its transmission power, I_c is the interference before, ΔT is the increase of the user experienced throughput after a UE increases its transmission power, T_c is the throughput before, then a metric M can be defined that evaluates gain in terms of throughput, and loss in terms of generated interference:

$$M = \frac{\Delta T}{\Delta I} \cdot \frac{I_c}{T_c} \quad (3)$$

This metric is used in the algorithm of the IBPC method. The implementation of the algorithm follows the following basic rules. One UE is allowed to increase it's transmit power p_{tx} only with fixed steps Δp at a time up to its power limitation. Each UE willing to increase its transmission power must send a request to the serving BS. The BSs that have received requests calculate the metrics M for all the UEs and the UE with the highest metric is permitted to increase its power [6, 7, 8].

III. ROLE VS. RULE GAMES IN POWER CONTROL

This current treatment is an attempt to appropriate a different perspective to the application of game theory models in achieving higher efficiency of LTE Uplink Power Control. Let's ask the question which approach - role-based or rule-based policy LTE Uplink Power Control - is more efficient. The currently known methods could be successfully listed in the rule-based game theory field because UE are based on rules which result from the main target function to provide the highest possible service quality (higher speed, lower bit error rates, continuity and congruity of quality) by measuring the distance between the user and the eNB, the incidence in view of the eNB system and the active UEs in the neighboring service area. This approach suggests that each UE is "obliged" to benefit from the maximum speed, to be by all means active in using the services.

In the rule-based methods the user's profile is not taken into account, the policies are applied to all users, no matter what type their end terminal devices are and no matter what average upstream they are generating at the moment or is statistically measured for a certain past time slot. Thus the role of the user in the communication process, whose quality we are trying to manage effectively, is not reviewed or measured. For the specific application the role as a concept could be defined as follows: *the roles are collections of properties and permissions to use resources appropriate to a UE's communication needs.* This assumes that all properties and

permissions needed to perform an individual role-connected communication function can be neatly encapsulated.

The primary purpose of establishing roles is to provide an easy way to manage power control rules for groups of users. In such a way the actual management of the „uplink power control" process is limited to a dynamic appropriation of roles to UEs. For each role a power control method is previously defined, which is specifically oriented towards the characteristics of the role.

The challenges of role-based LTE uplink power control will continue to be the contention between the user's satisfaction, in the sense of required throughput and service quality, and easier administration. For performance improvement to be achieved, it is better for each role to be more granular, to have multiple roles per UE, to try to allow each role to improve as much as possible its own performance. For easier administration, it is better to have fewer roles to manage. For each role, depending on its communication profile and the specific number of individual characteristics of UE, a method of LTE uplink power control will be applied which provides the effective execution of this role, i.e. the UE will receive servicing characteristic of the role that has been applied to it.

The idea of making a coalition is also pressing, though viewed from another perspective. Two types of coalitions could be defined:

- ✓ *Flat coalitions*, which are temporary associations of UEs to which one and the same role had been appropriated.
- ✓ *Subject coalitions*, consisting of different by type but compatible, collaborating UEs.

In this sense the role could be applied as a parameter for recalculating the weight function in a coalition's graph model and thus "resetting" the coalition partnership and discharging the process of uplink power control from the task of appropriating UE with roles which are not matching the current communication profile and the specific sum of individual properties of a specific UE [9].

IV. ROLE GAME APPROACH TO UPLINK POWER CONTROL

The proposed approach is a slot-by-slot power control scheme as shown in Fig. 1. The users are assigned different roles determined by their type, distance from eNB, activity, etc. The role appropriation is possible to be executed by two groups of criteria:

- ✓ *Static criteria* – they are related to the equipment limitations imposed by the UE type;
- ✓ *Dynamic criteria* – they are related to the current communication profile of the user [10].

In both cases the assigned role should aim at maximum of their *utility functions*. Utility is *the level of satisfaction that a user gets from consuming a good or undertaking an activity*. Applying this concept of utility to wireless mobile communications, the user is now the mobile subscriber and the good is the energy stored in the battery of the mobile terminal device. The subscriber consumes the battery energy to gain information throughput or respective quality of performance or service. The utility now measures how much information is delivered or what throughput is achieved by consuming a basic unit of energy. Based on this, the utility of a wireless data user is the total number of correct bits that a user can transmit per unit of its battery energy [bits/Joule] [11, 12, 13].

Coming back to Fig. 1, the different roles are selected so as to achieve the optimum level of user satisfaction defined by their utility functions. Thus the role can be modeled by:

$$Role_i(UPCA_i) = \{UET_i, UEA_i, UEPos_i, MAX[uf_i(ST_i, p_i, P_n)]\} \quad (4)$$

where UET_i is UE_i type, UEA_i is UE_i activity, $UEPos_i$ is UE_i position alteration in regard of the service eNB, uf_i is utility function derived by user i, ST_i is service type, such as voice, data, text, fax, image and video, p_i is power level chosen for UE_i and P_n is maximum power for service cell.

	Time slot 1	Time slot 2	…	Time slot K
UE_1	Role $_1$	Role $_1$	…	Role $_1$
UE_2	Role $_2$	Role $_3$	…	Role $_N$
UE_3	Role $_3$	Role $_N$	…	Role $_1$
…	…	…	…	…
UE_m	Role $_N$	Role $_2$	…	Role $_3$

Fig. 1. Slot-by-slot power control scheme based on roles

In the proposed approach the role appropriation is dynamic and the definition of the duration of the overall validity slot for each role (number of time slots from Fig. 1) is a function of the role itself as well as of the role structure of the users registered in the current eNB and the type of currently existing coalitions.

From a methodological perspective the application of the Role-play game theory in the LTE uplink power control suggests the application of the procedure from Fig. 2.

During its registration in the eNB the User Equipment (UE) - UE_i is assigned an initial role R1 with its corresponding algorithm for the LTE uplink power control - $UPCA_1$. For the duration of the first UE_i slot, role R1 shall be assigned and the eNB shall gather statistics at Key Performance Indicators, which are Utility Functions parameters allowing assessment of the quality of performance acquired by UE_i during the first time slot and within the assigned role R1. Prior to the completion of the first slot the value of the Utility Functions for UE_i shall be calculated and a decision shall be taken whether or not to change the currently acquired role – R_1. If the value of the Utility Functions lies within the frames of the confidence interval for the respective quality of performance, the role shall be preserved for the next slot; if it goes beyond the interval, a change of the role from R_1 to R_j shall be performed thus leading to the algorithm appropriate for the role R_j - $UPCA_j$.

For the effective implementation of the proposed approach it will be necessary to define the relation between the value of the Utility Functions for the current slot and the selection of an appropriate role (algorithm) for the next slot for the specific UE_i. Another issue to be considered relates to the slot duration, i.e. the time during which a role (algorithm) for the LTE uplink power control is not changing for the specific UE_i in regard of the time of calculation of Utility Functions and

the dynamics of the position change or the current UE_i activity. Thus each UE_i shall be awarded an individual management profile, formed by a dynamic roles succession (algorithms for LTE uplink power control), which are acquired with the time according the UE_i type, UE_i activity and the UE_i position alteration in regard of the service eNB. In this sense LTE uplink power control turns into a process, which is in close relation to and depending on the current activity of UE_i, as the interrelation between UE_i is assessed indirectly through the actual values of the Utility Functions calculated for each time slot and each UE_i.

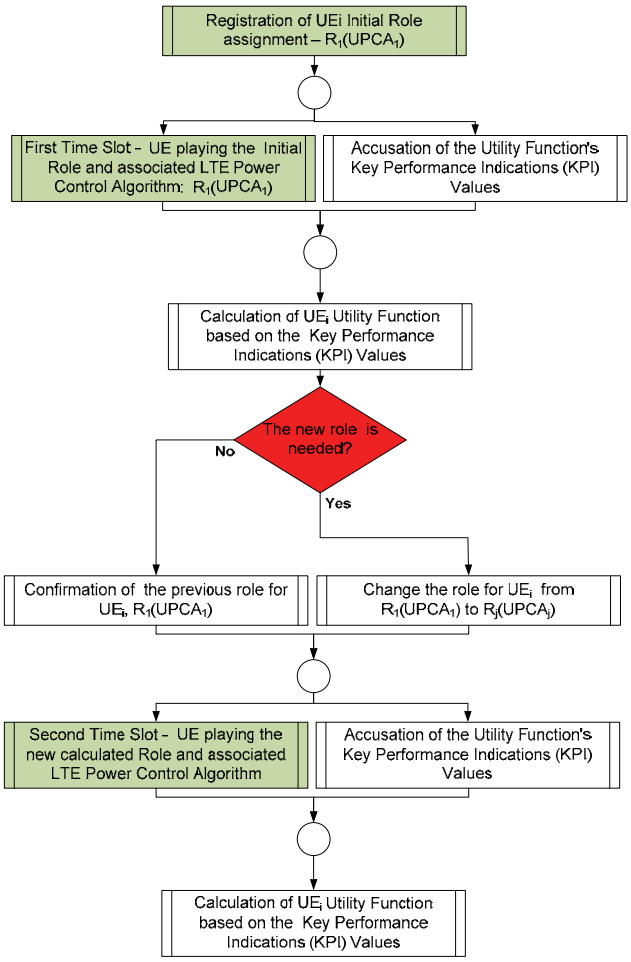


Fig. 2. Procedure for application of the Role-play game theory in the LTE uplink power control

V. Conclusion

In this paper an approach based on defining different roles of the subscribers within a cell of the LTE network for uplink power control is proposed. Comparison between Rule and Role Game Theory is presented. The concept of applying different methods and algorithms for uplink power control, depending on the classified by role subscriber is introduced. The role appropriation is dynamic based on a time slot-by-slot basis. The duration of the overall validity time slot for each role is a function of the role itself as well as of the role structure of the users. This gives the possibility of better utilization of the advantages and application of the existing uplink power control methods. The efficiency of this kind of a role approach is largely dependent on the effective role engineering, such as proper choice of appropriation criteria, characteristics and means of the role dependent method selected for LTE uplink power control.

Acknowledgement

This work was supported by the Technical University of Sofia under Grant 102NI193-7/2010 “Power Control and Control of the Inter-cell Interference in Next Generation Mobile Networks” of the University Research Fund.

References

[1] S. Sesia, I. Toufik, M. Baker, *LTE – The UMTS Long Term Evolution: From Theory to Practice*, John Wiley & Sons, Ltd, 2009.

[2] M. Ergen, *Mobile Broadband Including WiMAX and LTE*, Springer, 2009.

[3] Z. Li, Y. Wang, D. Yang, “A Novel Power Control Scheme in OFDMA Uplink”, *9th International Conference on Signal Processing ICSP*, pp. 2880 – 2883, 2008.

[4] J. Zyren, *Overview of the 3GPP Long Term Evolution Physical Layer*, White Paper, 2007.

[5] C. Castellanos, D. Villa, C. Rosa, K. Pedersen, F. Calabrese, P. Michaelsen, and J. Michel, “Performance of Uplink Fractional Power Control in UTRAN LTE”, *Proc. VTC Spring 2008*, pp. 2517–2521, Singapore, 2008.

[6] M. Boussif, N. Quintero, F. Calabrese, C. Rosa, and J. Wigard, “Interference Based Power Control Performance in LTE Uplink”, *Proc. IEEE International Symposium on Wireless Communication Systems*, pp. 698–702, Reykjavik, 2008.

[7] R. Mullner, C. Ball, K. Ivanov, J. Lienhart, and P. Hric, “Contrasting Open-Loop and Closed-Loop Power Control Performance in UTRAN LTE Uplink by UE Trace Analysis”, *Proc. IEEE International Conference on Communications*, pp. 1-6, Dresden, 2009.

[8] B. Muhammad, A. Mohammed, “Performance Evaluation of Closed Loop Power Control for LTE System”, *Proc. VTC Fall 2009*, pp. 1-5, Anchorage, 2009.

[9] W. Saad, *Coalitional Game Theory for Distributed Cooperation in Next Generation Wireless Networks*, Dissertation Thesis, Department of Informatics Faculty of Mathematics and Natural Sciences University of Oslo, Oslo 2010.

[10] A. Ubisse, N. Ventura, “Sub-Optimal Resource Allocation in the LTE Uplink under ICI Influence and limited Power”, *SATNAC2010 Working Progress*, Department of Electrical Engineering University of Cape Town, South Africa.

[11] B. Y. Bedzhev, M.P. Iliev. *A General Method for Synthesis of Uniform Sequences with Perfect Periodic Autocorrelation. Novel Algorithms and Techniques in Telecommunications and Networking*, Springer, USA, 2010.

[12] R. Mullner, C. Ball, M. Boussif, J. Lienhart, P. Hric, H. Winkler, K. Kremnitzer, and R. Kronlachner, “Enhancing Uplink Performance in UTRAN LTE Networks by Load Adaptive Power Control”, *European Transactions on Telecommunications*, pp. 459-468, 2010.

[13] Z. Lei, *Utility-Based Power Control for Packet-Switched Wireless Networks*, New Brunswick, New Jersey, 2009.

Game Theory Based Competitive Pricing in Next Generation Networks

Vesna Radonjić [1], Aleksandra Kostić-Ljubisavljević[2], Vladanka Aćimović-Raspopović[3]

Abstract – **In this paper we research possibilities of using game theoretic approaches for computing competitive prices in next generation networks. Possible applications and comparison of various game theory models, such as Cournot, Bertrand and Stackelberg game models are presented. We emphasize the advantages and importance of game theory based competitive pricing.**

Keywords – **Next generation networks, Competitive pricing, Stackelberg model, Bertrand duopoly, Cournot duopoly.**

I. Introduction

High speed Internet access and advances in network technologies provide incentives for service providers (SPs) to become more efficient and competitive. Many authors explore the possibilities of applying game theory for solving challenges in next generation networks (NGN) [1-4]. Game theory encompasses a set of mathematical tools adressing complex interactions among rational players, which can be used to explain the operation of various complex telecommunication systems. Among wide range of problems in telecommunications covered by game theory, pricing and competition issues in NGN environment are very important ones. Moreover, the aim of achieving efficient pricing policy can be obtained using the appropriate mechanisms of game theory.

In this paper we discuss the possible applications of game theory models for competitive pricing in NGN. We focus our research on several particular models: Stackelberg model with price and quantity leadership, Bertrand model and Cournot model. We analyzed all the models for the same case of two competing SPs in NGN market.

This paper is organized in the following way. After the brief overview of pricing issues in NGN given in Section 2, game theory based competitive models are discussed in Section 3. In Section 4 examples of possible applications of those models for competitive pricing NGN services are explained. Concluding remarks are given in Section 5.

II. Pricing in Next Generation Networks

Pricing greatly affects the usage of services and the resources consumption. Likewise, competition can be highly influenced by the architecture of a network and the ability of players to control scarce resources in access network.

NGN should be designed to provide an open competition environment for SPs and to provide new possibilities for providers and users to exchange economic signals on fast time scales. A wide range of different pricing schemes is likely to be applied for competitive pricing in NGN [1, 4] and is expected that competition will force service providers to rapidly create and deploy different pricing concepts.

It is required that pricing in NGN enables both off-line and on-line charging, open mechanisms for charging and billing management, various charging and billing policies (e.g. fixed rate charging and usage based per-session charging and billing). It is also expected that accounting functions support services with multicast functionality and to enable all possible types of accounting arrangements, including transfer of billing information between SPs and e-commerce arrangements.

III. Game Theory Based Competitive Pricing Models

A. Basic game theory components and assumptions

Game theory is a field of applied mathematics that describes and analyzes interactive decision making situations and consists of a set of analytical tools that predict the outcome of complex interactions among rational players [5, 6]. Basic components of a game are players, the possible strategies of the players and consequences of the chosen strategies. The players are decision makers and their strategy choices result in a consequence or outcome. They try to ensure the best possible consequence according to their preferences. The preferences of a player can be expressed either with a utility function, which maps every consequence to a real number, or with preference relations, which define the ranking of the consequences.

The most fundamental assumption in game theory is rationality. It is assumed that rational players try to maximize their payoff. If the game is not deterministic, the players maximize their expected payoff. It is also assumed that the players know the rules of the game well.

In game theory, a solution of a game is a set of the possible outcomes. A game describes what strategies the players can take and what the consequences of the strategies are. The solution of a game is a description of outcomes that may emerge in the game if the players act rationally and intelligently. Generally, a solution is an outcome from which no player wants to deviate unilaterally.

[1] Vesna Radonjić, e-mail: v.radonjic@sf.bg.ac.rs.
[2] Aleksandra Kostić Ljubisavljević, e-mail: a.kostic@sf.bg.ac.rs
[3] Vladanka Aćimović Raspopović, e-mail: v.acimovic@sf.bg.ac.rs
All with the University of Belgrade – The Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia.

In telecommunications game theory can be applied for solving a wide range of problems such as congesting control, resource allocation, routing, QoS provisioning, network security, sharing radio-communication spectrum and pricing telecommunication services. Various game models are proposed for pricing telecommunication networks. Some of the most appropriate game models for solving problems of pricing telecommunication services are Nash bargaining game, Stackelberg leader-follower game, Bertrand game and Cournot game.

B. The Nash equilibrium concept

Nash equilibrium is the well known concept for determining solutions in game theory. It implies that each player chooses the best strategy analyzing all possible strategies of all other players in the game.

In two players game, the couple of strategies (s_1^*, s_2^*) represents a Nash equilibrium if s_1^* is the best strategy for the first player when the other player uses strategy s_2^* and if at the same time s_2^* is the other player's best strategic choice for s_1^*. Mathematically expressed, the couple of strategies (s_1^*, s_2^*) represents a Nash equilibrium under following conditions:

$$U_1(s_1^*, s_2^*) \geq U_1(s_1, s_2^*) \text{ for all } s_1 \subset S_1 \text{ and} \quad (1)$$

$$U_2(s_1^*, s_2^*) \geq U_2(s_1^*, s_2) \text{ for all } s_2 \subset S_2. \quad (2)$$

Nash equilibrium doesn't have to be represented by a single best strategy for each player in a game. It can be represented by a set of strategies for each player, such that none is interested in choosing a strategy from any other set that is different from the Nash equilibrium.

Subgame perfect Nash equilibrium is defined as a solution such that players' strategies form Nash equilibrium in every game that is a part of the original game.

Nash equilibrium doesn't exist in every game. An opposite case is a game with several Nash equilibrium points. In both cases the single optimal solution should be chosen. The most commonly used criteria for finding the optimal solution are Pareto efficiency and social optimality.

The solution of the game is Pareto efficient if unilaterally deviating from that solution can not lead to higher payoff for any player in the game without reducing payoff of at least one of the other players in the game. The goal of applied game theory is to form a game with Pareto efficient outcome.

The solution of the game should also satisfy the criterion of social optimality. In more complex games with a great number of players optimal solution from individual players' point is not necessarily the optimal solution from point of system in which the game is implemented. The optimal solution from system's standpoint is actually the socially optimal solution. It can be determined using appropriate optimization techniques. In order to match the optimal solution in terms of individual players with socially optimal solution, pricing concepts with main objectives of system's revenue optimization and encouraging the efficient use of resources are widely used in computer and telecommunications networks. The solution that combines the goals of individual players and the system can be determined by means of appropriate pricing scheme. Pricing mechanism is considered to be incentive if it accomplishes both objectives. For that purpose usually dynamic or hybrid pricing schemes are considered. In those schemes users are charged in accordance to actual use of resources.

C. Stackelberg model

Stackelberg game is two-level optimization model in which at least one player is defined as the leader who chooses a strategy before other players defined as followers. Stackelberg model is an example of a dynamic or sequential game of perfect information. It is a game in which each player knows both the pay-off structures and the history of the game and can observe the actions of others before deciding upon his own optimal response [7].

In Stackelberg game there is a certain order in the decision-making process. Followers' decisions about their strategic choices are based on the strategy that was previously chosen by leader. Stackelberg game can be played with either price leadership or quantity leadership. They can provide a good base for defining prices in NGN especially in case of network congestions.

Regardless of whether the model is played with price leadership or quantity leadership, the problem of finding the optimal strategy for a leader has to be solved, considering that followers as rational players tend to optimize their utility functions with given leader's action.

In Stackelberg game interaction between leader and followers tend to be dynamic [8, 9]. The leader may choose a strategy with aim of maximizing his revenue assuming that the followers will choose their strategies to maximize their own utility functions i.e., best answers. The solution obtained in this way is called the Stackelberg equilibrium and can be analyzed by a backward induction method, which firstly considers the best answers of followers. The best responses of the followers in this game can be obtained as follows: with a given leader's action, total users' demand can be determined based on followers' utility functions. Then these best responses are used to calculate the leader's revenue and the leader chooses a strategy that maximizes his revenue. The equilibrium is achieved at the point of intersection of all the best responses.

D. Bertrand model

Bertrand model is a static or simultaneous game of complete information. In such a game players simultaneously choose their strategies and each player's pay-off structure is common knowledge during the game.

In Bertrand duopoly two players compete in terms of the price they charge users, rather than quantity levels. Thus, in Bertrand duopoly, the strategic variable is the price charged in the market. Players simultaneously decide their pricing

structures and market forces then decide about demand share for each player.

In Bertrand duopoly model, stability depends on whether or not services or products sold by competing firms are identical. In case of identical services/products a Nash equilibrium exists only if both competing firms charge the same prices and make normal profits [7]. The situation where players (i.e.firms) make a competitive outcome without any collusion to increase profits above the normal, is known as Bertrand paradox. One way to overcome the Bertrand paradox is to have firms sell distinguishable services/products. In that case, Nash equilibrium with different equilibrium prices can be obtained [7, 10].

E. Cournot model

Like Bertrand duopoly, Cournot duopoly is a static game, but one in which two players compete in terms of quantity levels, rather than the price they charge users. In Cournot model players choose their strategies independently and simultaneously and each player announces as his strategic choice the quantities of services that he intends to supply.

Cournot duopoly describes how two players selling the same service/product can settle on their respective output levels so as to maximise their own profits. With aim of determining Nash equilibrium, a reaction function has to be defined for each player. Reaction function of a player in Cournot game is a curve that shows his every optimal quantity level for every possible quantity level of the other player. Prices adjust in response to the aggregate supply, so that all the quantity can be sold, and each player obtains a proportional amount of outlay.

IV. Possible Applications of Game Theory Based Competitive Pricing Models

Here we consider possible applications of game theoretic approaches described in previous Section on determining competitive prices of NGN services. We give examples of Stackelberg, Bertrand and Cournot games:

1. Stackelberg duopoly model with price leadership,
2. Stackelberg duopoly model with quantity leadership,
3. Bertrand duopoly model and
4. Cournot duopoly model.

Players in games are two SPs and assumptions are as follows:

- set of possible prices is $[0,1]$,
- quantity level in function of service price is: $x_i(M_i)=1-M_i$ and
- marginal cost of provided services in function of quantity level is: $c_i(x_i)=x_i^2$.

Stackelberg model can be convenient for determining competitive prices at the telecommunication market where some SPs have a competitive advantage over the others (because of technological, historical or legal reasons, or just because their entry was not possible at an earlier stage). SPs with a competitive advantage will act as leaders in Stackelberg game and therefore they will be able to choose their strategies before other SPs who will act as followers.

In Stackelberg game with price leadership, a leader (or leaders) proposes the service price. Based on that price followers decide about their prices. In this paper we explain an example of Stackelberg game with price leadership with two SPs: SP_1 is the leader who commits to price M_1 and SP_2 is the follower. SP_2 will take the leader's price as given, decreasing it by an infinitesimal amount and choosing his quantity level x_2, to maximize $M_2\ x_2 - x_2^2$, giving $x_2 = M_2/2$. Nash equilibrium is obtained for $M_1=M_2=8/15$, $x_1=3/15$ and $x_2=4/15$. This result confirms the fact that in price leadership model a second mover has advantage over the leader. In this game leader obtains less profit than its follower. That is the reason Stackelberg game with price leadership is less frequently used compare to quantity leadership model.

In Stackelberg game with quantity leadership, a leader (or leaders) proposes the service quantity level. After that, followers decide about quantity level they will provide to their users. Here, we use the same example as for the previously explained price leadership model with two SPs. Initial assumption is: SP_1 commits to supply a quantity x_1. After SP_2 observe this, he chooses to supply x_2 with aim of maximizing $x_2(1-x_1-x_2)$, i.e. $x_2(x_1) = 1/2(1-x_1)$. Hence, SP_1 should choose x_1 to maximize $x_1(1 - x_1 - x_2(x_1))$, which gives $x_1=1/2$. Thus Nash equilibrium in this game is: $(x_1, x_2)=(1/2,1/4)$. This result indicates that the leader, i.e. SP_1 has advantage over the follower, i.e. SP_2, which holds in all Stackelberg games with quantity leadership.

In Stackelberg pricing problem with a provider as leader and users as followers there are two levels of decision making process: tariff setting by a provider, and then selection of the best alternatives by users. This game can be formed with a large number of followers, i.e. users and it can be based on both price and quantity leadership. For determining the Stackelberg equilibrium in such a game there is no unique procedure. Various authors propose different solutions [3, 8, 9, 11] based on Stackelberg model of interaction between leader and followers.

The solution of Bertrand game with two SPs and different marginal costs: $c_1<c_2$, which are known to both providers, can be defined by Nash equilibrium. Nash equilibrium in Bertrand duopoly is represented by the prices SPs charge their users for using the service: $M_2 \in (c_1, c_2]$ and $M_1 = M_2 - \varepsilon$ (M_1 and M_2 are the service prices provided by SP_1 and SP_2, respectively and ε is infinitesimally small positive value). This Nash equilibrium confirms the fact that if SP_1 charges his users with price M_1, such that $M_1 > c_1$, SP_2 has no incentive to deviate from M_2, because any reduction in price below marginal cost would mean a loss for him. SP_1 maximizes his profit by taking M_1 only slightly lower than M_2. Hence, SP_1 will always win with a net benefit equal to approximately M_2-c_1 per unit sold. Since none will offer the price below the marginal cost, for SP_2 the strategy $M_2=c_2$ dominates the strategy $M_2=M'_2$ for all $M'_2 < c_2$, i.e. the first strategy is as good or better than the second, for all values of M_1. Thus, by imposing the constraint $M_2 \geq c_2$, we conclude that $(M_1, M_2) = (c_2 - \varepsilon, c_2) = (x_2^2 - \varepsilon, x_2^2)$ is the Nash equilibrium of this Bertrand game. Following the same assumptions as in the previously explained Stackelberg

game models, Nash equilibrium in the Bertrand game can also be interpreted in terms of quantity levels: (x_1 , x_2)=(1/4,1/4).

Generally, in Bertrand game with more than two players with equal and constant marginal cost, the optimal solution does not dependent on the number of players and corresponding price for each player is equal to marginal cost.

In Cournot model with two SPs offering the same service with total quantity level $x = x_1 + x_2$, where quantity level of SPi is x_i for i =1,2, the resulting price in the market will be $M(x)$. Each provider must choose an amount of output to be produced, and then, as a function of both choices, receive a payoff (that is his net benefit). SPi's net benefit can be written as: $\pi_i(x_1,x_2)=M(x_1+x_2)x_i-c_i(x_i)$, where $c_i(x_i)$ is his cost of providing the quantity x_i. The Nash equilibrium in this game is the pair of outputs (x_1*, x_2*) with the property that if SPi chooses x_i*, then there is no incentive for SPj to choose other than x_j*, where $i, j \in \{1,2\},\ i \neq j$.

To obtain reaction functions x_i (x_j) for each SPi, the first order conditions are defined:

$$\partial \pi_1(x_1, x_2)/\partial x_1 = M(x_1 + x_2) + M'(x_1 + x_2)x_1 - c'_1(x_1) = 0 \quad (3)$$

$$\partial \pi_2(x_1, x_2)/\partial x_2 = M(x_1 + x_2) + M'(x_1 + x_2)x_2 - c'_2(x_2) = 0 \quad (4)$$

The Nash equilibrium point can be found in the intersection of these curves. It can be shown that with the adoption of appropriate assumptions for the reaction curves (such as concavity), the optimal solution is always stable. This means that if the game is played in several rounds and if the players sequentially choose their strategies depending on the previous moves of other players, then their outputs converge to a Nash equilibrium point.

In our example of Cournot game x_i is chosen within the interval [0,1] and the inverse demand curve is: $M(x_1 + x_2) = 1 - (x_1 + x_2)$. Then: $x_i(x_j) = 1/2(1 - x_j)$ and the Nash equilibrium is: (x_1 , x_2)=(1/3,1/3).

The solution of Cournot game compare to Stackelberg model with price leadership provides higher quantity level for both SPs. On the other hand, Stackelberg model with quantity leadership provides lower quantity level for SP_2, but higher quantity level for SP_1 compare to Cournot model. The given solution in Bertrand game is only better for SP_1 in comparision with the one obtained in Stackelberg model with price leadership.

V. Conclusion

Analysis of recent research in the field of pricing telecommunications services has shown that research efforts are increasingly directed to the use of game theory mechanisms for defining the tariff system in telecommunication networks.

For the analysis of the tariff system and the relationship between competitive service providers, Stackelberg, Bertrand and Cournot, models can be used. In this paper we analyzed those game theoretic approaches for the same case of two competing SPs in NGN market. Bertrand model is suitable for solving the problems of determining service prices. Cournot model is convenient for modeling strategic choices of service providers that focus on the services quantity levels. Stackelberg model can be applicable to solve the problem of determining prices and/or service quantity level a provider offers to his users.

Acknowledgement

This work is partially supported by the Ministry of the Science and technological development of the Republic of Serbia with the project TR32025.

References

[1] 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 A. Girard, B. Sanso and F. Vazquez Abad, Springer, 2005.

[2] Z. Ji, W. Yu, and K. J. R. Liu, "A Game Theoretical Framework for Dynamic Pricing-Based Routing in Self-Organized MANETs", *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1204-1217, 2008.

[3] P. Briest, M. Hoefer, and P. Krysta, "Stackelberg Network Pricing Games", *Proc. of the STACS 2008*, vol. 1, pp. 133-142, 2008.

[4] V. Radonjić, V. Aćimović-Raspopović, "Responsive Pricing Model with Fixed Bandwidth Usage for the Next Generation Internet", *Proc. of the ICEST 2008*, vol. 2, pp. 425-428, Nis, Serbia, 2008.

[5] M. J. Osborne, *An Introduction to Game Theory*, Oxford University Press, 2004.

[6] J. N. Webb, *Game Theory - Decisions, Interaction and Evolution*, Springer, 2007.

[7] A. Kelly, *Decision Making Using Game Theory - An Introduction for Managers*, Cambridge University press, New York, USA, 2003.

[8] P. Y. Nie, "Dynamic Stackelberg Games under Open-Loop Complete Information", *Journal of the Franklin Institute-Engineering and Applied Mathematics*, 342, pp. 737-748, 2005.

[9] S. V. Hoesel, "An Overview of Stackelberg Pricing in Networks", *European Journal of Operational Research*, vol. 189, no. 3, pp. 1393-1402, 2008.

[10] C. Courcoubetis, R. Weber, *Pricing Communication Networks*, John Wiley & Sons Ltd, 2003.

[11] T. Basar, R. Srikant, "A Stackelberg Network Game with a Large Number of Followers", *Journal of Optimization Theory and Applications*, vol.115, no.3, pp.479–490, 2002.

Fuzzy Evaluation of Service Level Management Metrics

Aleksandar Tsenov[1], Galina Yoncheva[2], Elena Stoyanova[2] and Aleksandar Pavlov[2]

Abstract – **SLM (Service Level Management) metrics, according the ITIL (Information Technology Infrastructure Library) framework, are the building blocks of the Service Level Management procedures. These metrics are the quantitative measure of the successful implementation of the SLM by the IT service providers. The evaluation of the whole SLM process is a complex problem that could not be fulfilled in regular way, such visual evaluation, because of the huge amount of process metrics. Moreover these metrics are different by the nature and are measured in different units (%, time, amount etc.) This paper introduces an appropriate fuzzy based approach for evaluating the SLM process metrics.**

Keywords – **Information Technology Infrastructure Library, Service Level Agreement, Service Level Management, Fuzzy logic, Customer Experience.**

I. Introduction

The quick development of new standards for delivery and support of IT services challenges the growth of the role of SLM generated and proposed according to the higher customer demand. From the other side that challenges the enterprises to look for new efficient approaches for delivering high quality service and higher level of ROI (Return of investments). Most of the enterprises have implemented the standards for design, delivery and support of managed services to their customer and now they search the most appropriate business conditions for itself in order to reduce cost and enhance the revenue [1],[2],[3].

In the field of the Information Technologies such an approach is the mapping of ITIL Procedures over known business process frameworks – such as eTOM [4]. Within that mapping the ITIL procedure metrics are well known, but the eTOM framework is being depicted only in three levels and only a small part of the business process metrics are conclusively defined. Until now the metrics according the SLM (Service Level Management) are not clarified. The difficulty ensues from the huge amount of services delivered and from the different points of view for the quality of these services – the technological characteristics (from the enterprise side) and perception level (from the customer side). Both views on one and the same service characteristics should be met on an interface between the enterprise and the customers where the technical parameters of the quality could be easy adopted to the customer perception and vice versa – the SLA.

[1]Aleksandar Tsenov is with Telecom Department at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: akz@tu-sofia.bg

[2] Former Master-Students in the FDIBA at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: galia_btg@abv.bg; lendra@abv.bg, alek_pln@abv.bg

The main goal of this work and next works as well, is to propose a fuzzy similarity based approach for evaluation and clustering the Service Level Management parameters of the services in mobile networks. The approach moves toward two threads – service centric (evaluation) and customer centric (clustering) of the SLM metrics. This work will represent the first of these processes – the fuzzy evaluation of the SLM metrics.

The nature of the requirements and the defined metrics for evaluation of the IT Service management is the precondition for choosing the Fuzzy set theory for solving such complex problems.

II. Evaluation of SLM Metrics

A. The SLM process

The SLM Process is responsible for negotiating Service Level Agreements, and ensuring that these are met. SLM is responsible for ensuring that all IT Service Level Agreements, Operational Level Agreements, and Under Pinning Contracts, are appropriate for the agreed Service Level targets [1],[2]. SLM monitors and reports on Service levels, and holds regular Customer reviews – Fig 1.

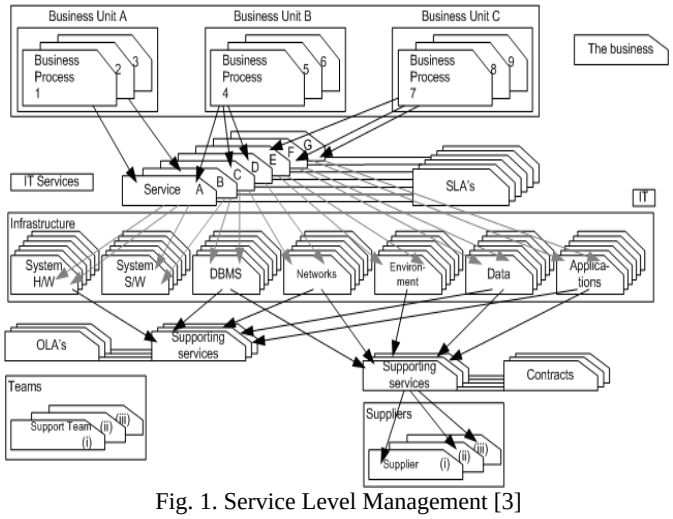


Fig. 1. Service Level Management [3]

Service Level Management (SLM) negotiates, agrees and documents appropriate IT service targets with representatives of the business, and then monitors and produces reports on the service provider's ability to deliver the agreed level of service. SLM is a vital process for every IT service provider organization in that it is responsible for agreeing and documenting service level targets and responsibilities within SLA's and SLR's (Service Level Requirements), for every activity within IT Service Delivery Process.

B. The SLM process metrics

The SLM process metrics are captured in the form of Critical Success Factors (CSFs), Key Performance Indicators (KPIs) and activity metrics for the service management processes [2], [3]. CSF would use these metrics as input in identifying improvement opportunities for each process. Some important metrics used for the evaluation process proposed in this work, are shown in Table 1.

TABLE 1
SLM EVALUATION METRICS FOR THE MOBILE NETWORKS SERVICES

No.	Characteristics	Description	Code
1	% of services covered by SLA	SC = ServCovered/AllServ x 100%	SL1
2	OLA's and UC's to support all SLA's	Contracts = OLA's + UC's	SL2
3	Service improvement measures	SIM = technical improvements + process&procedures+ "Quick wins"+ Training + dialogue	SL3
4	% of SLA targets being met	SLAm = SLAtargMet/ AllSLAtarg x 100%	SL4
5	Number of SLA targets on risk	SLAr = SLAtargRisk/ AllSLAtarg x 100%	SL5
6	Number of breaches per period	BR=1/T x ∑Br	SL6
7	Customer perceptions improvement	CustPercCurent = $CustPerc_{i+1}$ – $CustPerc_i$	SL7
8	Number of Service Improvement Plans (SIP's) Opened	SIPOp = ∑OpenedSIP	SL8
9	Number of SIP's closed	SIPCl = ∑ClosedSIP	SL9

The characteristics mentioned in the above Table 1 have the meaning as follows: "% of services covered by SLA" gives the coverage of the services with SLA. By most of the mobile network operators there are services which are delivered to the customers by default and aren't a part of the contract between the operator and the customer; "OLA's and UC's to support all SLA's" are the internal contracts (for the enterprise) and the contracts with external suppliers; "Service Improvement Measures" represents a number of activities fulfilled in order to evaluate the service improvement degree of the management process; "% of SLA targets being met" is a criterion for the accomplishment of the SLA have being contracted; "Number of SLA targets on risk" these are SLA targets not fulfilled for a definite time period and look like not feasible within the corresponding service delivery process; "Number of breaches per period" represent all breaches aroused in a period of time; "Customer perceptions improvement" gives the improvement of the satisfactory degree for each customer during the use of the service; "Number of Service Improvement Plans (SIP's) Opened" and "Number of Service Improvement Plans (SIP's) Closed" are measures for the reaction of the service provider in order to change the service characteristics in response to the customer demand.

In the following analysis, we simplify the description of the formula of the metrics – in most cases each characteristic includes a number of sub-characteristics. This simplification will not affect the method and conclusions of fuzzy evaluation method.

III. FUZZY EVALUATION OF SLM METRICS

A. The Fuzzy evaluation method

The following section describes the basic steps of the proposed fuzzy evaluation approach [5],[6]:

1. In an element set $U=\{u_1,u_2,u_3,\ldots,u_n\}$ each element denotes each characteristics in evaluation metrics of SLM process. The n means the number of characteristics in evaluation metrics.
2. Suppose that Evaluation set is $V=\{v_1,v_2,v_3,\ldots,v_m\}$. The m means the number of the evaluation levels. The value of evaluation is the degree of u_i to v_j.
3. Suppose that fuzzy matrix is $R=(r_{ij})_{mxn}$. The element r_{ij} means that the membership degree of the i-th element in set U to the j-th element in set V.
4. Suppose that set $A=\{a_1,a_2,a_3,\ldots a_n\}$ is the fuzzy matrix set of the weight of each characteristics – a_i means the weight of the element u_i in the set U. Set A is the calculation of the membership degree. This paper chooses the Pair-wise Comparison [5] method to calculate set A. The elements in set A satisfy normalization condition:

$$\sum_{i=1}^{n} a_i = 1 . \quad (1)$$

5. Suppose that the effect degree of evaluation set V element – a fuzzy matrix set is $B=\{b_1,b_2,b_3,\ldots b_m\}$. b_j means the weight of the evaluation decision in set V. The calculation of set B is by the equation: $b=a\circ R=\{b_1,b_2,b_3,\ldots,b_m\}$. "$\circ$" means compound algorithm operator (Adamar multiplication [7]).
6. The evaluation results of S is calculated in the following steps: In order to get the accurate calculation, the weight of each evaluation is calculated first by the following equation:

$$\delta_j = \frac{b_j}{\sum_{i=1}^{m} b_i}, j = 1,2\ldots m . \quad (2)$$

Then for each evaluation level v_j the total evaluation S is got from the following Equation:

$$S = \sum_{j=1}^{m} \delta_j v_j . \quad (3)$$

B. A practical example

The characteristics in Table 1 are evaluated according to the above fuzzy evaluation method. The key is how to get the membership degree set *A*. There are several known methods to calculate *A* [5]. The proposed approach concerns to a very new problem and therefore the Pair-wise Comparison method is chosen. That means – every two characteristic's membership degree is being compared to create the membership sequence with which fuzzy set *A* is abstracted.

Sequence relation in Table 2 is got from element set $U=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7,u_8,u_9\}$. The result of the comparison of every two characteristics is shown. Because the most characteristics have different values for different services, Table 2 gives just an instance of relationships. The same method may be used for different situations as well.

TABLE 2
MEMBERSHIP DEGREE OF SLM EVALUATION METRICS

Item	SL1	SL2	SL3	SL4	SL5	SL6	SL7	SL8	SL9
SL1	=	≥	≥	≤	≥	≥	≤	≤	≤
SL2	≤	=	≤	≤	≥	≥	≤	≤	≤
SL3	≤	≥	=	≤	≥	≥	≤	≤	≤
SL4	≥	≥	≥	=	≥	≥	≤	≤	≤
SL5	≤	≤	≤	≤	=	≤	≤	≤	≤
SL6	≤	≤	≤	≤	≥	=	≤	≤	≤
SL7	≥	≥	≥	≥	≥	≥	=	≥	≥
SL8	≥	≥	≥	≥	≥	≥	≤	=	=
SL9	≥	≥	≥	≥	≥	≥	≤	=	=

In the table "=" means that membership degrees to A of two characteristics are the same; "≥" means that the membership of SL_i to A is higher than that of SL_j to A; "≤" means that the membership degree of SL_i to A is lower than that of SL_j to A.

Set A is calculated from Table 2: $\forall u_i$, $u_i \in U$, let p_{ij} denotes the membership degree of u_i/A to u_j/A.

The rules to get p_{ij}:

$$0 \le p_{ij} \le 1; i, j = 1,2,... n; \quad (4)$$

$$p_{ij} + p_{ji} = 1; \forall i \ne j. \quad (5)$$

The fuzzy relation matrix (Eq. 6) is calculated according to Table 2 and Equations (4) and (5):

$$\begin{bmatrix} 1 & 0.6 & 0.6 & 0.4 & 0.6 & 0.6 & 0.4 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 & 0.4 & 0.6 & 0.6 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.6 & 1 & 0.4 & 0.6 & 0.6 & 0.4 & 0.4 & 0.4 \\ 0.6 & 0.6 & 0.6 & 1 & 0.6 & 0.6 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 & 1 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 & 0.6 & 1 & 0.4 & 0.4 & 0.4 \\ 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 1 & 0.6 & 0.6 \\ 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.4 & 1 & 1 \\ 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.4 & 1 & 1 \end{bmatrix} \quad (6)$$

The fuzzy set *A* is calculated by average method and for the given values for the p_{ij} the following values are obtained:

A = {0.556, 0.511, 0.533, 0.578, 0.467, 0.489, 0.644, 0.667, 0.667}.

The following set *A'* is the result to satisfy the normalization method:

A' = {0.108, 0.099, 0.105, 0.114, 0.091, 0.096, 0.125, 0.131, 0.131}.

According to the steps 1-3 of Section III.A and to the above calculations for the fuzzy set *A*, the following three SLM metric sets are evaluated:

U1 = {98, 97, 97, 98, 97, 97, 100, 99, 99};
U2 = {98, 98, 98, 98, 98, 98, 98, 98, 98};
U3 = {97, 96, 97, 98, 95, 96, 99, 98, 98}.

Evaluation set is *V* = {*Excellent, Very good, Good, Poor*} = {*100, 80, 60, 40*}. In Eq. 7, 8 and 9, the *R* matrices of the three systems, defined above by their SLM metrics, are described. Each matrix *R* is got from a method of expert's evaluation. We suppose that 100 experts give scores for each system. The number of experts for each evaluation level works as the element of matrix *R*. For instance: 0.99 in the first column means that 99 of the 100 expert gave "Excellent" to the corresponding metric.

$$\begin{bmatrix} 0.98 & 0.01 & 0.01 & 0.0 \\ 0.97 & 0.01 & 0.01 & 0.01 \\ 0.97 & 0.02 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.97 & 0.01 & 0.01 & 0.01 \\ 0.97 & 0.01 & 0.01 & 0.01 \\ 1 & 0.0 & 0.0 & 0.0 \\ 0.99 & 0.01 & 0.0 & 0.0 \\ 0.99 & 0.0 & 0.01 & 0.0 \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.02 & 0.0 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \end{bmatrix} \quad (8)$$

$$\begin{bmatrix} 0.97 & 0.01 & 0.01 & 0.01 \\ 0.96 & 0.02 & 0.01 & 0.01 \\ 0.97 & 0.02 & 0.01 & 0.0 \\ 0.98 & 0.02 & 0.0 & 0.0 \\ 0.95 & 0.02 & 0.02 & 0.01 \\ 0.96 & 0.01 & 0.02 & 0.01 \\ 0.99 & 0.0 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \\ 0.98 & 0.01 & 0.01 & 0.0 \end{bmatrix} \quad (9)$$

The calculation results of the corresponding sets B are as follows:

B1 = {0.98121, 0.00849, 0.00744, 0.00286};
B2 = {0.98000, 0.01096, 0.00904, 0.00000};
B3 = {0.97429, 0.01284, 0.01073, 0.00394}.

The three systems evaluation gene δ is calculated by set *B* and Eq. (2), then the total evaluation score ***S*** is calculated according the Eq. (3):

S1= 0.98121*100+0.00849*80+0.00744*60+0.00286*40 =
= **99,361**
S2= 0.9800*100 +0.01096*80+0.00904*60+0.0000*40 =
= **99,4192**
S3= 0.97429*100+0.01284*80+0.01073*60+0.00394*40 =
= **99,0776**

The conclusion from evaluation result of S is {**System-2, System-1, System-3**} in the sequence of service management quality from the best to the worst.

The same conclusion may be obtained through a visual analysis, but only if the score of metrics is not very complex. In this example we show that when the score of metrics is more complex, it is easier to get the evaluation result of service management quality using fuzzy evaluation algorithm than using direct analysis. In addition – in this example we have used only nine Service Level Management metrics. Actually the number of the SLM metrics, according the ITIL best practices, is much higher.

In this work we didn't made any classification of the metrics being evaluated. The fuzzy based approaches allow clustering of the metrics according common characteristics of the metrics (for example: unit of measurement). Such clustering will improve the complexity of the evaluation and therefore the use of fuzzy based handling is almost obligatory.

The proposed approach is open for future enhancement – firstly: there is no limitation of the number of the metrics to be evaluated, and secondly – it is possible to implement other methods for producing the relation matrices, such as cosine distance, Euclidean distance etc.

IV. Conclusion

Service management is very important for service provision in correct, effective and economic way. This paper analyzes the increasing service management requirement and points out that the evaluation metrics of service management quality is valuable and useful to improve and evaluate service management.

Based on the analysis and reference of SLA service management content, the evaluation metrics of SLA-oriented service management quality is presented. Then fuzzy evaluation method is used to evaluate service management quality. The process of fuzzy relation matrix *A'* calculation based on Pair-wise Comparison and the calculation of evaluation result is also illustrated. A simple example is explained to illuminate that fuzzy evaluation method is valid and correct. The metrics evaluation is a foundation to extend and to improve for different requirements and the evaluation algorithm can be amended to adapt multiple complexities.

Because the research to the evaluation of Quality of Service Management just begins, a lots of research work is required to do. Moreover the customer – centric evaluation is also required in order to "translate" the customer satisfaction into service quality parameters. So the service providers will be able to monitor and manage the service quality in response to the customer experience.

Acknowledgement

This work is established within a frame-contract between the Faculty of Telecommunications at the Technical University of Sofia and Mobiltel AD – Bulgaria.

References

[1] "SLA Management Handbook – Volume 4: Enterprise Perspective", TM Forum, Published by The Open Group, October 2004.
[2] "ITIL process procedure templates", available on http://www.metocube.com.
[3] ITIL V3 - Service Design - Management guide, Office of Government Commerce, first issue, 2008.
[4] ITU-T – Series M3050x, Enhanced Telecom Operations Map (eTOM), 2004 – 2007.
[5] H. Jiang, "Fuzzy Evaluation on ERP System Implementing Risk Based on Membership Degree Transformation New Algorithm", *Second International Symposium on Electronic Commerce and Security*, pp 409-416, Nanchang, China, 2009.
[6] K. -M. Chao, M. Younas, C. –C. Lo, and T.- H. Tan, "Fuzzy Matchmaking for Web Service", *Proceedings of 19 IEEE Conference on Advanced Network and Information Application, IEEE CS*, pp 721-726, 2005.
[7] S. Jordanova and V. Mladenov, "Fuzzy sets and computational intelligence", Lecture notes, Sofia University, 2005.

Modeling ITIL-SLM Process Flows with eTOM Level 3 Process Elements

Todor Georgiev[1] and Aleksandar Tsenov[2]

Abstract – **The paper will demonstrate how the frameworks of the Information Technology Infrastructure Library (ITIL) and eTOM (Enhanced Telecommunications Operation Map) can be combined to face the challenge to expand from pure telecom business to the converged information technology and telecommunication market. Furthermore it validates concepts and recommendations outlined in GB921U, GB921V, TR143, and other TM Forum documents. The goal of this work is to provide a model of the ITIL - SLM (Service Level Management) Process Flow with eTOM Level 3 process elements.**

Keywords – **Information Technology Infrastructure Library, Enhanced Telecommunications Operation Map, Service Level Management.**

I. Introduction

ITIL is a guideline for IT Service Management [1], [2]. ITIL was originally developed as a guideline to serve a single customer, though many Users, delivering many IT Services. The customer in this case is generally the organization or company that is using the IT Services to support its business whether they are provided using internal resources or external outsourcers. Below this single point of delivery any number of Operational Level Agreements (OLAs) and Underpinning Contracts (UC's) can be used to support the delivery of IT Services to the customer and their users. An OLA is an internal Service Level Agreement while an Underpinning Contract is a Service Level Agreement with a third party supplier.

ITIL is primarily non-prescriptive - it offers advice/ guidance on the implementation and continued delivery of Service Management. It contains extensive advice covering the construction of cases for the systematic implementation of Service Management, overcoming objections raised to those proposals, planning their implementation and resolving typical problems likely to be encountered during the implementation process [3],[4],[5].

eTOM is a catalogue of process element categories [6]. eTOM is a business process framework to guide the development and management of key processes within an Infor mation and Communications Service Provider. It provides this guidance by offering a catalogue of industry-standard names, descriptions and scope, at multiple hierarchical levels, of all the Business Activities (or process elements) within an information and communications Service Provider [6], [7].

The current version of eTOM is customer centric viewing business processes in terms of their contribution (whether directly or indirectly) to customer service. At that level, it is difficult to discern the shape of some of the internal support processes dealing with the infrastructure, resource and business needs of the organization [8], [9]. Those processes supporting IT systems fall into this category. Future iterations of eTOM may illustrate the shape of these processes, as well as develop the linkages needed between these processes and customer-related processes.

To accelerate these enhancements to eTOM, reference to specialized process recommendations such as ITIL will bridge these gaps with proven, reliable process models. Further, the application of eTOM to the development of ITIL processes is likely to improve the scope and design of these processes, and will ensure their successful integration into a Service Provider's overall process environment.

II. The Service Level Management Process

The SLM process is responsible for negotiating Service Level Agreements, and ensuring that these are met. SLM is responsible for ensuring that all IT Service Management Processes, Operational Level Agreements, and Underpinning Contracts, are appropriate for the agreed Service Level targets. SLM monitors and reports on service levels, and holds regular customer reviews.

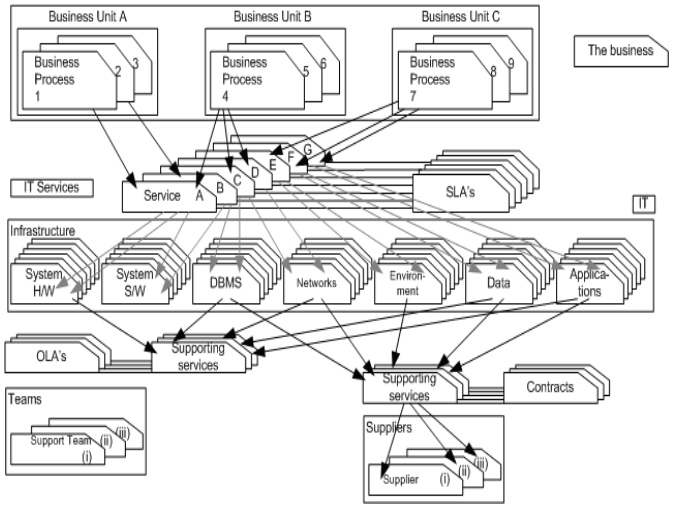


Fig. 1. Service Level Management [10]

[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

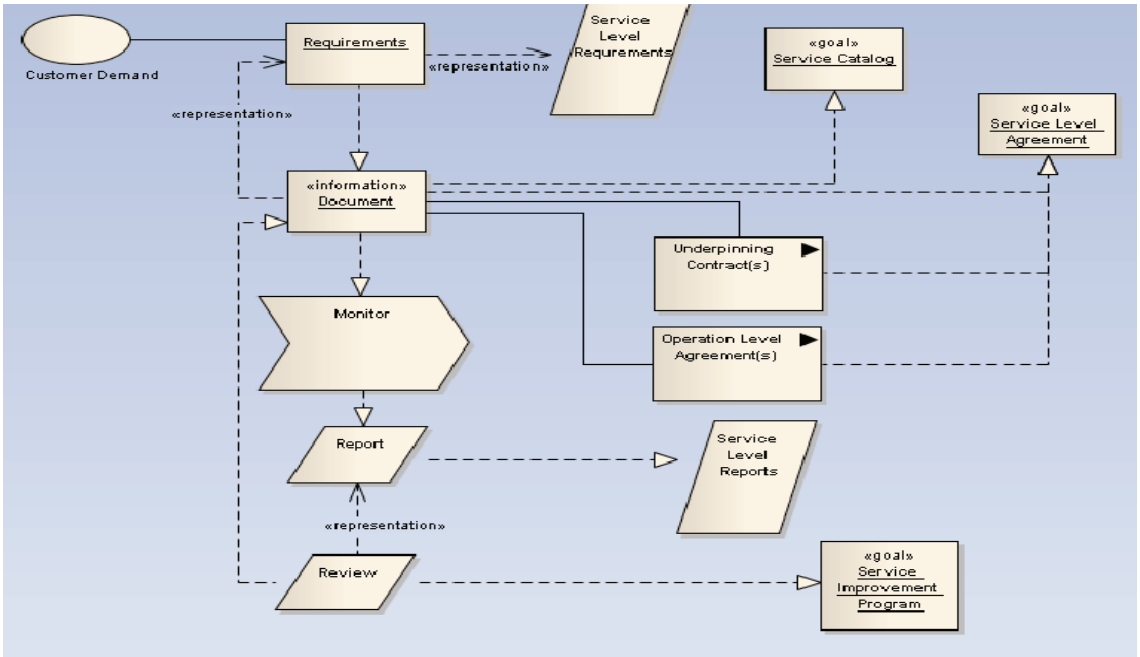


Fig. 2. The Service Level Management process (build and validated with Enterprise Architect® v8.0)

Service Level Management (SLM) negotiates, agrees and documents appropriate IT service targets with representatives of the business, and then monitors and produces reports on the service provider's ability to deliver the agreed level of service. SLM is a vital process for every IT service provider organization in that it is responsible for agreeing and documenting service level targets and responsibilities within SLAs and SLRs, for every activity within IT. If these targets are appropriate and accurately reflect the requirements of the business, then the service delivered by the service providers will align with business requirements and meet the expectations of the customers and users in terms of service quality. If the targets are not aligned with business needs, then service provider activities and service levels will not be aligned with business expectations and problems will develop. The SLA is effectively a level of assurance or warranty with regard to the level of service quality delivered by the service provider for each of the services delivered to the business. The success of SLM is very dependent on the quality of the Service Portfolio and the Service Catalogue and their contents, because they provide the necessary information on the services to be managed within the SLM process [10].

Figure 1 shows the relationship between the business and its processes and the services, and the associated technology, supporting services, teams and suppliers required to meet their needs. It demonstrates how important the SLAs, OLAs and contracts are in defining and achieving the level of service required by the business.

Although Figure 2 illustrates all the main activities of SLM as separate activities, they should be implemented as one integrated SLM process that can be consistently applied to all areas of the businesses and to all customers. These activities are described in the following sections. The model is developed and validated with Enterprise Architect® v8.0.

III. Mapping and Modelling between ITIL and eTOM

A. Scope and Type of Modelling

While eTOM and ITIL continue to provide value in their original fields of application, a number of factors are pushing the adoption of a combination of the two frameworks:

business drivers: Increase process efficiency by consolidating processes and technologies through standardization using the combined frameworks; Organizational consolidation between Operations Support Systems (OSS) and IT (driven by operational expenditure/capital expenditure [OpEx/CapEx] reduction pressure); Efficient management of convergent services requires more efficient end-to-end process management across Business Support Systems (BSS), OSS and IT; Interfacing with external or internal organisations that have adopted ITIL; Improvement of specific eTOM process segments by the adoption or reuse of ITIL best practices; Need for a consistent governance framework to govern strategy and align it to business objectives;

technology drivers: Service convergence on IP networks leads to the implementation of next-generation services, such as Voice over IP (VoIP), IP Television (IPTV), IP Multimedia Subsystem (IMS), and Service Delivery Platform (SDP) technologies. These services encompass pure network OSS layers and Value-Added Service (VAS) layers. End – to - end service management implies a fully

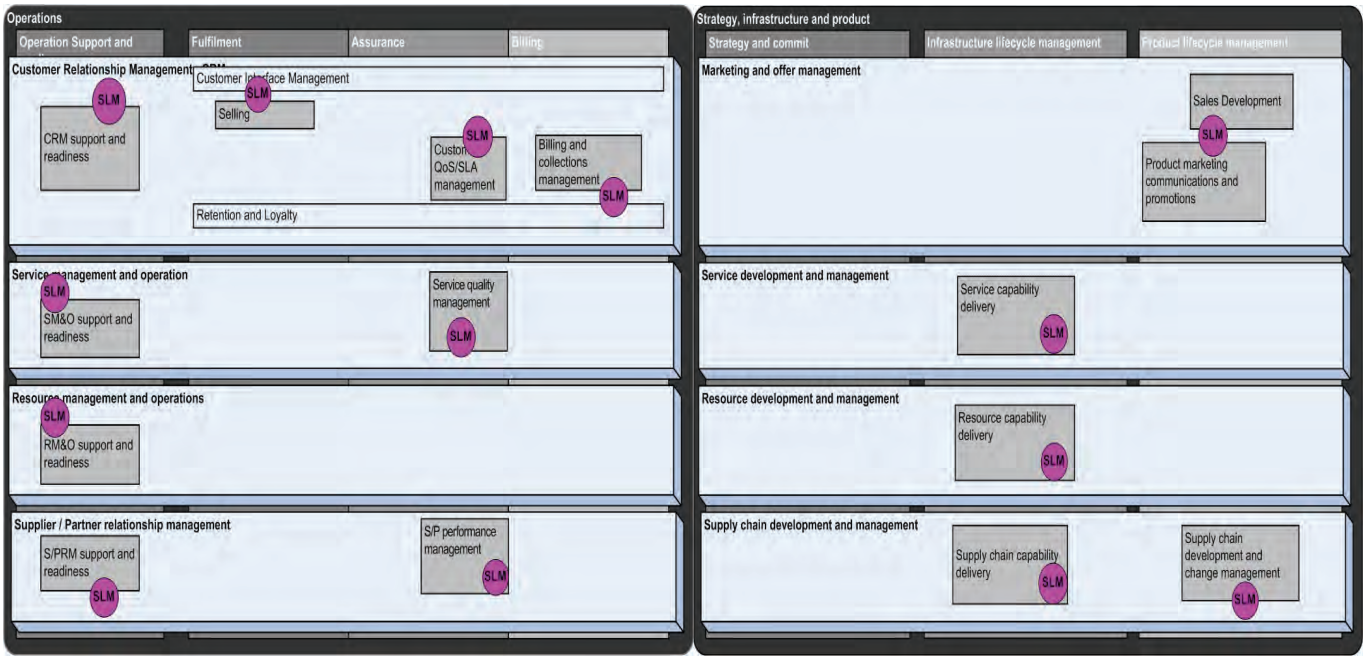


Fig. 3. Correlation between ITIL SLM processes and eTOM Operation process elements and SIP process elements

integrated service view that leads to end-to-end processes and interfaces across the OSS and IT organization.

There are several methods of mapping between the ITIL processes and the eTOM framework. One possible approach might be to decompose the ITIL processes into lower-level processes and end-to-end flows can be constructed from a combination of both eTOM and ITIL processes, alternatively ITIL processes can be constructed solely from lower-level eTOM process elements. The problem here is that the eTOM is currently defined to Level 2 for most processes and partially to Level 3.

The second way to map ITIL and eTOM is to define the correlation level between the major ITIL processes and the eTOM process elements.

The Fig. 3 represents the eTOM process elements and the position of the ITIL process elements toward to these elements. The different overlapping represents the correlation degree between them. The places and the correlations degree of the SLM Process are shown as well. The corresponding process flows that have to be modeled can be clearly seen on both figures.

The correlation degree is being fixed analytically by studying the SLM Process metrics and the predefined eTOM process elements metrics. Both metric types are the lowest-level decomposition of both ITIL and eTOM. According this study the Service Level Management Process corresponds to the eTOM Level 2 process elements shown on Fig. 3.

The eTOM Level 3 process elements were defined during the study as well. Because of the limited space Table I includes only some important eTOM process elements from the whole eTOM framework defined, that are relevant to the SLM process:

TABLE 1
ETOM PROCESS DECOMPOSITION

eTOM Level 2 Process	eTOM Level 3 Process
CRM, SM&O, RM&O and S/PRM Support and Readiness	Support Customer Interface Management Support Order Handling Support Problem Handling Support Billing & Collections - Support Bill Invoice Management - Support Bill Payment & Receivables Management - Support Bill Inquiry Handling Support Retention & Loyalty Support Marketing Fulfilment Support Selling Support Customer QoS/SLA
Customer QoS/SLA Management	Assess Customer QoS/SLA Performance Manage QoS/SLA Violation Report Customer QoS Performance - Create Customer QoS Performance Degradation Report - Track & Manage Customer QoS Performance Resolution - Close Customer QoS Performance Degradation Report
Service Quality Management	Monitor Service Quality Analyse Service Quality Improve service Quality Track & Manage service Quality Report Service Quality Performance
Service capability delivery	Map & Analyse Service Requirements Capture Service Capability Shortfalls Gain Service Capability Investment Approval Design Service Capabilities Enable Service Support & Operations Manage Service Capability Delivery Manage Handover to Service Operations
Technology Scanning	Scan for Emerging Technology from External Sources for the Enterprise Assess of Emerging Technology from External Sources for the Enterprise

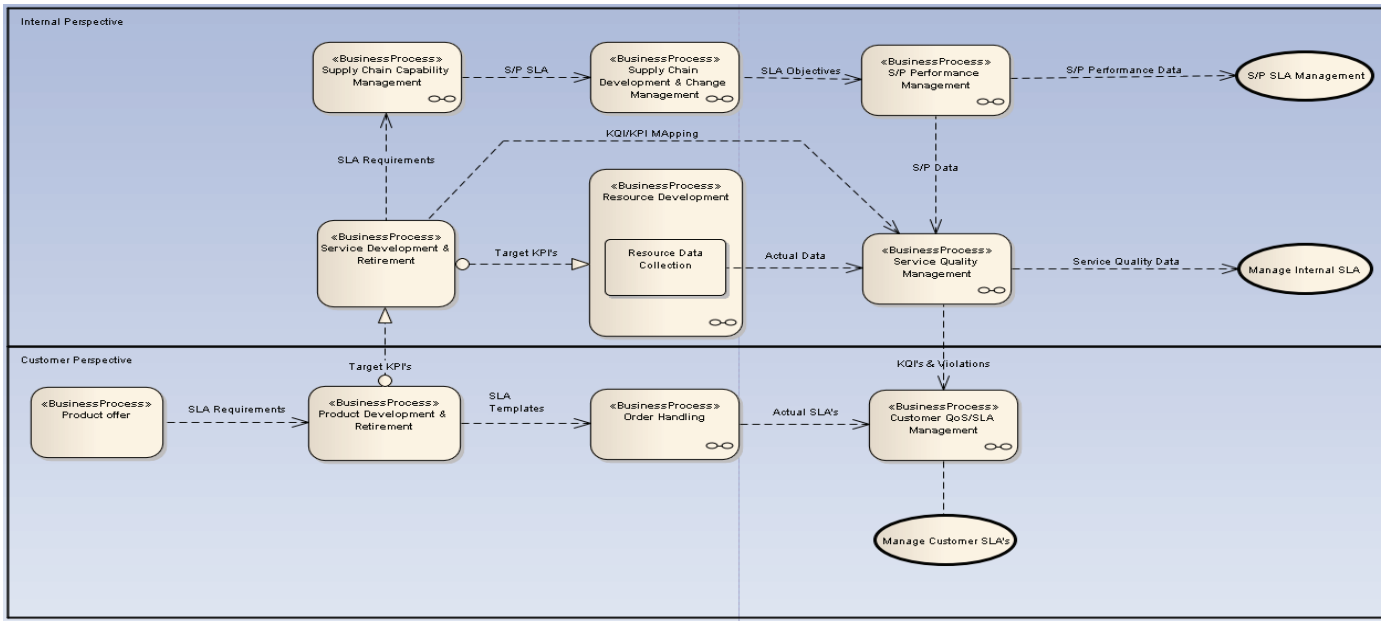


Fig. 4. The ITIL SLM Processes modeled with eTOM SIP Process Elements

B. Modelling ITIL Processes using eTOM Process Elements

Information and Communication Service Providers need to ensure that many different processes that they operate will work together effectively. The staff can easily understand the processes without having to learn a new language and they must be effectively implemented in IT. To do this they need to model all their operational process in a standard way and that can be done by modelling the processes using the Process Elements of the eTOM process framework.

Fig. 4 represents the goal of this work – the ITIL - SLM Process modelled with eTOM Process Elements. According to the eTOM Framework two areas where defined – the Customer- and the Internal Perspective. Each of them includes the appropriate business processes and activities, defined in eTOM and being mapped through the SLM Process. From the customer perspective the process starts with establishing the product offer and the leading requirements for the customer SLA. The goals of the customer centric process are the appropriate management activities and metrics for an optimal SLM process that includes the customer SLA's. The Service level monitoring process is internally fulfilled according to the predefined KPI's and KQI's of the Service Delivery and the Service Support processes. These processes involve the establishment, the management and the continual improvement of the corresponding UC's, internal SLA's and the relationship with the suppliers and the partners during the whole process.

Based on this model the optimal enterprise organisation can be developed in order to fulfil the processes needed without a redundancy or lack of workforces.

IV. Conclusion

This work is one of the series papers regarding the modeling of the ITIL Processes with eTOM Process elements. The IT industry needs not only to deliver high quality services to their customers but also to find the optimal business flows of the process and the optimal structure of the enterprise. This could have an adverse effect of the service prices and on the revenue as well.

The approach used in this work was applied also for modeling other ITIL processes such as Incident Management, Problem Management etc. and have led to positive results regarding the business goals of the enterprise that implemented these models.

References

[1] The Official Introduction to the ITIL Service Lifecycle, 2007.

[2] A. Nabiollahi, "Considering Service Strategy in ITIL V3 as a Framework for IT Governance", *The IT Service Management Forum*, UK , 2007.

[3] "ITIL Das Munich Institute for IT Service Management – mITSM", available on http://www.mitsm.de/.

[4] "ITIL Forum 2008 - Neupositionierung und Wertorientierung", available on http://www.itil-kongress-iir.de/.

[5] "ITIL process procedure templates", available on http://www.metocube.com.

[6] ITU-T – Series M3050x, Enhanced Telecom Operations Map (eTOM), 2004 – 2007.

[7] TMF GB921 v.4.0 – Addendums C,D,F,L – Enhanced Telecom Operation Map (eTOM) – The Business Process Framework, 2004 – 2007.

[8] TM Forum Business Benchmarking, Broadband Business Performance Study, Period: Q4 2007 through Q2 2008.

[9] TM Forum – GB929 v.3.2, Applications Framework (TAM) – The BSS/OSS Systems Landscape, 2008.

[10] ITIL V3 - Service Design - Management guide, Office of Government Commerce, first issue, 2008.

[11] eTOM and ITIL: A Powerful Combination for End-to-End Service Management, White paper, Hewlett-Packard Development Company, 2008.

Efficiency of NGN Interconnection Charging Methods

Aleksandra Kostić-Ljubisavljević[1], Vesna Radonjić[2], Vladanka Aćimović-Raspopović[3] and Snežana Mladenović[4]

Abstract – **Telecommunications networks are being upgraded from current generation circuit switched technology to Next Generation all-IP networks. These new networks will have lower operating costs and offer opportunities for new services. This paper reviews whether current interconnect charging principles or Bill and Keep will be more likely to promote dynamic and static efficiency gains when applied to voice and messaging services on NGNs.**

Keywords – **NGN interconnection, Efficiency, interconnection charging**

I. Introduction

Traditional electronic communications networks used for voice and related services employ circuit- switched technology and guarantee end-to-end quality of service. By contrast, the Internet employs packet-switched technology which, although less expensive to operate, does so on a "best efforts" basis with no guarantee of quality. To capture the efficiency benefits of the Internet, but also offer the quality benefits of traditional networks, the communications industry is developing Next Generation Networks (NGNs) capable of carrying voice and data to acceptable levels of quality depending on the consumer service.

While the transition to NGNs is at its early stages, regulators have already begun to assess its consequences for regulating interconnection charges between network operators for traffic which passes between their networks. A focus of the regulatory debates has been whether the bill-and-keep (BAK) interconnection charging model (whereby no interconnection payments are made for either termination or origination) is preferable to other charging models, and in particular to the initiating party's network pays model (IPNP, under which the originating network pays a termination fee to the terminating network). BAK is seen as having a key advantage over IPNP: by eliminating termination charges, because it saves regulators from the resource-intensive and often contentious task of setting termination charges. IPNP, on the other hand, is viewed by many regulators as leading to excessive termination charges requiring price caps.

This paper is organized as follows. After short introduction in section I, the literature review on NGN interconnection is presented in section II. Section II is dealing with "hot potato" problem of interconnection that is relating to termination of call/message/connection in NGN networks. In section IV are factors which determine the efficiency of an interconnection model presented and analyzed. Section V is related to welfare consequences of inefficient interconnection. At the end of paper conclusion is placed.

II. Literature Review on NGN Interconnection

There is little research on the charging principles most appropriate to NGN, though there has been some debate on the merits of BAK regardless of the underlying technology.

One of the first papers to study the economics of interconnection of all-IP networks is [1]. Author describes the Internet as having two distinct groups of subscribers, consumers and websites, who each gain value from more of the other side being present. Internet surfers gain more value if there are more websites of interest and websites gain value if there are more surfers. If a new surfer or website joins the network, there would be a positive benefit to those already connected to the network, but this benefit is not factored into the buying decision of the individual website or surfer. Therefore, in order to achieve the highest level of welfare across both sides of the market, one needs a pricing structure that encourages the highest number of users from both sides, which may be quite different for the two sides and need not necessarily relate to the costs caused by each party. The more flexible approach than simply charging at cost is required in all-IP network environments since it is extremely hard to calculate costs when multiple services are provided within a common network.

DeGraba in [2] proposed a BAK regime he termed Central Office Bill and Keep (COBAK). Central Office is the American equivalent of a local exchange. There was two rules proposed. First, the receiving party's carrier cannot charge an interconnecting carrier to terminate a call. Secondly, the calling party's carrier is responsible for the cost of transporting a call to the called party's central office. COBAK is proposed as a default rule for interconnection if carriers cannot agree on alternative terms in commercial negotiations.

The COBAK proposal is premised on three observations. First, that both parties generally benefit from a call, secondly that competition is more effective when carriers recover costs from their own customers and thirdly, that an arbitrage opportunity exists when regulation results in different charges being assessed for the same facility. The principle current

[1]Aleksandra Kostić-Ljubisavljević is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: a.kostic@sf.bg.ac.rs
[2]Vesna Radonjić is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: v.radonjic@sf.bg.ac.rs
[3]Vladanka Aćimović-Raspopović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: v.acimovic@sf.bg.ac.rs
[4]Snežana Mladenović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: snezanam@sf.bg.ac.rs

benefit of COBAK is that it "significantly reduces" the terminating monopoly problem. Other benefits are that it will lead to more efficient pricing and therefore more efficient usage and that it reduces the need for regulatory intervention.

Responding to [2], Wright sets out two problems [3]. First, that COBAK fails to internalise network externalities between calling parties and secondly its failure to apply Ramsey principles. Arguments on the first objection is that the calling party receives a direct benefit as a result of the called party being willing to accept the call and that this benefit is likely to be larger than that flowing in the opposite direction. If the calling party pays for the costs of the receiving the call by called party, this will result in an efficient transfer between the two types of callers. By imposing BAK, this transfer will be eliminated.

DeGraba responds to [3] rejecting his criticisms of the COBAK proposal [4]. The first criticism is rejected on the basis that artificially increasing origination charges by having the caller pay for termination may cause inefficient substitution of low cost wireline call by higher cost wireless calls. The second criticism is rejected on the grounds that the COBAK proposal will still impose a higher cost on the calling party even if the benefits of the call are shared equally by both parties.

III. The "Hot Potato" Problem

An argument often used against BAK is the "hot potato" problem which arises because communications providers have an incentive to hand over traffic to another network for termination as close to the point of origin as possible, thereby reducing their own costs and maximising the costs of the terminating network. If the terminating network is not able to recover these costs then, it is claimed, the terminating network will under invest.

The European Regulatory Group (ERG) suggests that the problem could be overcome by requiring operators to have a reasonable minimum number of interconnection points for BAK to be applicable to that operator. As discussed above this would involve the regulator in determining the topology of points of interconnection. The ERG then points out that if operators had to increase their network size to be BAK partners for other networks, the investment involved could be inefficient if infrastructures are unnecessarily duplicated [5].

The hot potato problem can be removed by making the originating network responsible for the costs of transport all the way to the terminating network's central office, including the cost of transit if involved [2]. This proposal maintains the incentive for the originating network to build an efficiently sized network and also maintains the customer-supplier relationship between the originating and terminating network and thus the incentive for investment as costs can be recovered from the originating network.

The choice of interconnection charging models will be critical to achieving the efficiency gains available from shifting to IP technology in networks that enable the network operator to manage Quality of Service (QoS). In particular, interconnection charges are important factors that impact the efficiency of service provision and network investment incentives.

IV. Factors which determine the efficiency of an interconnection model

There are a number of key factors which determine whether a given interconnection model is efficient, and most relevant are: externalities, network costs, stability of market conditions, and traffic balance [6].

Externalities and network costs

Messages are jointly consumed and thus their costs are jointly caused by both the receiving as well as the initiating party. The economic role of interconnection fees arises out of this fundamental characteristic. By initiating a message the initiating party causes an externality for the receiving party ("message externality"). Similarly, by deciding to join a network, a consumer may create an externality (a "network externality" or "subscriber externality") for other subscribers.

These externalities can be privately compensated between the end parties (e.g. by taking turns to call each other). If they arise "on-net" between customers on a single network, then retail charges can internalize them, with the effect of increasing total demand. However, where traffic flow "off-net" (i.e. between networks) achieving the right structure of retail prices to internalize externalities relies on payments at the wholesale level between network operators, because these payments signal retail pricing incentives. Through this mechanism interconnection fees enable externalities between retail customers to be internalized - although it is useful to bear in mind that this occurs at an aggregate level rather than for every individual message or subscription decision. Efficient interconnection charges can be derived from two elements: the nature of the externalities and the distribution of costs between the interconnecting networks.

IPNP will also provide a greater disincentive than the other major interconnection models to nuisance messages being sent. This results from the fact that IPNP can be implemented so that the full cost of the message being sent is recovered from the party initiating the message whereas alternative models recover some or all of this cost from the party receiving the message. This suggests that while all of the major interconnection models can support efficient message exchange in some circumstances, there are clear circumstances (that may arise relatively frequently in practice) in which IPNP will be superior to the other models. The academic and regulatory interconnection discussion has given little weight to the implications of the duality between messages for which, on average, compensation for externalities between retail parties is possible and messages where this is not the case.

Efficient compensation for externalities is also the basis for the efficient structure of transit charges. Where transit interconnection is required, consumers do not have a direct relationship with the transit network operator and the efficient balance of charges between the retail parties must be achieved

through transit payments between transit networks - which might take the form of cascading payments along the route taken by the message. Thus, even if transit is required, interconnection fees can be used to induce an efficient allocation of retail charges (subject to any practical restrictions on transit agreements such as those arising from the technical limitations of today's internet).

Traffic balance

If traffic between "peers" (i.e., networks which have the same cost structure and customer profile) is "balanced", then it might seem that the direction of payments is not relevant to efficiency, because any choice of interconnection fee would result in exactly the same (zero) net payment between networks. By avoiding the transactions costs of making offsetting payments, BAK could be efficient under these circumstances.

However, even if traffic between some peer networks appears balanced at a point in time, network operators typically have scope to influence this balance and their costs. Furthermore, the balance of traffic and/or costs is likely to be disturbed by evolving market conditions. Thus, whether BAK is indeed efficient in a situation of traffic balance depends on whether market factors could change the balance or peer status.

Stability of market conditions

Efficiency requires that the interconnection fee can be adjusted to respond to market changes. These changes can be due to exogenous factors, for example, the introduction of new services, which alter the typical distribution of benefits between the initiating and the receiving party. Change in market conditions can also be endogenous - that is, caused by the incentives established by the adoption of a specific interconnection model. In particular, network operators will have strategic incentives to avoid costs or to favourably alter the traffic balance where those actions do not change their interconnection payments.

It is clear that an efficient interconnection model requires flexibility to respond to both exogenous market changes as well as strategic actions by network operators. BAK, which by definition implies an interconnection fee which is always equal to zero, can only be efficient if the market conditions are such that the efficient interconnection fee is equal to zero (i.e. either when traffic is balanced or along the diagonal in Fig. 1) and if these conditions are stable.

V. Welfare consequences of inefficient interconnection

Interconnection charges can cause inefficiencies due to a mismatch between an operator's incremental revenues and incremental costs associated with interconnection. One way that an operator can react to a mismatch is by adjusting its retail pricing model in order to increase its revenues through retail payments, which would reduce consumer welfare to the extent that the change in the retail prices departs from the efficient retail model. Alternatively (or in addition), an operator may react to inefficient interconnection charges through a number of cost-avoidance strategies.

Business bias results from a customer-group-specific mismatch between a network's incremental interconnection costs and incremental retail revenues. For example, a terminating network that is not able to cover at least its marginal costs from termination would have an incentive to target customers who initiate more traffic than they receive (e.g., outbound telemarketers) and avoid customers who mainly receive messages (e.g., inbound call centres).

Freely negotiating networks would deter inefficient business bias by linking interconnection fees to overall traffic profiles. However, customer targeting can be socially harmful. First, targeting tends to leave some customer demand unserved or underserved: each message requires origination and termination, and a disincentive to provide one of these services (by avoiding customers that have a relatively high propensity to either originate or terminate messages) would tend to suppress traffic below its socially optimal level. Second, as a result of the targeting, traffic might be carried by the networks that are better able to bias their business, rather than the most efficient networks.

Network structure bias (the "hot-potato problem") occurs when investments in particular network elements are not fully rewarded through interconnection fees and where operators have the ability to determine the point of interconnection. For example, if interconnection fees do not respond to the costs that networks incur, then an operator would have the incentive to reduce its own costs by locating points of interconnection close to its own customers. This could then lead to inefficient network design (e.g., underinvestment in the trunk network) or inefficient network operation (e.g., inefficient vertical separation between network elements, so that more traffic is carried as transit).

All the distortions discussed above result from the same principle - a mismatch between incremental costs and incremental revenues associated with specific interconnection services. As IPNP, BAK, and RPNP represent a continuum of interconnection fees, similar (at least in a qualitative sense) inefficiencies can arise under each model where they are applied in circumstances that are not consistent with efficient message exchange for that model.

BAK is efficient in two specific situations: where traffic is evenly balanced between peers and this balance cannot be disturbed by either strategy or change in market conditions or where traffic is not evenly balanced but the distribution of retail benefits exactly matches the distribution of costs between the originating and terminating networks. If BAK is applied outside these circumstances, it will lead to inefficient traffic and subscription decisions and/or to strategic cost-avoidance behaviour [7].

The direct effect of BAK is to change the way in which an operator's costs are recovered. In particular, BAK prevents the terminating network from receiving revenues from the initiating party's network and thus the only source of revenue to the terminating network is its own retail customers. The operator of the terminating network has several options to

respond to the implementation of BAK. It could seek to recover its costs from the receiving party through a fee for receiving the message. While this option solves the cost recovery problem from the perspective of the network operator, it will lead to inefficient retail prices except in situations mentioned above. For example, if the receiving party usually rewards the initiating party for initiating messages from which it benefits (e.g. through monetary transfers, consumption decisions, or social interaction) but has less scope to punish parties who initiate nuisance messages, then imposing BAK will not improve the incentives to send socially beneficial messages. However, BAK will encourage the distribution of socially undesirable nuisance traffic as is illustrated by the experience in the USA where termination rates are set to zero or at relatively low rates and where customers are charged for receiving as well as sending text messages. As we discussed earlier, in this situation IPNP would achieve a more efficient traffic mix.

Another option to compensate for the absence of termination fees is to increase the price for bucket plans, flat rates ("all-you-can-eat" plans), or fixed access fees. The effects of such price adjustments on traffic depend on whether the price increase explicitly or implicitly affects the volume of messages initiated and received [8].

If the price increase affects a customer's incremental costs of receiving messages (e.g., through paying a higher fixed access charge in order to receive more messages) then traffic distortions are similar to those that would result from directly charging for receiving messages. If the price adjustment increases the customer's incremental costs of initiating messages (e.g., through tightening of bucket limits on initiating messages), this would distort price signals, because the parties who bear the termination cost of a particular message are involved in that message neither at the initiating nor the receiving end. As a consequence, this model would generally lead to inefficient traffic and subscription decisions.

In an NGN environment, BAK may also lead to operators terminating off-net traffic only at low-quality levels and reserving high-quality capacity for on-net traffic -that is, for traffic that generates revenues for QoS provision. Regulators might attempt to prevent this reaction by prescribing quality levels at which interconnection must occur. However, operators could try to mitigate such a requirement by reducing or delaying their efforts to prioritize traffic or to invest in other QoS capabilities in the first place. In a QoS-differentiated scenario, the inflexibility of interconnection fees under BAK would also aggravate network structure bias: operators would have an incentive to offer higher quality services while avoiding associated additional costs by locating interconnection points closer to their own customers [9].

IPNP will tend to be efficient in the following scenarios:

- all benefits of the message accrue to the initiating party;
- the share of the benefits accrued to the receiving party is small compared to share of network costs incurred by the terminating network; and/or
- benefits to the receiving party are mostly relevant in situations where individuals interact repeatedly or where a monetary transaction between the initiating and receiving party accompanies each message.

The scope of these scenarios is broader than the specific nature of circumstances in which a zero interconnection fee (i.e. BAK) is efficient. This is simply a result of the fact that IPNP encompasses a range of interconnection fees while BAK represents a single interconnection fee. Outside of the circumstances listed above and ignoring externalities (e.g. network externalities), IPNP is not efficient.

VI. Conclusion

There is no single interconnection model that is efficient across all circumstances. In order to enhance economic efficiency regulators should intervene at the network layer that is closest to the market failure. The charging principles applied to any new service should be decided upon taking into account which party is likely to derive most benefit from a call/message and may include IPNP, RPNP and BAK. Interconnection settlements depend on externalities, network costs, stability of market conditions, and traffic balance.

Acknowledgement

This work is partially supported by the Ministry of the Science and technological development of the Republic of Serbia with the project TR32025.

References

[1] K. Yoon, "Interconnection Economics of All-IP Networks", *Review of Network Economics*, vol. 5, no. 3, 2006.

[2] P. DeGraba, "Bill and Keep at the Central Office as the Efficient Interconnection Regime," *OPP Working Paper Series* no. 33, FCC, 2000.

[3] J. Wright, "Bill and Keep as the Efficient Interconnection Regime?", *Review of Network Economics*, vol. 1, pp. 54-60, 2002.

[4] P. DeGraba, "Bill and Keep as the Efficient Interconnection Regime?: A Reply", *Review of Network Economics*, vol. 1, pp. 61-65, 2002.

[5] Project Team on IP-Interconnection and NGN, ERG (07)09, (2007), available at: http://www.cmt.es/es/publicaciones/anexos/ERG%2807%2909_rept_on_ip_interconn.pdf.

[6] M. Dodd, A. Jung, B. Mitchell, P. Paterson, and P. Reynolds, "Bill-and-Keep and the Economics of Interconnection in Next-Generation Networks", *Telecommunications Policy*, vol 33, pp. 324–337, 2009.

[7] V. Radojičić, A, Kostić-Ljubisavljević, "Inteconnection of Next Generation Networks", *Proceedings of XXVII PosTel*, pp 341-350, 2009.

[8] R. Cadman, *NGN Interconnection: Charging Principles and Economic Efficiency*, NGNuk, London, 2007.

[9] J. S. Marcus, *Interconnection on an IP-based NGN Environment, Global Symposium for Regulators*, Dubai, United Arab Emirates, 2007, available at: http://www.itu.int/ITU-D/treg/Events/Seminars/GSR/GSR07/Documents_presentations/Session_III%20Scott%20Marcus_interconnect.pdf.

Review of Some Interconnection Charging Models

Aleksandra Kostić-Ljubisavljević[1], Vesna Radonjić[2], Vladanka Aćimović-Raspopović[3] and Valentina Radojičić[4]

Abstract – **In this the review of two most used interconnection charging models is presented. We focus on two representatives of paper bottom-up and top-down approach in creating model for interconnection calculation. As a typical bottom-up model the World Bank and European Commission created models based on Long run average incremental cost. On the other hand ITU created COSITU model based on top-down approach.**

Keywords – **interconnection charging, bottom-up, top-down**

I. INTRODUCTION

The economic cost of interconnection is generally the starting point in establishing economically efficient interconnection prices. In many jurisdictions, regulators set interconnection prices based on Long Run Incremental Costs (LRIC). There are numerous methods of estimating LRIC. Approaches to modelling LRIC can be broadly categorized as bottom-up and top-down modelling approaches. Bottom-up models include scorched earth or scorched node methods. The whole question of tariffs is crucial to the development of telecommunications, since it is tariffs that will mercilessly make or break anyone setting out in this sector. Negotiating tariffs or rates is hence a delicate matter, whether it is for a new operator entering a liberalized market or a regulator wishing to set affordable tariffs for national calls without compromising competitiveness among operators. Many other questions may be raised in this regard. Various cost concepts exist and are formulated in models (LRIC, LRAIC, FLEC, TELRIC, TSLRIC, CCA, FDC, etc.). Each concept presupposes the availability of a quantity of data without which the results obtained would be no more than vague estimates, however complex the models used.

This paper is organized as folows. After the introduction in section II is Word Bank/European Comission model presented. The major sections of that model are presented in section III. Section IV and V are related to ITUs model COSITU.

[1]Aleksandra Kostić-Ljubisavljević is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: a.kostic@sf.bg.ac.rs

[2]Vesna Radonjić is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: v.radonjic@sf.bg.ac.rs

[3]Vladanka Aćimović-Raspopović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: v.acimovic@sf.bg.ac.rs

[4]Valentina Radojičić is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: valentin@sf.bg.ac.rs

II. WORLD BANK/EC MODEL

The models produced by the World Bank and European Commission (EC) are very similar concerning costing principles and functionality. They both present an adaptable bottom-up model built under a "scorched node" assumption. A "scorched node" assumption requires that the model be constrained by the incumbent's existing switching centres [1]. The model created by the EC was freely available at [2].

Bottom-up models rely on a series of economic/engineering assumptions to dimension a network capable of meeting the busy hour demand in a network. Demand includes existing billed minutes and call attempts, but also other calls on the network such as call set-up time, unsuccessful call attempts, margins for growth and an allowance for capacity utilisation.

The costs produced from a bottom-up LRIC model should approximate those costs that would be incurred by a new entrant investing in a network today that could carry the incumbent operator's traffic. In other words, the costs produced are those that would be incurred were an operator to rebuild the incumbent operator's network, using existing switching centres.

As bottom-up models assume that the capital and operating expenditure required to meet demand is new, there should not be any inefficiencies in the network. This is strength of the approach, since the efficient levels of costs used in the model should provide an incentive to operators to operators to increase their own efficiency [3].

A bottom-up LRIC model provides considerable scope for debate about the type of network that would evolve if an incoming operator were re-build the incumbent's network (subject to the "scorched node" constraint). This debate is most evident in two areas — the optimisation of switching nodes and the configuration of the transmission network.

Illustrative results are provided in [4] for those counties where the model was tested. These results are based on data provided to the project team and, where these data were incomplete, on assumptions made by the project team. The results have not been subjected to detailed scrutiny by the incumbent operators in the relevant Member States and are provided as an illustration of the outputs of the model rather than as fully tested estimates of efficient levels of interconnection cost.

This model could be tested according the robustness of the results in two main ways. First, by identifying the variables in the model which have the greatest influence on the results. Sensitivity analysis is then conducted on these variables to provide the reader with a sense of how the results vary with changes in key assumptions. Second, by reconciling the results of the model against the results produced by a top-down model and by an independent bottom-up model

III. WB/EC MODEL SECTIONS

The model is made up of four closely inter-related sections — optimisation, network assumptions, costs, and calculations. These sections then produce results, both in terms of the unit cost of network elements and the interconnection services defined by the European Commission.

A. Optimisation

This model adopts the "scorched node" assumption, which means that we model the incumbent's current switching centres. This assumption, however, can be modified by changing the nature of the equipment at an existing location. The starting point in the model, therefore, is to identify the total number of "nodes" or switching centres and then to determine the optimal mix of those nodes. This process is referred to as "optimisation" and can be undertaken in two different ways in the model [5]. First, NRAs can request a node database from their incumbent operator, which lists each node and the total number of PSTN equivalent lines connected to that node. Once a threshold size of a remote concentrator is identified, the model will assign the most appropriate equipment to each node. Second, NRAs can use a ratio-based approach which determines the number of nodes required to serve a Member State with a given surface area and population density. The model then builds up the number of remote concentrator units (RCUs), local switches and tandem switches. These approaches can only approximate the optimal mix of nodes, since this will depend partly on specific geographical features. The actual mix will also depend on a number of factors such as the extent of digitalisation, whether incumbents offer other services such as cable television, and the view of the incumbent of the optimal length of the local loop. A third option for the users of the model is to ignore the optimisation options provided and to include in the model their view of the efficient number of nodes, based no doubt on discussions with incumbents and other operators as well as their assessment of the level of optimisation that has already taken place.

B. Network Assumptions

Once the approach to optimisation has been determined, the network needs to be dimensioned to meet the demand for the network. This is determined by measuring the existing demand for the network and adding the unbilled share of traffic such as call set-up time and unsuccessful call attempts, and margins for growth. Capacity utilisation is also taken into account at this stage.

Once existing demand has been adjusted to include the above factors, the total demand is attributed to each switching and transmission elements using "routeing factors". Routeing factors show how intensively each network element is used for each type of call. For example, a local call may, on average use less than one RCU, between one and two local switches, and less than one tandem switch. The usage of the transmission network is determined in a similar manner or can be derived from switching routeing factors.

The network needs to be dimensioned in order to carry the conventional "busiest hour" level of traffic. It is not dimensioned to carry an unrepresentative surge.

C. Calculations

The dimensioned traffic in the network is then used to determine the number and mix of equipment required. The number of busy hour call attempts, for example, is used to determine the variable cost of the processor. Busy hour minutes are used to determine the number of ports and size of the switchblock for switches, as well as the demand for circuits in the transmission network [6]. Other major determinants of network equipment in the model are the following:

- the number of nodes (which determine the site costs and the fixed cost of the processor);
- the number and size of leased lines (which affects demand for circuits in the transmission network); and
- the length of duct in the network (which affects the costs of infrastructure).

Other factors that will affect the results calculated through the model include routeing factors, the extent of duct sharing in the network, and the depreciation methods used by incumbents. Model users can conduct sensitivity analyses to determine the robustness of the model to changes in many of those variables.

D. Costs

The model then applies a unit cost to each category of equipment required in the network to meet the dimensioned demand. This provides the total forward-looking investment costs which are then annualised to determine the capital charge for the year. The costs in the model should been collected from a number of sources including telecom operators in the EU and from publicly available models. In some cases, we have been required to use own judgement backed by engineering advice.

It is not intended that the costs in the model be viewed by NRAs as any form of best practice costs. They have been provided in the model as defaults that could be used in the absence of specific data to provide a broad indication of individual costs in the network. There are likely to be significant differences as operators in some countries will be able to access volume related and other discounts and we would urge NRAs to engage interested parties in a consultation process to refine the cost estimate provided as defaults.

Other costs included in the model are:

- network related operating costs;
- non-network capital and operating costs;
- cost of capital;
- working capital.

IV. COSITU

COSITU is an example of a top-down cost model. COSITU is based on enhanced fully distributed costing principles, as adopted in the ITU-T D series of recommendations [7].

COSITU is a practical tool from ITU's Financing Strategies Unit to automate: the calculation of costs, taxes related to the exchange of international traffic (accounting, settlement and termination rates), interconnection rates between local operators, and tariffs for national and international telephone services taking into account the impact of Universal Service Obligations decided by public authorities.

This software can be applied to both fixed and mobile services. COSITU requires the following input data: investment and expense data from accounting systems, current cost data to convert historical capital asset costs to current costs. For example, when accounting records report the purchase price of a switch, the model calculates the cost of the switch at current purchase prices, inputs for depreciation and cost of capital. Where the inputs needed to estimate the cost of capital are not available, COSITU benchmarks these to countries or firms of comparable risk, and traffic demand and routing data.

COSITU produces unit costs and prices for international, sub regional, and regional calling. COSITU can account for the effects of universal service funding, taxes and any access deficit as mark-ups over current unit costs, to calculate interconnection prices.

To the extent that accounting and demand data are available, COSITU's basic modelling framework can be used to model interconnection costs for both fixed and mobile networks. COSITU embodies the following principles:

Transparency: Information used in the cost derivation process should be openly available, so that external analysts can comprehend the final rate.

Practicality: The demands of the costing methodology with respect to data availability and data processing should be reasonable, to keep the costing exercise economical yet still useful.

Causality: The model should demonstrate a clear cause-and-effect relationship between service delivery, on the one hand, and the network elements and other resources used to provide the service, on the other hand, taking account of relevant cost determinants (cost drivers).

Contribution to common costs: The cost calculation should provide for a reasonable contribution to common costs.

Efficiency: The cost calculation should provide a forecast of cost reductions that are likely to result from more efficient use of resources over time.

V. CLASSIFICATION AND DEFINITION OF SERVICES FOR WHICH COSITU CALCULATES COSTS

The classification of the services for which COSITU calculates the cost is given in Figure 1.

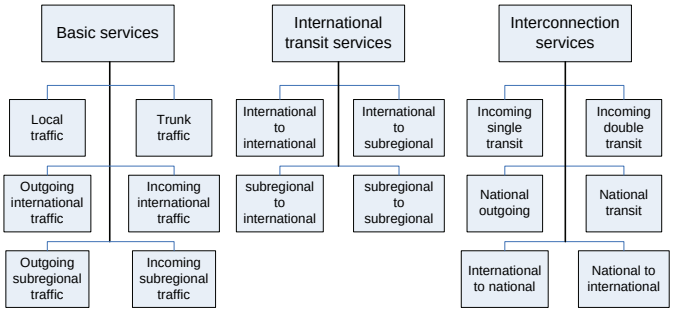


Figure 1 Classification of the services in COSITU model

Local/Urban: Traffic carried solely within the network of the operator for which the calculations are made, between users located in the same local charging area.

Trunk/Interurban: Traffic carried solely within the network of the operator for which the calculations are made, between users located in different local charging areas.

Incoming international: A call from a user located outside the national boundaries to an end user connected to the network of the operator using the international gateway.

International outgoing: A call from an end-user connected to the network of the operator using the international gateway to a correspondent located outside the national boundaries.

Outgoing sub regional: A call from an end-user connected to the network of the operator using the international gateway to a correspondent located outside the national boundaries, in a country which can be accessed by terrestrial media that are also used for trunk calls.

Sub regional incoming: A call from a user located outside the national boundaries, in a country, which can be accessed by terrestrial media also used for trunk traffic, to an end- user connected to the network of the operator using the international gateway.

International to international: A call between two non-sub regional international correspondents via the international gateway of the operator for which the calculations are made.

International to subregional: A call from a non-sub regional international correspondent to a sub regional correspondent via the international gateway of the operator for which the calculations are made.

Subregional to international: A call from a sub regional correspondent to a non-sub regional international correspondent via the international gateway of the operator for which the calculations are made.

Subregional to subregional: A call between two sub regional correspondents via the international gateway of the operator for which the calculations are made.

International to national: A call from an international correspondent to an operator without an international gateway located within the same political borders as the operator using the international gateway for which the calculations are made,

National to international: A call from an operator without an international gateway located within the same political borders as the operator using the international gateway for which the calculations are made, to an international correspondent.

Outgoing national: A call from an end-user of the network of the operator for which the calculations are made to another operator located within the same political borders as the first operator.

Incoming national, single transit: A call coming from the network of another national operator to an end-user located in the charging area of the interconnection point and connected to the network of the operator for which the calculations are made.

Incoming national, double transit: A call coming from the network of another national operator to an end-user located outside the charging area of the interconnection point and connected to the network of the operator for which the calculations are made.

National to national: Transits call between two national operators via the network of the operator for which the calculations are made.

VI. THEORETICAL ASPECT OF COSITU

COSITU can accommodate both Bottom Up and Top Down approach of calculating the cost of network components, the initial stage for the bottom-up method being completed outside the model.

Whatever the methods used to determine costs and traffic, the COSITU model can accommodate them.

COSITU has, however, been optimized for use of real information from the accounts and technical data of real network operators with a view to equitable allocation of costs to the services that generate them, collectively or separately. COSITU is unaffected by technological choice, addressing directly the services sold - retail or wholesale.

Adjusted depreciation

- Linear depreciation is the rule most widely applied in the accounts of telecommunication operators.
- It is nevertheless possible to take account of the natural evolution of the price of equipment in the specific market and adjust the depreciation accordingly.
- Currency depreciation must also be taken into account:

Efficiency is calculated by combining the installed capacity; utilized capacity; average annual growth rate in number of subscribers; replenishment period.

COSITU is able to calculate the Cost of Capital, assuming a preponderant risk of inflation for telecommunication companies in developing countries, the essential components of the cost of capital as adjusted to local conditions.

The routing table is an essential instrument for cost-orientated charging. It allows allocation to every service, according to the intensity of demand it places on each one, part of the resources needed for its production. COSITU uses traffic volume (adjusted by the geographical correction coefficient) for network component cost allocation. Based on the routing table, COSITU allocates to services their share of each cost component. The corresponding real traffic volume divides the resulting cost of a service in order to obtain the unit cost of the service. At this stage, the COSITU server allows an online comparison with other telephone network operators.

In addition to calculating per minute service and network element costs, COSITU computes tariffs based on cost data, taking into consideration the following factors:

- Corporation tax;
- Contribution to a Universal Service Obligation (USO) fund;
- Effect of Universal Service Obligation (USO) policies on Access Deficit.

COSITU fosters consensus building among policy makers, national regulatory authorities and operators with respect to tariffs. Both cost-based and cost orientated tariffs can be calculated. COSITU offers market actors a practical means to settling disputes.

VII. CONCLUSION

In this paper the short review of two most used models for interconnection charging are presented. First one, with two possible forms, was created by the World Bank and European commission. It is built based on bottom-up strategy. The second, COSITU, is created by the International Telecommunication Union and it is based mostly on top-down approach.

ACKNOWLEDGEMENT

This work is partially supported by the Ministry of the Science and technological development of the Republic of Serbia with the project TR32025.

REFERENCES

[1] Study on The Preparation of an Adaptable Bottom-Up Costing Model for Interconnection and Access Pricing in European Union Countries: A Final Report for Information Society Directorate-General of the European Commission by Europe Economics, April 2000.

[2] http://www.ispo.cec.be/infosoc/telecompolicy/en/Cost_model_2000.xls.

[3] P. Noumba, C. Rudelle, L. Simon, L. Gille , *A Model for Calculating Interconnection Costs in Telecommunications*, World Bank, 2004.

[4] V. Radojičić, A. Kostić-Ljubisavljević: "Odabrani modeli obračuna troškova interkonekcije", *XXIV Simpozijum o novim tehnologijama u poštanskom i telekomunikacionom saobraćaju, PosTel 2006*, Zbornik radova, Beograd, str.335-344, 2006.

[5] http://ec.europa.eu/archives/ISPO/infosoc/telecompolicy/en/lricexsum.pdf.

[6] A. Kostić-Ljubisavljević, V. Radojičić, and V. Aćimović-Raspopović: "An Implementation of Adaptable Bottom-up Model for Calculation Interconnection Costs", *Proceedings of XLII ICEST 2007*, vol.1, pp. 177-180, Ohrid, Macedonia, 2007.

[7] Question 6-1/1, Report on Interconnection, ITU-D Study group 1, 3rd study period, (2002-2006).

Analyzing the Network Real-Time Multimedia Traffic Profile Based on Content

Antoaneta A. Popova, Ivo R. Draganov, Vladimir K. Poulkov, Alexander A. Krupev[1]

Abstract – **In this paper an analysis is presented concerning the network real-time multimedia traffic profile based on content. The throughput, round trip time and packets' size are analyzed when slow, medium and fast changing videos with low, medium and high resolution are being streamed over an Internet channel. The results obtained reveal number of useful directions for enhancing video streaming systems' operation.**

Keywords – **Multimedia Traffic, Real-Time Video Streaming, Throughput, Round Trip Time, Packet Size.**

I. Introduction

Real-time video and audio streaming became extremely popular in the last 15 years. At first low bit rates were incorporated for the separate flows – based on low resolutions and frame rates with higher quantizing factors. With the increase of the network bandwidths more and more multimedia flows become available for simultaneous reception along with the wide variety of other services. This leads to a situation where the network traffic profile even at the very user entry point becomes very complicated and in number of cases prevents some of the services to be used in accordance with some preliminary expectations.

In 2004 Pinson and Wolf [1] suggested a standard way for measuring video quality of video which could be streamed among other ways for transmission. Based on such measurements a lot of useful techniques for enhancing the quality of the receiving video were proposed such as dynamically resource allocation based on traffic prediction [2], synchronization control schemes [3] for video conferencing and preliminary defined QoS constraints approach [4].

So far most of the research for video and audio quality degradation when streamed over a network is based assuming a fixed size video usually at low resolutions and slow changes in the scenes [5-7]. Only recently some really extensive research on incorporating simultaneous video and audio streams with vast dynamic ranges along with other services was made [8, 9]. In this paper we try to propose an investigation on the network traffic profile based on different multimedia content. The paper is organized as follows. In part 2 a description of the test environment is given, in part 3 – the experimental results are shown and part 4 is a conclusion.

[1]Antoaneta A. Popova, Ivo R. Draganov, Alexander A. Krupev, and Vladimir K. Poulkov are with the Faculty of Telecommunications, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria, E-mail: {antoaneta.popova, idraganov, vkp}@tu-sofia.bg, asfalot@abv.bg

II. Test Environment Description

A. Network Set Up Configuration

The network set up for the performance evaluation of the video streaming throughput and other statistical network data is given in Fig. 1.

The local network connection from the gate to the host is 100 Mbps Ethernet over which the Internet traffic is passing at speed no less than 64 Mbps without any other clients on the same sub-network after the local router.

The local host is equipped with 100/1000 Mbps network interface card over ASROCK P43 Motherboard (CPU Quad Core® 2.8 GHz, 4 GB RAM) with 750 GB HDD running at 7400 rpm and VGA ATI Radeon 4800 Series. The operating system is MS® Windows ® XP® SP2 64-bit. As a local multimedia player VLC v.1.1.7 is used, as a bandwidth limiter NetPeek v.1.1 and as a network monitor Wireshark v.1.4.4.

All the traffic is measured (captured) at two dialing speeds which are proven to be standard in practice – 768 and 2500 kbps respectively.

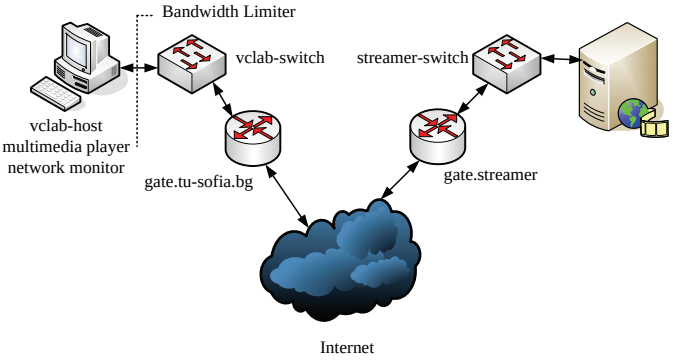


Fig. 1. Network Set Up Configuration

B. Test Videos

There are 3 types of videos used in the experimentation – a slow motion video with a news reader being filmed in a studio with no sharp moves in the scenes, a medium motion video showing football players moving with no too fast moves and a fast motion video – representing car racing in a city with lots of sharp moves, zooming in and out in large amount of details in the picture, large scale pan and tilt moves of the camera. Each video is in low, standard and high resolution.

The low resolution videos (400 x 226 pixels) are H.263 coded, while for the standard (854 x 480 pixels) and high resolution (1280 x 720 pixels) videos MPEG-4 AVC (Advanced Video Codec) coder is used all with variable bit rate option. The frame rate for all the videos is fixed to 29.970

frames per second. The color space is YC_bC_r 4:2:0 with 8 bits per component representation and the scan is progressive.

The audio interleaved along with the video for each clip is AAC (Advanced Audio Codec) coded consisting of 2 channels (stereo) sampled at 44.1 kHz in variable bit rate mode.

C. Test Procedure

The test procedure follows the steps given below:

1. Set up a speed limit of 768 kbps.
2. Receive slow motion video at low resolution.
3. Capture all the traffic at the local host.
4. Calculate the throughput, the round trip times and the packet lengths distribution.
5. Change the resolution to standard (medium).
6. Repeat steps from 2 to 4.
7. Change the resolution to high.
8. Repeat steps from 2 to 4.
9. Change the video to medium motion.
10. Repeat steps from 2 to 8.
11. Change the video to fast motion.

Repeat steps from 2 to 8.

III. EXPERIMENTAL RESULTS

It is obvious and expected that with the decrease of the speed the temporal change of the throughput (Fig. 2 – Fig. 5) is more and more wide meeting a certain limits that lead to transitory interrupting in the video being received. For the dialing speed of 2500 kbps (Fig. 3) this doesn't happen although for the fast motion video with high resolution certain saturation in the top limit is visible (Fig. 5).

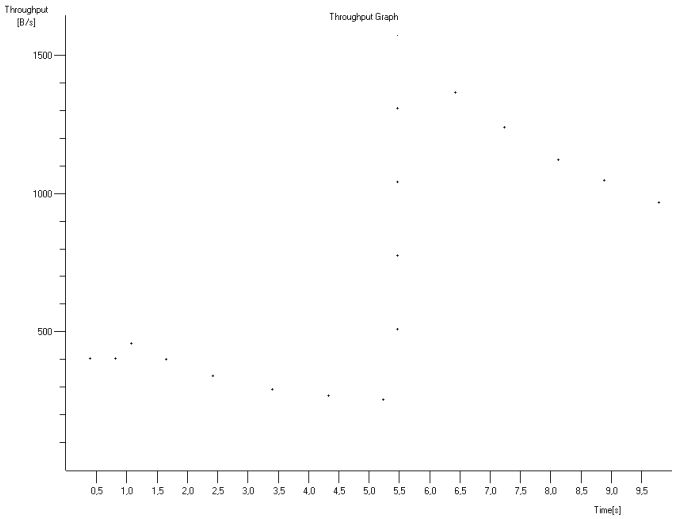


Fig. 2. Throughput for slow motion video with low resolution at 768 kbps limit

Interestingly there are number of cases in which for standard resolution and slow and medium motion the throughput is virtually constant but changing its average value from video to video.

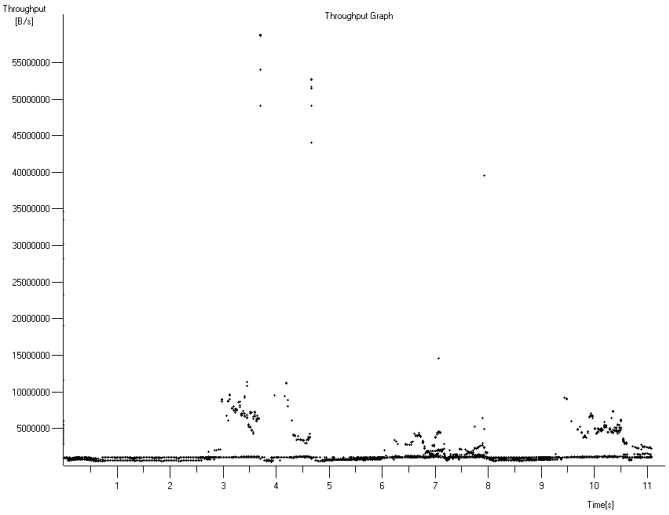


Fig. 3. Throughput for slow motion video with high resolution at 2500 kbps limit

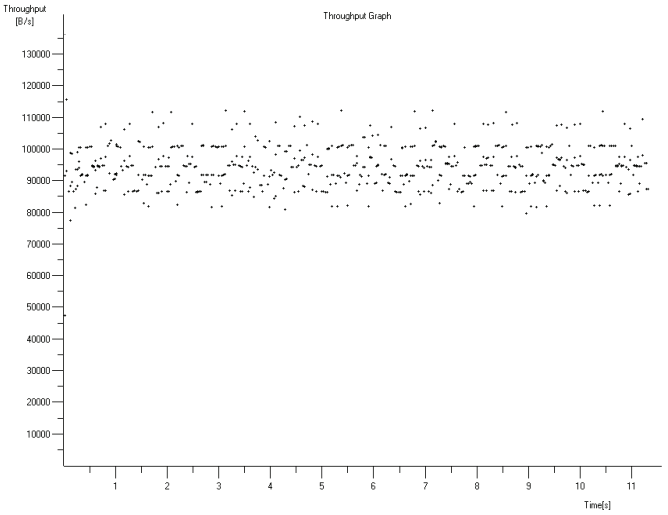


Fig. 4. Throughput for fast motion video with low resolution at 768 kbps limit

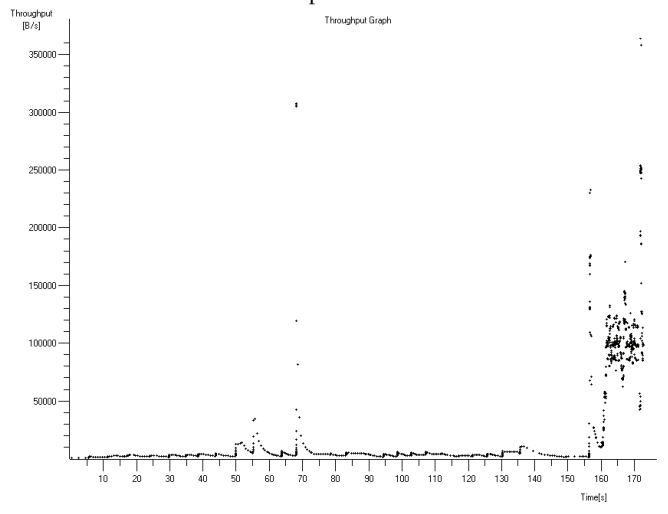


Fig. 5. Throughput for fast motion video with high resolution at 768 kbps limit

This property can be utilized in a wide range of practical cases where simultaneous transmit is needed for more than one multimedia streams and proper statistical scheme of time division could be applied saturating the channel to an extent such that interrupting can be minimized.

As for the round trip times (Fig. 6 – Fig. 9) the same behavior in the dynamic range change is observed for the slow and fast motion videos. This change ia a little bit unexpected for the medium motion video no matter of the resolution – a narrow wide range reveals the flat law for the distribution times for the packets. This could be of great importance when again multiple multimedia streams are going to be transmitted and statistically it is much easier to calculate the total amount of the bandwidth reserved for each of them in some timely manner. Actually the uniform distribution in such cases leads to simple linear equations for the time division intervals being estimated for each video to be streamed down.

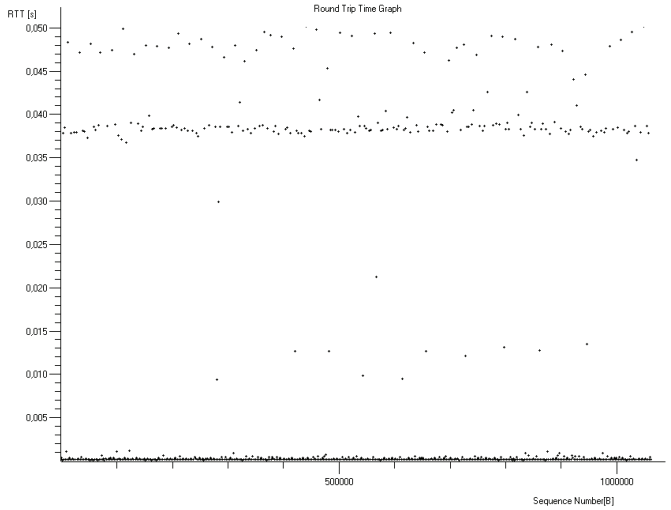


Fig. 6. Round trip time for slow motion video with low resolution at 768 kbps limit

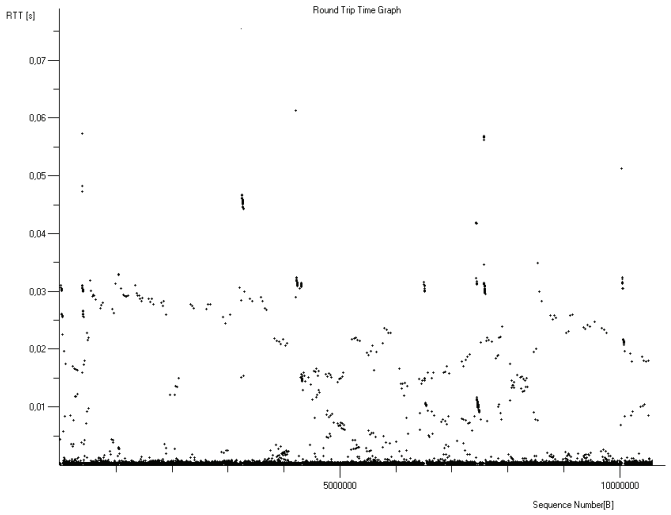


Fig. 7. Round trip time for slow motion video with high resolution at 2500 kbps limit

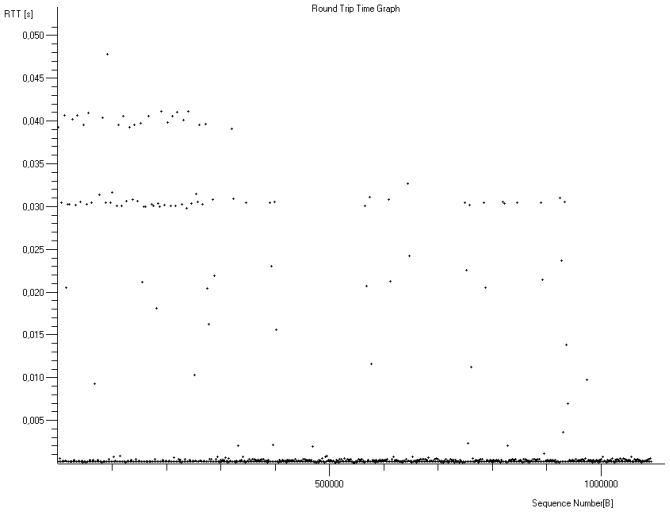


Fig. 8. Round trip time for medium motion video with standard resolution at 768 kbps limit

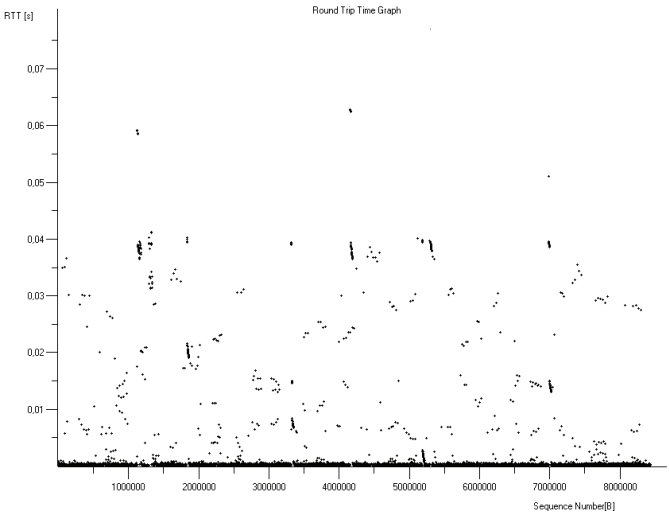


Fig. 9. Round trip time for fast motion video with high resolution at 2500 kbps limit

The packet lengths (Fig. 9 – Fig. 11) vary from about equally divided, in the ranges from 80 to 160 Bytes and from 160 to 320 Bytes for the low resolution videos (Type 1). This tendency occurs no matter of the speed of the scene changes which is then slightly changed, in the direction to change this ratio from 1:1 to 1:2 for both the ranges for the medium (standard) resolution videos.Then it goes up to even 1:3 for the high resolution videos (Type 6). A number of packets – between 7 and 12 %, for the high resolution video especially at faster scene changes, fall into the range from 40 to 80 Bytes which probably is a consequence of the saturations being observed for this particular cases in the bandwidth and due to the protocol specifications some considerable amount of data is being split into smaller packets for transmission. Further study of such behavior could be undertaken which might prove useful in cases where a small amount of users are being supplied with fast videos through narrow links perhaps in the last mile to their locations.

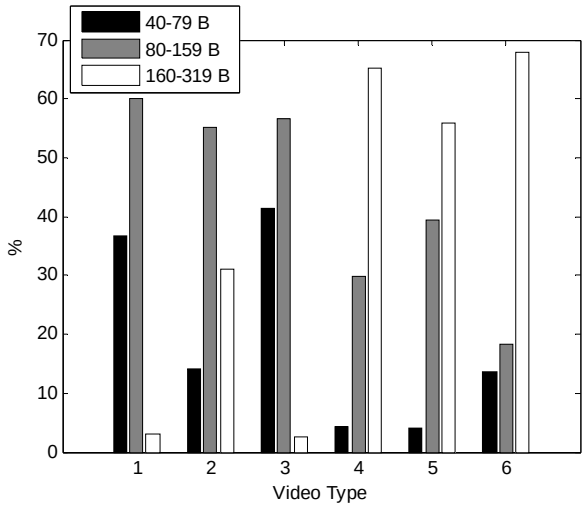


Fig. 9. Packet lengths distribution for slow motion video

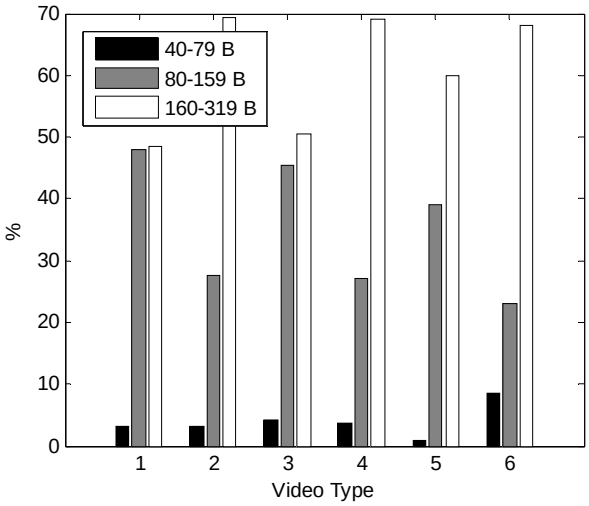


Fig. 10. Packet lengths distribution for medium motion video

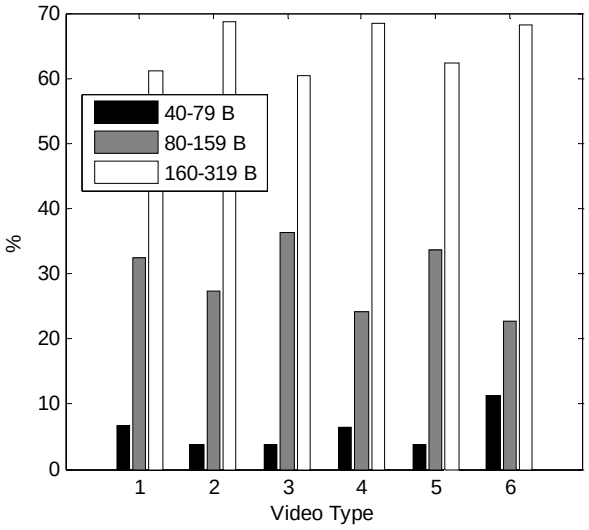


Fig. 11. Packet lengths distribution for fast motion video

IV. Conclusion

In this paper an analysis concerning the network real-time multimedia traffic profile based on content changing from slow to fast motion and at different resolutions is presented. The throughput, round trip time and packets' size are analyzed which reveal useful directions for network bandwidth planning when multiple clients should be supplied with high quality multimedia resources. Specifically the packet size distributions show what kind of other types of services could be incorporated along with the multimedia streaming in such a way that not exceeding the bandwidth limits no interference will occur based on this size in the process of packet flow control.

Acknowledgement

This paper was supported by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science (Contract – DDVU 02/13 – "Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization").

References

[1] M. Pinson, S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality", *IEEE Transactions on Broadcasting*, vol. 50, no.3, pp. 312-322, 2004.

[2] P. Koutsakis, "On Providing Dynamic Resource Allocation Based on Multimedia Traffic Prediction in Satellite Systems", *Computer Communications*, vol. 30, issue 2, pp. 404-415, 2007.

[3] I. Bartoli, G. Iacovoni, and F. Ubaldi, "A Synchronization Control Scheme for Videoconferencing Services", *Journal of Multimedia*, vol. 2, no. 4, 2007.

[4] I. Awana, S. Ahmad, and Bashir Ahmad, "Performance Analysis of Multimedia Based Web Traffic with QoS Constraints", *Journal of Computer and System Sciences*, vol. 74, issue 2, pp. 232-242, 2008.

[5] L. De Cicco, S. Mascolo, and V. Palmisano, "Skype Video Responsiveness to Bandwidth Variations", *In Proceedings of NOSSDAV'08*, pp. 81-86, Braunschweig, Germany, 2008.

[6] S. Dasguptaa, J. de Oliveiraa, and J.-P. Vasseur, "Dynamic Traffic Engineering for Mixed Traffic on International Networks: Simulation and Analysis on Real Network and Traffic Scenarios", *Computer Networks*, vol. 52, issue 11, pp. 2237-2258, 2008.

[7] Y. Lu, B. Fallica, F. Kuipers, R. Kooij, and P. Van Mieghem,"Assessing the Quality of Experience of SopCast", *International Journal of Internet Protocol Technology*, vol. 4, no. 1, pp. 11-23, 2009.

[8] K. Sharma, H. Sharma, and A. K. Ramani, "Towards Performance Analysis of Ad hoc Multimedia Network", Chapter in Machine Learning and Systems Engineering, *Lecture Notes in Electrical Engineering*, Springer, vol. 68, pp. 495-506, 2010.

[9] J. Fan, D. Wu, A. Nucci, R. Keralapura, and L. Gao, "Protocol Oblivious Classification of Multimedia Traffic", *Security and Communication Networks*, vol. 4, issue 4, pp. 357–371, 2011.

Session TN II:

TELECOMMUNICATION NETWORKS AND SERVICES II

Traffic Measurements and Flow Analyses in 3G Network

Rossitza Goleva[1], Seferin Mirtchev[2], Dimitar Atamian[3], Ljubomir Khadjiivanov[4] and Kiril Kassev[5]

Abstract – **This paper presents traffic measurements and statistical analyses performed in 3G network. The aim of the work is to estimate packet flow characteristics. Data, voice and video sessions are captured and analyzed. As a result statistical distributions of the pure and mixed traffic are specified and simulated. The simulation is used to analyse difficult behaviour of the network.**

Keywords – **3G UMTS networks, Traffic sources, Traffic measurements, Quality of Service.**

I. Introduction

Wireless technologies and their rapid convergence are a matter of fast development recent decades. Because of the new services involved in the networks the Quality of Service (QoS) and Quality of Experience (QoE) estimation is actual topic of research all the time. Their measurements and management procedures are not developed as fast as the market requires and it is a reason for complains from the service quality worldwide.

This paper is focused on traffic measurements and its analyses in 3G UMTS network. The derived results will allow authors to specify proper traffic models for further network design. They are also used to analyse the network behaviour in difficult conditions. The measurements and simulations are based on data, voice over IP (VoIP) and video over IP traffic sources. Attention is paid to the real-time services and services with high packet dispersion.

Active and passive traffic measurements [1] are explained in many papers. All of them concerns performance of the wireless access links that are a key points when considering the performance of the entire Internet connection. Laboratory based measurements of similar type can be seen in [1]. The traffic source in a multi-point videoconferencing application and data for One-Way-Delay, IP-Delay-Variation and Packet Loss Ratio are shown. The asymmetric nature of the UMTS link is demonstrated. There are no rules for traffic analyses that will allow further network behaviour forecast.

The foundation of QoS and QoE analyses can be seen in [2]. Clear mapping between QoE and QoS parameters is still missing.

[1, 2, 3, 5]Rossitza Goleva, Seferin Mirtchev, Dimitar K. Atamian and Kiril Kassev are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, e-mails: rig@tu-sofia.bg, stm@tu-sofia.bg, dka@tu-sofia.bg, kmk@tu-sofia.bg

[4]Ljubomir Khadjiivanov is with the Mobiltel EAD, Kukush str. 1, Sofia, 1309, Bulgaria, e-mail: Khadjiivanov.L@mobiltel.bg

In [3] QoS importance is explained. Measurements of IPTV and VoIP are presented to prove the explained ideas. One-way end-to-end QoS performance statistics in live HSDPA (High Speed Downlink Packet Access) is shown. The phenomenom of CBR-type (Constant Bit Rate) VoIP source traffic affecting with 3G network timing mechanisms is analyzed. The nature of distributions at packet layer and call layer is not explicitly specified.

The nature of TCP session and IP packets levels that affects the packet inter-arrival time is shown on [4]. Clear cross layer approach to the call level is not visible. An overview of existing measurements technologies is done in [5] and [6]. Common guidelines for measurements do not exist.

QoS and QoE parameters and its mapping with WiMAX – DiffServ platform are explained in [7]. Delay and throughput end-to-end budget is shown. There is no way to guarantee end-to-end traffic parameters in DiffServ domains because they work with aggregated traffic.

A very complicated correlation between video, voice quality and loss is published in [8]. The authors also show that the compression algorithm is important. The effect of codec bit rate and losses for video traffic sources is analysed in [9]. Round trip time and how it is influenced by different segments of the network is shown in [10]. Direct mapping to the call layer is missing.

The aim of this paper is to show passive traffic measurements in 3G network domain that will allow mapping of call, session and packet layers with traffic parameters. The distributions of the flows seen in the measured domain are applied in simulation model for further network behaviour analyses and QoS to QoE mapping.

II. Traffic Measurements in 3G Network

Passive traffic measurements are performed in 3G network (Fig. 1) for ftp, Voice over IP (VoIP) and Video over IP services. On the figure the Node-B serves the antenna and transceivers in 3G network, RNC (Radio Network Controller) schedules the resources for base stations, MSC-S (Mobile Services Switching Center Server) manages the network, MGW (Media Gateway) connects GSM and 3G networks, GMSC (Gateway Mobile Services Switching Center) connects the mobile network with Public Switched Telephone Network or Integrated Services Digital Network, HLR (Home Location Register) contains user data including QoS specific parameters, VLR (Visitor Location Register) contains data for all servicing terminals in the network, SGSN (Serving GPRS Support Node) supports packet switching and routing in UMTS network, GGSN (Gateway GPRS Support Node)

interconnects UMTS network with other packet switched networks, AUC (Authentication Center) contains all necessary data for user access.

The experiment is configured between Mobile Connect Card connected to a laptop for packet access via the 3G network. The data flow passes through Node B – RNC –

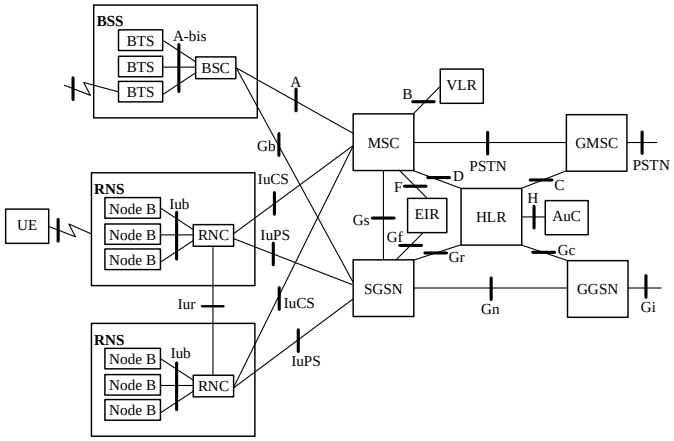


Fig. 1. Packet switched part of the 3G network

SGSN – GGSN –IP backbone transmission network – SBC (Session Border GW Controller) forward and backward following the same path.

Calling-party and Called-Party are UMTS users in the network. Different codecs are applied:

- For VoIP - GSM 6.10, G711a Law, G.729
- For video calls - H.263, H.263+ (1998), H.264 (low end), H.264 (high end)

The “low end” and “high end” are notations for mobile phones with low and high bandwidth accordingly. For tests with ftp traffic a server that is internal in the network is applied.

On Fig. 2 the VoIP packet flows with H.263 and H.264 codecs are shown. The observation is done in one minute interval. The codec characteristics influence the occupancy of the channel significantly. The dispersion of the signals is also significant. On Fig. 3 payload data rates are shown. The main result here concerns the limited capability of the mobile to transmit video with proper quality.

Fig. 4 shows the same video flows in packets per second. On the right side of the graph comparison with data payload can be seen. Low end mobile phones are not good enough to carry video traffic because of their low bandwidth and high difference between overhead and payload.

The video/voice packet sizes observed are mostly between 100 and 600 bytes. Because the traffic is mixed with signalling and data small packets of 20 bytes for acknowledgement and big packets of 1500 bytes from ftp are also visible in the channel.

Observation of packet inter-arrival times (Table I) shows three different groups of packets in the channels. Most of the packets are at packet layer. Some of the packets are at UDP/TCP session layer and few of the packets are at call layer. There is long range uncommon dependence between the session duration and number of packets in session, the distribution of number of packets and inter-arrival time. The traffic sources are modelled as ON/OFF sources with calls, sessions and packets layers. The long range dependence of the packets in buffers requires the use of special distributions like Pareto as well as a cross-layer approach.

Fig. 5 and Fig. 6 show the difference between low end and high end mobiles. The nature of the mobile phone bounds the

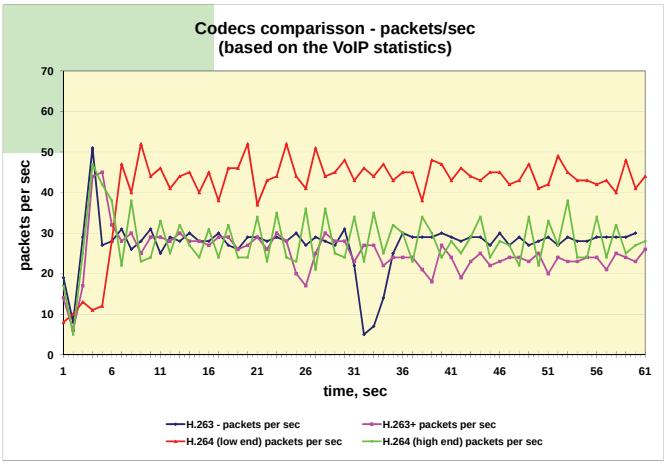


Fig. 2. VoIP flow for H.263 and H.264 codecs

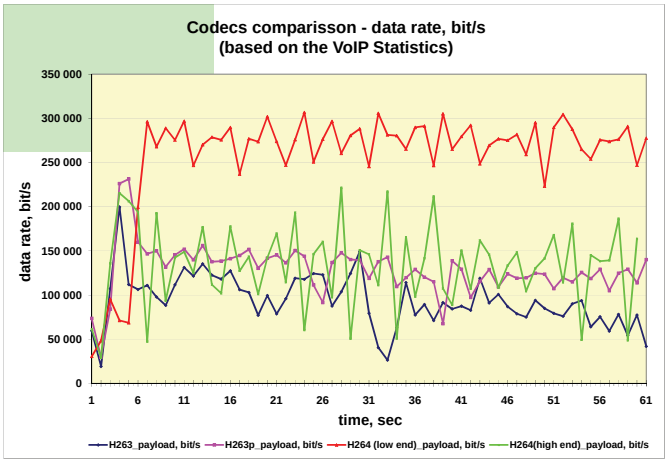


Fig. 3. VoIP flow rates for H.263 and H.264 codec

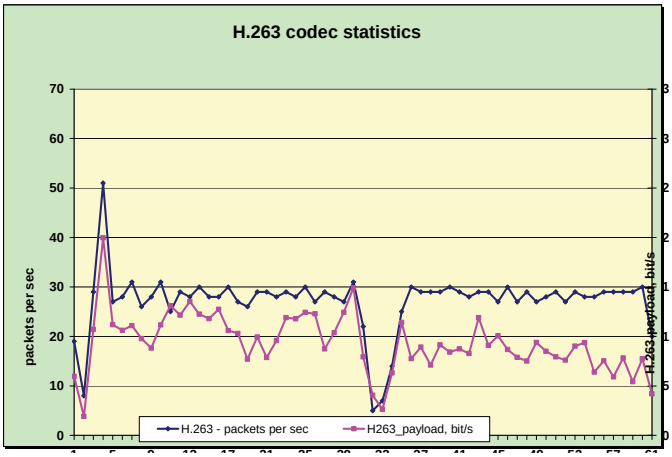


Fig. 4. Video service flows in packets per second and comparison to the payload

payload that might be transmitted in the networks in low end case. So, high end mobile or mobiles with better compression should be applied in order to satisfy the QoS and QoE requirements.

TABLE I
PACKET INTER-ARRIVAL TIMES

Inter-arrival time, ms	Number of packets observed	Layer of consideration
0,5 ms	250	Packet layer
1 ms	457	Packet layer
3 ms	50	Packet layer
5 ms	187	Session layer
8 ms	30	Session layer
11 ms	126	Session layer
13 ms	30	Call layer
15 ms	45	Call layer

III. TRAFFIC SIMULATION IN 3G DOMAIN

Based on the observed data from measurements a simulation model is built in order to analyse the behaviour of the 3G network in the case of mixed traffic. The applied QoS management serves in different ways the real-time multimedia services and the non real-time services. Usually, the real-time services are classified as conversational and streaming QoS classes in 3G network [7]. Because these traffic classes are not supported well by the equipment they are converted into interactive and best effort QoS classes in the simulation model (Table II).

TABLE II
SIMULATION PARAMETERS IN 3G NETWORK

Parameter	VoIP	Video	WWW	FTP
Packet size, bytes	168	208	300	540
Maximal rate, kbps	128	192	384	384
Transport protocol	UDP	UDP	TCP	TCP
Traffic source	ON/OFF with exponent-tial distribu-tion	Determi-nistic traffic	ON/OFF with Pareto distribution	Determi-nistic traffic
Ton, sec.	1.00	-	1.60	-
Toff, sec.	1.50	-	12.00	-

The QoS parameters are implemented in PDP (Packet Data Protocol) context and depend on HLR. The video streams are considered to be symmetric.

In this simulation the Pareto distribution is proposed for WWW traffic. WWW flows have long range dependence

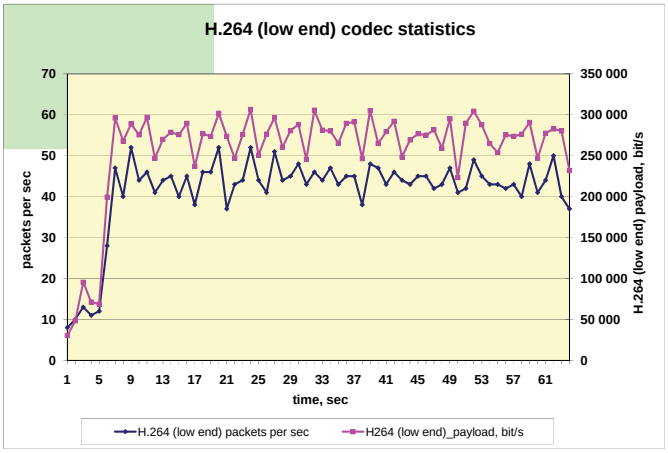


Fig. 5. Low end H.264 codec vs. payload

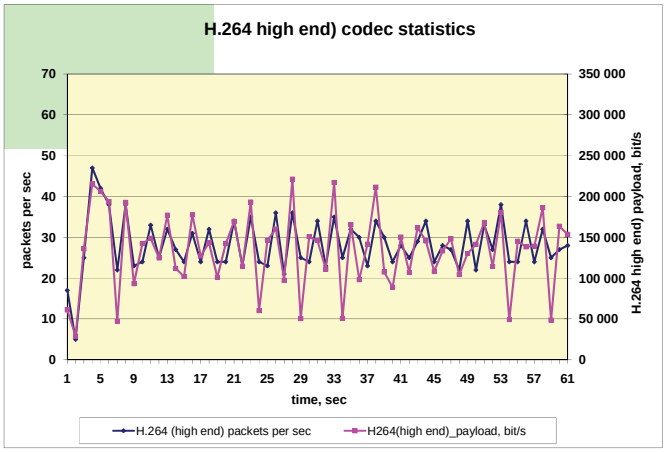


Fig. 6. High end rates vs. payload

between the intervals of web page load and web page reading. This feature increases the delay of the packets and influences the behaviour of the buffers. The family of Generalized Pareto Distributions (GPD) has three parameters: the location parameter μ, the scale parameter σ and the shape parameter ξ (Eq. 1).

The cumulative distribution function of the GPD is in Eq. 1:

$$F(x) = 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} . \quad (1)$$

We choose substitutions shown on Eq. 2:

$$\eta_0 = \frac{\xi}{\sigma}; \quad \frac{1}{\xi} = 1 + \frac{\lambda}{\eta_o}; \quad \mu = 0 . \quad (2)$$

Therefore, we receive another form of the generalized-Pareto distribution as in Eq. 3:

$$F(t) = 1 - (1 + \eta_o t)^{-\left(1 + \lambda/\eta_o\right)} . \quad (3)$$

The mean value of the generalized-Pareto distribution is on Eq. 4:

$$m_o = 1/\lambda . \quad (4)$$

The mean value is the average inter-arrival time for packets in our study. The parameter λ is the packet arrival intensity [11].

The core 3G packet switched network uses Diffserv for Quality of Service management. SGRN and GGSN are end nodes in the core DiffServ network. The DiffServ to 3G QoS mapping is different in radio interface and in core network. The operators may map 3G QoS classes and DiffServ for uplink in SGSN. Downlink mapping in performed in GGSN [12]. This functionality is part of the network configuration. On Table III possible mapping scenario is shown. Traffic Handling Priority (THP) is used to differentiate 3G UMTS interactive services.

The QoS observed during simulations shows that the error control conforms to the required 2% (Table IV). Low priority queues with THP=3 may introduce high delay. The jitter is less that 3 ms uplink and downlink and is considered acceptable. Bigger values are hardly expected not to exceed 10 ms.

TABLE III
DIFFSERV TO 3G QOS MAPPING IN CORE NETWORK

Service	3GPP UMTS QoS class	IETF Diffserv QoS class
VoIP	Interactive with THP=1	Explicit forward
Video	Interactive with THP=2	Assured Forward 21
WWW	Interactive with THP=3	Assured Forward 11
FTP	Best effort	Best effort

TABLE IV
3G NETWORK SIMULATION RESULTS

Service	Uplink	Downlink
VoIP	Rate 50 kbps	Rate 50 kbps
	Loss < 2%	Loss < 1%
	End-to-end delay = 150 ms	End-to-end delay = 80 ms
	Jitter = 1 ms	Jitter = 1 ms
Video	Rate 175 kbps	Rate 190 kbps
	Loss < 2%	Loss < 2%
	End-to-end delay = 200 ms	End-to-end delay = 100 ms
	Jitter = 3 ms	Jitter = 3 ms
WWW	Rate 20 kbps	Rate 30 kbps
FTP	Rate 40 kbps	Rate 60 kbps

IV. CONCLUSION

In this paper the voice and video service measurement is presented and analyzed. Proper traffic specification of traffic sources obtained is applied for simulation. QoS class mapping between 3G and DiffServ network allows better network design and configuration. During the work the mostly used codecs are considered. We conclude that the recent IP applications will work on a 3G or any wireless domain with small packet sizes. This will decrease congestion probability. The long-range dependence in buffers at packet, session and call layers will require overdimentioning of the network or other compression algorithms. The work continues with more precise measurements and analyses based on active and passive algorithms and additional statistical analyses.

ACKNOWLEDGEMENT

This paper is sponsored by project _102ni180-07 "Development of the Tools for Traffic Monitoring and Analyses in IP Networks" funded by the internal research program at Technical University of Sofia, Bulgaria. It also cannot be done without measurements performed by the mobile operator Mobiltel EAD. The analyses and simulations have been performed by valuable support of diploma thesis student Bojidar Georgiev.

REFERENCES

[1] R. Cosmad, A. Cabellos-Aparicioa, J. Domenech-Benllochb, J. Gimenez-Guzmanb,.J. Martinez-Bausetb, M. Cristiand, A. Fuentetajac, A. Lópeza, J. Domingo-Pascuala, and J. Quemadac, "Measurement-Based Analysis of the Performance of Several Wireless Technologies", *Proceedings of the 16th IEEE workshop on Local and Metropolitan Area Networks*, pp.19-24, 2008.

[2] R. Lioyd-Evans, *QoS in Integrated 3G Networks*, Artech House, 2002.

[3] J. Prokkola, M. Hanski, M. Jurvansuu and M. Immonen, "Measuring WCDMA and HSDPA Delay Characteristics with QoSMeT", *Proceedings from ICC*, pp. 492-498, 2007.

[4] F. Ricciato, P. Svoboda, J. Motz, W. Fleischer, M. Sedlak, M. Karner, R. Pilz, P. Romirer-Maierhofer, E. Hasenleithner, W. Ja¨ ger, P. Kruger, F. Vacirca, and M. Rupp, "Traffic Monitoring and Analysis in 3G Networks: Lessons Learned from the METAWIN Project", *Elektrotechnik & Informationstechnik* (2006) 123/7/8: pp. 288–296. DOI 10.1007/s00502-006-0362-y.

[5] F. Ricciato, "Some Remarks to Recent Papers on Traffic Analysis", *ACM SIGCOMM Computer Communication Review 99*, vol. 36, no. 3, pp. 99-102, 2006.

[6] F. Ricciato, "Traffic Monitoring and Analyses for the Optimization of a 3G Network", *IEEE Wireless Communications*, December, pp. 42-49, 2006.

[7] R. S. Ravindran, F. Blouin, "QoE Driven Framework for High Performance Multimedia Delivery Over WiMax", *ITC*, pp. 91-103, 2008.

[8] P. Casas, P. Belzarena and S. Vaton, "End-2-End Evaluation of IP Multimedia Services, a User Perceived Quality of Service Approach", *ITC*, pp. 13-25, 2008.

[9] L. Thorpe, F. Speranza, T. Rahrer, R. Renaud and R. Meier, "Preparing High-Quality Subjective Datasets For the Evaluation of Objective Video Quality Metrics", *ITC*, pp. 25-33, 2008.

[10] A. Barbuzzi, G. Boggia and L. Alfredo Grieco, "DeSRTO: an effective algorithm for SRTO detection in TCP connections", *Proc. of Traffic Monitoring and Analysis Workshop*, *TMA*, Zurich, Switzerland, 2010.

[11] H. Wang, D. Prasad, O. Teyeb, H. P. Schwefel, *Performance Enhancements of UMTS networks using end-to-end QoS provisioning*, Aalborg University, Alborg, Denmark, 2005.

[12] 3GPP, TS 23.107 V6.4.0 - Quality of Service (QoS) concept and architecture, 2006, [http://www.3gpp.org/ftp/Specs/archive/23_series/23.107/23107-640.zip].

Properties of Two Traffic Models with Changed Serving Intensities in Alternative Groups

Bojan Bakmaz[1] and Miodrag Bakmaz[2]

Abstract – **In this paper the two systems with overflow possibilities and changing serving intensities are investigated. A primary group with two Poisson traffics is considered whereupon the rejected calls from one of them are directed to the alternative groups and served with changed serving intensities. Generating function technique is used for the analytical solving model with secondary and ternary group and model which treated separately the channels with ordered selecting in secondary group. The analysis of the importance for using this models and comparison with the model with same serving intensities of overflow traffics is carried out.**

Keywords – **Changed serving intensity, Generating function, Ordered selecting, Overflow traffic, Traffic parameters**

I. Introduction

This paper deals with problems of traffic parameters analytical determination in the systems with overflow traffic which have the possibility of changing serving intensity. The models with overflow traffic became relevant in the second half of the last century by introducing the alternative routing in telecommunication networks. For the fundamental work in this field [1] can be considered and in [2] a substantial mathematical tool for the analytical solving of steady state equations with the generating function technique can be found.

Alternate path routing of overflow traffic has been widely used to improve the performance of hierarchical networks, including traditional circuit-switched networks [5], networks with dynamic routing [6], broadband networks, IP networks and hierarchical cellular networks [8]. Similar methods can also be used to improve traffic performance in heterogeneous wireless networks which support internetworking overflow. Traffic loss analysis in such networks is a challenging issue due to intensive computations involved, particularly in the cases where various services with multi-rate and multi-quality of service are supported [8]-[11].

The problem of the changed serving intensity is treated in several works as in [3] and [4]. Paper [7] seeks to provide the analysis of a model with overflow traffic components that originate from several Poisson traffics served by a common primary group. The overflow parts of the two of them are served in a secondary group but with changed and different serving intensities. The general solution of the system was obtained by generating function techniques and compared with the simpler model.

The interest for the analysis of the models with this characteristics results from the real needs. Besides the cases described in [3], [4] and [7], it should be noted that in some cases, this phenomenon can be reiterated, too. Characteristical examples are call centers, with operators using different average time for performing the same works, or the case of combined operator's and interactive voice response systems engagement. These models have special importance in the case of heterogeneous networks, when the alternating resources with different basic bandwidth unit, i.e. different serving intensity, are going to be chosen.

Call-level multi-rate teletraffic loss models aim at assessing the call-level QoS of IP based networks with resource reservation capabilities. This assessment is important for the bandwidth allocation among service classes (guaranteeing QoS), the avoidance of too costly over-dimensioning of the network and the prevention, through traffic engineering mechanisms, of excessive throughput degradation. Despite of its importance, the call-level performance modeling and QoS assessment is a challenge in the highly heterogeneous environment of next generation networks, due to the presence of elastic traffic, or complicated call arrival process. With the case of multi-rate traffic, the basic method for determining traffic characteristics is the application of the so-called Kaufman-Roberts formulas and their modifications [12].

This paper is organized as follows. We start with the analytical solution of the first model with secondary and ternary group where the intensity of serving is changed and the process of determining the unknown coefficient is proposed. Presence of the ternary group represents the extension of the model proposed in [7]. The second is the model where the intensity of serving is changed in each channel of the secondary group with ordered selecting. A general solution is also proposed and the procedure for determining the unknown coefficients is explained. Two Poisson traffics in primary group are observed in both models since that generalization does not complicate the solution. Both models are illustrated with simple unique numerical example and the influence of serving intensity change on losses is analyzed.

II. Model with Secondary and Ternary Group and Changed Serving Intensity

The serving system with two (or more) Poisson traffics, a_1 and a_2, with serving intensity $\mu_0 = 1$, offered to the primary group having c channels, whereupon the rejected calls of one of the traffics are offered to the secondary and ternary group,

[1]Bojan Bakmaz is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: b.bakmaz@sf.bg.ac.rs.

[2]Miodrag Bakmaz is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: bakmaz@sf.bg.ac.rs.

with s_1 and s_2 channels and with changed serving intensities, μ_1 and μ_2, is of interest for the analytical research (Fig. 1).

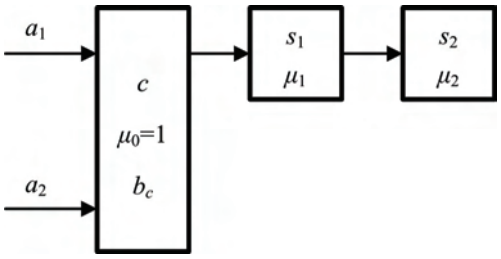


Fig. 1. Model with secondary and ternary group.

The model is described by $(c + 1)(s_1 + 1)(s_2 + 1)$ steady state equations. General solution for the state probabilities is reached by generating function technique, similar with [2], [7]

$$p(m,n_1,n_2) = (-1)^{n_1+n_2} \times \sum_{k_1=n_1}^{s_1}\binom{k_1}{n_1}\sum_{k_2=n_2}^{s_2}\binom{k_2}{n_2} R_{\mu_1 k_1+\mu_2 k_2}(m) C(k_1,k_2), \quad (1)$$

where (m, n_1, n_2) are states in groups with capacities (c, s_1, s_2). Here states represent the number of occupied channels. Also, relations

$$R_0(m) = \frac{(a_1+a_2)^m}{m!} \quad (2)$$

and

$$R_{(\mu n)}(m) = \sum_{i=0}^{m}\binom{(\mu n)+i-1}{i}\frac{(a_1+a_2)^{m-i}}{(m-i)!}, \quad n>0. \quad (3)$$

are obtained. It should be noted that $(\mu n) = \sum \mu_i n_i$.

With further $R_{(\mu n)}(c) = R_{(\mu n)}$, from the boundary steady state equations ($m = c$), using marginal generating functions technique, a recurrent formula for coefficients $C(n_1, n_2)$ is derived:

$$\begin{aligned} &-(\mu_1 n_1+\mu_2 n_2) R_{\mu_1 n_1+\mu_2 n_2+1} C(n_1,n_2) \\ &= a_1 R_{\mu_1(n_1-1)+\mu_2 n_2} C(n_1-1,n_2) \\ &+ a_1(-1)^{s_1+n_1}\binom{s_1}{n_1-1} R_{\mu_1 s_1+\mu_2 n_2} C(s_1,n_2) \\ &+ a_1(-1)^{s_1+n_1}\binom{s_1}{n_1} R_{\mu_1 s_1+\mu_2(n_2-1)} C(s_1,n_2-1) \\ &+ a_1(-1)^{s_1+n_1+s_2+n_2}\binom{s_1}{n_1}\binom{s_2}{n_2-1} R_{\mu_1 s_1+\mu_2 s_2} C(s_1,s_2). \end{aligned} \quad (4)$$

With this procedure the problem is reduced to solving $(s_1 + 1)(s_2 + 1)$ equations for unknown $C(n_1, n_2)$. However, having in mind the fact that from the normalized condition for state probabilities $C(0, 0) = 1/R_1$ there is a possibility to find solution for $C(s_1, s_2)$ from the recurrent formula (4), then $C(s_1, n_2)$, and finally to express all the rest $C(n_1, n_2)$.

It is possible to define more parameters for this model and primary is the call loss of traffic a_1 that is in the form

$$b_{cs12} = p(c,s_1,s_2) = (-1)^{s_1+s_2} R_{\mu_1 s_1+\mu_2 s_2} C(s_1,s_2). \quad (5)$$

The call loss of traffic a_1 in the secondary group is

$$b_{s1} = p(c,s_1)/b_c = (-1)^{s_1} R_{\mu_1 s_1} C_{s1} R_1 / R_0, \quad (6)$$

where $b_c = B(a_1+a_2, c) = R_0 / R_1$ is Erlang loss in primary group, while in ternary we have $b_{s2} = p(c,s_1,s_2)/p(c,s_1)$. Coefficient C_{s1} can be obtained from explicit solution for system without ternary group [7]

$$C_{s1} = \frac{\dfrac{(-a_1/\mu_1)^{s_1}}{s_1! R_{\mu_1 s_1}}\prod_{i=0}^{s_1}\dfrac{R_{\mu_i i}}{R_{\mu_i i+1}}}{\displaystyle\sum_{n=0}^{s_1}\frac{(a_1/\mu_1)^{s_1-n}}{(s_1-n)!}\prod_{i=n+1}^{s_1}\frac{R_{\mu_i i}}{R_{\mu_i i+1}}}, \quad \prod_{i=s_1+1}^{s_1}\frac{R_{\mu_i i}}{R_{\mu_i i+1}} = 1. \quad (7)$$

The following expressions can be used for determining the time loss in secondary

$$b_{ts1} = \sum_{m=0}^{c}\sum_{n_2=0}^{s_2} p(m,s_1,n_2) = (-1)^{s_1} R_{\mu_1 s_1+1} C_{s1} \quad (8)$$

and ternary group

$$b_{ts2} = \sum_{m=0}^{c}\sum_{n_1=0}^{s_1} p(m,n_1,s_2). \quad (9)$$

Traffic a_1 is offered to primary group, its calls formed traffic a_{1o} offered to the system

$$a_{1o} = a_1(1-b_c) + a_1 b_c (1-b_{s1})/\mu_1 + a_1 b_c b_{s1}/(\mu_1\mu_2), \quad (10)$$

while the carried traffic is

$$a_{1c} = a_1(1-b_c) + a_1 b_c (1-b_{s1})/\mu_1 + a_1 b_c b_{s1}(1-b_{s2})/(\mu_1\mu_2). \quad (11)$$

Now, the traffic loss in the system for a_{1o} can be obtained in the form

$$b_{a1o} = 1 - \frac{a_{1c}}{a_{1o}}. \quad (12)$$

It is obvious that the traffic offered to the system increases during the decreasing of serving intensity and vice versa.

In the bitrate and capacity domains, if arrival rate λ_i, $i = 1$, 2, is expressed in terms of calls per serving time $t_0 = 1/\mu_0$, equality $a_i = \lambda_i$ holds. The primary resources are of the capacity C, the mean value of information units per call r, while the basic bandwidth is b. Thus, the primary resource

represents a group of $c = C/b$ channels, while serving intensity is $\mu_0 = b/r$.

In secondary and ternary resources with capacities S_i, i = 1, 2, basic bandwidths are b_i, while the corresponding groups has got $s_i = S_i/b_i$ channels and the relative serving intensities are $\mu_i = b_i/b$.

III. Model with Changed Serving Intensities in Secondary Group Channels

The previously used technique can also be applied for the case when the channels in secondary group are ordered selected and own serving intensity exists for each of them (Fig. 2).

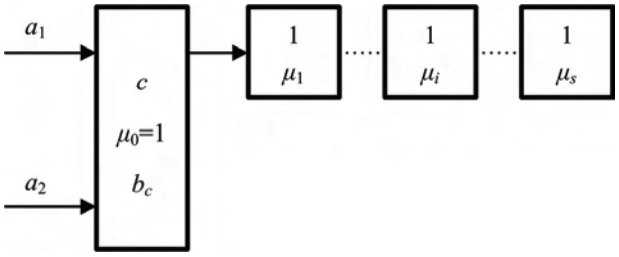


Fig. 2. Model with changed serving intensities in secondary group channels

Using generating function technique, we obtained the solution for state probabilities

$$p(m,n_1,...,n_s) = (-1)^{n_{1s}} \sum_{k_1=n_1}^{1} ... \sum_{k_s=n_s}^{1} R_{(\mu k)_{1s}}(m) C(k_1,...,k_s), \quad (13)$$

where $n_{ij} = n_i + \ldots + n_j$ and $(\mu n)_{ij} = \mu_i n_i + \ldots + \mu_j n_j$ (= (μn) for using (3)).

The equation system for unknown coefficients $C(n_1,\ldots, n_s)$ obtained from the boundary equations with marginal generating function technique is

$$\begin{aligned} &-(\mu n)_{1s} R_{(\mu n)_{1s}+1} C(n_1,...,n_s) = a_1 \sum_{i=1}^{s} ((-1)^{i-1+n_{1(i-1)}} \\ &\times R_{\mu_{1(i-1)}+\mu_i(n_i-1)+(\mu n)_{(i+1)s}} C(1,...,1,n_i-1,n_{i+1},...,n_s) \\ &+(-1)^{i+n_{1i}} \binom{1}{n_i-1} R_{\mu_{1i}+(\mu n)_{(i+1)s}} C(1,...,1,n_{i+1},...,n_s)), \end{aligned} \quad (14)$$

where $n_{10} = 0$, $n_{(s+1)s} = 0$ and $C(0,\ldots, 0) = 1/R_1$. The number of equations now is 2^s - 1 instead of $(c + 1)2^s$ in the case of solving system steady state equation.

Based on the fact that

$$p(c,n_1,...,n_{j-1}) = p(c,n_1,...,n_{j-1},0) + p(c,n_1,...,n_{j-1},1) \quad (15)$$

and using (13), it can be concluded that the equality $C(n_1,\ldots, n_{j-1}, 0) = C(n_1,\ldots, n_{j-1})$ holds. For j = 2, 3,…, s the system of 2^{j-1} equations

$$\begin{aligned} &-(\mu n)_{1j} R_{(\mu n)_{1j}+1} C(n_1,...,n_{j-1},1) = a_1 \sum_{i=1}^{j} ((-1)^{i-1+n_{1(i-1)}} \\ &\times R_{\mu_{1(i-1)}+\mu_i(n_i-1)+(\mu n)_{(i+1)j}} C(1,...,1,n_i-1,n_{i+1},...,n_j) \\ &+(-1)^{i+n_{1i}} \binom{1}{n_i-1} R_{\mu_{1i}+(\mu n)_{(i+1)j}} C(1,...,1,n_{i+1},...,n_j)), \end{aligned} \quad (16)$$

can be solved iteratively. This iterative procedure requires solving 2^{s-1} equations for unknown coefficients $C(n_1,\ldots, n_s)$ in the last step which, compared to the former case, makes only half of the equations. That is important for calculating time and convergence.

In the case of the same serving intensity in secondary group with j channel, $\mu_j = \mu$, j = 1,…,s expression for C_j from [7] can be used, so $C(1,\ldots, 1_{(j)}) = C_j$ and we must solve the system with $2^{j-1} - 1$ equation, i.e. $2^{s-1} - 1$ in the last step, if we wish to determine microstate probabilities. For c = 1, a_2 = 0 and μ = 1 we could have a classical system with (s + 1) channel and ordered selecting. If we use the steady state equations for determining the microstate probabilities, we need to solve 2^{s+1} equation which is around four times more than in the last step of developed procedure. The losses for these cases have explicit solutions [7].

IV. Numerical Example

As an example, simple but enough illustrative case is analyzed when c = 10, s_1 = 1, s_2 = 1 channel and that results from the first and second model. Then, (4) or (14) will be in the form

$$\begin{aligned} &-(\mu_1 n_1 + \mu_2 n_2) R_{\mu_1 n_1 + \mu_2 n_2 + 1} C(n_1,n_2) = a_1 (R_{\mu_1(n_1-1)+\mu_2 n_2} \\ &\times C(n_1-1,n_2) - (-1)^{n_1} \binom{1}{n_1-1} R_{\mu_1+\mu_2 n_2} C(1,n_2) - (-1)^{n_1} \\ &\times R_{\mu_1+\mu_2(n_2-1)} C(1,n_2-1) + (-1)^{n_1+n_2} \binom{1}{n_2-1} R_{\mu_1+\mu_2} C(1,1)) \end{aligned} \quad (17)$$

From the relation (17) for n_i the system of tree equations is formed, while $C(1, 0)$ is dependent only on coefficient $C(0, 0)$ and is equal with C_1 from (7).

Fig. 3 illustrates call losses in alternate groups with one channel for a_1 = 6 Erl., a_2 = 2 Erl. and different values of μ_1 and μ_2. The mutual relation between the losses curve b_{s2} for μ_2 = μ_1 and b_{s1} is characteristical and correspond to the simpler model from [7]. For different μ_2 there exist the values $\mu_1 > \mu_2$ when b_{s2} becomes greater then b_{s1}. For μ_1 = 0 ternary group is going on secondary.

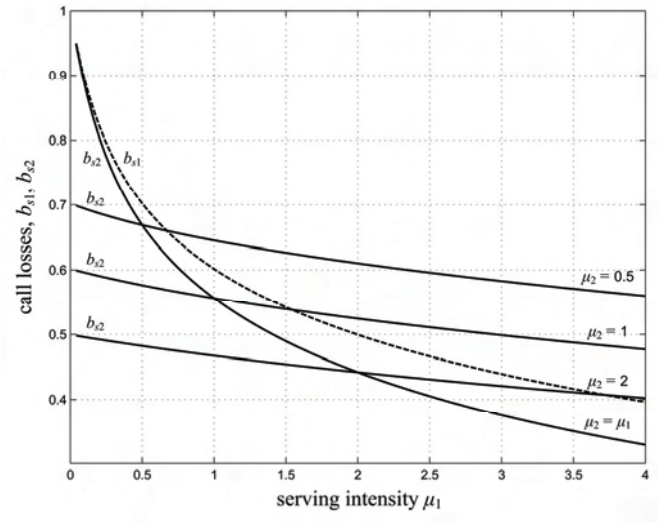


Fig. 3. Call losses in secondary and ternary group

Fig. 4 illustrates significant losses discrepancy concerning the difference of serving intensities. In this case, traffic loss in the system b_{a1o} is more sensitive to serving intensities changes than call losses b_{cs12}.

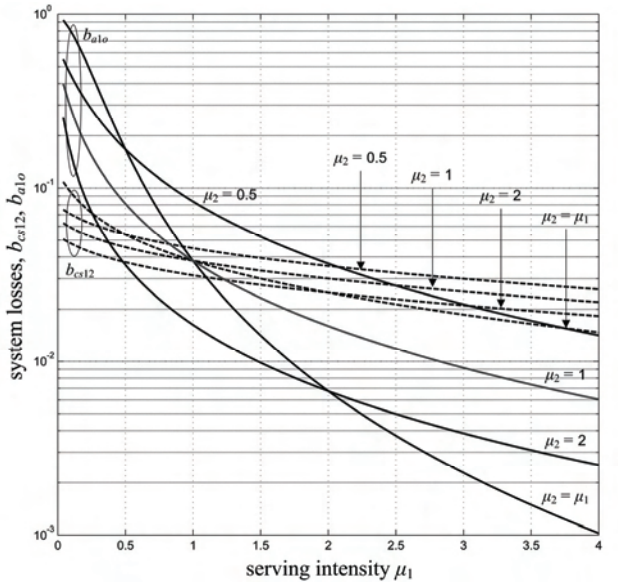


Fig. 4. Call and traffic losses in the system.

V. Conclusion

Models with overflow traffic have been relevant for over half a century and the need for their application exists in modern communication networks. In spite of the great number of papers concerning this topic, there are a few explicit solutions, when dealing with obtaining steady state equation system. Among the different models, the two are emphasized for which analytical solution is obtained. Analytical solution of the steady state equations presented in this work is a challenge which gives as a result fast calculation and possible system volume increment. For the simpler case representative enough for both models graphical representation and traffic parameters analysis is given.

Attained knowledge about parameters characteristics is of importance for traffic analysis of modern telecommunication networks in which, when using alternate routing, we can observe cases which these models are going to simulate. Results obtained using the analysis of traffic parameters behavior, justify the needs to dedicate special care to the case where phenomenon with changed serving intensity is remarked. The proposed solutions will be a motivation for further researches in accordance with the cases where various services with multi-rate and multi-quality of service are supported.

Acknowledgement

This research is partially supported by the Ministry of Education and Science, Republic of Serbia, Project No. TR-32025.

References

[1] R. I. Wilkinson, "Theories for Toll Traffic Engineering in the U.S.A.", *Bell Syst. Tech. J.*, vol. 35, no. 2, pp. 421–514, 1956.

[2] S. Neal, "Combining Correlated Streams of Non-random Traffic", *Bell Syst. Tech. J.*, vol. 50, no 6, pp. 2015-37, 1971.

[3] B. Wallström, "Loss Calculations in Certain Overflow Systems Where the Holding Times in Successive Groups Have Different Means", *Proc. ITC-7*, Stockholm, Sweden, p. 417/1-5, 1973.

[4] R. G. Schehrer, "A Two Moments Method for Overflow Systems with Different Mean Holding Times", *Proc. 15th International Teletraffic Congress*, pp. 1303–1314, Washington, DC, USA, 1997.

[5] H. Akimaru, K. Kawashima, *Teletraffic: Theory and Applications*, Springer-Verlag, 1993.

[6] G. Ash, *Dynamic Routing in Telecommunications Networks*, McGraw-Hill, 1998.

[7] M. Bakmaz, "Analysis of a Model of Two Overflow Traffic Components with Different Serving Intensities", *International Journal of Electronics and Communications* (*AEUE: Archiv fuer Elektronik und Uebertragungstechnik*), vol. 60, no 1, pp. 65-70, 2006.

[8] X. Wu, B. Mukherjee, D. Ghosal, "Hierarchical Architectures in the Third-Generation Cellular Network", *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 62-71, 2004.

[9] A. Ghosh, et. al., "Broadband Wireless Access with WiMax/802.16: Current Performance Benchmarks and Future Potential", *IEEE Commun. Magazine*, vol. 43, no. 2, pp. 129-136, 2005.

[10] S-P. Chung, J-C. Lee, "Performance Analysis and Overflowed Traffic Characterization in Multiservice Hierarchical Wireless Networks", *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 904-918, 2005.

[11] Q. Huang, K. T. Ko, and V. B. Iversen, "Approximation of Loss Calculation for Hierarchical Networks with Multiservice Overflows", *IEEE Trans. Commun.*, vol. 56, no. 3, pp. 466-473, 2008.

[12] M. Głabowski, K. Kubasik, and M. Stasiak, "Modeling of Systems with Overflow Multi-rate Traffic", *Telecommunication Systems*, vol. 37, no. 1-3, pp. 85-96, 2008.

Requirements to Mobile Telemetry Application Protocol

Elitsa Gospodinova[1], Ivaylo Atanasov[1], Evelina Pencheva[1]

Abstract – **In this paper an analysis on functional requirements to application protocol for mobile telemetry is presented. The protocol is used for communication between mobile agents and a central control unit. It relies on Internet protocol connectivity and is access network independent. Along with the domain specific functionality, security issues are discussed such as authentication, ciphering and integrity check at application level.**

Keywords – **Mobile agents, Internet protocol connectivity, Reliable transport, Authentication, Ciphering, Integrity check.**

I. Introduction

The mobile telemetry uses mobile data for remote information measurement and reporting. Special mobile agents containing modules for data transfer and positioning combined with sensors can be used to implement a wide variety of outdoor and indoor applications. Some example applications of mobile telemetry include traffic monitoring, resource distribution, medicine, home security, energy monitoring etc. Care needs to be taken when assessing requirements to monitoring data and data costs. Furthermore, possible external attacks have to be taken into account, especially when mobile agents use internet for data transmission.

Security aspects of mobile telemetry are subject of intensive research. The proposed security mechanisms include authorization through digital certificates in virtual private networks [1], security policy assigning new identification numbers [2], location privacy and pseudo-direct communication, which can be incorporated into the distributed authentication protocol [3], authentication and privacy by using public keys to certify the identity of parties [4] and other. Different aspects of telemetry functions and their implementation in mobile agents are discussed in [5], [6] and [7].

The proposed solutions are applicable to mobile hosts like personal computers having relatively high computational power and onboard energy supply. Smart solutions require reduction of energy budget and autonomous operation with battery powered.

In this paper, we study the requirements to application layer protocol that can be used for mobile telemetry. The protocol is based on Internet protocol (IP) connectivity that may be obtained by means of any cellular or wireless network. The protocol has to be access-independent so that the data transfers may take place over Global System for Mobile Communications (GSM), General Packet Radio Service (GPRS), Wideband Code Division Multiple Access (WCDMA), Enhanced Data Rates for Global Evolution (EDGE), Wireless Local Area Networks (WLAN), WiMAX and cdma2000 networks. The application layer protocol follows the pattern of requests going from client to server, and responses sent back from server to client.

The paper is structured as follows. In Section II, services that have to be provided by mobile telemetry application protocol are investigated. The security issues at application level are discussed in Section III. Section IV describes a mathematical model for evaluation of protocol efficiency. Before concluding the paper some remarks on future work ate outlined.

II. Common Protocol Requirements

The main requirements to application layer mobile telemetry protocol include access independency which calls for IP connectivity, operation with low energy consumption, and secure data transfer.

A mobile telemetry system consists of a central control unit that handles information gathering from multiple mobile sources and a set of mobile agents capable of domain data monitoring and measurement reporting.

The control unit plays an important role in registration of mobile agents. It is responsible for configuration of the operation modes of mobile agents regarding measurements and reports transmission. The control unit has to store all mobile agent and operation-related data of the telemetry system. The main data stored in the control unit include identities of the mobile agents, registration information and parameters related to the monitoring and reporting modes. It also stores a secret key for each mobile agent used to generate dynamic security data for each mobile agent. In addition to functions related to security, the control unit needs to store the measurement database or to provide interfaces to a measurement data repository if standalone solution is chosen.

The mobile agent is an embedded device equipped with sensor(s), positioning module, data transmission module and power supply module. The main requirement to mobile agent is to operate using as low energy as possible which calls for usage of energy efficient transmission methods [8].

Being an application layer protocol, the mobile telemetry protocol makes use of transport-layer protocols such as User Datagram Protocol (UDP) or Transmission Communication Protocol (TCP) in order to send and receive messages. UDP as a transport does not offer a reliable message-delivery service. Therefore, the mobile telemetry protocol, when forwarding a message to the UDP layer for transmission, has to expect that the message may not reach the destination. In order to cope with this limitation when using UDP, the mobile telemetry protocol must implement, as part of itself logic that guarantees the reliable delivery of messages. This logic basically utilizes retransmissions of messages upon expiry of timers in order to

[1]The authors are with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mails: ed_gospodinova@tu-sofia.bg; iia@tu-sofia.bg; enp@tu-sofia.bg

increase the probability of successful message delivery. The mechanism that may be used in order to implement this reliability at application layer is like the transaction concept defined for Session Initiation Protocol (SIP).

SIP possesses all required functionality for registration and management of sessions but it is not as suitable as one for mobile telemetry proposes. The main reasons are: the SIP session is a complex process, including many messages exchange; SIP messages are text-based which means considerable number of bytes; implementation of SIP compression in the mobile is undesirable because of the limited power budget. In order to decrease communication overhead, the mobile telemetry protocol must be binary.

Registration is performed after the mobile agent is switched on. The aim is to bind the IP address that is currently used by the mobile agent and mobile agent identity. Prior to registration which allows the mobile agent to operate in the telemetry system e.g. to perform measurements and send reports, the mobile agent must obtain IP connectivity. The registration contains two phases: first, the control unit challenges the mobile agent and then the mobile agent responds to the challenge and completes the registration as shown in Fig.1. In case of successful authentication, a new temporary identity is assigned to the mobile agent to avoid exposure of the unique mobile identity during transmission.

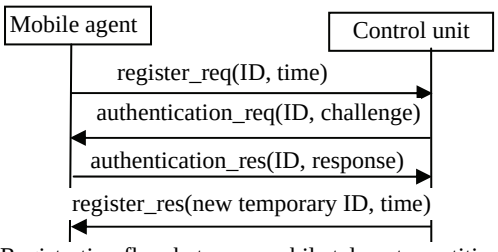


Fig.1 Registration flow between mobile telemetry entities

If the registration fails, for example, due to unsuccessful authorization, the control unit rejects the mobile agent's registration. It is the mobile agent's responsibility to keep its registration active by refreshing it periodically. If the mobile agent does not refresh its registration, then the control unit will silently remove the registration when the registration timer expires. The time for active registration may depend on the application domain. If the mobile agent is switched off, it performs de-registration setting the registration timer to zero. Sometimes the control unit needs to deregister the user, e.g. for administrative reasons. In this case, the control unit sends notification request to the mobile agent indicating that mobile agent has been deregistered as shown in Fig.2. If the mobile agent is unable to recognize the notification request it reports an error condition. Authentication procedure is directly coupled to registration and configuration of operation modes.

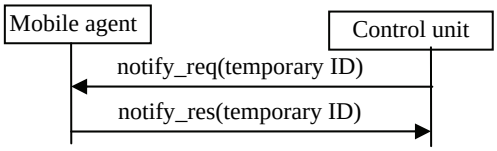


Fig.2 Control unit initiated de-registration of a mobile agent

The mobile agent may be configured to operate in different operation modes. Periodic monitoring and reporting require the mobile agent to make measurements and send reports at given rates. The measurement periods for the domain-related sensor function, the periods of gathering positioning data, and the periods for sending reports via the communication network may differ. This requires the mobile agent to possess functions for local data storage and processing. The frequency of reports is application dependent. The control unit requests for periodic reports after successful authentication as shown in Fig.3. The parameters of the request to start periodic reporting include the following: current time, needed for synchronization with the mobile agent, mobile agent temporary ID, dialogue ID, period of report transmissions, period of domain measurements, period of location measurements, accuracy of domain measurements, and accuracy of location measurements.

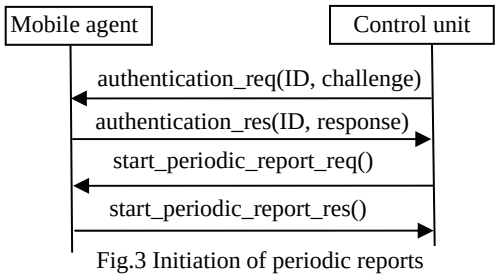


Fig.3 Initiation of periodic reports

In case of intensive reporting e.g. every minute, no mobile agent authentication may be required with each report but the first. The signaling flow for periodic reports is shown in Fig.4.

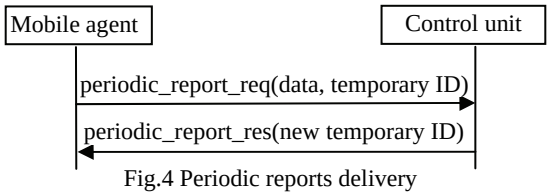


Fig.4 Periodic reports delivery

At any time the control unit may request for modification of the periodic reports. Any modifications of the reporting should take place after successful authorization. It is the responsibility of the control unit to stop the periodic reports which is also bound by authentication as shown in Fig.5.

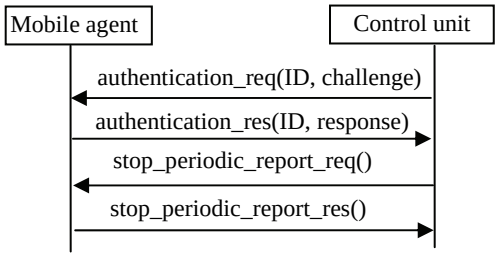


Fig.5 End of periodic reporting

Triggered reporting operation mode requires monitoring for certain criteria and submitting reports on their occurrence. The control unit starts the triggered reports by setting the reporting criteria. The parameters of the start trigger reporting request include current time, new temporary ID of the mobile agent, dialogue ID, trigger type (location or data values), location data (longitude, latitude, radius), threshold values of the measured data. The control unit may modify the trigger criteria after successful authentication as shown in Fig.6.

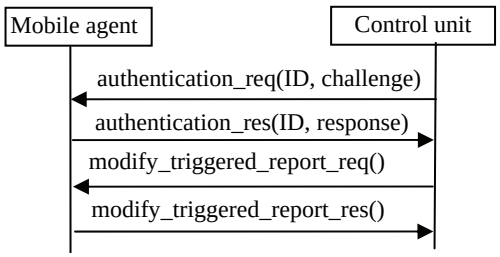


Fig.6 Modification of trigger criteria

When the criteria for triggered reporting are met, the mobile agent sends reports as shown in Fig.7.

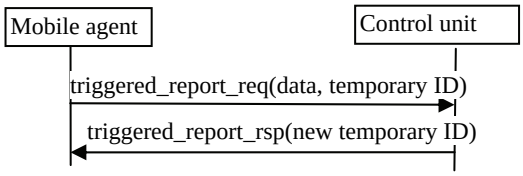


Fig.7 Triggered reporting

The termination of the triggered reporting is initiated by the control unit and it is also bound by authentication.

The third mode of operation is reports upon request. The control unit induced reporting requires mobile agent to perform measurements and to send reports on demand as to Fig.8. The mobile agent response in Fig.8 just acknowledges that the request is accepted. Fig.9 shows the signaling flow for reporting the measurement data upon requests.

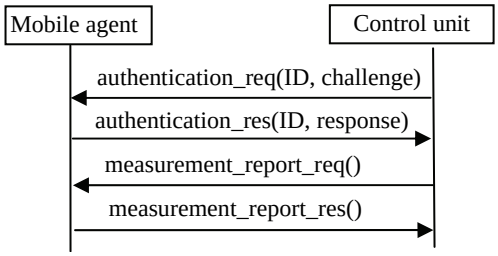


Fig.8 Measurement on demand

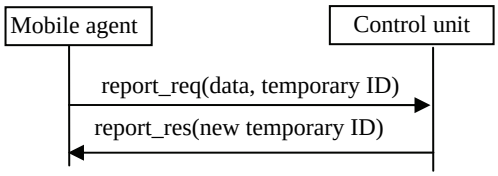


Fig.9 Delivering report on request

Any errors concerning problems in configuration or modifications of the operation modes are reported. In case of error, the mobile agent indicates the error type. The possible error types include: unsuccessful request for start/ modification/stop of periodic reports, unsuccessful request for start/modification/stop of triggered reports, unsuccessful request for report on demand. Receiving report the control unit also may response by indicating an error. Possible reasons for request rejection may be format error, invalid identification, invalid parameter values etc. Fig.10 shows an error case when the mobile agent responses to a start request for periodic reports caused by wrong message code. Fig.11 shows the signaling flow when the control unit indicates that the triggered report fails.

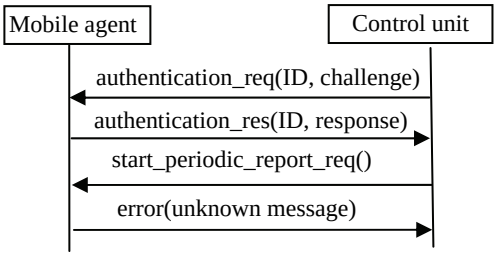


Fig.10 Rejection of request to start periodic reports

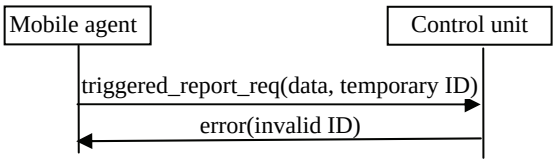


Fig.11 Rejection of triggered report

III. Requirements to Security

The access security is responsibility of the network that provides IP connectivity. In addition to access security, the security mechanisms at upper layers are mandatory.

The basic tool used for protection of IP network traffic is the IP Security (IPSec) protocol suite. It provides confidentiality and integrity of communication in the IP layer. Communicating parties can also authenticate each other using IPsec [9]. The critical issue is key management: how to generate, exchange and distribute the various keys needed in algorithms that are used to provide confidentiality and integrity protection. This security mechanism is unsuitable for low class embedded mobile devices.

At application layer, the security mechanisms for mobile telemetry need to include the following:

- mutual authentication;
- integrity check;
- temporary identifications;
- optional ciphering.

The authentication is required on registration, configuration and modification of operation modes. The aim of authentication is twofold: the control unit needs to authenticate the mobile agent and the mobile agent needs to verify if the control unit is authorized to request authentication. The authentication procedure relies on shared secret stored in the mobile agent and in the control unit. Each time an authentication is required the control unit computes an authentication token and sends it in the authentication request together with the challenge. The mobile agent uses the authentication token to check the authority of the request. If the authentication token does not coincide with the mobile agent's expectation the authentication request is rejected. If the control unit is authorized to request authentication the mobile agent computes its response using the challenge and the shared secret.

The integrity check has to be mandatory for all requests and responses. Messages are checked to ensure they have been received intact.

The ciphering is optional. The control unit may decide to activate the ciphering mode as shown in Fig.12. The control unit also decides when to stop ciphering.

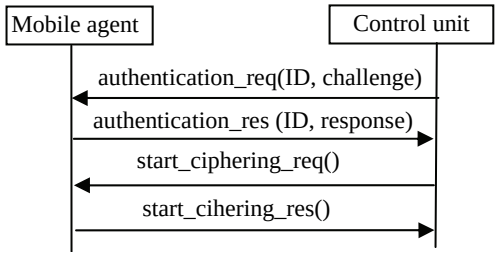


Fig.12 Activation of ciphering mode

The mobile agent has a unique identity (ID). This identity is used for authentication and registration. Temporary identity is used to protect the unique ID from disclosure. New temporary ID is assigned on registration, configuration or modification of operation modes.

IV. Protocol Mathematical Model

To evaluate efficiency of the mobile telemetry protocol, the metric used is message arrival delay. The mathematical model is characterized by the following assumptions. The message loss rate is α. The number of maximum message transmissions is N. The initial timeout before the first re-transmission is T. We assume that the timeout between re-transmissions increases exponentially.

The probability of message arrival exact after i-th retransmission is:

$$P_i = \alpha^{i-1}(1-\alpha) \quad (1)$$

Then the probability of message arrival within N retransmissions is:

$$P_N = \sum_{i=1}^{N} \alpha^{i-1}(1-\alpha) \quad (2)$$

The overall message loss probability is:

$$Q_N = 1 - P_N \quad (3)$$

In exponential back off case, the expected protocol delay for message arrival at the corresponding entity (mobile agent or control unit) is expressed by:

$$W = (1-\alpha).T.\sum_{i=1}^{N}(\alpha^{i-1}(2^{i-1}-1)). \quad (4)$$

Having $\alpha = 10^{-3}$, $N = 4$ and $T = 500$ *ms*, the expected message arrival delay is $W = 0.501$ *ms*.

V. Conclusion

In this paper we study the requirements to application level mobile telemetry protocol. We argue that the protocol needs to be based on IP connectivity to provide access independency and define the basic protocol functions. In order to be transport independent, the main protocol requirement is to provide a mechanism for reliable message delivery. The requirements to security functions include mutual authentication, integrity check to protect against any manipulation, and confidentiality. Our future work will be focused on different schemes for message re-transmission in order to decrease the protocol latency considering message processing time. At the protocol design phase, we will evaluate the protocol overhead.

Acknowledgement

The work is funded by the project grant 112pd004-7/2011 at Research and Development Sector, TU-Sofia, Bulgaria.

References

[1] M. Huerta, T. Viva, R. Clotet, R. Gonzalez, R. Alvizu, A. Perez, D. Rizas, F. Lara, and R. Esclante, "Implementation of a Open Source Security Software Platform in a Telemedicine Network Advances in E-Activities", *Information Security and Privacy*, pp. 72-76, 2009.

[2] V. Sanjeevi, Veluchandhar, S. Sakthivel, and M. Supriya, "Security Policy for Deducting Unauthorized IP Based Mobile Host Inside the Network", *7th WSEAS Int. Conf. on Electronics*, Hardware, Wireless and Optical Communications, Cambridge, UK, pp. 177-181, 2008.

[3] A. Georgiades, Y. Luo, A. Lasebay, and R. Comley, "Location Privacy in Mobile IPv6 Distributed Authentication Protocol Using Mobile Home Agents", *8th WSEAS Int. Conf. on Electronics*, Hardware, Wireless and Optical Communications, Cambridge, UK, pp.243-248, 2009.

[4] M. Worris, "Single-chip ZigBee for Indoor Mobile Telemetry", *IEE Seminar on Telemetry and Telematics 2005*, (Refl No 2005/11009), pp.10/1-10/4, 2005.

[5] S. Carrellas, "Mobile Telemetry as a Data Gathering Tool for the Advanced Mobile Phone Service Field Trial", *32nd IEEE Vehicular Technology Conference*, pp. 327-330, 2006.

[6] G. Horler, S. Hindle, and D. McGorman, "Including Coupled Telemetry and Actuation", *IEE Seminar on Telemetry and Telematics 2005*, (Refl No 2005/11009), pp.5/1-5/6, 2005.

[7] H. Nassar, H. Al-Mahdi, M. El-Gabali, and S. Aziz, "Design and Analysis of a One-Step Addressing Protocol for a Ad Hoc Networks", *7th WSEAS Int. Conf. on Electronics*, Hardware, Wireless and Optical Communications, Cambridge, UK, pp. 140-145, 2008.

[8] H. Gochev, V. Poulkov, and G. Iliev, "Uplink Power Control for LTE Improving Cell Edge Throughput", *Int. Conference on Telecommunications and Signal Processing, TSP 2010*, Baden near Vienna, Austria, pp. 465-467, 2010.

[9] C. Wu, S. Wu, and R. Narayan, "IPsec/PHIL (packet header information list): Design, Implementation and Evaluation", *10th International Conference on Computer Communications and Networks*, Scottsdale, AZ, USA, pp. 206-211, 2001.

Third Party Policy Management in Multimedia Networks

Dora Marinska[1], Ivaylo Atanasov[1] and Evelina Pencheva[1]

Abstract – **In this paper the capabilities for open access to policy management functions in all Internet protocol based multimedia networks are investigated. The deployment of Open Service Access Policy Management application programming interfaces in architectures that support Policy and Charging Control and User Data Convergence is studied. A mapping of Policy Management interfaces onto Lightweight Directory Access Protocol, Hypertext Transfer Protocol and Diameter protocol is suggested. Examples of policy creation and policy evaluation are presented.**

Keywords – **Open service access, Policy management, User Data Convergence, Policy and Charging Control.**

I. Introduction

Internet Protocol Multimedia Subsystem (IMS) is access independent service control architecture that enables all kind of multimedia applications over a common IP based backbone. One of the main requirements to IMS is to provide Quality of Service (QoS) control which means capabilities for authorization and usage control on bearer resources intended for IMS traffic. That is why a Policy and Charging Control (PCC) functionality is defined. The PCC encompasses policy control decisions based upon subscription information and flow based charging control functionalities [1].

It is expected that more and more multimedia services will use policies to express operational criteria [2], [3]. It is also expected that network providers will host policy-enabled services that have been created by 3rd party application service providers [4]. In order to manage policy information, to control access to it and to request evaluation of policies, a policy management service is needed.

The necessity of open access to QoS control is substantiated in [5]. A protocol-based approach to QoS management, where the collected information about the network is provided to external applications in a structurally standardized format is presented in [6]. Standardized application programming interfaces (API) for open access to policy management functions are defined in [7]. The Open Service Access (OSA) Policy Management service offers provisioning services and facilitates interactions between 3rd party applications and the policies of any policy enabled service. Implementation of OSA API requires translation between OSA API and network control protocols. The motivation behind the research is the lack of mapping of OSA Policy Management API onto network control protocols.

The paper investigates the capabilities for open access to offer provisioning services in IMS networks. In Section II, architecture for deployment of OSA Policy Management API in IMS networks with User Data Convergence (UDC) and PCC functionality is suggested. The proposed mapping of OSA Policy Management API onto UDC and PCC control protocols is presented respectively in Section III and Section IV. Examples of policy creation and policy evaluation are designed to illustrate the deployment of 3rd party Policy Management applications in IMS.

[1]The authors are with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mails: dmarinska@gcn.bg; iia@tu-sofia.bg; enp@tu-sofia.bg

II. Architecture for Third Party Policy Management

The OSA Policy Management API may be used to create, update or view policy information for any policy enabled service as well as to subscribe to policy events, to request evaluation of policies, and to request the generation of policy events. The API may be used for policy provisioning and policy evaluation in the context of QoS management in IMS networks with UDC and PCC functions.

The suggested architecture for OSA Policy Management API deployment is shown in Fig.1. OSA Application server runs OSA applications which use Policy Management API to create policy provisioning services for QoS management and to evaluate the provided QoS. OSA gateway is a special type of application server that provides OSA API toward applications and control protocols toward the network. User Data Repository (UDR) stores subscription data as to the architectural model of 3GPP [8]. In the suggested architecture, it plays a role of policy server that stores policy information related to QoS. The OSA gateway communicates with UDR over the Ud reference point. At the application level, the protocol for Ud data access messages is Lightweight Directory Access Protocol (LDAP) and the protocol for Ud subscriptions/notifications is Simple Object Access Protocol (SOAP)/HTTP(S) [9]. Policy and Charging Rule Function (PCRF) encompasses policy control decisions and flow based charging control functionalities. In the suggested architecture, the PCRF plays a role of policy engine which evaluates QoS related policy rules provisioned by 3rd party applications. The Ud reference point is used by the PCRF to access PCC related subscription data when stored in the UDR [8]. The Gx reference point is used by the PCRF to provide network control regarding the service data flow detection, gating control, QoS and flow based charging (except credit management) towards Policy and Charging Enforcement Function (PCEF). The PCEF encompasses service data flow detection, policy enforcement and flow based charging functionalities. The Rx reference point enables transport of application level session information from OSA gateway to PCRF. The Rx reference point also enables the OSA gateway subscription to notifications on bearer level events in the IP

connectivity access network. The protocol over the Rx and Gx reference points is Diameter [10].

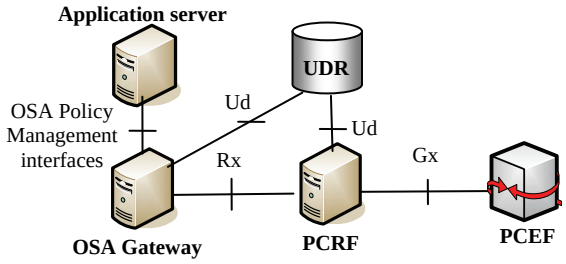


Fig. 1. OSA Policy Management API deployed in IMS

III. Policy Management over Ud Reference Point

This section presents the suggested mapping of OSA Policy Management API onto UCD control protocols.

Applications that perform administrative tasks of behalf of a policy enabled service, e.g. create, update or delete policy information must obtain access to IpPolicyManager. Administrative tasks may then be performed using methods supported by IpPolicyManager.

All procedures related to querying or deleting data from the UDR and creating or updating data within the UDR are controlled by LDAP protocol. The subscription/notification messages related to changes in user data stored within the UDR are transferred by HTTP in SOAP envelopes.

The OSA method startTransaction opens a transaction and all modifications to the policy information base up to the call to either method commitTransaction or abortTransaction will be treated as part of this transaction. To initiate a LDAP session the OSA gateway first establishes a transport connection with the UDR and then initiates a LDAP session by sending a LDAP BindRequest message. In order to allow the application to relate a number of operations and to have them performed in one unit of interaction, a transaction is used as shown in Fig.2.

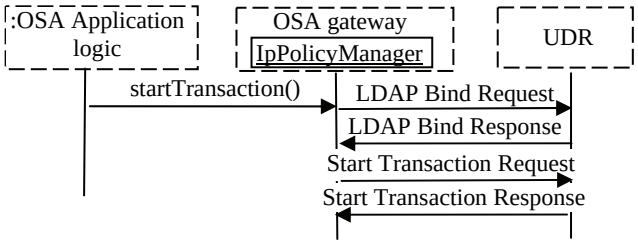


Fig. 2. Opening a link for LDAP session and transaction for modifications of policy information

The OSA method commitTransaction makes all modifications to the policy information base made since the last call to startTransaction to be committed. The OSA gateway ends the transaction within the UDR and if it is the last transaction a termination of the LDAP session will be done as shown in Fig.3.

Applications that wish to participate in Policy Management process use the methods of the IpPolicyManager interface to access a policy domain of interest, iterate through the names of all policy domains, create a new policy domain, or remove an existing one.

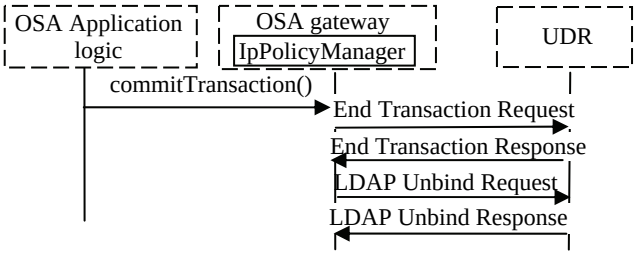


Fig. 3. Closing a transaction and the link for LDAP session

Applications may also use these methods to access policy repository, iterate through the names of all policy repositories, create a new policy repository or remove an existing one. The OSA gateway translates the method invocations in LDAP Query, Create, Delete and Update messages for UDC. Fig.4 shows a top-level policy domain creation and how a reference to a specified top-level repository is got.

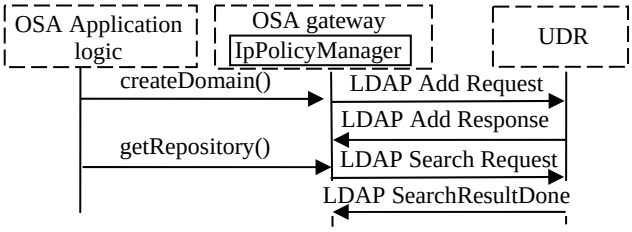


Fig. 4. Creation of policy domain and getting a reference to a specified top-level repository

The IpPolicy is an OSA interface from which are derived all of the Policy interfaces except IpPolicyManager. Using the methods of IpPolicy interfaces the application may manage policy domains, policy groups, policy rules, or policy event definitions. Fig.5 shows the usage of IpPolicy interface to set a generic attribute of a policy object.

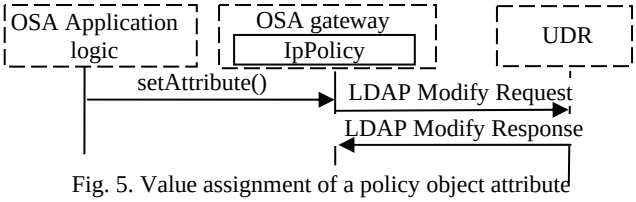


Fig. 5. Value assignment of a policy object attribute

Fig.6 shows how the application removes a specified rule.

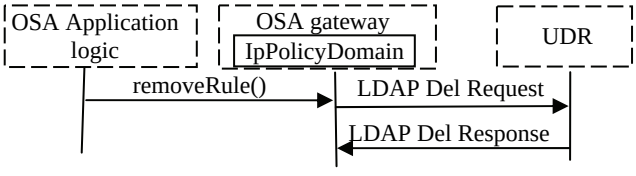


Fig. 6. A policy rule deletion

The OSA Policy evaluation interfaces address evaluation of policy rules upon request of a client application, subscription to and receiving notification of policy events, and ability for authorized clients to generate events. Using the

IpPolicyEvalManager interface an authorized application may request evaluation of policy rules, subscribe to and receive notification of events and to generate policy events. The createNotification method allows applications to specify a set of events that they are interested in receiving, e.g. any changes in policy information, adding policy rules and so on. The distroyNotification method allows the application to indicate that it is no longer interested in receiving events that it previously subscribed to. On receiving subscription request the OSA gateway translates it in HTTP Post method contained in a SOAP message envelope as shown in Fig.7.

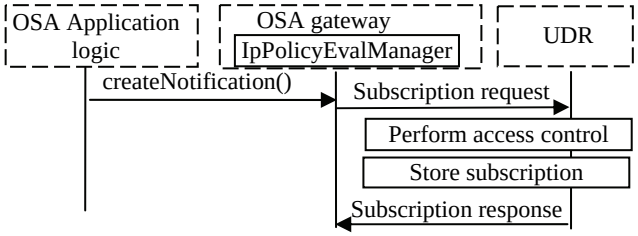


Fig. 7. Subscription to policy information changes

Using the OSA IpAppPolicyDomain interface applications may receive values or error messages resulting from asynchronous method calls sent to this interface. The reportNotification method notifies about the specified event. The UDR notifies the OSA gateway about any changes in policy information using Notify request messages sent by the HTTP Post method in a SOAP message envelope. Notify response messages are coded as HTTP response message and contain a SOAP message envelope as shown in Fig.8.

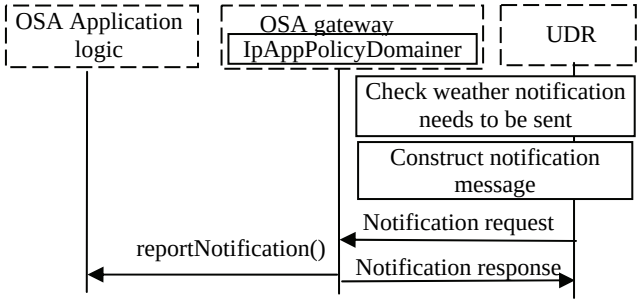


Fig. 8. Notification about policy information change

IV. Policy Management over Rx Reference Point

This section presents the suggested mapping of OSA Policy Management API onto PCC control protocol Diameter.

Using operator defined policy rules the PCRF is able to apply QoS control on a per service data flow basis in the PCEF, to support control of QoS reservation procedures for bearer resources, and to apply usage monitoring for the accumulated usage of network resources. The Rx reference point defines transport of application level session information from OSA gateway to PCRF. The Rx reference point enables the application server subscription to notifications on bearer level events (e.g. signaling path status of multimedia session) in the IP connectivity access network.

Using the evalPolicyReq method of the OSA IpPolicyEvalManager interface the application may invoke an evaluation of policy rules. The evalPolicyRes or evalPolicyErr methods of the OSA IpAppPolicyDomain interface are invoked to pass back the results of an evalPolicyReq invocation. The Diameter messages AA-Request (AAR) and AA-Answer (AAA) over the reference point Rx may be used to request notifications of specific IP connectivity access network session events. In case of successful subscription for notifications, the PCRF sends a Re-Auth-Request (RAR) command when a corresponding event occurs. The OSA gateway acknowledges by Re-Auth-Answer (RAA) command. Applications may use the createEventDefinition method of the OSA IpPolicyDomain interface to define a new event type, specifying the definition's name and the required and optional attributes that appear in an instance of that event. Having a subscription for notifications of policy events, the application is notified when an event occurs. The application may also invoke an evaluation of policy rules related to the specified event type. Fig.9 shows the information flow when the application requests an evaluation of policies associated for example with QoS events related to a given session in the IP connectivity access network.

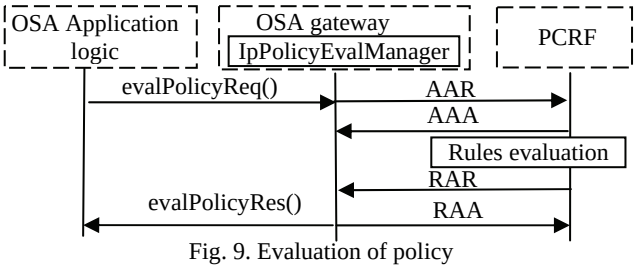


Fig. 9. Evaluation of policy

V. Example Application

To illustrate the applicability of the suggested mapping we design an application that uses OSA Policy Management API.

Let us assume that an enterprise operator has used the OSA Connectivity Manager interfaces to establish QoS parameters for enterprise network packets travelling through the provider network. Using these interfaces the enterprise operator can create virtual provisioned pipes in the provider network to carry the enterprise traffic and support it with pre-specified QoS attributes. The OSA Connectivity Manager interfaces have been used to set and negotiate QoS parameters between the enterprise operator and provider network operator but these interfaces do not support means for evaluation of the QoS provided. For the purpose OSA Policy Management interfaces may be used. The enterprise operator may decide to enable a policy rule to be hosted in the policy engine (PCRF) of the provider network operator and thus to evaluate the QoS parameters of the virtual provisioned pipes created in the provider network. The enterprise operator then subscribes for notifications of events related to the actual QoS parameters with respect to the negotiated one monitored by the subscription mechanism. The sequence diagram is shown in Fig.10. First the policy event is defined and then the event is triggered when the action part of a rule fires. This may happen when, e.g. a pre-defined threshold (say, a guaranteed bit rate)

is reached causing the conditions of a policy rule to be satisfied thus resulting in execution of the action part.

VI. Conclusion

The paper investigates the capabilities for open access to policy management functions in all IP-based multimedia networks. The behavior of OSA gateway that supports Policy Management interfaces is studied. A functional mapping of OSA Policy Management interfaces for policy provisioning onto User Data Convergence protocols LDAP and HTTP/SOAP is suggested. Further, OSA Policy Management interfaces for policy evaluation are mapped onto Diameter protocol in Policy and Charging Control architecture. An example application is designed to illustrate the applicability of open access to policy creation and evaluation.

The results may be used in implementation of open access to policy management functions in multimedia networks.

Acknowledgement

The work is conducted under the grant of the Project 102ni107-7 funded by Research and Development Sector, TU-Sofia, Bulgaria.

References

[1] 3GPP TS 23.203 “Policy and Charging Control Architecture”, Release 11, v11.0.1, 2011.

[2] S. Selvakumar, S. Xavier, and V. Balamurugan, “Policy Based Service Provisioning System for WiMAX Network: An Approach”, *ICSCN’08 Conference Proceedings*, issue 4-6, pp. 177-181, 2008.

[3] F. Zhao, L. Jiang, and C. He, “Policy-Based Radio Resource Allocation for Wireless Mobile Networks”, *Neural Networks and Signal Processing 2008 Conference Proceedings*, pp. 476-481, 2008.

[4] K. Haddadou, S. Ghamri-Doudane, Y, Ghamri-Doudane, and N. Agoulmine, “Designing Scalable On-demand Policy-Based Resource Allocation in IP Networks, *IEEE Communications Magazine*, vol. 44, issue 3, pp. 142-149, 2006.

[5] M. Elkotob, “Autonomic Resource Management in IEEE 802.11 Open Access networks”, Dissertation, Lules University of Technology, Sweden, http://epubl.ltu.se/1402-1757/2008/38/LTU-LIC-0838-SE.pdf, 2009.

[6] M. Katchabaw, H. Lutfiyya, and M. Bauer, “Usage Based Service Differentiation for End-to-End Quality of Service Management”, *Computer Communications*, vol. 28, issue 18, pp. 2146-2159, 2005.

[7] 3GPP TS 29.198-13 “Open Service Access (OSA) Application programming Interfaces (APIs); Part 13: Policy Management Service Capability Feature (SCF)”, Release 9, v9.0.0, 2009.

[8] 3GPP TS 23.335 “User Data Convergence (UDC); Technical realization and information flows”, Release 9, v9.3.0, 2010.

[9] 3GPP TS 29.335 “User Data Convergence (UDC); User Data repository Access Protocol over the Ud interfaces”, Release 9, v9.2.0, 2010.

[10] 3GPP TS 29.214 “Policy and Charging Control Architecture over Rx Reference point”, Release 10, v10.1.0, 2010.

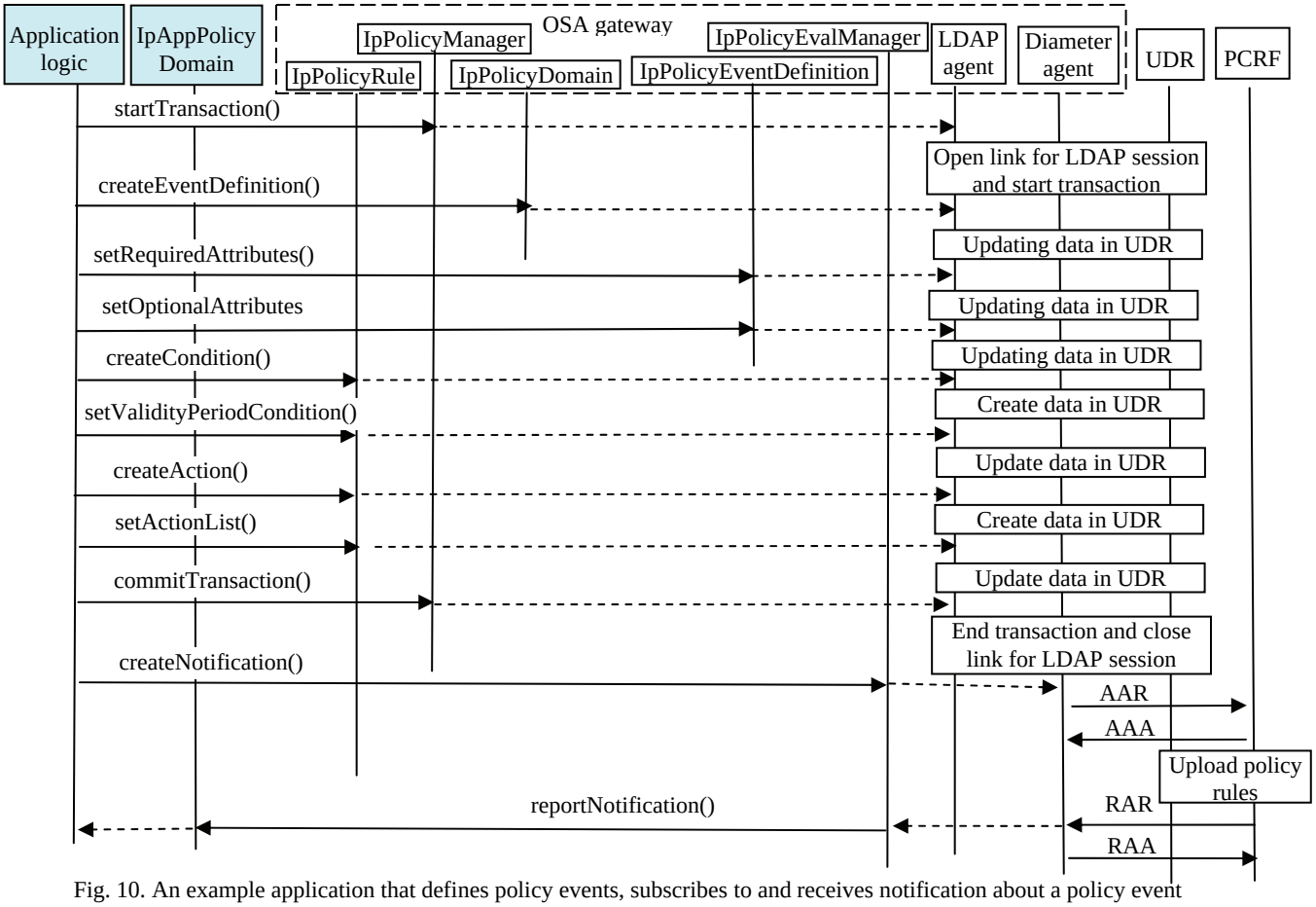


Fig. 10. An example application that defines policy events, subscribes to and receives notification about a policy event

Estimation of Optical Receiver Sensitivity in HFC Network

Krasen Angelov[1], Stanimir Sadinov[2], Kiril Koitchev[3]

Abstract – **An essential parameter in determining the system power budget in an optical part of the hybrid fiber coaxial (HFC) network is optical receiver sensitivity, defined as the minimum average optical power for a given bit-error rate (*BER*). When designing a good optical receiver, it is critical to understand the different parameters that will impair overall receiver sensitivity. This article provides an analysis of receiver optical sensitivity. The analysis is based on normal receiver sensitivity, assuming an ideal input signal with negligible impairment from factors like inter-symbol interference (ISI), rise/fall time, jitter, and transmitter relative intensity noise (*RIN*).**

Keywords – **optical receiver sensitivity, bit error rate, inter-symbol interference, transimpedance and limiting amplifier.**

I. INTRODUCTION

In optical communication systems, sensitivity is a measure of how weak an input signal can get before the bit-error ratio (*BER*) exceeds some specified number. The standards body governing the application sets this specified *BER* [4,6]. For example, Gigabit Ethernet and Fibre Channel specifications require a *BER* of 10^{-12} or better. This BER is the foundation for determining a receiver's sensitivity.

In the design of an optical receiver, such as a HFC optical node, it is vital that the module be capable of converting and shaping the optical signal while meeting or surpassing the maximum *BER*. Ultimately, the influence of noise on the signal will determine the sensitivity of the system [1-3,9-12]. The portion of the receiver that contributes the most noise is the optical-to-electrical conversion provided by the photodetector and the transimpedance amplifier (TIA) [5,8,11].

Sensitivity can be expressed as average power (P_{AVG}) in *dBm* or as optical modulation amplitude (*OMA*) in W_{P-P} (peek-to-peek). Each gives a figure of merit for the receiver. The sensitivity is the minimum *OMA* or P_{AVG} at which the maximum (worst tolerable) *BER* can be maintained [1,2,14]. Optical transmission system designers use sensitivity to determine the maximum distance or link margin available in their system. Expressing the sensitivity in terms of average power is useful, because the average power of a laser is more easily measured than peak-to-peak power. Measuring the peak-to-peak power of a laser at high data rates requires expensive equipment that is error-prone due to the amount of operator intervention. Average optical power can be measured easily and reliably with a relatively inexpensive optical power meter.

[1]Krasen Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@tugab.bg

[2]Stanimir Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

[3]Kiril Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University of Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

II. DETERMINING THE Q-FACTOR

A typical optical receiver is composed of an optical photo detector, a transimpedance amplifier (TIA), a limiting amplifier (LA), and a clock-data recovery (CDR) block. Fig. 1 shows a simple block diagram of the front end of an optical receiver module.

The dominant noise sources in this section are the linear components that provide the optical-to-electrical conversion, namely, the photodiode and the TIA [10,11]. Transimpedance amplifiers are used to amplify and convert the photodiode current into a voltage.

The received optical signal is first converted into photocurrent and amplified by the TIA. The limiting amplifier (LA) acts as a "decision" circuit, where the sampled voltage $v(t)$ is compared with the decision threshold V_{TH}. At this data decision point, the signal is significantly degraded by the accumulation of random noise and inter-symbol interference (ISI), resulting in erroneous decisions due to eye closure [11,12].

To know the relationship between *BER* and eye opening at data decision, the statistical characteristics of the amplitude noise need to be determined. Usually, as a figure of merit, it can be used signal Q-factor to measure the signal quality for determining the BER [12]. If the ISI distortion does not exist and the dominant amplitude noise has Gaussian distribution, the signal Q-factor is defined as:

$$Q = \frac{V_1 - V_0}{\sigma_1 - \sigma_0} \tag{1}$$

In a practical receiver implementation, ISI exists due to receiver bandwidth limitation, baseline wander, or non-linearity of the active components. If the signal eye diagram is monitored before the data decision, it is evident that in addition to random noise, the signal has a certain amount of bounded amplitude fluctuation caused by ISI, which exhibits strong pattern dependence. To estimate the ISI penalty on optical sensitivity, a simple solution is to consider a worst-case amplitude-noise distribution. This is done separately by shifting the mean value of the Gaussian distribution from V_1

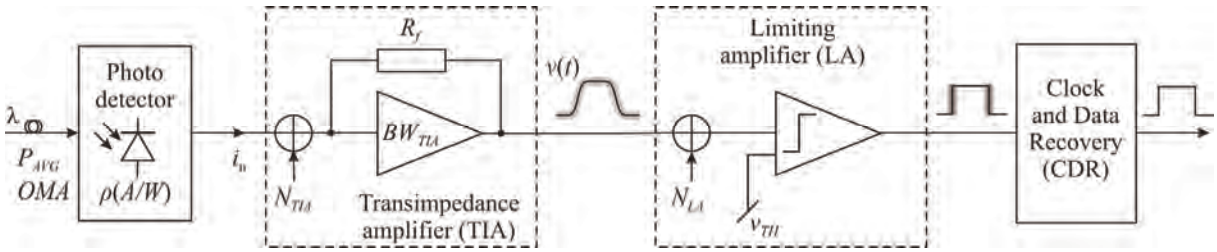


Fig. 1. Simplified block diagram of the optical receiver module

and V_0 to the lower amplitude boundary ($V_1 - V_{ISI}$) and ($V_0 + V_{ISI}$) – Fig. 2. It is assumed that V_{ISI} is the vertical eye closure caused by ISI.

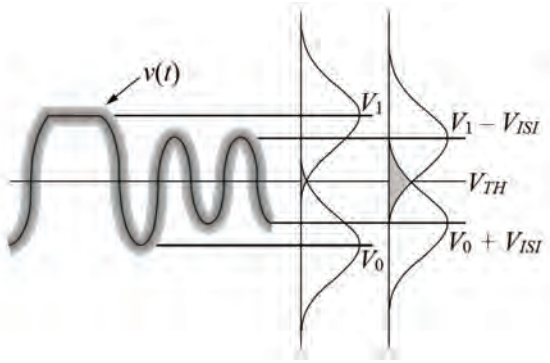


Fig. 2. Worst-case amplitude-noise distribution in the presence of ISI

Under this condition, the signal *Q*-factor can be obtained by calculating the *BER* from the worst-case noise distribution. Assuming the decision threshold is optimized for minimum *BER*, the *Q*-factor is related to vertical eye closure V_{ISI} as follows:

$$Q = \frac{V_1 - V_0 - 2V_{ISI}}{\sigma_1 + \sigma_0}. \quad (2)$$

$$BER = \frac{1}{2} erfc\left(\frac{Q_{BER}}{\sqrt{2}}\right). \quad (3)$$

where $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-v^2} dv$. Q_{BER} is the minimum required *Q*-factor for a given *BER*. Based on Eq. (3), the relationship between Q_{BER} and *BER* is plotted in Fig. 3 [12,13].

Usually it is measured the signal peak-to-peak differential swing ($V_{P\text{-}P} = V_1 - V_0$) and assuming $\sigma_1 = \sigma_0 = N_{RMS}$, so the *Q*-factor becomes:

$$Q = \frac{V_{P-P} - 2V_{ISI}}{2N_{RMS}}. \quad (4)$$

where N_{RMS} is the equivalent RMS noise at the input of the limiting amplifier.

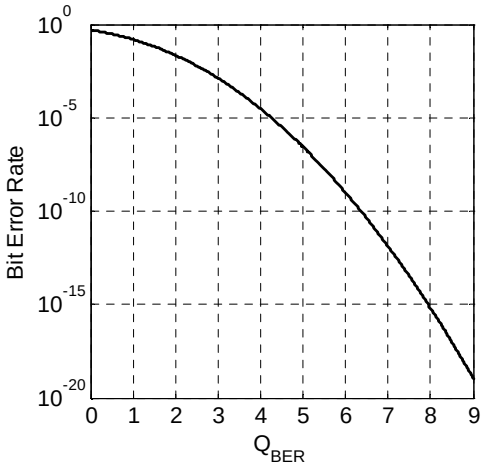


Fig. 3. *BER = f(Q_{BER})*

III. Optical Receiver Sensitivity Estimation

To achieve the best optical sensitivity, it is important to maximize the signal *Q*-factor before data decision. Here is demonstrated an estimation of the optical receiver sensitivity with practical device implementations, when overall receiver random noise and ISI are taken into account.

The equation for calculating sensitivity is as follows [13,14]:

$$P_{AVG} = 10\log\left(\frac{i_n SNR(r_e + 1)}{2\rho(r_e - 1)} 1000\right), \quad dBm. \quad (5)$$

i_n is the noise of TIA; ρ – responsivity flux (conversion efficiency) of the photodetector, in A/W; r_e – the ratio of a logic-one power level (P_1) relative to a logic-zero power level (P_0) and it can be expressed as $r_e = P_1 / P_0$ or $r_e = 10\log(P_1 / P_0)$, dB.

Eq. (5) assumes that all of the noise in the system is due to the TIA. It also assumes that the LA following the TIA has a decision threshold of zero ($V_{TH} = 0$).

The noise of the TIA, i_n, is expressed as "input referred noise" in RMS current (A_{RMS}) or "input referred noise density" in ($A_{RMS} / \sqrt{Hz}$). This is the inherent noise of the amplifier. Input referred noise is directly proportional to the value of the photodiode capacitance and bandwidth of the TIA [10].

The process in estimating the minimum peak-to-peak swing of the optical signal begins with the choice of the maximum *BER* [13]. This determines the signal-to-noise ratio (*SNR*). Next, the RMS input referred noise (i_n) of the TIA and the responsivity (ρ) of the photodetector must be found from the vendor's data sheets. These are related by:

$$OMA_{\min} = \frac{i_n SNR}{\rho}. \quad (6)$$

This relationship assumes that the noise is Gaussian.

A. Overall receiver penalty

To estimate the receiver total RMS noise impact on optical sensitivity, it is necessary to know the minimum required peak-to-peak current at the TIA input (noted as $I_{P\text{-}P}$) that will result in a specified *BER*. For this random noise analysis it is assumed $V_{ISI} = 0$, and $I_{P\text{-}P}$ can be obtained by substituting $V_{P\text{-}P} = I_{P\text{-}P}.R_f$ and $N_{RMS} = N_{TOTAL}.R_f$ in Eq. (4), resulting in:

$$I_{P-P} = 2Q_{BER} N_{TOTAL}, \quad \mu A_{RMS}, \quad (7)$$

where N_{TOTAL} is the total equivalent RMS noise at TIA input, which is determined by the TIA input-referred noise N_{TIA} (in μA_{RMS}), the limiting-amplifier input-referred noise N_{LA} (mV_{RMS}), and the TIA small-signal transimpedance gain R_f ($k\Omega$). The relationship is shown as:

$$N_{TOTAL} = \sqrt{N_{TIA}^2 + \left(\frac{N_{LA}}{R_f}\right)^2}. \quad (8)$$

In practice, the input-referred noise N_{LA} may not be given, but it can be estimated from the limiting-amplifier input-sensitivity V_{LA}, a measure of the minimum differential peak-to-peak signal swing to achieve a given *BER*. In general, the limiting-amplifier sensitivity could result from the input-referred noise N_{LA}, DC-offset, or ISI due to bandwidth limitation [10,14]. The random noise is the dominant factor for limiting amplifier sensitivity. Under this condition, N_{LA} can be estimated from the following equation:

$$N_{LA} = \frac{V_{LA}}{2Q_{BER}}. \quad (9)$$

Substituting the Eq. (7) in Eq. (6) and assuming $Q_{BER} = 7{,}1$ for $BER = 10^{-12}$ (Fig. 3), the $OMA_{(N)}$ is obtained as:

$$OMA_{(N)} = \frac{I_{P-P} SNR}{\rho}, \quad \mu W. \quad (10)$$

B. Intersymbol interference penalty

In an optical receiver, ISI can result from the following sources: high-frequency bandwidth limitation; insufficient low-frequency cutoff caused by AC-coupling or DC-offset cancellation loop; in-band gain flatness; or multiple reflections between the interconnection of a TIA and a LA [10,14]. ISI results in eye closure in both amplitude and timing.

The *ISI* due to vertical eye closure is defined as:

$$ISI = \frac{2V_{ISI}}{V_{P-P}}. \quad (11)$$

The minimum-required TIA input current is related to ISI according to:

$$I_{P-P(ISI)} = \frac{2Q_{BER} N_{TOTAL}}{(1 - ISI)}, \quad \mu A_{RMS}. \quad (12)$$

Substituting the Eq. (12) in Eq. (6) and assuming $Q_{BER} = 7{,}1$ for $BER = 10^{-12}$ (Fig. 3), the OMA_{ISI} is obtained as:

$$OMA_{(ISI)} = \frac{I_{P-P(ISI)} SNR}{\rho}, \quad \mu W. \quad (13)$$

The ISI penalty is defined as the difference in optical sensitivity in the presence of ISI, as compared to an ideal case when *ISI* = 0. The calculated result is shown in Fig. 4.

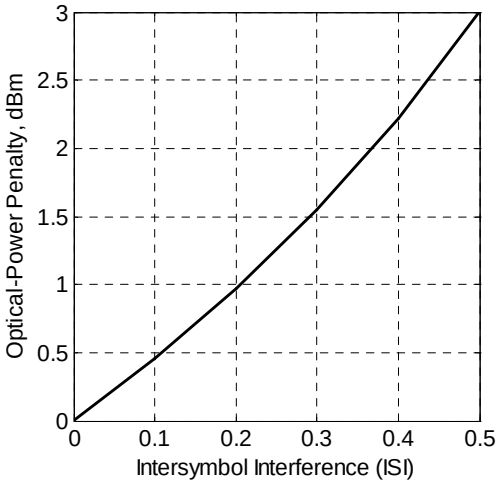


Fig. 4. Optical-power penalty versus ISI for an ideal case

C. Optical receiver sensitivity

Finally, the total optical power penalty in dB is the sum of the ISI penalty and the overall random-noise penalty.

For estimation of optical receiver sensitivity limited by the receiver noise impact and ISI impact, Eqs. (10) and (13) are substituted in Eq. (5).

Examples are given for optical receiver using MAXIM devices MAX3277 TIA and MAX3272 LA for HFC fiber channel applications. The datasheet parameters are as follows: $N_{TIA} = 0{,}35\mu A$, $R_f = 3{,}3k\Omega$, $N_{LA} = 0{,}22mV$ [15,16]. Assuming $r_e = 6{,}6$ and two values of $\rho = 0{,}85A/W$ and $0{,}65A/W$, the calculated optical sensitivity (P_{AVG}) versus Signal-to-Noise Ratio (*SNR*) is shown in Fig. 5.

The results on Fig. 5 are based on Eqs. (5), (10) and (13). It is shown the minimum required optical sensitivity P_{AVG} for a

given *SNR*. For example, when the *SNR* = 14,1*dB* (which is equivalent to *BER* = 10^{-12} or Q_{BER} = 7,1 [12]) the optical sensitivity is –21,78*dBm* when ρ = 0,85*A/W* and –19.45*dBm* when ρ = 0,65*A/W*. For the receiver with higher responsivity of the photodetector the optical sensitivity will be lower.

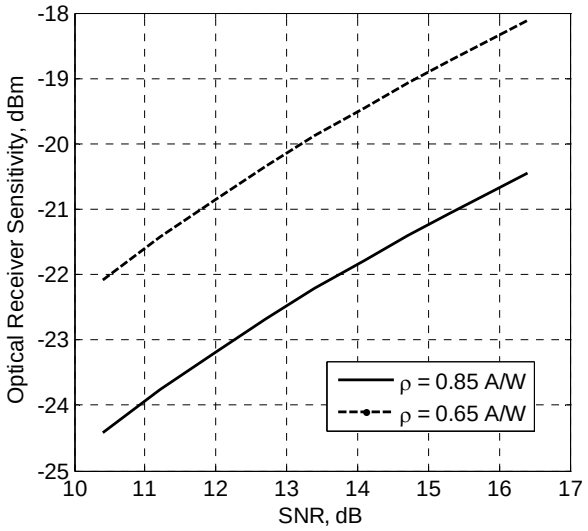


Fig. 5. Minimum required optical receiver sensitivity versus signal-to-noise ratio

Another useful representation of minimum required optical sensitivity P_{AVG} is the dependence of *BER* needed. The results are shown in Fig. 6.

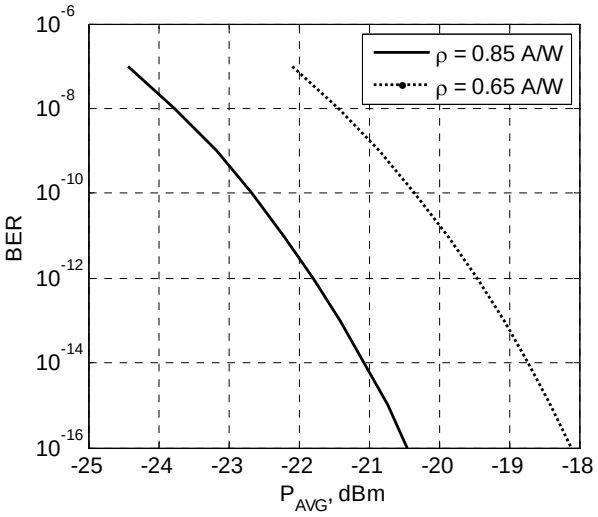


Fig. 6. Minimum required optical receiver sensitivity (P_{AVG}) versus Bit Error Rate

IV. Conclusion

By applying the technique presented in this paper, it is easy to estimate and predict optical receiver sensitivity. Relying on a vendor specifications accurate estimate of optical sensitivity can be made.

To estimate optical-receiver sensitivity, it is necessary to consider error sources in amplitude. It is shown how the amplitude error sources separately affect the overall receiver BER with practical device implementations. Optical receiver performance can now be accurately predicted to choose the proper TIA, limiting amplifier, and CDR. In reality, the optical input is not an ideal signal, because it suffers random noise from the transmitter as well as ISI from fiber dispersion. When a stressed optical signal is received, the same approach presented in this article can be used for estimating the signal *Q*-factor and, therefore, determining the *BER*.

Acknowledgement

This paper has been sponsored under the auspices of the "Increasing Efficiency and Quality of Service in Broadband Hybrid Cable Television Networks" project – a part of the University Center for Research and Technology (UTzNIT) at the Technical University of Gabrovo, contract E904/2009.

References

[1] Е. Фердинандов, Ц. Мицев, В. Кърдев. Световодни комуникационни системи. Част 3 Приемна част, Сиела, София, 2003.
[2] Ст. Рабов, Л. Христов, Оптични комуникации, Нови знания, София, 2002.
[3] Accurately Estimating Optical receiver Sensitivity, Application note HFAN-3.0.0, Rev.1, MAXIM, April 2008.
[4] CENELEC EN 50083-6. Cable Networks for Television Signals, Sound Signals and Interactive Services. Part 6: Optical Equipment. December 2002.
[5] B. Kanack, C. Goldsmith, "Improved Semiconductor Based Analog Fiber Optic Link Performance Through Reactive Matching," *Proc. DOD Fiber Optics*, *AFCEA*, pp. 391—396, 1994.
[6] eDOCSIS™, eDOCSIS Specification, CM-SP-eDOCSIS-I11-070223, CableLabs, 2007.
[7] Fiber Optic Design Guide, BLONDER TONGUE Laboratories Inc., 2007.
[8] I. Nedelchev, I. Nemigenchev, "RF Power Amplifier Class E with a Capacitive Impedance Converter". *UNITECH 2007*, vol. 1, pp.I-220–I-225, Gabrovo, Bulgaria, 2007.
[9] J. Barry, E. Lee, "Performance of Coherent Optical Receivers", *Proceedings of IEEE*, vol. 78, no. 8, 1990.
[10] K. Angelov, K. Koitchev, and N. Varbanova, "Estimating Losses from Transient and Intersymbol Distortions in Hybrid Fiber-Coaxial Television Network", *ICEST 2009*, vol. 1, pp.113-116, Sofia, Bulgaria, 2009.
[11] K. Angelov, K. Koitchev, and S. Sadinov, "An Investigation of Noise Influences in Optical Transmitters and Receivers in Cable TV Networks", *ICEST 2006*, pp.102-105, Sofia, Bulgaria, 2006.
[12] K. Koitchev K., K. Angelov, and S. Sadinov, "Determining Bit Error Rate in Digital Optical Transmission Network Using the Q-Factor", *ICEST 2010*, pp.53-56, Ohrid, Macedonia, 2010.
[13] Optical Modulation Amplitude (OMA) and Extinction Ratio, Application note HFAN-02.2.2, Rev.1, MAXIM, April 2008.
[14] R. Freeman, *Fiber-Optic Systems for Telecommunications*, John Wiley & Sons, New York, 2002.
[15] http://para.maxim-ic.com/en/search.mvp?fam=ti_amp&tree=master – Transimpedance Amplifiers (TIA).
[16] http://para.maxim-ic.com/en/search.mvp?fam=limit_amp&tree=master – Limiting Amplifiers.

Teletraffic Analysis of Spectrum Handover in Cognitive Radio Networks

Yakim Mihov[1] and Boris Tsankov[2]

Abstract – **This paper presents a well-known and general model of a serving system with primary and secondary users. A novel approximate but simple and computationally efficient analytical approach for solving the state probabilities of the state-transition diagram of the system is developed. A new precise formula for evaluation of the call dropping probability of the secondary users is derived. Channel limitation as a method for reducing the call dropping probability is analyzed. The proposed analytical approach and formulae are validated through extensive simulation experiments. Numerical results are presented and conclusions are drawn.**

Keywords – **call blocking probability, call dropping probability, channel limitation, cognitive radio network, spectrum handover.**

I. Introduction

Cognitive radio (CR) is a type of radio in which communication systems are aware of their environment and internal state and can make decisions about their radio operating behavior based on that information and predefined objectives [1]. It is capable of adjusting automatically its behavior or operations to achieve desired objectives. A brief overview of CR is presented in [2].

One of the most prominent applications of CR is in dynamic spectrum access (DSA) networks. DSA is a new paradigm for spectrum regulation which is expected to alleviate the artificially created scarcity of spectrum resources caused by the traditional static command-and-control approach for spectrum regulation. Hierarchical spectrum overlay is a promising method for DSA. It allows secondary (unlicensed or cognitive) users (SUs) to temporarily utilize spectrum resources assigned to primary (licensed or incumbent) users (PUs) if these resources are not currently being used for PU transmission. SUs have to vacate the occupied resources as soon as PUs start using them for transmission, i.e. PUs have preemptive priority over SUs. The cognitive network utilizes opportunistically the available unoccupied spectrum of the primary network on a non-interference basis. Spectrum handover is an essential function of CR since it enables and assures link maintenance and service resilience and thus facilitates the quality of service (QoS) provisioning for SUs.

[1]Yakim Mihov is with the Faculty of Telecommunications, Technical University of Sofia, K. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: yakim_mihov@abv.bg.

[2]Boris Tsankov is with the Faculty of Telecommunications, Technical University of Sofia, K. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: bpt@tu-sofia.bg.

II. Performance Analysis

In our study we analyze a well-known general model ([3]-[5]) of a serving system with PUs and cognitive SUs in accordance with the hierarchical spectrum overlay approach for DSA. We assume that the primary network and the cognitive network provide multimedia services with different bandwidth demands and that the bandwidth of a PU call is k times greater than the bandwidth of a SU call. Let us define the term *channel* as the necessary mean bandwidth for a PU multimedia call to be served and the term *subchannel* as the necessary mean bandwidth for a SU multimedia call to be served. It is obvious that one channel comprises k subchannels. The total bandwidth of the serving system is assumed to comprise n channels and hence nk subchannels. We denote with i ($i = 1, \dots, n$) and j ($j = 1, \dots, nk$) the number of PU and SU calls in the system, respectively. The offered PU and SU traffic is modeled by two Poisson random processes with arrival rates λ_p and λ_s, respectively. The PU and SU call durations follow a negative exponential distribution with mean $1/\mu_p$ and $1/\mu_s$, respectively.

PUs have a preemptive priority over SUs. If a PU starts transmitting on a channel, all subchannels occupied by SUs within that channel have to be vacated immediately. If a channel is being used by a PU, the subchannels within that channel are unavailable to SUs. The service of PU calls is absolutely independent of the service of SU calls.

In the model perfect spectrum sensing and spectrum handover procedures are assumed. Under these conditions, SU call blocking occurs only if there is not an unoccupied subchannel in the system to serve a new SU call. Similarly, SU call dropping occurs only if there is not an unoccupied subchannel in the system to continue the service of a SU call during spectrum handover. Since one channel comprises k subchannels, up to k SU calls may be dropped simultaneously at the arrival of a new PU call.

The described model of the serving system can be presented by a 2-D continuous time Markov chain (Fig. 1). Let us denote with $P_{i,j}$ the probability that the system is in state (i,j), i.e. the steady state probability that there are i PU calls and j SU calls in the system. Based on the state-transition diagram in Fig. 1, we can derive the global balance equations:

$$\left(\lambda_p + \lambda_s\right) P_{0,0} = \mu_p P_{1,0} + \mu_s P_{0,1}; \quad (1)$$

$$\left(\lambda_p + \lambda_s + i\mu_p\right) P_{i,0} = \lambda_p P_{i-1,0} + \left(i+1\right)\mu_p P_{i+1,0} + \mu_s P_{i,1}, \quad (2)$$

where $0 < i < n$;

$$n\mu_p P_{n,0} = \lambda_p \sum_{j=0}^{k} P_{n-1,j}; \quad (3)$$

$$\left(\lambda_p+\lambda_s+j\mu_s\right)P_{0,j}=\lambda_s P_{0,j-1}+\mu_p P_{1,j}+(j+1)\mu_s P_{0,j+1}, \quad (4)$$

where $0<j\le(n-1)k$;

$$\left(\lambda_p+\lambda_s+j\mu_s\right)P_{0,j}=\lambda_s P_{0,j-1}+(j+1)\mu_s P_{0,j+1}, \quad (5)$$

where $(n-1)k<j<nk$;

$$\left(\lambda_p+nk\mu_s\right)P_{0,nk}=\lambda_s P_{0,nk-1}; \quad (6)$$

$$\left(\lambda_p+i\mu_p+j\mu_s\right)P_{i,j}=\lambda_s P_{i,j-1}+\lambda_p\sum_{m=j}^{j+k}P_{i-1,m}, \quad (7)$$

where $i>0; j>0; ik+j=nk$;

$$\left(\lambda_p+\lambda_s+i\mu_p+j\mu_s\right)P_{i,j}=\lambda_p P_{i-1,j}+\lambda_s P_{i,j-1}+ +(j+1)\mu_s P_{i,j+1}, \quad (8)$$

where $i>0; j>0; (n-1)k<ik+j<nk$;

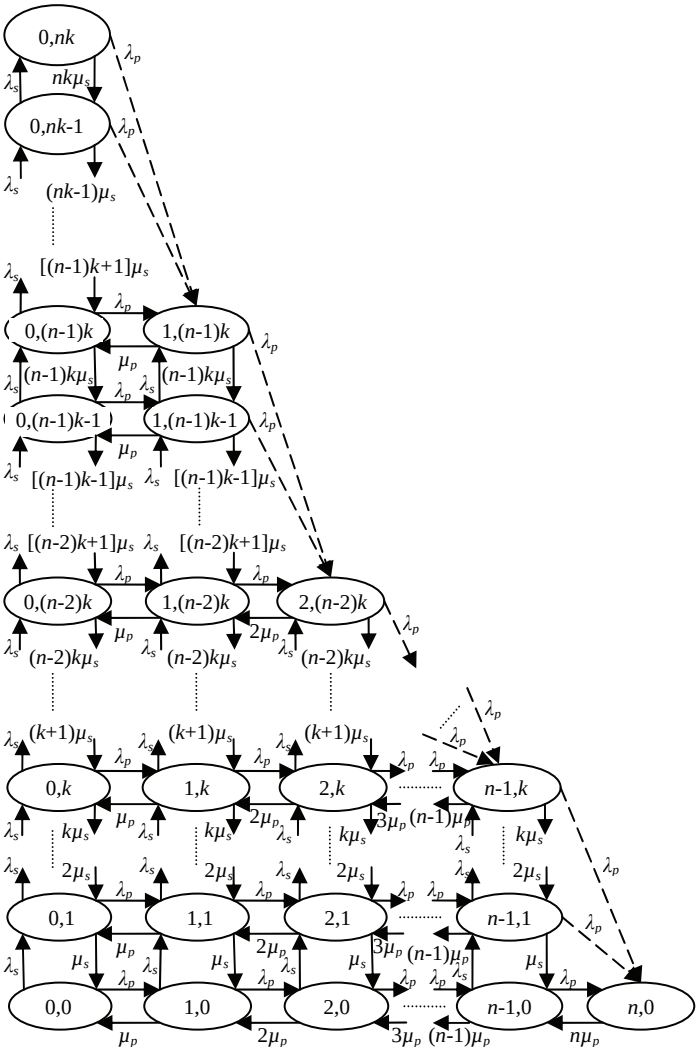


Fig. 1. The state-transition diagram of the teletraffic serving system

$$\left(\lambda_p+\lambda_s+i\mu_p+j\mu_s\right)P_{i,j}=\lambda_p P_{i-1,j}+\lambda_s P_{i,j-1}+ +(i+1)\mu_p P_{i+1,j}+(j+1)\mu_s P_{i,j+1}, \quad (9)$$

where $i>0; j>0; ik+j\le(n-1)k$;

$$\sum_{i=0}^{n}\sum_{j=0}^{(n-i)k}P_{i,j}=1. \quad (10)$$

The system of Eqs. (1) - (10) contains $\frac{(n+1)(nk+2)}{2}$ unknown state probabilities and can be solved using an appropriate iterative method, such as the *Gauss-Seidel* method or the method of *successive over-relaxation* (SOR) [5],[6]. However, due to the high computational complexity of these iterative methods, their implementation in CR may be infeasible with respect to the real-time and power consumption design requirements on CR performance.

We develop and validate by simulation a novel approximate approach for solving the state probabilities of the diagram (Fig. 1) with relatively low computational complexity. Moreover, our approach is also applicable in case that channel limitation is used in the cognitive network for reducing the SU call dropping probability and thus guaranteeing the CR QoS provisioning. Channel limitation sets an upper bound on the maximum admissible number of calls in the system, i.e. it sets the call admission control (CAC) threshold. Let us denote with l the maximum allowable number of SU calls in the system. If channel limitation is applied, $0\le j\le l<nk$; otherwise: $0\le j\le l=nk$.

Because of the preemptive priority of PUs over SUs, our state-transition diagram (Fig. 1) differs from an ordinary multidimensional state-transition diagram and a trivial solution based on *state-based algorithms* or the *convolutional algorithm* [7] cannot be applied. Since the service of PUs is independent of the service of SUs and the PU and SU call arrival processes are i.i.d, we have:

$$P_{i,j}=P_i P_j^i=\frac{\frac{A_p^i}{i!}}{\sum_{m=0}^{n}\frac{A_p^m}{m!}}P_j^i, \quad i<n; \quad (11)$$

and

$$P_{n,0}=\frac{\frac{A_p^n}{n!}}{\sum_{m=0}^{n}\frac{A_p^m}{m!}}=B; \quad (12)$$

where P_j^i is the conditional probability that there are j SU calls in the system provided that the number of PU calls is i; B is the PU call blocking probability.

Now we proceed to derive P_j^i by inspecting the columns of the diagram in Fig. 1 and considering limitation if applied. Let us introduce the notations:

$$t=(n-i-1)k, \quad 0\le i<n; \quad (13)$$

and

$$\rho=\frac{\mu_p}{\mu_s}. \quad (14)$$

In states (i,j), where $i < n$ and $j > t$, SU call dropping occurs if a new PU call arrives, which is designated with the dashed transitions λ_p in Fig. 1. Since dropping decreases the number of SU calls in the system, the effective departure (service) rate of the SU calls in state (i,j) $\{i<n; j>t\}$ is assumed to be $j\mu_s + \frac{\lambda_p}{j-t}$. Based on this assumption and taking into account the optional use of limitation ($0<l\leq nk$), we solve the balance equations about column i of the state-transition diagram for SU traffic only. Thus we obtain:

$$P_j^i = \frac{\frac{A_s^j}{j!}}{\sum_{m=0}^{l} \frac{A_s^m}{m!}}, \quad t \geq l; j \leq l; i < n; \tag{15}$$

or

$$P_j^i = \frac{\frac{A_s^j}{j!}}{\sum_{m=0}^{t} \frac{A_s^m}{m!} + \frac{1}{t!} \sum_{x=t+1}^{\min[l,(n-i)k]} \frac{A_s^x}{\prod_{m=t+1}^{x} \left(m + \rho \frac{A_p}{m-t}\right)}}, \tag{16}$$

where $t < l; j \leq t; i < n;$

or

$$P_j^i = \frac{\frac{A_s^j}{t! \prod_{m=t+1}^{j} \left(m + \rho \frac{A_p}{m-t}\right)}}{\sum_{m=0}^{t} \frac{A_s^m}{m!} + \frac{1}{t!} \sum_{x=t+1}^{\min[l,(n-i)k]} \frac{A_s^x}{\prod_{m=t+1}^{x} \left(m + \rho \frac{A_p}{m-t}\right)}}, \tag{17}$$

where $t < l; t < j \leq \min\left[l,(n-i)k\right]; i < n.$

Substituting Eq. (15) or (16) or (17) into Eq. (11), P_{ij} is obtained.

SU call blocking occurs if all subchannels are occupied:

$$P_b = \sum_{i=0}^{n} P_{i,\min[l,(n-i)k]}. \tag{18}$$

SU call dropping occurs only if $j > t$ and a new PU call arrives. We propose a new precise formula which evaluates the SU call dropping probability by the ratio of the mean number of dropped SU calls to the mean number of SU calls in the system (instead of just summing up the probabilities of the states in which SU call dropping occurs):

$$P_d = \frac{\sum_{i=0}^{n-1} \sum_{j=t+1}^{\min[l,(n-i)k]} (j-t)\left(1 - e^{-\frac{\lambda_p}{j\mu_s}}\right) P_{i,j} \Delta(l,t)}{\sum_{i=0}^{n-1} \sum_{j=1}^{\min[l,(n-i)k]} j P_{i,j}}, \tag{19}$$

where $\Delta(l,t)$ =1 if $l > t$ and $\Delta(l,t)$ = 0 otherwise.

The mean system service rate μ_{ss} for SU calls is:

$$\mu_{ss} = \sum_{i=0}^{n-1} \sum_{j=1}^{\min[l,(n-i)k]} j\mu_s P_{i,j}. \tag{20}$$

III. NUMERICAL RESULTS

A simulation model for performance evaluation of the serving system under consideration is developed to validate the proposed analytical approach. Various simulation experiments are performed and in all cases there is a good coincidence between analytical and simulation results, as shown in Figs. 2, 3, 4 and 5, which verifies and validates the developed new approximate approach and formulae.

We first analyze the effect of PU traffic load on the performance of the cognitive network. As A_p increases, P_b and P_d also increase, as shown in Figs. 2 and 3, which means that the cognitive traffic capacity decreases. Therefore, it is reasonable to deploy CR only in primary networks whose transmission resources are underutilized in order to provide significant cognitive capacity.

Next, we analyze the effect of applying channel limitation. Figs. 2 and 3 illustrate that due to limitation it is possible to reduce P_d at the price of increasing P_b, i.e. there is a trade-off relationship between these two parameters. In general, the use of limitation is desirable since the reduction in P_d improves and facilitates the QoS provisioning in the CR network. However, the effect of limitation is significant only if A_p is relatively small. If A_p is relatively large, the performance of the CR network is no longer determined by l but by A_p since in this case the number of PU calls (not l) limits the maximum possible number of SU calls that the system can serve. Consequently, in order to apply limitation efficiently and to guarantee the QoS provisioning in the CR network, A_p must be relatively small, i.e. we again conclude that the primary network has to be underutilized in order to provide service with QoS guarantee over CR.

Figs. 4 and 5 show that channel limitation decreases both the throughput and the capacity of the CR network since the maximum number of admissible SU calls is reduced.

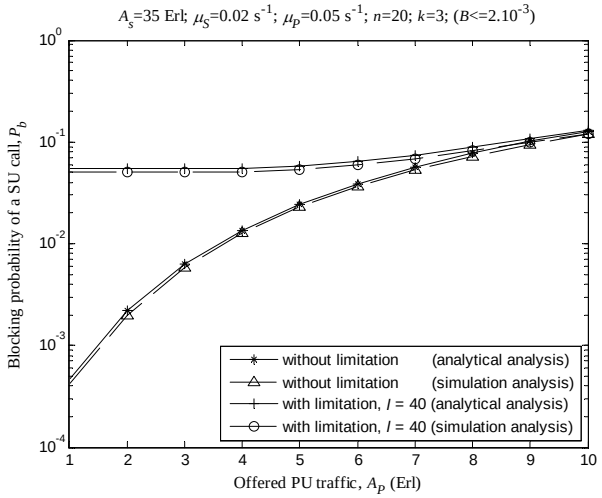


Fig. 2. SU call blocking probability versus the offered PU traffic

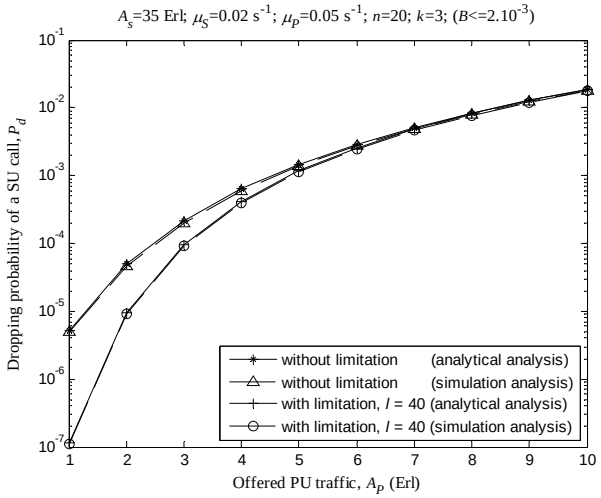


Fig. 3. SU call dropping probability versus the offered PU traffic

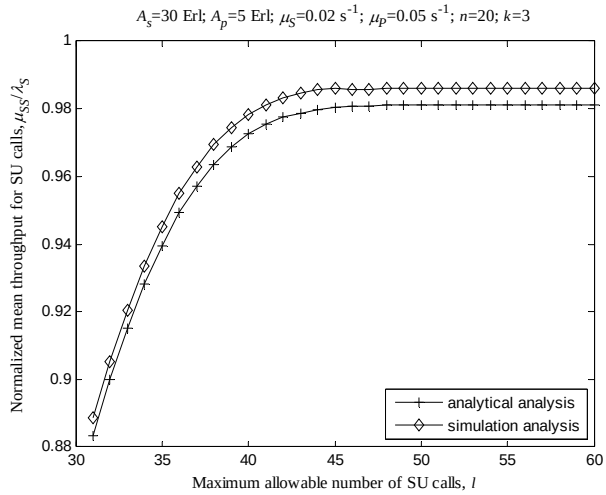


Fig. 4. Normalized mean service rate of SU calls versus the SU CAC threshold

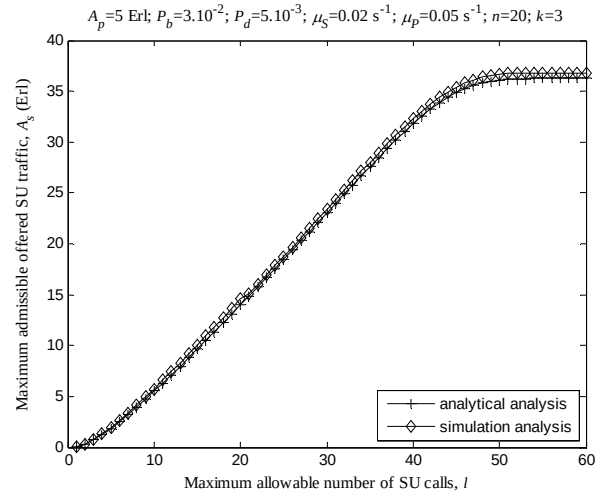


Fig. 5. Cognitive traffic capacity versus the SU CAC threshold

IV. CONCLUSION

In our paper we analyze a general model of a serving system with PUs and cognitive SUs in accordance with the hierarchical spectrum overlay approach for DSA, which makes the scope of our study generic and determines its wide applicability and theoretical significance.

A novel approximate but computationally efficient approach for solving the steady state probabilities of the state-transition diagram of the system and applicable in case of channel limitation is developed and validated by simulation. The main advantage of our approximate approach over the precise iterative methods for solving the state probabilities of the state-transition diagram of the system is its computational simplicity which facilitates its application and implementation in CR with respect to satisfying the real-time and power consumption design requirements on CR. A new precise formula for the SU call dropping probability is also proposed.

Our study corroborates that the effect of PU traffic on the capacity and performance of CR should always be considered. Channel limitation decreases the CR throughput and capacity but facilitates the SU QoS provisioning if prudently applied.

For future research work, we plan to use the analytical model presented in this paper for the design of a cognitive CAC algorithm capable of providing QoS for heterogeneous multimedia services over CR.

ACKNOWLEDGEMENT

This research was supported by the Bulgarian Ministry of Education and Science under Grant DVU01/0109 (DO-02-135/2008).

REFERENCES

[1] IEEE Std. 1900.1-2008, "IEEE Standard Definitions and Concepts for Dynamic Spectrum Access: Terminology Relating to Emerging Wireless Networks, System Functionality, and Spectrum Management", pp. c1-48, September 2008.

[2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A Survey on Spectrum Management in Cognitive Radio Networks", *IEEE Communications Magazine*, vol. 46, no. 4, pp. 40-48, 2008.

[3] X. Zhu, L. Shen and T.-S. P. Yum, "Analysis of Cognitive Radio Spectrum Access with Optimal Channel Reservation", *IEEE Commun. Letters*, vol. 11, no. 4, pp. 304-306, 2007.

[4] W. Ahmed, J. Gao, H. A. Suraweera and M. Faulkner, "Comments on "Analysis of Cognitive Radio Spectrum Access with Optimal Channel Reservation"", *IEEE Trans. on Wireless Commun.*, vol. 8, no. 9, pp. 4488-4491, 2009.

[5] Y. Zhang, "Dynamic Spectrum Access in Cognitive Radio Wireless Networks", *IEEE ICC*, pp. 4927-4932, Beijing, China, 2008.

[6] R. B. Cooper, "*Introduction to Queueing Theory*", 2nd edition, Elsevier North Holland, USA, 1981.

[7] V. B. Iversen, "*Teletraffic Engineering and Network Planning*", COM department, Technical University of Denmark, 2009.

Performance Analysis of an Intra-cell Handover Management Policy in Wireless Access Networks

Kiril Kassev

Abstract – **The scarce nature of the bandwidth of wireless links is a key issue for QoS provisioning in next generation wireless systems. In the presence of mobile users willing to access to a rich variety of services, advanced mobility management is needed such that these services are provided seamlessly. From teletraffic point of view, bandwidth demands of active connections could fluctuate due to movement of mobile stations within a cell coverage. The transmission rate adjustment is based on the adaptive modulation and coding (AMC) technique employed. The resulting intra-cell handover has a major impact on the system performance. In the context of resource management, the problem is directly related to the issue of resource reservation and admission control. This paper aims at developing a threshold-based bandwidth management policy for QoS guarantee in next generation wireless access networks. The policy prioritizes connections according to their QoS requirements and the cell area of a call origin.**

Keywords – **Adaptive modulation and coding, Bandwidth allocation, Call admission control, Intra-cell handover**.

I. INTRODUCTION

For efficient usage of limited resources, wireless communications systems employ the cellular concept. In such a multicell environment new problems arise. The following two aspects could be distinguished – inter-cell interference management and handover management. In this paper, the attention is paid to a handover management policy, which maintains target QoS of active connections, over new ones. It aims at balancing the QoS of both handover and new calls (connections) arrivals. This is done by an appropriate resource reservation and admission control.

Wireless resource management policies should be capable of determine the optimal use of limited resources, according to the wireless channel state information and specific QoS requirements. Since bandwidth is a fundamental wireless resource, from teletraffic point of view, it may be considered as a transmission capacity of the wireless medium, which is shared by multiple users.

In order for the emerging wireless access systems to overcome the limitation of wireless communications environment, significant research efforts have been put towards development of technologies for enhancing the spectral efficiency. AMC has been well adopted as an advanced physical layer technology [1]. In legacy wireless access systems each active connection is assigned a fixed amount of bandwidth independently of the mobile station (MS) location within a cell. The quantity of resource consumption depends on the traffic source characteristics.

Initially, handover management for single-service wireless access systems had been extensively studied. A number of prioritization policies that handle the handover traffic has been proposed and analyzed [2], [3]. The basic mechanism is to suitably partition the available resource in a cell to a different traffic types (new and handover call arrivals). The investigations are made under the assumption of conversational services, such a voice, occupying a fixed number of resource units (channels) per call.

A common feature of admission policies is their ability to use the target cell information only (i.e., the number of occupied channels). It would be more beneficial some additional information of the adjacent cell to be used. Since handover decision is solely based on the availability of a free channel without taking into consideration the signal quality, [4] extends the well-known "guard channel" (GC) policy. This is done by combining the mobile assisted handover technique (available at GSM cellular system) and GC. A more efficient handover management scheme is proposed in [5]. The future behavior of an active call can be estimated more precisely, based on the mobility (location and velocity) of each MS. A channel reservation message can be sent to the approaching cell, allowing adaptive resource reservation.

In [6], a novel analytical method for performance analysis of GC policy is proposed. The commonly accepted assumption for Markovian arrival and departure processes is omitted. Performance metrics of interest are derived in presence of self-similar traffic.

One of the challenges in handover management schemes is towards efficient sharing of limited resources among multiple traffic classes. In general, the following approaches can be distinguished – complete partition (*reservation* of a bandwidth exclusively for each traffic class); complete sharing (sharing of available bandwidth among all traffic classes). For a particular traffic class, it is possible a *limitation* level to be set; hybrid schemes. A comparative study of the first two schemes is carried out in [7]. It is shown the advantage of complete sharing to complete partition scheme for efficient use of scarce resource in case of two traffic classes with different bandwidth requirements. The potential of movable boundaries allocation strategies, that can adjust the number of channels for each traffic class, has been investigated in [8]. The proposed scheme extends the GC by introducing different admission thresholds for different traffic classes. It is assumed each traffic class requires one channel per connection. A further extension of bandwidth allocation schemes has been presented in [9]. Authors sketch the design tradeoffs of known solutions and propose as well as study the performance of a

Kiril Kassev is with the Faculty of Telecommunications, Technical University of Sofia, 8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria, E-mail: kmk@tu-sofia.bg

policy that supports traffic flows with variable bandwidth requirements.

Considering the importance of the problem, a great extent of research work has been devoted to development of handover management policies applied for the case of inter-cell handover only. Having interested in performance analysis of emerging wireless access technologies, the application of above-mentioned methods may be inaccurate. In AMC-enabled systems (cells) the bandwidth demands of active connections may fluctuate due to movement of mobile stations within cell coverage. This is true for services requiring constant transmission capacity. As a consequence, the resulting intra-cell handover and its influence on the system performance should be taken into consideration.

This topic is covered by [10] in terms of physical layer performance. An interference avoidance technique based on the use of intra-cell handover in OFDMA femtocells is proposed. Authors apply the concept of intra-cell handover in GSM networks to OFDMA subchannels selection, in order to mitigate the interference between macro- and femtocell tiers.

The rest of the paper is organized as follows: The traffic model and resource management strategy are described in Section 2. Numerical studies in Section 3 reveal the impact of the both intra-cell handover and applied resource management policy on the system performance. Some conclusion remarks and suggestions for future work are presented in Section 4.

II. System Model and Performance Analysis

We consider a cell with a fixed amount of capacity. Independently of the multiple access technology employed, the cell capacity could be represents in terms of effective bandwidth. Thus, the total cell capacity is C resource units (bandwidth units).

The threshold-based bandwidth management policy is based on the complete sharing approach (Fig. 1). Both new calls (NC1) and handover calls (HO1) incoming in the cell area served by the highest modulation and coding scheme (MCS), get the highest priority (full access to the available cell resources). These calls are more profitable for the network operator, because they are allocated the smallest amount of resources per active call. This area is referred to as "Ring 1", linked with the 64-QAM modulation scheme. The new calls (NC2) and handover calls (HO2) offered to "Ring 2" area compete for the remaining resource. This is done by setting the threshold levels T_0 and T_1, in terms of maximum number of resource units. Since dropping a handover connection (HO2) is not considered acceptable, $T_1 > T_0$. The policy aims at adjusting the threshold levels, in order to satisfy the handover dropping probabilities, keeping reasonable values of new calls blocking probabilities.

A call, with certain traffic characteristics, to (from) a MS in Ring l (l = 1, L) simultaneously requires d_l resource units in order to be guaranteed desired QoS.

A new call arrival in both rings follows a Poisson process with mean rate λ_l (l = 1, L). Both new call duration and Ring l dwell time are assumed to be negative exponentially distributed random variables with mean $1/\mu$ and $1/\delta_l$, respectively.

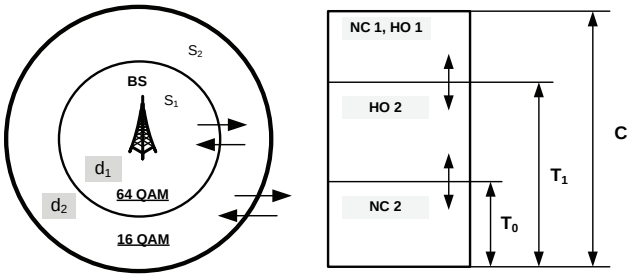


Fig. 1. An AMC-enabled cell with a handover management policy

Intracellular handover calling rate from Ring l to Ring $l + 1$ is $\lambda_l^{h+} = n_l \delta_l^+$. $1/\delta_l^+$ denotes the mean dwell time of handover calls outward of Ring l. The handover calling rate λ_l^{h-} from Ring l to Ring l - 1 can be estimated in a similar manner, where n_l denotes the number of both new and handover calls (active MSs) in Ring l.

In case of the most outer cell Ring L consideration, the incoming inter-cell handover rate from adjacent cells is defined by $\lambda_h = \lambda_L^{h+} = n_L \delta_L^+$. For cellular networks, the value of λ_h can be estimated by a relation derived from [11]. The average Ring l dwell time is in direct proportion to the cell radius and inverse proportion to the average speed v of MSs [12], [13].

Under the assumptions we stated above, the handover management policy can be modeled as a multi-dimensional Markov process. The multidimensional random variable depends on the number of rings L in a cell. Since we are interested in the steady state probabilities estimation, the challenge is to solve the system of linear equation for such a process. Based on the possible state transitions (Fig. 2), the topic under consideration faces the problem of irreversibility (i.e., the underlying Markov process is non-reversible). Thus, the steady states probabilities cannot be calculated by using product form solutions. A common approach is towards a direct method application, which requires a matrix inversion technique to be used. Depending on the structure of the coefficient matrix, an appropriate numerical method (algorithm) for solving simultaneous linear equations is required. This could limit the application of the method to relatively small state-spaces. The same could be true for a specific class of recursive methods for solving non-Markovian processes [14]. Although the method is well-adopted [9], [15], the limitation could arise in case a system with large state space and number of state transitions is investigated.

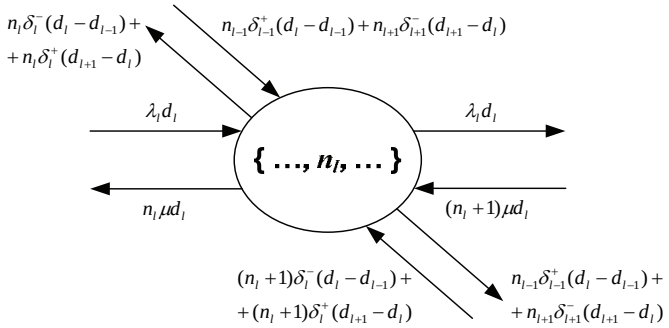


Fig. 2. System state transition for Ring l

In order to fully exploit the advantages of certain recursive methods applied to reversible Markov processes, a challenging work is to construct a reversible Markov chain that well approximate the non-reversible system under consideration (in a manner similar to [16]).

Since $L = 2$, the call (connection) level performance measures of interest are obtained by solving a two-dimensional non-reversible Markov process.

The state of the system (a cell) is defined by the set:

$$s = \{n_1, n_2, ..., n_l, ..., n_L\}.$$

Let the set of allowable states (determined by the resource sharing policy in use) is denoted by S. Therefore, $s \in S$ if and only if the following conditions are satisfied:

$$S = \{s : 0 \le n_2 \cdot d_2 \le T_0 \le T_1;\ 0 \le \sum_{i=1}^{2} n_i \cdot d_i \le C\}. \quad (1)$$

We introduce the notation $s_l^+ = \{n_1, n_2, ..., n_l + 1, ..., n_L\}$.

Both new and handover calls offered to Ring 1 share the entire available cell resource. The call blocking (NC1)/dropping (HO1) probabilities are equal and can be derived as:

$$P_{NC1,HO1} = \sum_{s \in BD^+_{NC1,HO1}} P(s), \quad (2a)$$

$$\begin{aligned} BD^+_{NC1,HO1} &= \{s \in S \mid s_1^+ \notin S\} \\ &= \{s \in S \mid d_1 + \sum_{j=1}^{2} n_j \cdot d_j > C\}. \end{aligned} \quad (2b)$$

Let us introduce the subset $\hat{S} \in S$, defined by the threshold level T_0. The cell resource is completely shared by all types of calls offered from both rings. Thus, the blocking probability P_{NC2} for new calls offered to Ring 2 is given by:

$$P_{NC2} = \sum_{s \in B^+_{NC2}} P(s), \quad (3a)$$

$$\begin{aligned} B^+_{NC2} &= \{s \in S \mid s_2^+ \notin \hat{S}\} \\ &= \{s \in S \mid d_2 + n_2 \cdot d_2 > T_0 \vee T_0 < n_2 \cdot d_2 \le T_1 \vee \\ &\vee\ d_2 + \sum_{j=1}^{2} n_j \cdot d_j > C\}. \end{aligned} \quad (3b)$$

The dropping probability P_{HO2} for handover calls offered to Ring 2 is given by:

$$P_{HO2} = \sum_{s \in D^+_{HO2}} P(s), \quad (4a)$$

$$\begin{aligned} D^+_{HO2} &= \{s \in S \mid s_2^+ \notin S\} \\ &= \{s \in S \mid d_2 + n_2 \cdot d_2 > T_1 \vee d_2 + \sum_{j=1}^{2} n_j \cdot d_j > C\}. \end{aligned} \quad (4b)$$

III. Numerical Results

We assume the total cell capacity of 20 resource units. Based on a particular resource management policy, it is shared by both new call arrivals from each ring, and intra-cell handover arrivals, as a result of MSs movement across inner cell boundaries (rings). The intra-cell handover rates δ_l are related to the MSs average speed, rings area, and the MSs distribution within the cell [12], [13]. We suppose uniformly distributed MSs.

It is assumed a cell radius $r = 2$ km. The relation of the outer radius of Ring 1, linked to a MCS of the highest order (64-QAM), to the cell radius is denoted by p. This proportion depends on the wireless environment propagation conditions and statistical characteristics. The distance covered by a MCS is also governed by a set of target performance measures at physical level. Based on a relation between MCSs and resources necessary for constant transmission capacity provisioning, we assume $d_1 = 1$, $d_2 = 2$ resource units per call [17]. The new call arrival rate λ_l at Ring l is tightly coupled with the ratio of the Ring l area S_l to the entire cell area. The average service rate of MSs distributed over the cell is assumed to be $\mu = 0.0167\ s^{-1}$.

For complete sharing scenario both type of traffic flows offered from Ring 2 area experience higher losses compared to traffic arrivals from Ring 1 (Fig. 3). Connections with lower bandwidth requirements have a better chance at occupying the bandwidth than those with higher bandwidth requirements. In case handover limitation threshold T_1 is set to 16 resource units, it is shown that significant improvement of the performance measures cannot be reached. It is more likely Ring 2 handover dropping probability to get increased. For this reason, the rest of investigations are carried out for $T_1 = 20$ resource units. Fig. 4 depicts the impact of users' (MSs) mobility on the system performance. The resource management policy has low efficiency for high mobility users, compared to low mobility ones, as a result of increased intra-cellular handover rates. The threshold level T_0 has to be carefully set, such that Ring 2 new call blocking meet QoS.

The influence of the Ring 1 region area is illustrated on Fig. 5 (in terms of the parameter p). For low mobility users, an efficient point of operation of the management policy can be distinguished at $p \approx 0{,}6$. A good resource management scheme has to balance the tradeoffs between new call blocking and handover dropping probabilities in order to meet target QoS requirements.

The policy operation under certain dynamic range of new call arrival rates in both rings is depicted on Fig. 6. Again, the policy takes effect on low mobility users. When the new call arrival rate is high, no matter how parameters are adjusted, the policy cannot guarantee QoS.

IV. Conclusion and Future Work

In this paper, a threshold-based intra-cellular handover management policy has been proposed and studied. Numerical analysis shows that important performance metrics can be improved as well as some tradeoffs have been identified.

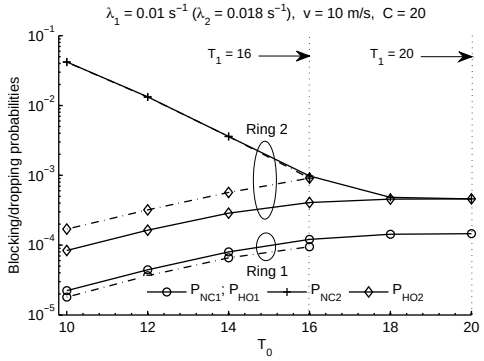


Fig. 3. Influence of Ring 2 limitation thresholds on QoS metrics

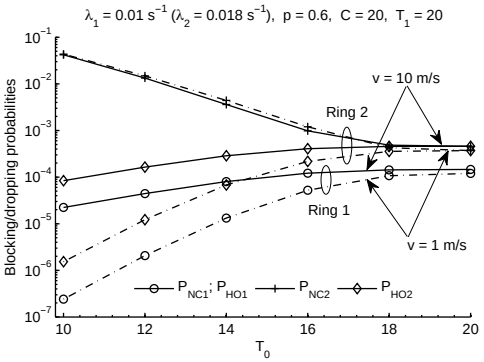


Fig. 4. Influence of MSs mobility on the system performance

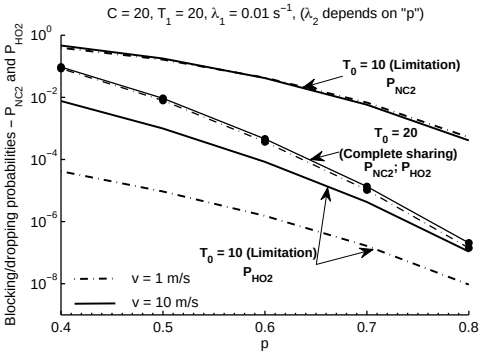


Fig. 5. Ring 1 region influence on system performance

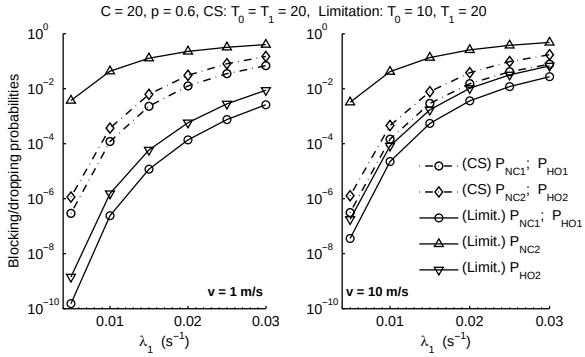


Fig. 6. Call-level QoS parameters vs. new call arrival rate dynamic

The future work will focus on methods for reversible Markov chains construction, which well approximate the non-reversible system investigated. This will enable us to use attractive recursive methods (e.g., Kaufman-Roberts) that can be applied to realistic systems with large state-space.

ACKNOWLEDGEMENT

This work was supported by the Bulgarian Ministry of Education, Youth and Science in the context of research project "Optimal telecommunication resource allocation considering cross-layer interaction", under Contract DO-02-135/2008.

REFERENCES

[1] M. S. Alouini, A. J. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels", *J. Wireless Communications*, vol. 13, no. 1-2, pp. 119-143, 2000.

[2] R. Ramjee *et al.*, "On Optimal Call Admission Control in Cellular Networks", *J. Wireless Networks*, Kluwer Academic Publishers, vol. 3, no. 1, pp. 29-41, 1997.

[3] J. G. Markoulidakis *et al.*, "Optimal System Capacity in Handof Prioritized Schemes in Cellular Mobile Telecommunication Systems", *J. Comp. Comm.*, vol. 23, pp. 462-475, 2000.

[4] B. B. Madan, S. Dharmaraja, and K. S. Trivedi, "Combined Guard Channel and Mobile Assisted Handoff in Cellular Networks", *IEEE Trans. Veh. Tech.*, vol. 57, no. 1, pp. 502-510, 2008.

[5] S. Choi, K. G. Shin, "Adaptive Bandwidth Reservation and Admission Control in QoS-sensitive Cellular Networks", *IEEE Trans. on Parallel and Distrib. Syst.*, vol. 13, pp. 882-897, 2002.

[6] X. Jin, G. Min, and J. Jiang, "Analytical Modeling of the GC-Based Handover Scheme with Heavy-Tailed Call Holding Times", *IEEE GLOBECOM'2009*, pp. 1-5, 2009.

[7] B. Epstein, M. Schwartz, "Reservation Strategies for Multimedia Traffic in a Wireless Environment", *IEEE Vehicular Technology Conference*, pp. 165-169, 1995.

[8] L. Yin, Z. Zhang, and Y.-B. Lin, "Performance Analysis of Dual-threshold Reservation Scheme for Voice/Data Integrated Mobile Wireless Networks", *IEEE WCNC*, pp. 258-262, Chicago, USA, 2000.

[9] B. Li *et al.*, "Call Admission Control for Voice/Data Integrated Cellular Networks: Performance Analysis and Comparative Study", *IEEE JSAC*, vol. 22, no. 4, pp. 706-717, 2004.

[10] D. L. Perez *et al.*, "Intracell Handover for Interference and Handover Mitigation in OFDMA Two-Tier Macrocell-Femtocell Networks", *EURASIP J. Wireless Commun. Networking*, vol. 2010, pp. 1-15, 2010.

[11] S. Nanda, "Teletraffic Models for Urban and Suburban Microcells: Cell Sizes and Handoff Rates", *IEEE Trans. Veh. Tech.*, vol. 12, no. 4, pp. 673-682, 1993.

[12] B. Jabbari, "Teletraffic Aspects of Evolving and Next-generation Wireless Communication Networks", *IEEE Personal Commun.*, pp. 4-9, 1996.

[13] R. A. Guerin, "Channel Occupancy Time Distribution in a Cellular Radio System", *IEEE Trans. on Veh. Tech.*, vol. 35, no. 3, pp. 89-99, 1987.

[14] U. Herzog, L. Woo, and K. M. Chandy, "Solution of Queuing Problems by a Recursive Technique", *IBM J. .Research and Development*, vol. 19, pp. 295-300, 1975.

[15] D. Xue, X. Wang, "Adoption of Cognitive Radio Scheme to Class-based Call Admission Control", *IEEE ICC'2009*, pp. 1-7, 2009.

[16] G. Fodor, M. Telek, "A Recursive Formula to Calculate the Steady State of CDMA Networks", *19th International Teletraffic Congress (ITC)*, pp. 1285-1294, Beijing, China, 2005.

[17] K. Kassev, B. Tsankov, "Intra-cell Handover in OFDMA-based Wireless Access Networks", *ICEST'2010*, vol. 1, pp. 43-46, Ohrid, Macedonia, 2010.

M/M/k Queues Modelled by Using of Petri Net Simulator

Zoran Gacovski[1] and Emilija Kamceva[2]

Abstract - **In this paper we present a Petri net simulator that will be used for modeling of queues. The concept of queue is ubiquitous in the everyday life. Queues are present at the airports, banks, shops etc. Many factors are important in the studying of queues – some of them are: the average waiting time in the queue, the usage of the server, expected length of the queue etc. The paper is organized as follows: first we give a brief overview of the queuing theory. After that – we describe the Petri nets, as a tool for modeling of discrete event systems. We have implemented Petri net simulator that is capable for queue modeling. Finally we are presenting the simulation results for M/M/3 queue and its different properties. We've also made comparison between theoretically obtained results and our simulation results.**

Keywords – **Queuing theory, Petri net Simulator, M/M/k queue, Simulink.**

I. INTRODUCTION

The concept of queue line is one of the most adequate modeling techniques for the class of discrete-event system. Even though they evolved as an outcome of pure pragmatic reasons, in the early 1940 in Great Britain, their appliances developed and grow in a variety of aspects of the modern society, particularly concentrated on the Information and Communications Technologies. The exact analytic solutions provide the response for each relevant parameter toward system performance from the structural, dynamic and behavioral aspects. According to various researches, on the annual bases are spent more than 37 billion of hours on some of the queue strings types. Due to these reasons, the optimization of the attitude on this class of system has not only academic, but also enormously practical meaning. The length of this kind of waiting in everyday life is a stochastic process. For this kinds of "waitings" - information science has developed special branch of learning called the theory of queue lines.

The theory of line queues is a mathematical study of the length of the line. The theory of line queues represents and provides a mathematical analysis of several related processes, including arrival at the end of the line, waiting in line, and servicing the users by the repairers in order. The theory of queue lines allows distribution and calculation of several measures of performance, including the average time waiting in line or system, the expected number of customers waiting to be served or the one who already have been served, and the probability to find the system in certain situations, such as empty, full, available service at the moment, or taking some time waiting to be serviced [1].

[1]Zoran Gacovski is with the Faculty of Information Technology, FON University, Bul. Vojvodina, bb 1000 Skopje, Macedonia, E-mail: zoran.gacovski@fon.edu.mk.

[2]Emilija Kamceva is with the Faculty of Information Technology, FON University, Bul. Vojvodina, bb 1000 Skopje, Macedonia, E-mail: emilija.kamceva@fon.edu.mk.

II. PETRI NETS IN MATLAB

Petri Nets offer profound mathematical background originating namely from linear algebra and graph theory. Various Petri Net tools offer convenient graphical environment and sometimes they provide complex simulation and analysis of various high level Petri Net classes. Petri Net (PN) is mathematical and graphical modeling tool well suited for describing and analyzing discrete events systems (DES). PNs allow to model and visualize systems, which contain concurrence, resource sharing or synchronization. These possibilities allow them to be used for various applications in areas including computer systems, communication protocols, flexible manufacturing systems and software verification [2,3].

Within the mentioned context, the initiative of developing instruments for simulation, analysis and design of PNs under MATLAB brought remarkable benefits for training and research, because Control Engineering people are familiar with the exploitation of *Graphical User Interfaces* (GUIs) [4] based on this popular software. Although a recent list of the programs developed for PNs includes many resources (Mortensen, 2003) running under different operating systems, our initiative was successful due to the large preferences shown for MATLAB.

It is worth separately mentioning that the overall design and implementation gives us the ***PN Toolbox***, which is well integrated with MATLAB, and allows further developments in the modern direction of studying hybrid dynamics involving both DES and ODE models [5,6].

After ending a simulation experiment, several *Performance Indices* are available to globally characterize the simulated dynamics. Some of the indices recorded for the transitions of the net refer to: the total number of firings during the simulation (*Service Sum*), the mean frequency of firings (*Service Rate*), the mean time between two successive firings (*Service Distance*), the fraction of time when server is busy (*Utilization*). For the places of the net, the recorded indices refer to: the total number of arrived (*Arrival Sum*) and departed (*Throughput Sum*) tokens, the mean time between two successive instants when tokens arrive in (*Arrival Distance*) and depart from (*Throughput Distance*) the place, the mean time a token spends in a place (*Waiting Time*), the average number of tokens weighted by time (*Queue Length*) [7].

Only for timed or (generalized) stochastic PNs, the time evolution for both current and global values of a *Performance Index* may be displayed dynamically while in the *Step* and

Run Slow simulation modes by means of the *Scope* command. Another facility available only for timed or (generalized) stochastic PNs is *Design*, which can be used for the synthesis of the models. One or two *Design Parameters* varying within intervals defined by the user can be included in the model. For each test-point belonging to this (these) interval(s) a simulation experiment is performed in the *Run Fast* mode. The dependence of a *Design Index* on the *Design Parameter*(*s*) can be visualized as a graphical plot (2-D or 3-D, respectively) [8].

III. Petri Net Queue Simulator

The M/M/3 queue has been observed and toward it we have built a Petri network, shown on Figure 1. We have been carried out theoretical and simulated surveys for these queues. We have used the PN Toolbox for Matlab for our simulation. Typical for these types of queues is that theoretical results are calculated without considering the execution time for the system, while in simulation - the execution (service) time is also considered. From this point - obviously follows a diminutive difference between the obtained theoretical and simulation calculations.

During reviews and research of the queues, the most important properties that are required to be obtained are:

- λ - arrival rate of users
- μ - customer service rate
- ρ - occupancy rate of the server
- $E(X)$ - expected length of the line of customers in systems that have come and have been served
- $E(S)$ - average time in the system
- $E(W)$ - average waiting time in the line
- $E(R)$ - the flow of the users in the system

The M/M/3 queue model is shown on Figure1. It can be used to model discrete event systems, e.g., machines or operators, processing orders or communication equipment processing information.

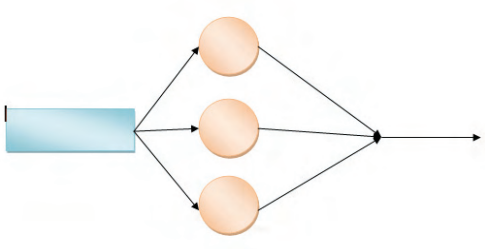


Fig. 1. Basic M/M/3 queue line

IV. Simulation Results

Case 1 - The Figure 2 shows the model of M/M/3 queue after performed simulation for population of 10000. The value of the customer arrival rate is $\lambda = 0{,}1$ and the value of the service rate is $\mu = 6{,}667$. We are presenting both theoretically obtained and simulation results. The time required to perform this simulation is 428,323 seconds.

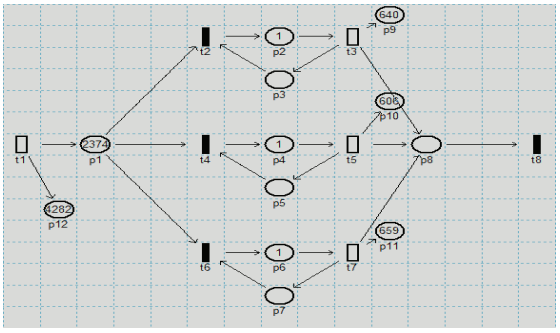


Fig. 2. Petri network for the M/M/3 queue - simulation for case 1

TABLE I
Theoretical and simulation results for the case 1

M/M/3	lamda	mi	ρ	po	p1
Theoretical	0,100	0,667	0,050	0,868	0,130
Simulation	0,100	0,675	0,049	0,870	0,129
Error %	0,003	1,250	1,231		

	p2	p3	E(U)	E(X)	E(S)	E(W)
Theor.	0,010	0,000	0,150	0,150	1,500	0,000
Simul.	0,010	0,000	0,148	0,148	1,481	0,000

Case 2 - Figure 3 shows the M/M/3 queue after performed simulation for population of 10000 for $\lambda = 0{,}1$ и $\mu = 0{,}392$. In Table 2 are shown the results obtained theoretically and by simulation. The time required to perform this simulation is 2573,458 seconds.

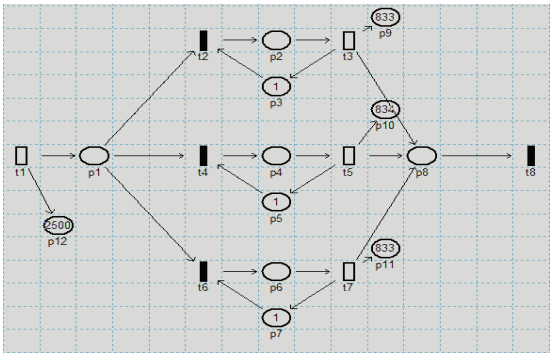


Fig. 3. Petri network for the M/M/3 queue - simulation for case 2

TABLE II
THEORETICAL AND SIMULATION RESULTS FOR THE CASE 2

M/M/3	lamda	Mi	ρ	po	p1
Theoretical	1,000	0,392	0,850	0,017	0,043
Simulation	0,971	0,324	1,000	0,000	0,000
Error %	2,858	17,428	14,646		

	p2	p3	E(U)	E(X)	E(S)	E(W)
Theor.	0,055	0,047	0,313	0,313	2,863	2,863
Simul.	0,000	0,000	0,333	0,343	3,431	3,333

Case 3 - Figure 4 shows the M/M/3 queue after the performed simulation for population of 10000, for $\lambda = 1$ и $\mu = 0,337$, then in Table 3 are shown, the results obtained theoretically and by simulation. The time required to perform this simulation is 2524,93 seconds.

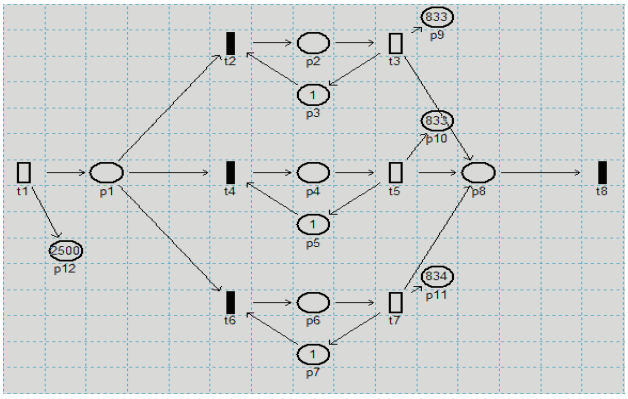


Fig. 4. Petri network for the M/M/3 queue - simulation for case 3

TABLE III
THEORETICAL AND SIMULATION RESULTS FOR THE CASE 3

M/M/3	lamda	mi	ρ	po
Theoretical	1,000	0,337	0,990	0,001
Simulation	0,990	0,330	1,000	0,000
Error %	0,987	1,990	1,023	

p1	p2	p3	E(U)	E(X)	E(S)
0,002	0,003	0,003	0,332	0,332	3,302
0,000	0,000	0,000	0,333	0,337	3,367

Case 4 - Figure 5 shows the M/M/3 queue after the performed simulation for population of 10000, for $\lambda = 1$ и $\mu = 0,337$, then in Table 4 are shown the results obtained theoretically and by simulation. The time required to perform this simulation is 2591,79 seconds.

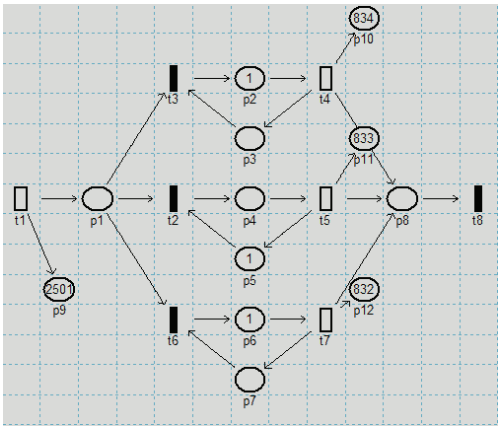


Fig. 5. Petri network for the M/M/3 queue - simulation for case 4

TABLE IV
THEORETICAL AND SIMULATION RESULTS FOR THE CASE 4

M/M/3	Lamda	Mi	ro	Po
Theoretical	1,000	0,303	1,100	-0,006
Simulation - 10000	0,965	0,321	1,001	0,000
Error %	3,503	6,061	9,017	

p1	p2	p3	E(U)	E(X)	E(S)
-0,019	-0,031	-0,034	0,342	0,342	3,642
0,000	0,000	0,000	0,334	0,346	3,457

After several simulations performed for the queue M/M/3 by using the concept of Petri nets, we have concluded that: by comparing the values of arrival rate and the service rate (theoretical and simulation research), we have obtained similar values with a small error, which expressed in percents - does not exceed 15%. This confirms the accuracy of our research of the M/M/3 queue.

If we compare the time of execution for different values of ρ, we can conclude that by increasing the value of ρ - the time of execution of the simulation is reduced (work of the system).

In the last case - for Petri network model of the M/M/3 queue, we've taken the value $\rho = 1.1111 > 1$. This contradicts the theoretical results, because of the definition $\rho < 1$. But for the same default values for the arrival and the

service rate, after performed simulation - we've received value $\rho < 1$, which confirms the accuracy of the simulator and the goal of our research. That shows a possible blockade in the system.

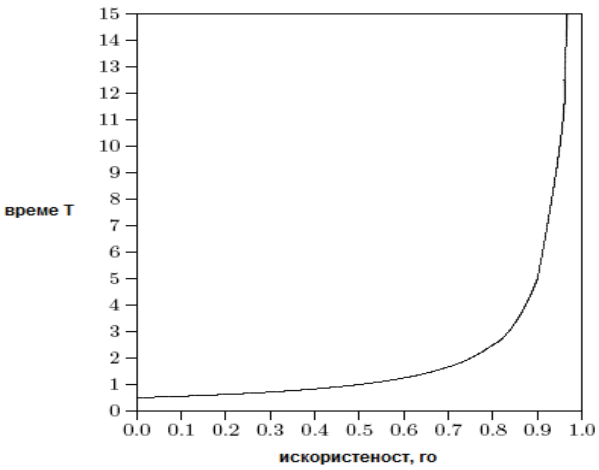


Fig. 6. The usage of the server.

V. CONCLUSION

In this paper we've presented a Petri net simulator that was used for modeling of queues. Many factors are important in the studying of queues – we have considered the following: the average waiting time in the queue, the usage of the server, expected length of the queue etc. First we gave a brief overview of the queuing theory. After that – we described the Petri nets, and a software tool for modeling of discrete event systems. We have implemented Petri net simulator (via PN Matlab toolbox) that is capable for queue modeling. Finally we have calculated the simulation results for M/M/3 queue and its different properties (and also its theoretical values). We've made comparison between our simulation results and theoretically obtained results and we have shown significant matching.

REFERENCES

[1] B.D. Bunday, *An introduction to queueing theory*, Arnold, London,1996.

[2] M.V. Iordache and P.J. Antsaklis, "Software Tools for the Supervisory Control of Petri Nets Based on Place Invariants", Technical Report ISIS-2002-003, University of Notre Dame, IN, USA, 003.pdf, 2002.

[3] M. Svádová and Z. Hanzálek, "Matlab Toolbox for Petri Nets", *22nd International Conference ICATPN 2001*, Newcastle, UK, pp. 32-36, 2001.

[4] The MathWorks Inc., *Building GUIs with MATLAB*, Natick, Massachusetts, 2000.

[5] C. Mahulea, L. Bârsan and O. Pastravanu, "Matlab Tools for Petri-Net-Based Approaches to Flexible Manufacturing Systems", In: F.G. Filip, I. Dumitrache and S. Iliescu (Eds.), *9th IFAC Symposium on Large Scale Systems LSS 2001*, Bucharest, Romania, pp. 184-189, 2001.

[6] M.H. Matcovschi, C. Mahulea, and O. Pastravanu, "Exploring Structural Properties of Petri Nets in MATLAB", *7th International Symposium on Automatic Control and Computer Science SACCS 2001*, Iasi, Romania, CD Rom, 2001.

[7] C.G. Cassandras, *Discrete Event Systems: Modeling and Performance Analysis*, Irwin, 1993.

[8] T. Murata, "Petri Nets: Properties, Analysis and Applications", *Proc.of the IEEE*, vol. 77, pp. 541-580, 1989.

Session RMA I:

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS I

Outage Probability of AF System with Interference-Limited Relay over Rayleigh/Rician Fading Channels

Mihajlo Č. Stefanović[1], Aleksandra M. Cvetković[2], Jelena A. Anastasov[3] and Goran T. Đorđević[4]

***Abstract* – This paper studies the performance of a dual-hop amplify-and-forward system where the source-relay and the relay-destination channels experience Rayleigh and Rician fading, respectively. The relay node is corrupted by Rayleigh faded multiple co-channel interferences. New closed-form expressions for outage probability of the end-to-end signal-to-interference and noise ratio (SINR) are derived. An influence of various parameters, such as Rician *K* factor, number of cochannel interferences and outage threshold on outage probability is considered.**

***Keywords* – amplify-and-forward relay, co-channel interference, outage probability**

I. INTRODUCTION

Multihop relaying technology in communication networks is often used as a reasonable solution in a case of deep faded direct link. Relayed transmission is based on several terminals between source and destination providing better communication and reliable signal traversing without using large transmitting power [1, 2]. This type of transmission has gained enormous interest in the context of user-cooperative communications. Regarding the nature and complexity of relays used in cooperative communication, there are two dominant categories of multihop transmission systems: regenerative and nonregenerative systems. In regenerative systems, referred as decode-and-forward (DF), relays decode the signal and retransmitting it to the next node [1]. On the other hand, in nonregenerative systems or amplify-and-forward (AF) systems, relays just amplify and forward the incoming signal without performing any decoding at all [2-4]. AF relaying mechanism imposes less power consumption and can be preferable in practice. Variable gain requires knowledge of the instantaneous channel realization at the relays. Those are so-called channel state information (CSI) - assisted relays, which use the instantaneous CSI from the previous hop to produce their gain leading to a power control of the retransmitted signal [4-7].

In many published papers, the research goal on this subject has been to derive the outage probability expression of dual-hop wireless communication systems in noise-limited environment [1-7]. Nevertheless, in many practical scenarios co-channel interference (CCI) is also an important issue and should be taken in consideration. Consideration of CCI is necessary because of the aggressive reuse of frequency channels for high spectrum utilization in cellular systems [8-10]. In few published works, the impact of interference on the AF and the DF relaying performance have been investigated either at the relay(s) or destination(s) [11-12]. In [11], the performance of a two hop CSI-assisted amplify-and-forward system, with co-channel interference at the relay was analyzed. The system's outage probability and the average bit error rate (ABER) in the presence of Rayleigh faded multiple interferers were investigated. In the paper [12], analytical closed-form expressions for outage probability of AF and DF relays in interference-limited environment were derived. In that paper, the performance of a Rayleigh dual-hop relay channel effected by an additive white Gaussian noise (AWGN) at the relay and multiple co-channel interferences at the destination was investigated. The paper [13] studies the outage probability of the non-regenerative relays over Rayleigh fading channels in an interference-limited environment (the relay and destination nodes are corrupted by co-channel interference).

In this paper, we focus on dual-hop AF relay transmission systems and study their performance over mixed Rayleigh and Rician fading channels in the presence of co-channel interferences. In practice, different links in relay networks can experience separate fading conditions. For example, WINNER II project [14] confirms the existence of different (mixed Rayleigh/Rician) fading conditions. Also in [15-16], a base station-relay link is considered as Rayleigh, while the relay-mobile link is observed as Rician link because of a strong line-of-sight (LoS) component. In the following analysis, we assume that there are multiple interferers at the relay node independent of the desired signal. In the observed scenario, co-channel interferences are subjected to Rayleigh fading.

Closed-form expressions for the outage probability of the end-to-end signal-to-interference and noise ratio (SINR) for CSI-assisted relayed systems are derived. In particular, we derive exact expressions in terms of an infinite series, which numerically converges quickly for a finite number of terms.

[1]Mihajlo Č. Stefanović is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: mihajlo.stefanovic@elfak.ni.ac.rs.

[2]Aleksandra M. Cvetković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: aleksandra.cvetkovic@elfak.ni.ac.rs.

[3]Jelena A. Anastasov is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: jelena.anastasov@elfak.ni.ac.rs.

[4]Goran T. Đorđević is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: goran.t.djordjevic@elfak.ni.ac.rs.

II. System and Channel Model

The system observed in this paper is depicted in Fig. 1. We consider the wireless communication system, from the source terminal *S* to the destination terminal *D* assisted by a nonregenerative gain relay transmission. The relay terminal is corrupted by co-channel interferences and AWGN while destination terminal is only perturbed by an AWGN.

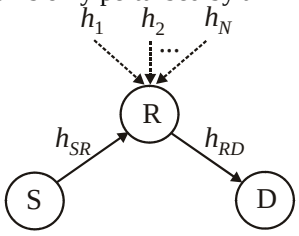


Fig. 1. System model

We assume that terminal *S* transmits a desired symbol, s_0, with an average power $\mathrm{E}\left[|s_0|^2\right] = P_0$ (E[.] is the expectation operator). The level of co-channel interference at the relay is high enough compared to the level of thermal noise, so the thermal noise can be neglected as in all interference-limited fading environments. The received signal at relay terminal *R*, in interference-limited fading environment, can be written as

$$r_R = h_{SR}s_0 + \sum_{i=1}^{N} h_i s_i \tag{1}$$

where h_{SR} is the fading amplitude of the channel between terminals *S* and *R*, $\{h_i\}_{i=1}^N$ are amplitudes of the interferers at the input of *R* and $\{s_i\}_{i=1}^N$ are interfering symbols with an average power P_i each of them. In nonregenerative systems, the signal r_R is then multiplied by the gain *G* of terminal *R* and retransmitted to terminal *D* effected by an AWGN. The received signal at terminal *D* can be presented as

$$r_D = h_{RD}G\left(h_{SR}s_0 + \sum_{i=1}^{N} h_i s_i\right) + n_D \tag{2}$$

where h_{RD} is the fading amplitude of the channel between terminals *R* and *D*, and n_D is the AWGN at the input of *D* with $\mathrm{E}\left[|n_D|^2\right] = \sigma_D^2$.

The overall SINR at the receiving end can be expressed as [11]

$$\gamma_{eq} = \frac{|h_{SR}|^2|h_{RD}|^2G^2P_0}{|h_{RD}|^2G^2\sum_{i=1}^{N}|h_i|^2P_i + \sigma_D^2} = \frac{|h_{SR}|^2P_0\dfrac{|h_{RD}|^2P_R}{\sigma_D^2}}{\dfrac{|h_{RD}|^2P_R}{\sigma_D^2}\sum_{i=1}^{N}|h_i|^2P_i + \dfrac{P_R}{G^2}} \tag{3}$$

where P_R is the power of the transmitted signal at the output of the relay.

When terminal *R* has available instantaneous CSI from the first hop, the gain *G* is given by

$$G_1^2 = \frac{P_R}{|h_{SR}|^2P_0 + \sum_{i=1}^{N}|h_i|^2P_i} \tag{4}$$

Therefore, the instantaneous equivalent end-to-end SINR in this case can be obtained by substituting (4) in (3) and has a form [11]

$$\gamma_{eq1} = \frac{\gamma_1\gamma_2}{\gamma_3(\gamma_2+1)+\gamma_1} \tag{5}$$

where $\gamma_1 = |h_{SR}|^2P_0$, $\gamma_2 = \dfrac{|h_{RD}|^2P_R}{\sigma_D^2}$ and $\gamma_3 = \sum_{i=1}^{N}|h_i|^2P_i$. A CSI-assisted nonregenerative relay requires a continuous estimation of the channel fading amplitude from the first hop and inverts the fading effect in order to limit the output power of the relay [5].

We consider an asymmetric fading scenario of the *S*−*R* and *R*−*D* links. Namely, the *S*−*R* link is subjected to Rayleigh fading and the *R*−*D* link is subjected to Rician fading. If a link experiences Rayleigh fading, γ_1 is exponentially distributed by probability density function (PDF) given by

$$p_{\gamma_1}(\gamma) = \frac{1}{\bar{\gamma}_1}\exp\left(-\frac{\gamma}{\bar{\gamma}_1}\right) \tag{6}$$

where $\bar{\gamma}_1 = \mathrm{E}\left[|h_{SR}|^2\right]P_0$ is the average signal power of *S*−*R* channel.

If a link experiences Rician fading, γ_2 is a noncentral χ^2 chi-squared distributed

$$p_{\gamma_2}(\gamma) = \frac{(K+1)\exp(-K)}{\bar{\gamma}_2}\exp\left(-\frac{(K+1)\gamma}{\bar{\gamma}_2}\right)I_0\left(2\sqrt{\frac{K(K+1)\gamma}{\bar{\gamma}_2}}\right) \tag{7}$$

where $\bar{\gamma}_2 = \mathrm{E}\left[|h_{RD}|^2\right]P_2/\sigma_D^2$ is the average SNR per hop of *S*−*R* channel, *K* is Rician factor and $I_0(.)$ is the zeroth order modified Bessel function of the first kind [17, eq. 8.406]. We assume that co-channel interference fading amplitudes are also modeled as Rayleigh random processes. When all *N* interferers are identical, γ_3 becomes a central χ^2 random variable with 2*N* degrees of freedom [12]

$$p_{\gamma_3}(z) = \frac{1}{\bar{\gamma}_3^N\Gamma(N)}\exp\left(-\frac{z}{\bar{\gamma}_3}\right) \tag{8}$$

where $\bar{\gamma}_3 = \mathrm{E}\left[|h_i|^2\right]P_i$, *i*=1,2,...*N*, is the average power of each CCI.

III. Outage Probability

One of the accepted performance measures, involving dual-hop systems in fading environments, is the outage probability. Roundly speaking for this case, outage probability is defined

as probability that the instantaneous equivalent SINR, γ_{eq}, falls below a predetermined protection ratio, γ_{th} [18]-[19].

When terminal R has available CSI from the first hop, the outage probability of the instantaneous equivalent SINR can be expressed as [6], [11]

$$\begin{aligned} P_{eq1}(\gamma_{th}) &= \Pr(\gamma_{eq1} < \gamma_{th}) \\ &= \int_0^\infty \int_0^\infty P_{\gamma_1}\left(\frac{\gamma_1\gamma_2}{\gamma_3(\gamma_2+1)+\gamma_1} \le \gamma_{th} \middle| \gamma_2, \gamma_3\right) p_{\gamma_2}(\gamma_2) p_{\gamma_3}(\gamma_3) \mathrm{d}\gamma_2 \mathrm{d}\gamma_3 \\ &= \int_0^\infty \int_0^\infty P_{\gamma_1}\left(\gamma_1 \le \frac{\gamma_{th}(y+1)z}{y-\gamma_{th}}\right) p_{\gamma_2}(y) p_{\gamma_3}(z) \mathrm{d}y\mathrm{d}z \end{aligned} \quad (9)$$

After some mathematical manipulations and simplifications, (11) can be rewritten as

$$P_{eq1}(\gamma_{th}) = 1 - \int_0^\infty \int_0^\infty P_{\gamma_1}\left(\gamma_1 \ge \frac{\gamma_{th}(x+\gamma_{th}+1)z}{x}\right) p_{\gamma_2}(x+\gamma_{th}) p_{\gamma_3}(z) \mathrm{d}x\mathrm{d}z \quad (10)$$

In order to evaluate (10), the complementary cumulative distribution function (CDF) of γ_1 and PDFs of the γ_2 and γ_3 are needed. The complementary CDF of γ_1 can be expressed as $P_{\gamma_1}(\gamma) = \exp(-\gamma/\bar{\gamma}_1)$ and PDFs of γ_2 and γ_3 are given by (7) and (8), respectively.

After applying these formulae, the outage probability has the following form

$$\begin{aligned} P_{eq1}(\gamma_{th}) = 1 &- \frac{(K+1)\exp(-K)}{\bar{\gamma}_2} \exp\left(-\frac{K+1}{\bar{\gamma}_2}\gamma_{th}\right) \frac{1}{\bar{\gamma}_3^N \Gamma(N)} \\ &\cdot \int_0^\infty \int_0^\infty \exp\left(-\frac{\gamma_{th}(x+\gamma_{th}+1)z}{\bar{\gamma}_1 x}\right) \exp\left(-\frac{K+1}{\bar{\gamma}_2}x\right) \\ &I_0\left(2\sqrt{\frac{K(K+1)(x+\gamma_{th})}{\bar{\gamma}_2}}\right) z^{N-1} \exp\left(-\frac{z}{\bar{\gamma}_3}\right) \mathrm{d}x\mathrm{d}z \end{aligned} \quad (11)$$

We are unaware of a closed-form analytical solution to this integral. Nevertheless, using the infinite-series representation of $I_0(.)$ [17, Eq. (8.447.1)], the integral in (11) can be solved

$$\begin{aligned} P_{eq1}(\gamma_{th}) = 1 &- \frac{(K+1)\exp(-K)}{\bar{\gamma}_2} \exp\left(-\frac{K+1}{\bar{\gamma}_2}\gamma_{th}\right) \left(\frac{\rho}{\gamma_{th}(\gamma_{th}+1)}\right)^N \\ &\cdot \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{\gamma_{th}^{i-j}}{i!\,j!\,(j-1)!} \left(\frac{K(K+1)}{\bar{\gamma}_2}\right)^i \left(\frac{\gamma_{th}(\gamma_{th}+1)}{\gamma_{th}+\rho}\right)^{N+j+1} \\ &\Gamma(N+j+1) U\left(N+j+1,\ j+2;\ \frac{K+1}{\bar{\gamma}_2}\frac{\gamma_{th}(\gamma_{th}+1)}{\gamma_{th}+\rho}\right) \end{aligned} \quad (12)$$

where $\rho = \bar{\gamma}_1/\bar{\gamma}_3$ and $U(a,b;z)$ is the confluent hypergeometric function of the second kind defined as $U(a,b;z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1} \mathrm{d}t$ [17, Eq. (9.211.4)].

Outage probability as a function of average SNR of S–R channel, $\bar{\gamma}_2$, for various values of ρ is presented in Fig. 2. Numerical results are presented for the case of Rayleigh fading (K=0) on both hops and obtained curves are consistent with those reported in [11, Fig. 1]. Also, the results for the considered model in this paper and K=10 dB are shown. As expected, when ρ increases, the outage probability decreases.

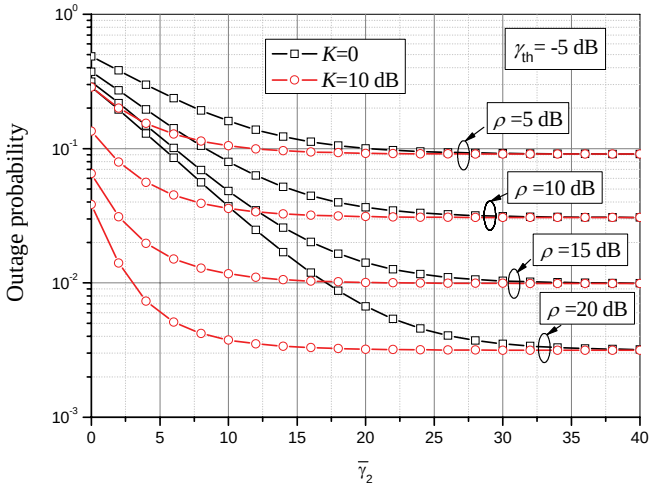


Fig. 2. Outage probability for various values of ρ

In Fig. 3. outage probability is represented for various values of Rician K factor. Rician K factor significantly affects the outage performance i.e. outage probability decreases as K increases. It is noticeable that for particular values of average SNR, values of outage probability tend to irreducible outage floor. We can also see that outage floor does not depend on K, but only on ρ.

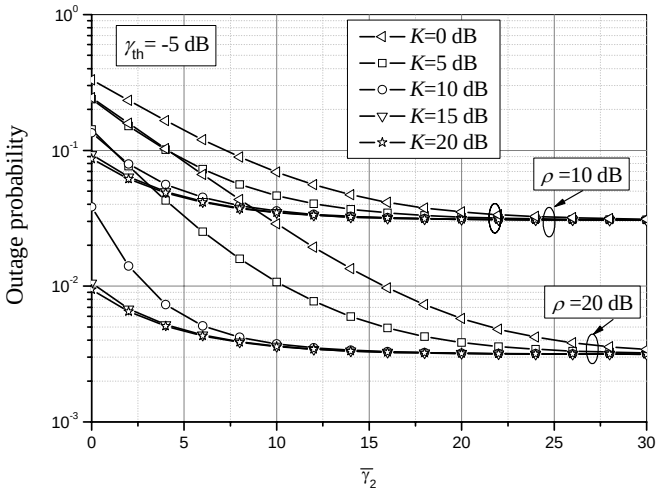


Fig. 3. Outage probability for various K factors

Fig. 4. depicts the outage performance for different outage threshold of CSI-assisted relay. It is observed that as γ_{th} increases, the outage probability also increases. When γ_{th} increases from -5dB to 5dB, outage probability changes about 10 times.

Fig. 5. shows the outage performance for different number of interferences. As the number of interference increases, outage probability increases, which degrades the system performance. The largest performance degradation is present when the number of CCI increase from one to two. It is evident that outage floors depend on number of CCI.

Figs. from 2 to 5 cleary show the existence of outage probability floor. It is evident that outage floors strongly depend on values of ρ.

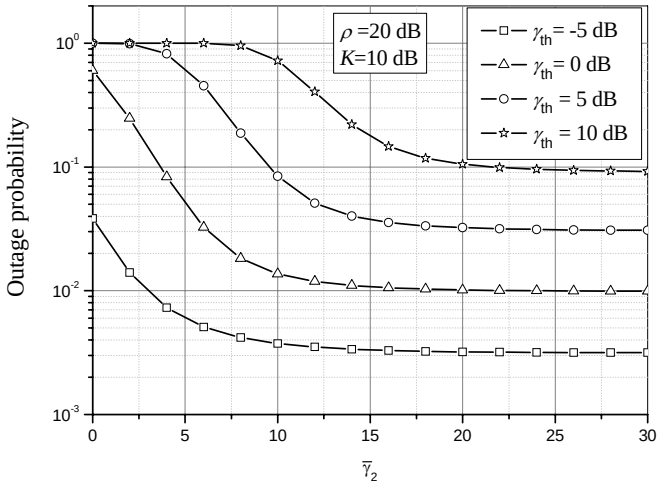


Fig. 4. Outage probability for various values of outage threshold

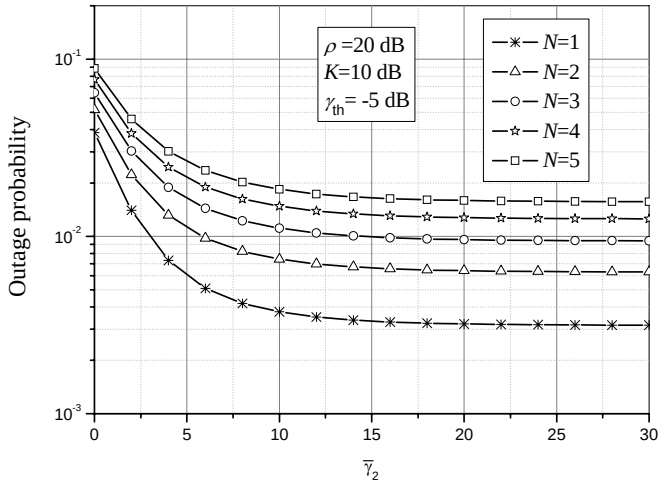


Fig. 5. Outage probability for various number of co-channel interferences

IV. Conclusion

The performance of dual-hop systems with amplify-and-forward relays under assumption that relay is affected by multiple co-channel interferences has been studied. We have derived closed-form expression for the outage probability of the instantaneous SINR for dual-hop transmission with CSI-based relays operating over asymmetric fading channels. The effects of Rician *K* factor, number of co-channel interferences, outage threshold on the system's performance, were interpreted. After all, proposed analysis can be used in design of a cellular mobile system to determine optimal values of system parameters to provide a reasonable indication of signal outage.

Acknowledgement

The work presented here was supported by the Serbian Ministry of Education and Science (project TR32052).

References

[1] M.O. Hasna, and M.-S. Alouini, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Communications Letters*, vol. 7, no. 5, pp. 216-218, 2003.

[2] M.O. Hasna, and M.-S. Alouini, "A performance study of dual-hop transmissions with fixed gain relays," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1963-1968, 2004.

[3] J. Boyer, D.D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," IEEE Transactions on Communications, vol. 52, no. 10, pp. 1820-1830, 2004.

[4] G.K. Karagiannidis, "Performance bounds of multihop wireless communications with blind relays over generalized fading channels," IEEE Transactions on Wireless Communications, vol. 5, no. 3, pp. 498-503, 2006.

[5] M.O. Hasna, and M.-S. Alouini, "End-to-end performance of transmission systems with relays over Rayleigh-fading channels," IEEE Transactions on Wireless Communications, vol. 2, no. 6, pp. 1126-1131, 2003.

[6] T.A. Tsiftsis, G.K. Karagiannidis, P.T. Mathiopoulos, and S.A. Kotsopoulos, "Nonregenerative Dual-Hop Cooperative Links with Selection Diversity," *EURASIP Journal on Wireless Communication and Networking*, vol. 2006, pp.1-8, 2006.

[7] M.O. Hasna, and M.-S. Alouini, "Harmonic mean an end-to-end performance of transmission systems with relays," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 130-135, 2004.

[8] W.C. Jakes, *Microwave Mobile Communications*, IEEE Press: Piscataway, New Jersey, 1994. 2nd edn.

[9] M.K Simon, M-S Alouini, *Digital Communication Over Fading* Channels, Wiley, New York, 2000.

[10] P. Stavroulakis, *Interference analysis and reduction for wireless systems*, Artech House, Boston, London, 2003.

[11] H. A. Suraweera, H. K. Garg, and A. Nallanathan, "Performance Analysis of Two Hop Amplify-and-Forward Systems with Interference at the Relay," *IEEE Communications Letters*, vol. 14, no. 8, pp. 692-694, 2010.

[12] C. Zhong, S. Jin, and K.-K. Wong, "Dual-hop systems with noisy relay and interference-limited destination," *IEEE Transactions on Communications*, vol. 58, no. 3, pp. 764-768, 2010.

[13] A. M. Cvetković, G. T. Đorđević, and M. Č. Stefanović, "Performance of interference-limited dual-hop non-regenerative relays over Rayleigh fading channel," *IET Communication*, vol. 5, no. 2, pp. 135–140, 2011.

[14] P. Kyosti, J. Meinila, L. Hentila et al., *WINNER II Interim Channel Models* (IST-4-027756 WINNER II D1.1.1 V1.1) [Online]. Available: http://www.ist-winner.org/WINNER2-Deliverables/D1.1.1.pdf.

[15] H. A. Suraweera, G. K. Karagiannidis, and P. J. Smith, "Performance Analysis of the Dual-Hop Asymmetric Fading Channel," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 2783-2788, 2009.

[16] H. A. Suraweera, R. H. Y. Louie, Y. Li, G. K. Karagiannidis, and B. Vucetic, "Two Hop Amplify-and-Forward Transmission in Mixed Rayleigh and Rician Fading Channels," IEEE Communications Letters, vol. 13, no. 4, pp. 227-229, 2009.

[17] I. S. Gradshteyn, and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, 5th ed., Orlando, 1994.

[18] H. Yang, and M.-S. Alouini, "Outage probability of dual-branch diversity systems in presence of co-channel interference," *IEEE Transactions on Communications*, vol. 2, no. 2, pp. 310–319, 2003.

[19] A. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, 2006.

Outage Probability of Correlated SC SIR-based Diversity System over *K* Fading Channels

Jelena A. Anastasov[1], Aleksandra M. Cvetković[2], Stefan R. Panić[3], Dejan N. Milić[4] and Dušan M. Stefanović[5]

***Abstract* – The paper presents brief analysis of selection combining (SC) diversity over correlated *K*–fading channels. The proposed system is considered as interference-limited system in correlated fading environment. The cumulative distribution function (cdf) of signal-to-interference ratio (SIR) at the output of SC receiver is derived in a form of Meijer*G* functions. According to this new expression, the outage probability is considered and the effects of shaping parameters and correlation coefficients on the performance gain are analysed. The proposed analytical approach is accompanied by numerical results.**

***Keywords* – SC diversity system, interference, *K*-fading channels, outage probability**

I. INTRODUCTION

The statistical characterization of signal propagation, in performance evaluations of wireless communication system, is frequently based on analysis of composite fading: large-scale fading due to multipath propagation and small-scale fading due to shadowing phenomenon [1]. This destructive combination of obstacles often occurs simultaneously. The multipath fading is mostly modeled by Rice, Rayleigh and Nakagami-*m* distributions and shadowing by lognormal distribution. This lognormal based fading model is analyzed in few papers [1-3]. However, the analytical analysis of this model is very complicated, so in some other papers the Gamma distribution is employed as an useful solution for describing shadow fading phenomena [4]. Furthermore, based on Gamma distribution, several generalized distributions have been proposed as composite fading channel models, the generalized Gamma , the *G* [5] and *K* [6-9] fading channel models.

An efficient technique to diminish the influences of fading and cochannel interference is space diversity reception. The principle of space diversity reseption is to provide the receiver with multiple faded replicas of the same information-bearing signal. The goal is to upgrade transmission reliability without increasing transmission power and bandwidth, as well as to increase channel capacity.

Among the well-known diversity schemes, selection combining (SC) space-diversity technique is with lower complexity nature opposed to maximal-ratio combining (MRC) and equal-gain combining (EGC) techniques, which require the amount of the channel-state information of transmitted signal.

In an interference-limited system (the thermal noise is ignored), the effect of interference needs to be taken into consideration of performance analysis. The most effective performance criterion, in that case, is to select the signal-to-interference ratio (SIR) [10]. Independently of the channel condition, the transmitted signals as well as the interfered ones could be correlated due to small distance between the diversity antennas.

The presented evaluations in [6] affirm that the Rayleigh-lognormal and *K* distributions are similar, but the latter has a simpler form and thus is more appropriate for analysis and design of wireless communication systems. The moment generating function (mgf) based performance analysis of generalized selection combining (GSC) diversity operating over independently distributed *K* fading channels is presented in [7]. In [8] the brief performance evaluations of the output signal-to-noise ratio (SNR) for SC diversity operating over correlated *K* fading channels is presented. The performance analysis of multiple branches SC diversity combining technique, this time over exponentially correlated *K* fading channels is considered in [9]. Again, the environment corrupted by additive white Gaussian noise (AWGN) is considered.

In this paper, the performance of SC diversity receiver in interference-limited environment is analyzed. Considering the composite correlated fading channel condition, infinite series expressions for the performance evaluation are derived. The detailed analysis of outage probability is presented. Based on analytical results, required numerical results are also presented.

II. SC PERFORMANCE OVER *K* FADING CHANNELS

In general, the SC combiner selects the branch with the highest signal-to-noise ratio (SNR), actually the branch with the strongest signal assuming equal noise power among the branches. In environments where the level of the cochannel interference is sufficiently high as compared to the thermal noise, SC picks the signal at the branch with highest SIR

[1]Jelena A. Anastasov is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: jelena.anastasov@elfak.ni.ac.rs

[2]Aleksandra M. Cvetković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: aleksandra.cvetkovic@elfak.ni.ac.rs

[3]Stefan R. Panić is with the University of Priština, Faculty of Natural Science and Mathematics, Lole Ribara 29, 38000 Kosovska Mitrovica, Serbia, E-mail: stefanpnc@yahoo.com

[4]Dejan N. Milić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dejan.milic@elfak.ni.ac.rs.

[5]Dušan M. Stefanović is with the University of Niš, High Technical college, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: dusan.stefanovic@itcentar.rs

(SIR-based selection diversity). The system model, we have analyzed in this paper is presented in Fig. 1.

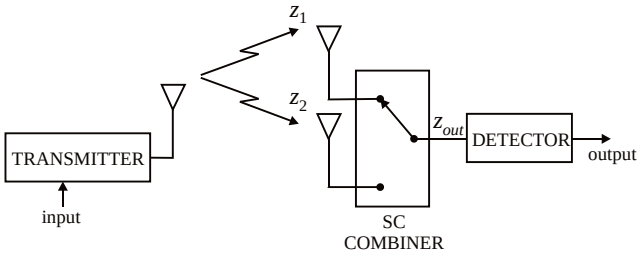


Fig. 1. System model

Assuming the composite fading channel model, the *K* distribution is convenient and mathematically tractable distribution for evaluating adequate performance criteria. The *K* distribution accurately approximates some of fading and/or shadowing models.

The received resultant of desired and interfering signal, considering *K* fading environment, can be presented as [11]

$$y(t)=x(t)\exp\left(j\left(2\pi f_c t+\psi(t)+\theta(t)\right)\right) \tag{1}$$

where f_c is a carrier frequency, $\psi(t)$ is the information signal, $\theta(t)$ is the random phase and $x(t)$ is *K* distributed random envelope process, given by [7, eq. (1)]

$$p(x)=\frac{4}{\Gamma(k)\Omega^{k/2}}x^k K_{k-1}\left(\frac{2x}{\sqrt{\Omega}}\right) \tag{2}$$

where $K_{k\text{-}1}(.)$ is the (k-1)th order modified Bessel function of the second kind [12, eq. (8.407)], $\Gamma(.)$ is the Gamma function [12, eq. (8.310^7.1)] and $\Omega=\mathrm{E}(x^2)/k$ with E(.) being expectation operator. The parameter k is a shadowing severity parameter $k\in(0,\infty)$.

The proposed model refers to the case of a single cochannel interferer. The performance of SC is carried out by considering the effect of only the strongest interferer. We assume that the remaining interferers are combined and considered as uncorrelated interference between antennas.

There is a need to derive the joint statistics for two *K* variables. Assuming [8, eq. (4)], the joint pdfs for both desired and interfering signal envelopes can be, respectively, expressed as

$$p_{R_1R_2}(R_1,R_2)=\frac{16}{\Gamma(k_d)}\sum_{a,b=0}^{\infty}\frac{\rho_{nd}^a\rho_{gd}^b}{\Gamma(1+a)\Gamma(k_d+b)} \times\frac{\prod_{l=1}^{2}\left(R_l/\sqrt{\Omega_{dl}}\right)^{\varepsilon_d}K_\psi\left(2R_l\sqrt{1/(1-\rho_{nd})(1-\rho_{gd})\Omega_{dl}}\right)}{a!b!(1-\rho_{nd})^{k_d+a+b}(1-\rho_{gd})^{1+a+b}R_1R_2} \tag{3}$$

and

$$p_{r_1r_2}(r_1,r_2)=\frac{16}{\Gamma(k_c)}\sum_{c,d=0}^{\infty}\frac{\rho_{nc}^c\rho_{gc}^d}{\Gamma(m_c+c)\Gamma(k_c+d)} \frac{\prod_{l=1}^{2}\left(r_l/\sqrt{\Omega_{cl}}\right)^{\varepsilon_c}K_\psi\left(2r_l\sqrt{1/(1-\rho_{nc})(1-\rho_{gc})\Omega_{cl}}\right)}{c!d!(1-\rho_{nc})^{k_c+c+d}(1-\rho_{gc})^{1+c+d}r_1r_2} \tag{4}$$

where, $\varepsilon_d=k_d+1+a+b$, $\varepsilon_c=k_c+1+c+d$, $\psi_d=k_d+b-1-a$, $\psi_c=k_c+d-1-c$; ρ_{nd} and ρ_{nc} are the power correlation coefficients between the envelopes of the desired signals and interfering signals, respectively; and ρ_{gd} and ρ_{gc} are correlation coefficients between the average fading power of the desired signals and interfering signals, respectively. Parameters k_d and k_c are shadowing severity parameters of desired and interfering signals, respectively.

Instantaneous values of SIR at the l-th input diversity branch can be defined as $z_l=R_l^2/r_l^2$, l=1, 2. The selection combiner chooses and outputs the branch with the highest SIR

$$z=z_{out}=\max(z_1,z_2). \tag{5}$$

Let $S_l=\hat{R}_l^2/\hat{r}_l^2$ be the average SIR at the l-th input branch of the selection combiner.

Furthermore, due to a scenario with closely placed diversity antennas, both desired and interfering signal envelopes experience correlated *K* fading. The joint pdf for input SIRs, regarding $\Omega_{dl}=\mathrm{E}(R_l^2)/k_d$, $\Omega_{c_l}=\mathrm{E}(r_l^2)/k_c$, can be evaluated as [13]

$$p_{z_1z_2}(z_1,z_2)=\frac{1}{4\sqrt{z_1z_2}}\int_0^\infty\int_0^\infty p_{R_1R_2}\left(r_1\sqrt{z_1},r_2\sqrt{z_2}\right)p_{r_1r_2}(r_1,r_2)r_1r_2dr_1dr_2 \tag{6}$$

Substituting (3) and (4) in (6) and representing $K_\psi(.)$ in a form of Meijer$G$ function as $K_\psi(x)=\frac{1}{2}G_{0,2}^{2,0}\left(\frac{-}{\frac{\psi}{2},-\frac{\psi}{2}}\middle|\frac{x^2}{4}\right)$, [14, eq. (03.04.26.0006.01)], we get the following double-fold integral form

$$I=\int_0^\infty\int_0^\infty\prod_{l=1}^{2}r_l^{\varepsilon_d+\varepsilon_c-1}G_{0,2}^{2,0}\left[\frac{r_l^2z_l}{(1-\rho_{nd})(1-\rho_{gd})\Omega_{dl}}\middle|\begin{matrix}-\\ \psi_d/2,\ -\psi_d/2\end{matrix}\right] \times G_{0,2}^{2,0}\left[\frac{r_l^2}{(1-\rho_{nc})(1-\rho_{gc})\Omega_{cl}}\middle|\begin{matrix}-\\ \psi_c/2,\ -\psi_c/2\end{matrix}\right]dr_1dr_2 \tag{7}$$

Now, making change of variables $t=r_1^2/(1-\rho_{nc})(1-\rho_{gc})\Omega_{c1}$ and $u=r_2^2/(1-\rho_{nc})(1-\rho_{gc})\Omega_{c2}$ and using [14, eq. (07.34.21.0011.01)], we get

$$p_{z_1z_2}(z_1,z_2)=\sum_{a,b,c,d=0}^{+\infty}\mathrm{A}\times z_1^{\varepsilon_d/2-1}z_2^{\varepsilon_d/2-1} \times G_{2,2}^{2,2}\left[\frac{z_1\sigma}{S_1}\middle|\begin{matrix}1-(\varepsilon_d+\varepsilon_c+\psi_c)/2,\ 1-(\varepsilon_d+\varepsilon_c-\psi_c)/2\\ \psi_d/2,\ -\psi_d/2\end{matrix}\right] \times G_{2,2}^{2,2}\left[\frac{z_2\sigma}{S_2}\middle|\begin{matrix}1-(\varepsilon_d+\varepsilon_c+\psi_c)/2,\ 1-(\varepsilon_d+\varepsilon_c-\psi_c)/2\\ \psi_d/2,\ -\psi_d/2\end{matrix}\right] \tag{8}$$

with

$$\mathrm{A}=\frac{\rho_{nd}{}^{a}\rho_{gd}{}^{b}\rho_{nc}{}^{c}\rho_{gc}{}^{d}}{\Gamma(k_d)\Gamma(k_c)\Gamma(1+a)\Gamma(k_d+b)\Gamma(1+c)\Gamma(k_c+d)S_1^{\varepsilon_d/2}S_2^{\varepsilon_d/2}} \times\frac{(1-\rho_{nc})^{k_d+a+b+2}(1-\rho_{gc})^{k_d+k_c+a+b+1}}{a!b!c!d!(1-\rho_{nd})^{k_d+a+b}(1-\rho_{gd})^{1+a+b}},$$

and $\sigma=(1-\rho_{nc})(1-\rho_{gc})/((1-\rho_{nd})(1-\rho_{gd}))$.

For this case, the joint cumulative distribution function (cdf) of z_l, l=1,2, can be evaluated as [13]

$$F_{z_1,z_2}(t_1,t_2)=\int_0^{t_1}\int_0^{t_2}p_{z_1,z_2}(x_1,x_2)dx_1dx_2 \tag{9}$$

Substituting (8) in (9) and following the same procedure and using the same equations as in solving double-fold integral in (6); the integral in (9) is solved in the form of infinite series. Finaly, equating arguments t_1=t_2=t, the derived cdf of output SIR has the following form

$$F(z)=\sum_{a,b,c,d=0}^{+\infty}\mathrm{A}\times z^{\varepsilon_d} \times G_{3,3}^{2,3}\left[\frac{z\sigma}{S_1}\middle|\begin{matrix}1-(\varepsilon_d+\varepsilon_c+\psi_c)/2,1-(\varepsilon_d+\varepsilon_c-\psi_c)/2,(1-\varepsilon_d)/2\\ \psi_d/2,-\psi_d/2,-\varepsilon_d/2\end{matrix}\right] \times G_{3,3}^{2,3}\left[\frac{z\sigma}{S_2}\middle|\begin{matrix}1-(\varepsilon_d+\varepsilon_c+\psi_c)/2,1-(\varepsilon_d+\varepsilon_c-\psi_c)/2,(1-\varepsilon_d)/2\\ \psi_d/2,-\psi_d/2,-\varepsilon_d/2\end{matrix}\right] \tag{10}$$

III. Numerical Results

The outage probability, P_{out}, is a standard measure of the communication system performance and is commonly used to control the noise or cochannel interference level in wireless communication systems.

In the interference-limited environment, outage probability is defined as the probability that combined-SIR falls below a given outage threshold λ, also known as a protection ratio. The outage probability can be evaluated using (10) as

$$P_{out}=\int_0^{\lambda}p(z)dz=F(\lambda) \tag{11}$$

Based on (11), we have analysed outage probability for different condition parameters. Fig. 2 shows the outage probability versus the normalized average SIR, S_1/λ, for various values of shaping parameter k_d. We have observed balanced S_1=S_2 and unbalanced S_2=S_1/2 case at the input branches. It is obvious that when parameter k_d increases, shadowing severity decreases and so the outage probability decreases. Also, it is noticable that the performance gain achieved when k_d increases from 1 to 2 is better compared to the performance gain achieved when k_d goes from 2 to 3. The better performance gain is achieved for balanced SIRs at the input for all randomly picked values of parameters.

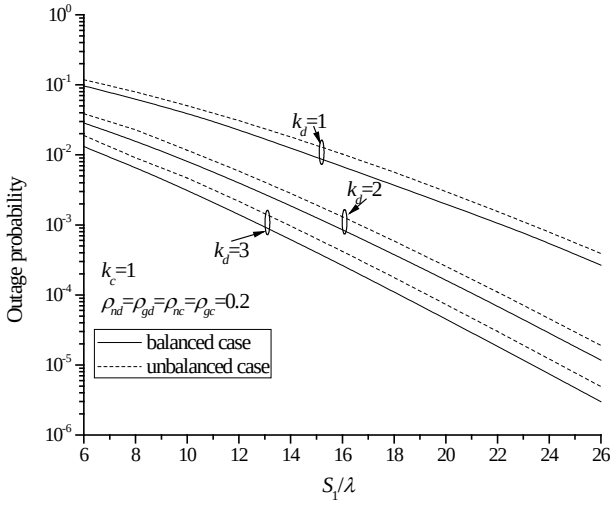


Fig. 2. Outage probability for various values of parameter k_d

The outage probability, as a function of correlation coefficients ρ_{nd}=ρ_{nc} for different values of shaping parameters, k_d and k_c, is shown in Fig. 3. We observed the balanced SIRs at the input S_1=S_2=S and different values for normalized average SIR S/λ. When shaping parameters increases, the shadowing severity decreases and performance is better. When correlation coefficients increase, the outage performance is degraded. The better outage performance is achieved in the case of S/λ=20dB.

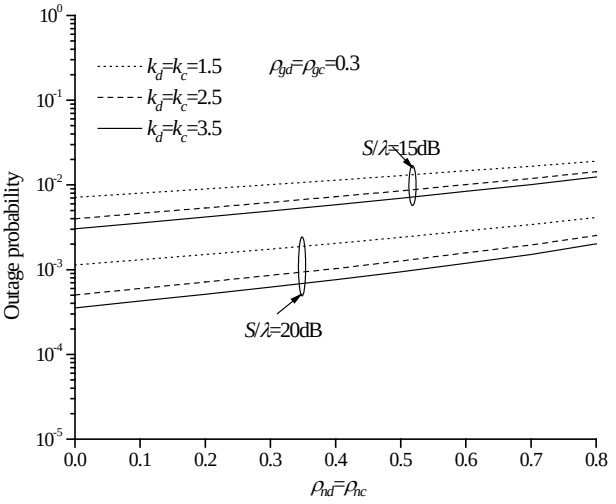


Fig. 3. Outage probability as a function of correlation coefficients

IV. Conclusion

In this paper, the performance analysis of system with SC based on SIR, over correlated composite fading channels, was carried out. The K distribution was observed for evaluating outage performances of proposed diversity system. The new expression for cdf of SIR at the output of SC receiver was derived. Using this new formula, the outage probability was efficiently evaluated. The effects of various parameters, such as the shadowing severity parameters and level of correlation

of desired signal and interference to the system's performance were also examined.

ACKNOWLEDGEMENT

This paper was supported in part by the Ministry of Science of Serbia within the project "Development of new information and communication technologies, based on advanced mathematical methods, with applications in medicine, telecommunications, power systems, protection of national heritage and education" (project III 0440066).

REFERENCES

[1] M. K. Simon and M. S. Alouni, *Digital Communication over Fading Channels.* New York, NY: John Wiley & Sons Inc., 2nd ed., 2004.

[2] A. A. Abu-Dayya and N. C. Beaulieu, "Micro- and macrodiversity NCFSK (DPSK) on shadowed Nakagami-fading channels" *IEEE Trans. Commun.*, vol. 42, no. 9, pp. 2693-2702, Sept. 1994.

[3] M. J. Ho and G. L. Stuber, "Co-channel interference of microcellular systems on shadowed Nakagami fading channels," in *Proc. IEEE Veh. Technol. Conf. (VTC93)*, pp. 568-571, May 1993.

[4] A. Abdi and M. Kaveh, "On the utility of the gamma PDF modeling shadow fading (slow fading)," in *Proc. IEEE 49th VTC*, vol. 3, May 16-20, 1999, pp. 2308-2312.

[5] A. Laourine, M. S. Alouni, S. Affes and A. Stephenne, "On the performance analysis of composite multipath/shadowing channels using the *G*-distribution," in *Proc. IEEE ICC 2008*, pp. 1333-1338, BEIJING, China, May 2008

[6] A. Abdi and M. Kaveh, "*K* distribution: An appropriate substitute for Rayleigh-lognormal distribution in fading-shadowing wireless channels" *Electron. Lett.*, vol. 34, pp. 851-852, 1998.

[7] P. Theofilakos, A. G. Kanatas and G. P. Efthymoglou, "Performance of generalized selection combining receivers in *K* fading channels," *IEEE Commun. Lett.*, vol. 12, no. 11, pp. 816-818, Nov. 2008.

[8] P. S. Bithas, N. C. Sagias, P. T. Mathiopoulos, S. A. Kotsopoulos, and A. M. Maras," On the Correlated *K*-distribution with arbitrary fading parameters,"*IEEE Signal Processing Lett*., vol.15, pp. 541-544, 2008.

[9] T. S. Bhaskar Reddy, R. Subadar and P. R. Sahu, "Outage probability of SC receiver over exponentially correlated *K* fading channels," *IEEE Commun. Lett*., vol. 14, no. 2, pp. 118-120, Feb. 2010.

[10] M. Č. Stefanović, D. M. Milović, A. M. Mitić and 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

[11] A. A. Abu-Dayya and N. C. Beaulieu, "Outage probabilities of diversity cellular systems with cochannel interference in Nakagami fading," *IEEE Trans. on Vehicular Technology*, vol. 41, no. 4, pp. 343-355, Nov. 1992.

[12] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and products*, 5th, Ed. New York: Academic Press, 1994.

[13] C. W. Helstrom, *Probability and Stohastic Processes for Engineer*s, 2nd ed. New York: MakMillian, 1991.

[14] The Wolfram Functions Site, 2008. [Online] Available: http:/functions.wolfram.com

The influence of multiple co-channel interferers on the Selection Diversity System Performance over Weibull fading channels

Ivana Petrovic[1], Stefan Panic[2], Petar Spalevic[3], Sinisa Minic[4], Bosko Radovanovic[5]

[1]***Abstract*** **— Selection combining (SC) diversity receiver experiencing an arbitrary number of multiple, independent, equal power co-channel interferers, in the presence of Rayleigh fading channels was analyzed. Closed form expressions are obtained for the output SIR's probability density function (PDF) and cumulative distribution function (CDF). In order to show the effects of the number of multiple interferers, diversity order and input SIR unbalance to the system performances, an outage probability (OP) analysis is derived. Another important measure of the system's performances, an average bit error probability (ABER) is efficiently evaluated for non-coherent modulation schemes such as binary frequency-shift keying (BFSK) and binary differentially phase-shift keying (BDPSK).**

Keywords **— co-channel interference, correlated fading channels, probability distribution, Weibull fading, selection diversity combining**

I. INTRODUCTION

In cellular communication systems, usually a large number of low-power transmitters broadcast a signal in relatively small geographic areas - cells. Commercial and military cellular systems tend to conserve the available spectrum by reusing allocated frequency channels in areas that are geographically located as close to each other as possible. Unfortunately, due to frequency reuse, signals from two or more channels operating at the same frequency, but from different locations, interfere. Amount of co-channel interference determines limitation in distance for reusing frequency channels.

In general, the power of any interfering signal diminishes with increasing distance between interfering users. A carrier frequency can be reused if the interference level is reduced sufficiently by separation between the co-channel calls.

In order to determine the practical system implementation which satisfies the predetermined minimum performance levels, it is necessary to analyze how the interference as a general distortion affects well-accepted criterions of performance of wireless systems, such as outage probability and average bit error probability [17]. The interference level can be measured through the signal-to-interference power ratio (SIR). The SIR ratio is the primary criteria used in designing frequency reuse plans.

Multipath fading can also seriously degrade performances of wireless communication systems. In cellular radio systems, various techniques for reducing fading effects and influence of co-channel interference are used. Space diversity reception is an effective remedy that exploits the principle of providing the receiver with multiple faded replicas of the same information-bearing signal. When multiple receiver antennas are used space diversity is an efficient method for amelioration system's quality of service (QoS). Two criteria are necessary for obtaining high degree of improvement from a space diversity system. First, the fading between the branches should have low cross-correlation. Second, the mean power available from each branch should be almost equal.

In fading environments as in cellular systems where the level of the co-channel interference is sufficiently high as compared to the thermal noise, SC selects the branch with the highest signal-to-interference ratio (SIR-based selection diversity) [5].

This type of SC in which the branch with the highest SIR is selected, can be measured in real time both in base stations and in mobile stations using specific SIR estimators as well as those for both analog and digital wireless systems (e.g., GSM, IS-54) [6-7]. Most of the recently the published papers assume independent fading between the diversity branches and also between the co-channel interferers.

The effect of co-channel interference on the performance metrics of wireless communication system has been extensively analyzed [8-12]. In [8-9] performance analysis of optimum combining with multiple co-channel interferers over Rayleigh fading channels were presented. In [10] closed-form expressions for outage probabilities of mobile radio channels experiencing multiple co-channel, independent Nakagami-*m* interferers are derived. SIR based analysis of dual branch SC was presented in [11-14], but for the case of single interferer over each channel.

In this paper an analytical study of SIR based selection combining involving assumed Weibull fading channels experiencing an arbitrary number of multiple, independent, equal power Weibull co-channel interferers, will be presented.

[1]Ivana Petrovic is with the High School of Electronics and Telecommunications, Vojvode Stepe nr.283, 11000 Beograd, Serbia, e-mail: petrovicvanja@yahoo.com

[2]Stefan Panic is with the Faculty of Natural Sciences and Mathematics, Lole Ribara nr. 29, Kosovska Mitrovica, Serbia, e-mail: stefanpnc@yahoo.com

[3]Petar Spalevic is with the Faculty of Technical Sciences Str. Kneza Milosa nr.7, Kosovska Mitrovica, Serbia, e-mail: petarspalevic@yahoo.com

[4]Sinisa Minic is with the Faculty of Teachers, Nemanjina bb, 38218 Leposavic, Serbia, e-mail: sinisa.minic@pr.ac.rs

[5]Bosko Radovanovic is with the Faculty of Technical Sciences Str. Kneza Milosa nr.7, Kosovska Mitrovica, Serbia, e-mail: boskoradovanovic@yahoo.com

For proposed system model, closed form expressions for cumulative distribution function (CDF) and probability distribution function (PDF) of the output SIR will be derived. Infinite-series expressions for important performance measures such as the outage probability will be obtained. Outage probability will be shown graphically for different system parameters. Effects of the number of multiple interferers, diversity order and input SIR unbalance to the system performances will be discussed. In designing a cellular mobile system, one may wish to determine optimal values of system parameters in order to achieve reasonable influence of interferers on the outage, so this discussion could have high level of significance. An average error probability will be also efficiently evaluated for several non-coherent modulation schemes such as binary frequency-shift keying (BFSK) and binary differentially phase-shift keying (BDPSK).

II. System Model

Let us assume M_i independent equal power Weibull distributed interferers over *i*-th branch of the SC diversity system with arbitrary number of branches. This assumption is a reasonable one when all interfering signals are at approximately the same distance from the mobile station. The independent instantaneous interfering signals are added together to produce the resultant instantaneous interfering signal at the *i*-th branch of diversity system can be written as:

$$I_i = I_{1i} + I_{2i} + I_{3i} + \cdots I_{M_i} \tag{1}$$

with the total probability density function given by [10] as:

$$p_{y_i}(y_i) = \frac{\alpha_i {M_i}^{M_i} {y_i}^{\alpha_i M_i - 1}}{{M_i}^{M_i} \Omega_{yi}^{M_i} \Gamma(M_i)} \exp\left(-\frac{y_i^{\alpha_i}}{\Omega_{yi}}\right) \tag{2}$$

where Ω_{yi} is the average power of each interferer

The desired signal envelopes on the *i*-th diversity branch also follow the Weibull fading distribution, whose probability density function is given by:

$$p_{x_i}(x_i) = \frac{\alpha_i x_i^{\alpha_i - 1}}{\Omega_{xi}} \exp\left(-\frac{x_i^{\alpha_i}}{\Omega_{xi}}\right) \tag{3}$$

In previous equations parameters α_i denote Weibull-fading parameters ($\alpha_i > 0$) which represent fading intensity measure.

Let $\lambda_i = x_i^2/y_i^2$ be the SIR at the *i*-th (i=1,2…..N) diversity branch of the SC receiver. The joint probability density function of independent random variables λ_1, λ_2… λ_N (since branches are not correlated), can be written as:

$$\begin{aligned} &p_{\lambda_1,\lambda_2\ldots\lambda_N}(\lambda_1,\lambda_2\ldots,\lambda_N) = p_{\lambda_1}(\lambda_1)\, p_{\lambda_2}(\lambda_2)\cdots p_{\lambda_N}(\lambda_N); \\ &p_{\lambda_i}(\lambda_i) = \int_0^\infty y_i\, p_{x_i}(y_i\lambda_i)\, p_{y_i}(y_i)\, dy_i \end{aligned} \tag{4}$$

After substituting (2) and (3) into (4) and some mathematical manipulation pervious expression can be written in the form of:

$$p_{\lambda_i}(\lambda_i) = \frac{1}{2} \frac{\alpha_i \lambda_i^{\frac{\alpha_i}{2}-1} {M_i}^{M_i} {S_i}^{M_i} \Gamma(M_i+1)}{\Gamma(M_i)\left(\lambda_i^{\frac{\alpha_i}{2}} + S_i M_i\right)^{M_i+1}} \tag{5}$$

with $S_i = \dfrac{\Omega_{xi}}{M_i \Omega_{yi}}$ being the average SIR's at the *i*-th input branch of the selection combiner system.

Similarly, joint cumulative distribution function can be written in the form of:

$$\begin{aligned} &F_{\lambda_1,\lambda_2\ldots\lambda_N}(\lambda_1,\lambda_2\ldots,\lambda_N) = F_{\lambda_1}(\lambda_1) F_{\lambda_2}(\lambda_2)\cdots F_{\lambda_N}(\lambda_N); \\ &F_{\lambda_i}(\lambda_i) = \int_0^{\lambda_i} p_{t_i}(t_i)\, dt_i \end{aligned} \tag{6}$$

After substituting (5) into (6) and some mathematical manipulation pervious expression can be written in the form of:

$$F_{\lambda_i}(\lambda_i) = \left(\frac{\lambda_i^{\alpha_i}}{\lambda_i^{\alpha_i} + S_i M_i}\right) \frac{\Gamma(M_i+1)}{\Gamma(M_i)} {}_2F_1\left(1, 2; 1-M_i; \frac{\lambda_i^{\alpha_i}}{\lambda_i^{\alpha_i} + S_i M_i}\right) \tag{7}$$

with ${}_2F_1(u_1,u_2;u_3;x)$, being the Gaussian hyper-geometric function [11].

The selection combiner chooses and outputs the branch with the largest SIR.

$$\lambda = \lambda_{out} = (\lambda_1, \lambda_2 \ldots \lambda_N) \tag{8}$$

The CDF of multibranch SIR-based SC output could be derived from (7) by equating the arguments $\lambda_1 = \lambda_2 = \ldots = \lambda_N$ as:

$$\begin{aligned} &F_\lambda(\lambda) = F_{\lambda_1}(\lambda) F_{\lambda_2}(\lambda)\cdots F_{\lambda_N}(\lambda) = \prod_{i=1}^{N} F_{\lambda_i}(\lambda); \\ &F_\lambda(\lambda) = \prod_{i=1}^{N} \left(\frac{\lambda^{\alpha_i}}{\lambda^{\alpha_i} + S_i M_i}\right) \frac{\Gamma(M_i+1)}{\Gamma(M_i)} {}_2F_1\left(1, 2; 1-M_i; \frac{\lambda^{\alpha_i}}{\lambda^{\alpha_i} + S_i M_i}\right) \end{aligned} \tag{9}$$

The probability density function at the output of the SC can be found as:

$$\begin{aligned} &p_\lambda(\lambda) = \frac{dF_\lambda(\lambda)}{d\lambda} = \sum_{i=1}^{N} p_{\lambda_i}(\lambda) \prod_{\substack{j=1 \\ j\neq i}}^{N} F_{\lambda_j}(\lambda); \quad p_{\lambda_i}(\lambda) = \frac{dF_{\lambda_i}(\lambda)}{d\lambda}; \\ &p_\lambda(\lambda) = \sum_{i=1}^{N} \left(\frac{1}{2} \frac{\alpha_i \lambda^{\frac{\alpha_i}{2}-1} {M_i}^{M_i} {S_i}^{M_i} \Gamma(M_i+1)}{\Gamma(M_i)\left(\lambda^{\frac{\alpha_i}{2}} + S_i M_i\right)^{M_i+1}} \right) \\ &\prod_{\substack{j=1 \\ j\neq i}}^{N} \left(\frac{\lambda^{\alpha_j}}{\lambda^{\alpha_j} + S_j M_j}\right) \frac{\Gamma(M_j+1)}{\Gamma(M_j)} {}_2F_1\left(1, 2; 1-M_j; \frac{\lambda^{\alpha_j}}{\lambda^{\alpha_j} + S_j M_j}\right) \end{aligned} \tag{10}$$

Fig. 1 shows the PDF of output SIR for various values of the number of multiple interferers and fading severity.

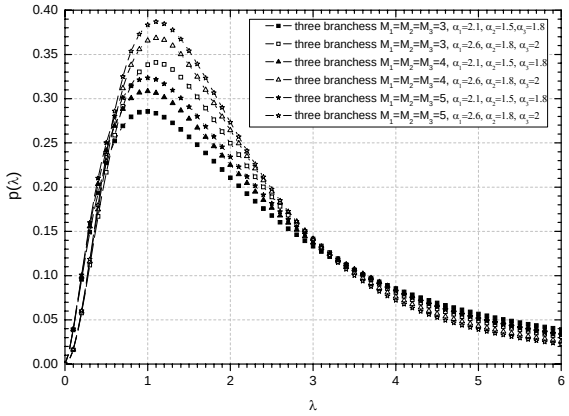


Fig. 1 PDF of output SIR for various values of the number of multiple interferers and diversity branches.

III. OUTAGE PROBABILITY

Outage probability P_{out} is one of the accepted performance measure for diversity systems operating in fading environments. P_{out} is a measure of the system performance, used to control the co-channel interference level, helping the designers of wireless communications system's in order to meet the QoS and grade of service (GoS) demands. This performance measure is very useful in wireless communication systems design especially for the cases when co-channel interference is present.

In the interference limited environment, P_{out} is defined as the probability that the output SIR of the SC falls below a given outage threshold γ_{th} also known as a protection ratio. Protection ratio depends on modulation technique and expected QoS.

$$P_{out} = P_R(\xi < \gamma) = \int_0^{\gamma} p_{\xi}(t)\,dt = F_{\xi}(\gamma) \qquad (12)$$

Outage probability versus normalized parameter S_1/γ for balanced and unbalanced ratio of SIR at the input of the branches and various values of the number of multiple interferers, diversity order is shown on Fig. 2.

From Fig. 2. we can see that for the constant number of co-channel interferers outage probability behavior improves as the diversity order (number of branches) increases. It can be observed that if we want to achieve the same quality of the transmission (for example outage probability of 10^{-4}), we need higher level of average input SIR for dual branch case than for the triple branch case (for example of above 6 dB).

IV. AVERAGE ERROR PROBABILITY

The average error probability at the SC output is derived for non-coherent and binary signaling according to following expressions:

$$P_e = \int_0^{\infty} p_{\zeta}(t)\frac{1}{2}e^{-gt}dt, \qquad (13)$$

Where g denotes modulation constant, i.e., g = 1 for BDPSK and g = 1/2 for BFSK. Substituting (11) in (13) numerically obtained average error probability is shown on Figs. 3 and 4 for balanced and unbalanced ratio of SIR at the input of the branches and various values of diversity order and the number of multiple interferers.

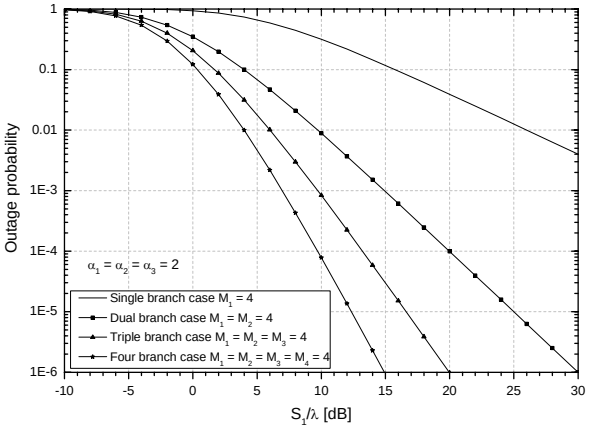


Figure 2. Outage probability versus normalized parameter S_1/γ for various numbers of diversity order

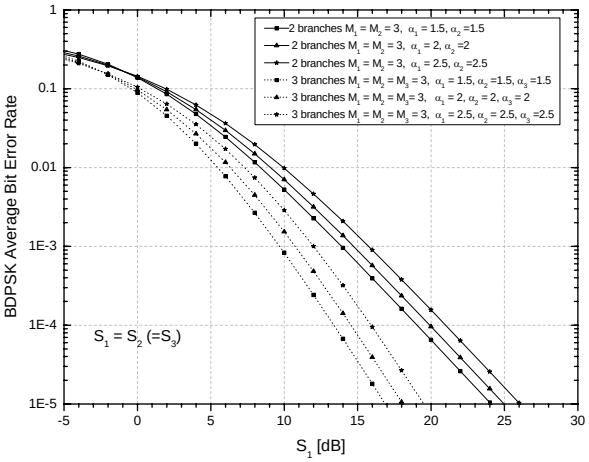


Figure 3. Average BER versus S_1 in non-coherent BDPSK versus SIR at the input of the branches and various numbers of diversity order and fading severity

It is evident how ABER increases at both figures when the number of multiple independent co-channel interferers increases due to growth of interference domination. ABER performance behavior improves for the constant number of co-channel interferers as the diversity order increases. If we want to achieve the same quality of the transmission (for example outage probability of 10^{-4}), we need higher level of average input SIR for dual branch case then for the triple branch case (for example of above 6 dB for BDPSK).

Also comparison of curves shows better performance of BDPSK modulation scheme against BFSK modulation scheme.

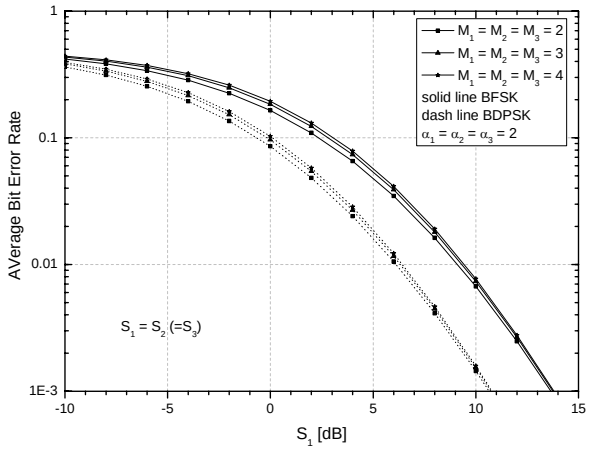


Figure 4. Average BER versus S_1 in BFSK and BDPSK at triple branch diversity for various numbers of multiple interferers

V. CONCLUSION

Multibranch SIR-based SC diversity receiver operating over Weibull fading channels where each channel experiences an arbitrary number of multiple, independent, Weibull co-channel interferers with equal average powers, was analyzed in this paper. Expressions for the first order statistics of the combiner output, namely SIR's probability density function (PDF) and cumulative distribution function (CDF) are presented in the closed form. Standard performance measures, OP and ABER for some non-coherent modulation techniques, are graphically analyzed, in order to show the effects of the number of fading severity, multiple interferers, diversity order. Obtained expressions, analysis and discussions, could find application in designing cellular mobile systems.

ACKNOWLEDGEMENT

This paper was produced under the project TR32023 and III 4406 funded by the Ministry for Science and Environment of the Republic of Serbia.

REFERENCES

[1] G. Tzeremes, C. G. Christodoulou, "Use of Weibull distribution for describing outdoor multipath fading", *Antennas and Propagation Society International Symposium*, vol. 1, pp. 232–235, 2002.

[2] F. Babich, G. Lombardi, "Statistical analysis and characterization of the indoor propagation channel", *IEEE Trans. Commun.*, vol. 48, pp. 455–464, Mar. 2000.

[3] M. K.Simon, M., S. Alouini, *Digital communication over fading channels*, 2nd edn., Wiley, New York , 2005.

[4] P. Stavroulakis, *Interference analysis and reduction for wireless systems*, Artech House, INC, London, 2003.

[5] D. Austin, L. Stuber, "In-service signal quality estimation for TDMA cellular systems", *Sixth IEEE international symposium on personal, indoor and mobile radio communications, PIMRC '95*, Toronto, ON, Canada, pp. 836–40, 1995.

[6] A. Brandao, L. Lopez, C. McLernon, "Co-channel interference estimation for M-ary PSK modulated signals". *IEEE Wireless Pers Commun*, vol. 1, no. 1, pp.23–32, 1994.

[7] N. Sagias and G. Karagiannidis, 'Gaussian class multivariate Weibull distributions: theory and applications in fading channels,' *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3608–3619, Oct. 2005.

[8] G. K. Karagiannidis, D. A. Zogas, N. C. Sagias, S. A. Kotsopoulos, and G. S. Tombras, "Equal-gain and maximal ratio combining over nonidentical Weibull fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 841–846, May 2005.

[9] M. Stefanović et al., "Performance analysis of system with selection combining over correlated Weibull fading channels in the presence of cochannel interference", *Int J Electron Commun (AEU)*, vol. 62 , pp 695-700, 2007.

[10] J. Filho, M. Yacoub, "Simple Precise Approximations to Weibull Sums", *IEEE Comm. Let.*, Vol. 10, No. 8, pp. 614-616, 2006.

[11] I. Gradshteyn, I. Ryzhik, *Tables of Integrals, Series, and products*, Academic Press, New Yourk, 1980.

Increasing the Reliability of Video Information Transmitted over Satellite Radio Channel

Lidia Jordanova[1], Dobri Dobrev[2] and Jordan Nenkov[3]

***Abstract* – The work deals with the results of a research work on the effectiveness of different methods to improve noise immunity of the satellite radio channel when digital TV signals are transmitted. An algorithm to determine the achievable value of the CNR parameter at the satellite receiver input is suggested that takes into consideration both the attenuation of signals, noise and interference in the radio channel and the parameters of the equipment applied.**

***Keywords* – DVB-S receiver, CNR, noise figure of the LNB, antenna gain, E_b/N_0 , BER, RS- and convolutional coding .**

I. INTRODUCTION

The work of the systems of satellite television is based on the active retranslation of signals from satellites, which are launched into geostationary orbit. For transmission of signals to the satellite the frequency band from 14 to 18 GHz is used, while 10,7-12,75 GHz frequency band is used for transmission from satellites to the Earth. The width of the satellite radio channel is most often from 26 MHz to 36 MHz, whereas the polarization of signals in adjacent channels is orthogonal. At present, the majority of satellite TV programs are digital and for shaping of their signals the DVB-S standard has been established.

One of the main problems in satellite TV systems is the high attenuation of signals in a radio channel Satellite - Earth (more then 200 dB) and comparatively high level of noise that adds to the useful signal. This leads to worsening of the Carrier to Noise Ratio (CNR) at the satellite receiver input, and from there to worsening of the quality of received picture and sound. The quality of received signals is measured by the parameter Bit Error Ratio (BER), and in order to achieve a studio quality of image and sound, BER need to be more than 10^{-7} [1-5].

For that reason the MPEG-2 transport stream, which is to be sent to the satellite as QPSK-modulated signal, must firstly be provided with error protection before being fed into the actual modulator. In DVB-S, two error protection mechanisms are used, namely a Reed-Solomon block code which is coupled with convolutional (trellis) coding [6-8].

In the RS-coder the MPEG-2 transport stream packets with a length of 188 bytes are expanded by 16 bytes. The RS (204,188) coding allows correction of up to 8 errors at the receiving end. After performing the convolutional coding, the digital data stream is expanded by a factor which depends on the selected code rate (1/2, 3/4, 2/3, 5/6 and 7/8). If the code rate is 1/2, the data stream is expanded by a factor of 2. The error protection is now at a maximum and the net data rate has dropped to a minimum. A code rate of 7/8 provides only a minimum overhead but also only a minimum error protection. The available net rate is then at a maximum. The good compromise is usually a code rate of 3/4 [9-12].

The aim of this work is to examine the influence of the satellite radio path parameters and the parameters of the receiver equipment used on the reliability of received video information at signal shaping of satellite TV programs under the DVB-S standard.

II. ATTENUATION OF SIGNALS AND NOISE IN SATELLITE RADIO CHANNEL

The attenuation of satellite TV signals is due both to the huge distance between the satellite and the Earth (about 36 000 km) and to other factors. The more important of these factors are: absorption of electromagnetic energy in the atmosphere and in the hydro meteors (rain, snow etc.) and loses due to inaccurate orientation of the receive antenna (receiver directional error) and to the rotation of polarization plane of radio waves (polarization error).

For calculation of signal attenuation along the radio path Satellite - Earth the following expression is used:

$$L = L_0 + \Delta L = 20\lg d + 20\lg f + 92.4 + \Delta L\,, \quad (1)$$

where L_0 is the free space attenuation, d is the Satellite-Earth distance in km, f is the frequency of the received signal in GHz, and ΔL gives loses in atmosphere (0.3 dB) and hydro meteorites (for our climatic zone – under 1 dB), the receiver directional error (0.5 dB) and the polarization error (0.2 dB). At the maximum transmitter-receiver distance (for the Polar Regions) $L_0 \approx 206$ dB, i.e. it grows with about 1 dB compared to the attenuation on the Equator.

The level of the received satellite signals C could be determined (in dBW) under the formula:

$$C = EIRP - L + G\,, \quad (2)$$

where EIRP is the satellite equivalent isotropic radiated power in dBW and G is the antenna gain in dB. For the satellite television systems the values of EIRP parameter are within the limits from 45 to 65 dbW.

The noise in a satellite radio channel is caused by external sources, for example the outer space (the Sun, the Moon, planets and stars), the atmosphere, the Earth etc. and from the

[1]Lidia Jordanova is with the Technical University, Faculty of Telecommunications, Kliment Ochridski 8, 1000 Sofia, Bulgaria, E-mail: jordanova@tu-sofia.bg

[2]Dobri Dobrev is with the Technical University of Sofia, Bulgaria, Faculty of Telecommunications, E-mail: dobrev@tu-sofia.bg

[3]Jordan Nenkov is with the Technical University - Sofia, Bulgaria, Faculty of Telecommunications, E-mail: jordan_n2002@yahoo.com

receiver itself (internal noise). For determination of the noise power at the receiver input the following expression is used:

$$N = 10\lg k + 10\lg B + 10\lg\left(T_0\right) + NF\,. \qquad (3)$$

where k = 1.38.10^{-23} W/Hz.K is the Boltzmann' constant, B is the channel bandwidth in Hz, T_0 is the temperature in ^{0}K (usually it is accepted that T_0 = 293 K), NF is the noise figure of the receiver in dB.

III. Influence of the Parameters of the Receiving Equipment on CNR

Taking into account expressions (2) and (3), the following formula is obtained for calculation of carrier to noise ratio at the receiver input in dB:

$$CNR = EIRP - L + G - 10\lg B - NF + 144\,, \qquad (4)$$

where B is substituted in MHz.

In order to compensate the huge attenuation of signals by the Satellite - Earth radio path, it is necessary to use antennas with comparatively large size (most often parabolic ones with about 40 dB gain factor) and receivers where double down-conversion of signals is applied. At the first down-conversion, witch is made in the low noise block (LNB), the received signals are shifted to the band of the first intermediate frequency (IF) f_{i1} = (950 … 2050) MHz, and are amplified with about 55-60 dB. The second down-conversion provides shifting of the selected program's signal to the second IF f_{i2} = 479,8 MHz, at which over 100 dB gain can be achieved.

The receiver's self-noise can be reduced through realization of its input stages (LNB) with low-noise elements (Ga-As and HEMT transistors). The noise figure of the modern LNBs varies within the limits from 1.0 to 0.6 dB. Penetrating of external noise in the receiver is determined mainly by the receive antenna radiation pattern. To obtain low levels of these noises it is necessary for the receive antenna to have a high directivity factor, i.e. very small beamwidth and low levels of side lobes in its receive antenna radiation pattern.

Generally parabolic antennas are used to receive satellite TV programs. Their gain factor *G* (in dB) is related to the diameter of the reflector *D* (in m), the frequency of received signal *f* (in GHz) and the aperture efficiency η through the following dependence:

$$\begin{aligned} G &= 10\lg\left[\left(2\pi/\lambda^2\right)\cdot\eta\cdot\left(\pi D^2/4\right)\right] = \\ &= 20\lg D + 20\lg f + 20.4 + 10\lg\eta\,. \end{aligned} \qquad (5)$$

The values of the parameter η for a circular parabolic antenna typically runs about 0.55, while values of 0.7 and higher are available for high performance antenna systems.

After taking into consideration the parameters of currently manufactured satellite antennas and low noise blocks and they are substituted in expression (4), the real idea about the possible limits of CNR parameter variations can be obtained. With the values of this parameter calculated in such a way, the bit rate ratio on the output of DVB-S tuner is about 10^{-2}. A significant reduction of BER is achieved after the channel decoding in the DVB-S receiver.

IV. Determining the Parameter E_B/N_0 at the Output of Satellite Receiver

Figure 1 presents the block scheme of a DVB-S receiver, named also DVB-S set-top box (without MPEG-2 decoder). In the satellite front-end the signal undergoes a second down-conversion to a second satellite IF. This down-conversion is performed with the aid of an IQ mixer which is fed by an oscillator controlled by the carrier recovery circuit. The I and Q signals are then A/D converted and supplied to a matched filter in which the same root cosine square filtering process as at the transmitting end takes place with a roll-off factor of 0.35. The baseband square root raised cosine filter has a theoretical function defined by the following expression:

$$H(f) = \begin{cases} 1, & f < f_N(1-\alpha) \\ \left\{\frac{1}{2} + \frac{1}{2}\cdot\sin\left(\frac{\pi}{2f_N}\right)\cdot\left[\frac{f_N - f}{\alpha}\right]\right\}^{1/2}, & f_N(1-\alpha) < f < f_N(1+\alpha) \\ 0, & f > f_N(1+\alpha)\,, \end{cases} \qquad (6)$$

where $f_N = 1/(2T_S)$ is the Nyquist frequency, T_S is the symbol duration and α is the roll-off factor.

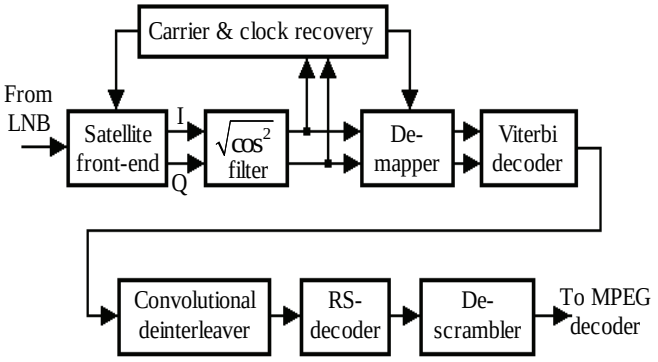


Fig. 1. DVB-S set-top box

The demapper again generates a data stream from which the first errors are removed in the Viterbi decoder. Then, the data stream is passed to convolutional de-interleaving block where any burst errors, i.e. multiple errors in a packet, are broken up into individual errors. The bit errors still present are then corrected in the Reed Solomon decoder. If the error protected TS packet with the length of 204 bytes contains more than 8 errors, the error protection will fail. TS packets flagged as errored must not be used by the MPEG-2 decoder and error concealment must be applied. After the Reed Solomon decoding, the energy dispersal is removed and the MPEG-2 transport stream is supplied to an MPEG-2 decoder.

For determination of signal to noise ratio (SNR) after the matched filter in dB the following expression is used:

$$SNR = CNR + 10\lg\left(1 - \alpha/4\right), \tag{7}$$

where α is the roll-off factor of the root cosine square filter (in DVB-S standard, α = 0.35).

The parameter SNR is related to the energy per information bit E_b referred to the normalized noise power N_0 thought the following dependence:

$$SNR = E_b / N_0 + 10\lg m, \tag{8}$$

where $m = \log_2 M$ is the number of bits-per-symbol and M is the number of constellation points (in this case m = 2).

After the channel decoding, the value of the parameter E_b/N_0 grows proportionally to the rate of the RS-code and the one of the selected convolutional code. The RS-code rate is $R_{RS} = 204/188 = 0.922$, while the convolutional code rate R_C can take one of the following five possible values: 1/2, 3/4, 2/3, 5/6 and 7/8. Considering the growth of the energy per information bit and expressions (6) and (7) are determined, the following formula is obtained for calculation of the parameter E_b/N_0 at the output of satellite receiver (in dB):

$$E_b / N_0 = CNR - 10\lg m + 10\lg\left(1 - \alpha/4\right) - \\ - 10\lg R_{RS} - 10\lg R_C . \tag{9}$$

From expression (8) it follows that as a result of the Reed-Solomon and convolutional coding an increase of E_b/N_0 from 0.93 dB (at $R_C = 7/8$) up to 3.36 dB (at $R_C = 1/2$) can be achieved, which is a prerequisite for substantial decrease of BER of the received digital information.

V. Requirements to CNR Ensuring the Achievement of Pre-set BER

The dependence of the bit error ratio (BER) from the parameter E_b/N_0 is used most often for evaluation of the reliability of the received digital information. When QPSK method is used for transmission of the signals on the radio channel, it can be described with the expression:

$$BER = Q\left[\sqrt{2 \cdot 10^{0.1(E_b/N_0)}}\right], \tag{10}$$

where $Q(x)$ is the Gauss error integral, and the value of E_b/N_0 is substituted in dB. The following approximate formula is worked out for determining of $Q(x)$:

$$Q(x) \approx \left(x\sqrt{2\pi}\right)^{-1} \cdot \exp\left(-0.5 \cdot x^2\right). \tag{11}$$

Expression (10) is used for evaluation of the reliability of received video information when channel coding is not applied in the satellite receiver. The decrease of the probability of error as a result of Reed-Solomon and convolutional coding can be determined when the argument of the function $Q(x)$ is substituted by x^*. The new value of the argument is calculated with the formula:

$$x^* = \left[\frac{2 \cdot 10^{0.1(E_b/N_0)}}{R_{RS} \cdot R_C}\right]^{1/2}, \tag{12}$$

where E_b/N_0 is in dB.

After (11) is substituted in expression (10), the following dependence is obtained for determining the Gauss error integral which takes into consideration the profit $\Delta Q(x)$ of using the Reed-Solomon and convolutional coding:

$$Q(x^*) = Q(x) \cdot \exp\left\{-0.5x^2\left[\left(R_{RS} \cdot R_C\right)^{-1} - 1\right]\right\} \text{ x} \\ \text{x } \left(R_{RS} \cdot R_C\right)^{1/2} = Q(x) \cdot \Delta Q(x) . \tag{13}$$

Figure 2 shows the dependences of BER from E_b/N_0 obtained at convolutional code rate of 1/2, 2/3, 3/4, 5/6, 7/8 and 1. From them, it is easy to determine the necessary values of the parameter E_b/N_0 at which the bit error ratio does not exceed 10^{-7}. The dependence referring to $R_C = 1$ takes into consideration the effect of application of RS coding only in the receiver. Having in mind that in absence of a channel coding the pre-set maximum value of BER (10^{-7}) is ensured at $E_b/N_0 = 11.3$ dB, a conclusion can be made that the DVB-S receiver requires with about 1 to 3.4 dB less values of the parameter E_b/N_0.

In the course of designing the satellite radio channel, the

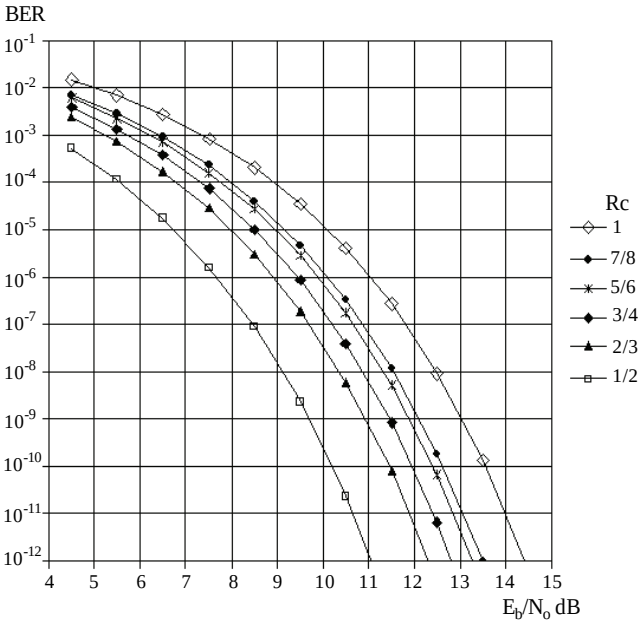


Fig. 2. BER in DVB-S as a function of E_b/N_0

parameter CNR is used, which is related to E_b/N_0 through dependence (8). By taking into account the regulated in the DVB-S standard values of the roll off factor α of the digital filter (α = 0.35) used and the rate of RS (204,188) code, the following formula for calculation of CNR is worked out:

$$CNR = E_b / N_0 + 2.26 + 10\lg R_C . \tag{13}$$

After taking into consideration the values of parameter E_b/N_0, which are required for achieving BER = 10^{-7} at R_C = 1/2, 3/4, 2/3, 5/6 and 7/8, it is found out that the actual CNR values can be expected between 7 … 12 dB. For example, if it is accepted that R_C = 3/4 (in this case BER = 10^{-7} is achieved at E_b/N_0 = 9.7 dB), the carrier to noise ratio at the satellite receiver input needs to be 10.7 dB.

By known values of CNR, satellite equivalent isotropic radiated power (EIRP), attenuation of the received signal (L) and channel bandwidth (B) it is easy to determine the diameter of the receive antenna and to select appropriate low noise block for the specific case, whereas using for that purpose expressions (4) and (5). For example, at EIRP = 52 dBW, L = 207 dB (this value is obtained taking into consideration the geographic location of Bulgaria), B = 27 MHz and selection of LNB with noise factor 1 dB, the gain of the receive antenna must be 37 dB. With the individual satellite receiving, where usually offset antennas are used, such antenna gain can be achieved with antenna diameter of 60 cm.

VI. Conclusion

The results presented in this work are used to develop algorithms for design of digital satellite TV channels of individual and of CATV systems. They allow taking into consideration the parameters of the satellite radio path and at the same time of the manufactured equipment, and they are the base for further investigations related to improvement of the quality of received digital satellite TV programs.

Acknowledgement

The research described it this paper is supported by the Bulgarian National Science Fund under the contract No ДДВУ 02/74/2010.

References

[1] J. Louis, Jr. Ippolito, *Satellite Communications Systems Engineering*, John Wiley & Sons, Inc., New York, 2008.
[2] D. Roddy, *Satellite Communications*, McGraw-Hill, New York, 2001.
[3] L.J. Ippolito, *Propagation Effects Handbook for Satellite Systems Design*, NASA Reference Publication 1082(5), 1999.
[4] ITU-R Rec. P.841-4, *Conversion of annual statistics to worst-month statistics*, Geneva, 2005.
[5] T. Pratt, C. Bostian and J. Allnutt, *Satellite Communications*, John Wiley & Son, Inc., New York, 2003.
[6] B. Sklar, *Digital Communications: Fundamentals and Applications*, Prentice Hall, New Jersey, 2001.
[7] W. Fisher, *Digital Video and Audio Broadcasting Technology*, Springer, Heidelberg, 2008.
[8] G. Corazza, *Digital Satellite Communications*, Springer, New York/ Heidelberg 2007.
[9] A. Burr, Modulation and Coding for Wireless Communications, Prentice Hall, New Jersey, 2001.
[10] R. Morelos-Zaragoza, *The Art of Error Correcting Coding*, JohnWiley & Sons, New York, 2002.
[11] L. Charles, *Error-Control Block Codes*, Artech House, Inc., Norwood, 2000.
[12] H. Lee, *Convolutional Coding: Fundamentals and Applications*, Artech House, Inc., Norwood, 2003.
[13] ETR 102 154, Digital Video broadcasting: *Implementation guidelines for the use of MPEG-2 Systems, Video and audio in satellite, cable and terrestrial contribution broadcasting application.*
[14] ETS 300 421, Digital broadcasting systems for television, sound and data services: *Framing structure, channel coding and modulation for 11/12 GHz satellite services*.

Throughput Maximization in Wireless Fading Channel based on Markov Decision Process

Zoran Veličković[1], Milojko Jevtović[2] and Vlastimir Pavlović[3]

***Abstract* – To guarantee the quality of service (QoS) in wireless communication systems is a very challenging task which can be solved through joint optimization of network parameters on many ISO/OSI layers. In this paper the optimizing cross-layer algorithm is presented which minimizes long-term average packet dropping, average packet delay and average transmission power of single user wireless channel. The optimization problem is defined as Markov decision process (MDP) and the composite state space of the communication system with transition matrices of probabilities is determined. For solution of MDP problem the cost function is formed whose minimization allows for getting the optimal transmission policy for given criteria. The structure of the obtained optimal policies for Rayleigh and independent and identically distributed (i.i.d.) fading channel is taken into consideration. Simulation results confirm that the proposed cross-layer algorithm provides adaptation of network parameters and makes it possible for satisfying QoS in a wide range according to application demands. Through simulations it has been confirmed that the obtained solutions are Pareto-optimal, so that through adaptation of one QoS parameter it can be influenced on all the others. Application of MDP in multicriteria optimization problems has proved to be justified.**

***Keywords* – Adaptive cross-layer adaptation, Markov Decision Process, Optimal transmission policy, Quality of Services (QoS).**

I. INTRODUCTION

Extraordinary interest of users for the access to wireless communication systems can be justified by attractive multimedia services they support. The support of wireless multimedia services implies provision of a great network throughput as well as certain level of QoS (Quality of Services). Independently of networks they are applied in, ITU-T has standardized key parameters which influence on QoS of multimedia applications [1]. Without taking technical aspects into consideration, in specification G.1010, ITU-T has standardized QoS parameters. These parameters are classified into eight categories according to the type of application and include voice, video, image and text transfer. Thus, delay up to 150 *ms* and delay variation up to 1 *ms* at conversational speech can be tolerated with maximum 3% packet los rate. On the other hand, Web-browsing can tolerate a delay up to *4 s* and a variation of the delay is not specified with information loss zero. It is evident that different applications require a wide range of QoS, from intolerant of delay to those more tolerant of packet loss. In wireless communication systems, along with the standard assembly of QoS, demand for minimization of transmit power is very important. In this way the efficient using of battery supplying the mobile users is secured. Considering the limited radio recourses, dynamic network structure, time changeable characteristics of a communication medium, appearance of fading, as well as dynamic network traffic, meeting QoS in wireless communication systems is not a simple task. For this reason in some wireless networks the *soft QoS* is defined or the *best effort* service where time period when QoS is not satisfied can be tolerated. On the other hand, the *hard QoS* is defined which is sometimes hard to satisfy even in the wired IP networks. Reason for this lies in the standard ISO/OSI stack of protocols which is not projected to guarantee QoS. In order to guarantee QoS and multimedia applications in wireless computer networks the standard 802.11e has been developed [2]. Under this standard, depending on the type of application is given priority before the packets delivery to MAC layer. In this way, better system performance can be got in low traffic, but the higher traffic they will degrade.

Optimization of network resources in wireless communication systems demands consideration of many networks ISO/OSI layers, in order to satisfy QoS defined by the application. Joint optimization of network parameters of many ISO/OSI layers is called *Cross-Layer* (CL) design [3]. Three basic architectures are usually applied for an undisturbed access to network parameters: the architecture with the direct communication between layers, the architecture with divided data base, and the modular architecture of CL design. CL design of stack of communication protocols in the wireless environment is a method which could improve satisfying QoS. Optimal CL protocols which fulfill the intended QoS can be realized by joint adaptation of parameters through many ISO/OSI layers. This actually means that network parameters which relate to all aspects of wireless communications can be adapted by CL design. In this way at the same time parameters of PHY layer (transmission power, modulation type, bit error rate - BER), MAC layer (access scheme, buffer size of transmitter) and APP layer (source coding, permitted delay) can be at the same time optimized. Generally, all layers from the stack of protocols can be included into CL design, but this is often not practical. Increase of the number of optimizing network parameters demands provision of considerable processing resources which complicates realization of these protocols. Optimization CL algorithms which consider network resources with PHY, MAC and APP layers are most often realized.

[1]Zoran S. Veličković is with the High Technical School, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.velickovic@vtsnis.edu.rs.

[2]Milojko Jevtović is with the Serbian Engineer Academy, L.Ribara 7, 11000 Beograd, Serbia, E-mail: vladmijev@ptt.rs

[3]Vlastimir D. Pavlović is with Faculty of Electronic Engineering Niš, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: vlastimir.pavlovic@elfak.ni ac.rs

CL algorithm which makes adaptation of network parameters in function of varying the quality of communication channel (fading level) [4], total capacity of the transmitting buffer and the type of network traffic in the network node is taken into consideration in this paper. A mathematical model based on Markov Decision Process (MDP) which optimizes network parameters is applied in order to satisfy QoS [5]. The communication system is modeled by the composite discrete Markov process which describes behaving of the wireless communication system in various working conditions. The system modeled in this way can be solved numerically if dynamic programming is applied [6], wherewith optimal transmission policy for the set objective function is determined. Various objective functions can as a solution have various optimal transmition policies. The aim of this paper is maximization of average long-term network throughput along with limitation of the average transmission power, constant BER satisfying, minimization of the number of rejected packets and minimization of packets delay. The composite model of the wireless single-user communication system which consists of Markov chains PHY and MAC-LLC layers is developed in Section II. The optimization problem and its solution are presented in Section III. The obtained results are presented in Section IV. Certain conclusions are given and directions for further investigations are suggested in Section V.

II. Model Of Wireless Communication System

A. Dynamics and Throughput

Considered wireless communication system consists of a transmitter, which incorporates a buffer of the limited capacity and encoder-modulator and of a receiver. Loading of the buffer is done with packets from the higher ISO/OSI layers and is modeled by Poisson distribution. Transmission rate out of the buffer is adaptive, and it sustains the realization of optimization CL algorithms. The connection between the transmitter and the receiver is realized in Rayleigh fading channel, while information about the quality of the channel and the buffer state are being interchanged through the control channel. The analysis of the work of the presented system has been considered through a series of successive time frames whose duration is T_f. Frame i comprises the time period of $[i \cdot T_f, (i+1) \cdot T_f]$ for which the number of incoming packets into the buffer A_i is specified. Only after the expiration of the time frame i to the existing state of the buffer, the number of the arrived packets A_i can be added. The mean value of the number of packets coming into the buffer while the frame lasts, λ, is calculated as $\lambda = E\{A_i\}$, where $E\{\bullet\}$ is the operator of the mathematical expectation. It has been supposed that all the packets are of equal length and that the buffer is of limited capacity and that it can receive only B packets. Each the packet that arrives at the moment when the buffer is completely loaded will be rejected and regarded as lost. However, in the case when the buffer is not completely loaded, it will be complemented with the arrived packets until the full capacity has been reached (B), and all the rest packets are rejected and considered lost. Another source of the rejected packets in the considered system is the wireless communication channel with fading. Each error created during packets transmission will have as a consequence rejecting of that packet and it will be regarded as lost. The average number of packets η which will be delivered to the application on a higher layer of the receiver can be determined in the following way:

$$\eta = \left(\lambda T_f\right) \cdot \left(1 - P_o\right) \cdot \left(1 - P_p\right), \tag{1}$$

where P_o is the buffer overflow and P_p is the packet error probability. The expression (1) binds two independent subsystems of the wireless communication system on two ISO/OSI layers: the subsystem of the transferring medium (PHY layer) and the subsystem of the buffer in the network node (MAC-LLC layer).

B. Composite Markov Model of the Communication System

The state space of the communication system (S) is composite, determined by states of the communication channel g and the buffer in the communication node $\mathcal{B}$:

$$S = g \times \mathcal{B} \tag{2}$$

where $\{\times\}$ represents an operator of Cartesian product. On the base of (2), the state space of Markov chain of the communication system is determined in the expression:

$$S = \left\{ s_1, s_2, ..., s_Q \right\}, \tag{3}$$

where

$$Q = K \cdot \left(B + 1\right). \tag{4}$$

Therefore, state of the communication system for frame i, S_i, is determined by two components: state of the buffer B_i and state of the communication channel G_i:

$$S_i = \left(B_i, G_i\right). \tag{5}$$

The packet number for empting the buffer when the communication system is in the state S_i is marked with U_i. CL optimization algorithm described in the following section implies that both transmitting and receiving sides have information about the state of the communication system for every frame i. To complete determination of Markov chain it is necessary to determine transition probability for the whole communication system. It is important to mention here that there is not only one probability for transition from the system state S_i into the system state S_{i+1}, but that the number of possible transitions determined by adaptive possibilities of unloading the buffer in the communication node. Therefore, set of actions A_s, which can be performed for every state of the system is determined by adaptive possibilities of coding and choice of modulation techniques in the communication system. For the observed communication system the set of

actions represents the number of packets through which the buffer in the communication node can be emptied during one frame. By fixing the symbol rate and adaptation of parameters of signal constellation in MQAM modulator the rate of emptying the communication buffer can be adapted. Transition probabilities of the communication system p_S in realization of the action u can be determined in the following way:

$$p_S(s,s') = P_r\{s_{i+1} = s' | s_i = s, U_i = u\} \quad (6)$$

$$= p_G(g,g') \cdot p_B(b,b',u), \quad (7)$$

where p_G and p_B are transition probabilities of the subsystem which describe the communication channel, that is, state of the communication buffer, respectively. The quadrate matrix of transition probabilities for the whole communication system for one value of the parameter u is given in (8).

$$P^{(S)} = \begin{bmatrix} p_{0,0}^{(S)} & p_{0,1}^{(S)} & \cdots & \cdots & \cdots & p_{0,Q}^{(S)} \\ p_{1,0}^{(S)} & p_{1,1}^{(S)} & \cdots & \cdots & \cdots & p_{1,Q}^{(S)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{Q-1,0}^{(S)} & p_{Q-1,1}^{(S)} & \cdots & \cdots & \cdots & p_{Q-1,Q}^{(S)} \\ p_{Q,0}^{(S)} & p_{Q,1}^{(S)} & \cdots & \cdots & \cdots & p_{Q,Q}^{(S)} \end{bmatrix} \quad (8)$$

III. Optimization Problem

If P_i marks the necessary transmit power of the i-th frame, it is extremely important for wireless communication systems to limit the maximum of the long-term average transmission power:

$$\lim_{T\to\infty} \sup \frac{1}{T} \mathrm{E}\left\{\sum_{i=0}^{T-1} P_i\right\} \le \overline{P}. \quad (9)$$

On the other hand, the important parameter in multimedia applications is the maximum of the permitted delay of packets. Delay of packets is defined as the buffer occupancy via Little's theorem:

$$D_i = \frac{B_i}{\lambda(1-P_p)} \quad (10)$$

where B_i is the number of packets in the buffer during the frame i. The maximum long-term average delay of packets can be defined in the following way:

$$\lim_{T\to\infty} \sup \frac{1}{T} \mathrm{E}\left\{\sum_{i=0}^{T-1} D_i\right\}. \quad (11)$$

One of the goals of this paper is minimization of the long-term average packets delay. An algorithm which is going to be presented in the next section will minimize the mean time of packets delay simultaneously with satisfying the demanded BER and limiting the average transmission power. In this paper, protocols which keep BER constant (P_b = const) will be considered, so that it will result in constant PER. Increasing of network throughput for the case of constant PER can be realized through minimization of the number of rejected packets produced by the buffer overflow in the communication node. The average number of rejected packets depends on the capacity, dynamics and fulfillment of the buffer:

$$L_o(b,u) = \mathrm{E}\{\max(0, A+b-u-B)\}, \quad (12)$$

where B is the capacity of the buffer in packets, b the number of packets in the buffer which is unloading with u packets and loading with A packets per frame. The expectation is with respect to the number of packets arriving in frame. The maximum long-term average number of rejected packets can be presented in the following way:

$$\lim_{T\to\infty} \sup \frac{1}{T} \mathrm{E}\left\{\sum_{i=0}^{T-1} L_o(B_i, U_i)\right\}. \quad (13)$$

Through previous expressions (9), (11) and (13) conflict network demands are defined regarding limiting long-term average transmit power, delay and the number of rejected packets. Namely, minimization of the average transmit power will influence on the average delay and average dropping packet. Therefore, optimization of one parameter is nor possible without influence on the other. Finally, the optimization problem can be defined as minimization problem in the following way:

Minimize the maximum of long-term average rejected packets L_o and average packet delay $\overline{D}$ of the wireless communication system subject to an average transmission power constraint $\overline{P}$.

We are interested in three objectives, minimizing the average packet delay, average packet loss over the average transmission power. All of these criteria can not be minimized at the same time. One of the solutions of the multi-objective minimization problem is, instead of the original problem, to minimize the weight function of all of three criteria. In this way the minimization problem is defined as the long-term average cost MDP.

$$\lim_{T\to\infty} \sup \frac{1}{T} \mathrm{E}\left\{\sum_{i=0}^{T-1} C_I(B_i, G_i, U_i)\right\}, \quad (14)$$

where C_I is immediate cost incurred in state (b, g) when control action (u, P) is taken, i.e.:

$$C_I(b,g,u) = P(u,g,P_b) + \beta_1 \cdot L_o(b,u) + \beta_2 \cdot D(b). \quad (15)$$

The weight factors β_1 and β_2 are positive numbers with the role of Lagrangian multipliers and indicate the relative importance of average packet lost and average buffer delay over the average power. The dynamic programming is recursion procedure for MDP solution by *value function*.

IV. Simulation Results

The results are obtained for the following simulation parameters. On APP layer packets are generated which arrive in buffer according to a Poisson distribution with λ=3000 packets/s. On MAC-LLC layer it is supposed that all the packets consist of $L=100$ bits and that the capacity of the transmitting buffer is limited to $B=15$ packets. On the physical layer the following parameters are supposed: the channel bandwidth is $W=100$ kHz, noise power density $N_o=2\cdot10^{-5}$ W/Hz and bit error rate is fixed on $P_b=10^{-6}$. MQAM modulator is used in the transmitter where through varying the constellation of the signal M adaptation of the transmission rate can be realized. Duration of the symbol is fixed to $Ts=1/W$ and the frame duration T_f is determined with 100 symbols. With the supposed parameters the buffer is to be emptied with u packets/frame when signal constellation is $M = 2^u$. Adaptation of transmission rate U_i is performed for every frame i.

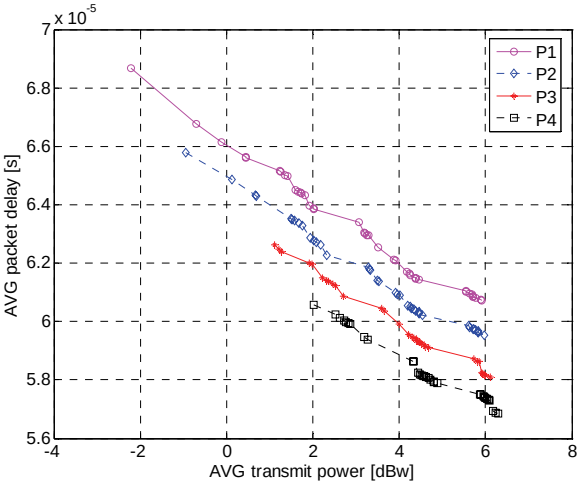


Fig. 1. Pareto-optimal AVG long-term packet delay as function of long-term AVG transmit power

Fig. 1 and Fig. 2 were obtained by variation of parameters β_1 and β_2 which is determined by the optimal policy. There are shown results for different optimal transmission policy labeled P1 to P4. Graph relating to policy P1 is obtained for constant values of β_1, while the parameter β_2 varied in the range $100\cdot10^4>\beta_2>1\cdot10^4$. Parameter β_1 corresponding to the considered policy is: $\beta_1 = [1\cdot10^4\ 20\cdot10^4\ 50\cdot10^4\ 100\cdot10^4]$. As a solution to the MDP, we get the vector U_i, by which it determines the number of packets dropped and the L_o level transmitting power P. Averaging obtained values of the transmitting power, number of dropped packets and packet delay for all frames are Pareto-optimal values. With the graphics shown in Fig. 1 can be concluded that the average packet delay increases with reduction of the AVG long-run transmit power.

On the graphs in Fig. 2, the results of long-term AVG number of rejected packets as a function of long-term AVG time delay of packet are shown. It is clear that there is great number of the optimal policy, and which of them will be chosen depends on the desired QoS parameters.

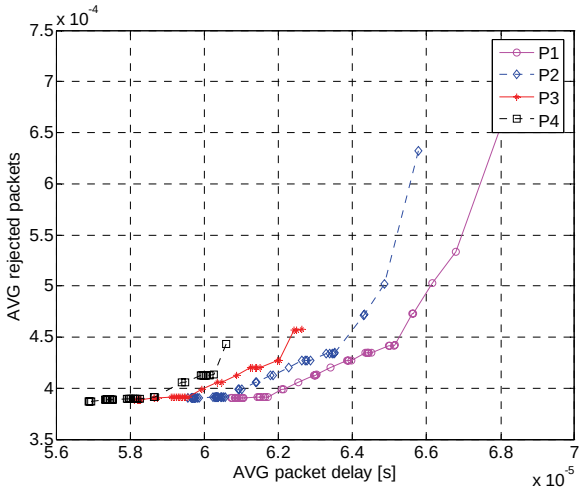


Fig. 2. Pareto-optimal AVG long-term rejected packets as function of long-term AVG packet delay

V. Conclusion

In this paper it was shown that simultaneous optimization of network parameters with many ISO/OSI layers allows for satisfying the set QoS demands for multimedia wireless applications. If wireless multimedia application set the demanded QoS in the form of MDP problem, it is possible to find effectively an optimal transmission policy. The application type, quality of the communication channel and resources of the network node determine the choice of the optimal transmission policy. The wireless communication channel with Rayleigh is considered in this paper. The channel with i.i.d. fading demands considerably less transmission power in relation to the wireless communication channel with Rayleigh fading. The cross-layer design provides joint optimization of the application layer with other layers from the protocol stack, which improves performances of the wireless communication system and provides support for the wide set of multimedia applications.

References

[1] ITU-T Recommendation G. 1010: End-user multimedia QoS categories – ITU-T, 2001.

[2] IEEE 802.11e WG. 2005. Wireless LAN MAC and PHY specifications amendment 8: MAC quality of service enhancements, http://standards.ieee.org.

[3] V. Srivastana, M. Motani, "Cross-layer design: a survey and the road ahead", IEEE Commun. Mag., vol. 43, no. 12, pp. 112–119, Dec. 2005.

[4] Z. Veličković, M. Jevtović, "Adaptive Cross-Layer Optimization in Wireless Fading Channel and Limited Buffer Capacity," ICEST Proceedings vol. 1, pp. 61–64, 2010.

[5] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, New York, NY: John Wiley & Sons, 1994.

[6] Z. Veličković, M. Jevtović, "Adaptive Cross-Layer Optimization Based on Markov Decision Process", Journal Electronics and Electrical Engineering, No. 2 (108), pp. 39–42, Feb. 2011.

Toward Adaptive Initialization of New Tracks in MTT Systems

Nemanja Mitrović[1] and Zeljko Djurović[2]

***Abstract* – Environment topology in multiple target tracking (MTT) systems has important influence on initialization of new tracks. This paper presents functional analysis of MTT system block responsible for initialization of new sources. Having in mind relief topology as a static radar environment, the aim of presented analysis is to reduce false targets tracks with simultaneous preserving of real tracks confirmation. This goal is achieved by estimation of new source density and adjusting the initialization procedure based on this estimate. The efficiency of proposed procedure is demonstrated through computer simulations. The obtained results may be used to improve the efficiency of radar or sonar systems or any other sensors network that performs multiple target tracking.**

***Keywords* – Multi Target Tracking (MTT), Radar System, Digital Signal Processing.**

I. Introduction

Multi target tracking (MTT) is essential requirement for surveillance systems employing one or more sensors together with computer subsystem to interpret the environment [1]. Sensors systems are radars, sonar, infrared thermal cameras, and etc [2]. Mainly, MTT has two tasks - to interpret the targets of interest and to interpret the background noise sources such as radar ground clutter, internal error sources (thermal noise), environment topology, and etc. Environment topology in MTT systems has important influence on initialization of new tracks. This paper presents functional analysis of MTT system block responsible for initialization of new sources. Having in mind relief topology as a static radar environment, the aim of presented analysis is to reduce false targets tracks with simultaneous preserving of real tracks confirmation.

Basically, sequential MTT consists of five components: sensor data processing and measurement formation, data association (correlation), track initialization, confirmation, and deletion, filtering and prediction, and gating [3]. Connections between components are very strong so there is no clear distinction between the boundaries. Sequential MTT as any other realization type of MTT require a lot of calculations and memory space [7]. Next section gives short description of each specific MTT component.

Reducing false targets tracks is achieved by estimation of new source density and adjusting the initialization procedure based on this estimate. The efficiency of proposed procedure is demonstrated through computer simulations. The obtained results may be used to improve the efficiency of radar or sonar systems or any other sensors network that performs multiple target tracking.

[1]Nemanja Mitrović is with the Institute for Microwave Technique and Electronic - IMTEL Komunikacije A D, Boulevard Mihajla Pupina 165b, 11070 Belgrade, Serbia,
E-mail: nemanja@insimtel.com.

[2]Zeljko Djurović is with the School of Electrical Engineering, Boulevard of King Alexander 73, 11000 Belgrade, Serbia, E-mail: zdjurovic@etf.bg.ac.rs.

II. MTT Architecture (Structure)

Basic MTT architecture could be distributed to five functional blocks presented on Fig. 1.:

1) Sensor data processing and measurement formation
2) Data association (correlation)
3) Track initialization, confirmation, and deletion
4) Filtering and prediction
5) Gating.

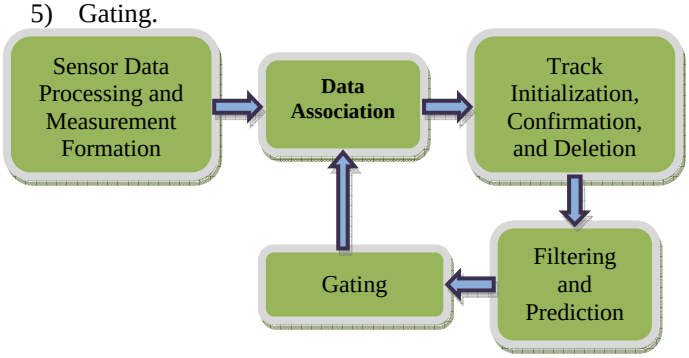


Fig. 1. Sequential MTT block diagram

Once confirmed observation initiated new track. MTT predicts new positions of observations for all active tracks and create gates around new positions. Correlation process performs association between new observation and existing track according to norm *d* as:

$$d = \left| y_i - \overline{y_i} \right|, \tag{1}$$

where y_i measured observation position at i^{th} sample time and $\overline{y_i}$ predicted observation position at i^{th} sample time. For 2D case relation (1) could be defined as:

$$d^2 = \frac{(R_0 - R_p)^2}{\sigma_R^2} - \frac{(\theta_0 - \theta_p)^2}{\sigma_\theta^2}, \tag{2}$$

where R_0 and Rp are existing and predicted target radial components, θ_0 and θp are existing and predicted target azimuths, and σ_R and σ_θ are residual variance of target radial component and azimuth. Also, it is useful to define residual standard deviation σ_r as:

$$\sigma_r = \sqrt{\sigma_0^2 + \sigma_p^2}, \tag{3}$$

where σ_0 and σ_p are standard deviation of existing target position and prediction error.

A. Data Association (Correlation)

Most sensitive part of MTT is data association – correlation [8]. The correlation function takes the inputs from of the gating function and makes final observation-to-track data assignment. Basically, data association problem has two approaches: all neighbors (AN) and nearest neighbor (NN). For each approach there are many techniques dedicated to improve performance data association. It is very difficult to obtain optimal solution for this problem, so most of techniques are suboptimal. Here, is presented one technique of NN approach, in [1] called suboptimal solution one.

Data association process starts with formation of assignment matrix. Assignment matrix fields are norms *d* of observations to centers of predicted gates of existing tracks, according to the relation (1). This way, one dimension of assignment matrix is equal to number of observation and other to number of tracks. Next step in data association is association matrix search and association by six rules.

1) An observation that validates with a singly validated track is rejected by multiply validated track.
2) A multiply validated observation is rejected by any track that validates with singly validated observation.
3) Whether or not a track is multiply validated is determined again after each application of Rule 1 affecting it. A track that becomes singly validated is again subject to Rule 1.
4) Whether or not an observation is multiply validated is determined again after each application of Rule 2 affecting it. An observation that becomes singly validated is again subject to Rule 2.
5) For each remaining multiply validated track, choose the observation with minimum distance.
6) For each remaining multiply validated observation, choose the track with minimum distance.

B. Track Initialization, Confirmation, and Deletion

The second step in MTT processing is track initialization, confirmation and deletion. One of many techniques that obtain it is sequential probability ratio test (SPRT). SPRT propose two hypotheses:

H_0 – no true target is present, so return is from false alarms or clutter.

H_1 – a true target is present.

SPRT accept one of three possible choices: (1) accept H_0 (2) accept H_1 (3) defer decision for some next sampling period. The likelihood functions of H_0 and H_1 after k subsequent scans with particular sequence of m detections are:

$$\begin{aligned} P_{0,k} &= P_F^m \cdot \left(1-P_F\right)^{k-m}, \\ P_{1,k} &= P_D^m \cdot \left(1-P_D\right)^{k-m}, \end{aligned} \tag{4}$$

where P_F and P_D are detection probabilities of false and true targets. Now, likelihood ratio U_k is defined as:

$$U_k = \frac{P_{1,k}}{P_{0,k}}, \tag{5}$$

SPRT test U_k likelihood ratio respect to C_1 и C_2 thresholds:

$$C_1 = \frac{1-\beta}{\alpha}, C_2 = \frac{\beta}{1-\alpha}, \tag{6}$$

where α is probability of accepting H_1 when H_0 is true and β is probability of accepting H_0 when H_1 is true. Formulation of SPRT is:

1) If $U_k \le C_1$ accept H_0.
2) If $U_k \ge C_2$ accept H_1.
3) If $C_1 \le U_k \le C_2$ continue testing.

C. Filtering and Prediction

Idea of reducing false targets tracks with simultaneous preserving of real tracks confirmation is achieved by estimation of new source density β_{NS} and adjusting the initialization procedure based on this estimate. This improvement of MTT is embedded in *α-β-γ* tracker. Standard target parameters that are subject of *α-β-γ* tracker estimation are position, speed, and acceleration are extended with estimation of new source density. Equations that describe extended α-β-γ tracker are:

$$\begin{aligned} x_s(k) &= x_p(k) + \alpha \cdot \left(x_0(k) - x_p(k)\right), \\ v_s(k) &= v_s(k-1) + T \cdot a_s(k-1) + \frac{\beta}{qT} \cdot \left(x_0(k) - x_p(k)\right), \\ a_s(k) &= a_s(k-1) + \frac{\gamma}{(qT)^2} \cdot \left(x_0(k) - x_p(k)\right), \\ x_p(k+1) &= x_s(k) + T \cdot v_s(k) + \frac{T^2}{2} \cdot a_s(k), \\ \beta_{NS,s}(k) &= \beta_{NS,p}(k) + \delta \cdot \left(\beta_{NS,0}(k) - \beta_{NS,p}(k)\right), \\ \beta_{NS,p}(k+1) &= \beta_{NS,0}(k), \end{aligned} \tag{7}$$

where T=1s is sampling period, α, β, and γ filter fixed coefficients, q parameter which value is 1 when at least one observation pass gating test, or value of number scans without target presents, x target position, v target speed, a target acceleration, and β_{NS} new source density. Indexes s indicate estimated state of vector, indexes p indicate predicted state of vector, and indexes 0 indicate measured state of vector.

Standard set of initialization parameters is:

$$\begin{aligned} x_s(1) &= x_p(2) = x_0(1), \\ v_s(1) &= a_s(1) = \alpha_s(2) = 0, \\ v_s(2) &= \frac{x_0(2) - x_0(1)}{T}, \\ a_s(3) &= \frac{x_0(3) + x_0(1) - 2 \cdot x_0(2)}{T^2}, \\ \beta_{NS}(0) &= 0. \end{aligned} \tag{8}$$

All calculations in this paper are performed for standard set of filter coefficients [1]:

$$\alpha \le 0.6,$$
$$\beta = 2\cdot\left(2-\alpha\right) - 4\cdot\sqrt{1-\alpha}, \quad (9)$$
$$\gamma = \frac{\beta^2}{2\cdot\alpha}.$$

D. Gating

The gates are forming around the predicted tracks positions. One simple gating technique is rectangular technique that is defining as:

$$d = \left|y_i - \overline{y_i}\right| \le K_{Gi}\cdot\sigma_r, \quad (10)$$

where σ_r target residual standard deviation given with relation (3), K_{Gi} the gating constant of i^{th} track. Assuming the Gaussian error model, it is possible to adopt same gating constant K_G for all tracks.

Define probability of correct decision P_{CD} as:

$$P_{CD} = P_D\cdot P_{CC/D} + \left(1-P_D\right)\cdot P_{NE/D}, \quad (11)$$

where P_D is detection probability, $P_{CC/D}$ is probability of correct correlation, $P_{NE/D}$ is probability of no correlation.

Probability of correct correlation $P_{CC/D}$ refers to state of detection when the true target occurs:

$$P_{CC/D} = \left[C\cdot erf\left(\frac{K_G}{\sqrt{2\cdot C}}\right)\right]^2, \quad (12)$$

where constant C is defined by next relation:

$$C = \frac{1}{\sqrt{1+\frac{\pi}{2}\cdot n_{TF}}} \quad (13)$$

Probability of no correlation $P_{NE/D}$ refers to state of detection when the false alarm satisfy gate:

$$P_{NE/D} = e^{-n_{TF}\cdot K_G^2}, \quad (14)$$

where n_{TF} is dimensionless variable that represent expected number of false returns within track gates. Quantity n_{TF} defines as:

$$n_{TF} = \beta_{NS}\cdot\left(2\cdot\sigma_{r,R}\right)\cdot\left(2\cdot\sigma_{r,\theta}\right) \quad (15)$$

Standard MTT systems assume new source density β_{NS} as a static parameter of radar system. Nature of new source density β_{NS} is more complex and except it depends on hardware platform it is also depends on relief topology. Variation of new source density respect to relief topology affects on gates that has influence on correlation and new tentative track initialization.

After several algebras transformations (11-15) with maximization probability of correct decision in equation (11), the gating constant is:

$$K_G = \sqrt{\frac{b^2}{4\cdot c^2} - \frac{a}{c}} - \frac{b}{2\cdot c}, \quad (16)$$

where a, b, c are given as:

$$a = \ln\left[\frac{2+n_{TF}-2\cdot P_D}{\left(2+n_{TF}\cdot\pi\right)\cdot P_D}\right],$$
$$b = 0{,}14\cdot\sqrt{1+\frac{n_{TF}\cdot\pi}{2}}\cdot a + n_{TF}^2 = \frac{4}{\pi}\cdot\sqrt{1+\frac{n_{TF}\cdot\pi}{2}}, \quad (17)$$
$$c = 0{,}07\cdot\left(1+\frac{n_{TF}\cdot\pi}{2}\right)\cdot\left(1-n_{TF}\right).$$

III. Results and Discussion

The efficiency of toward adaptive initialization of new tracks in MTT systems is demonstrated through computer 2D simulations. It is simulated crossing maneuver condition of four targets. Further, maneuver area were distributed to four areas with different new source density that are simulated relief in real condition. Reflected signals from inhomogeneous relief structure (stones, hills, woods, and water) produce false alarm signals in MTT systems.

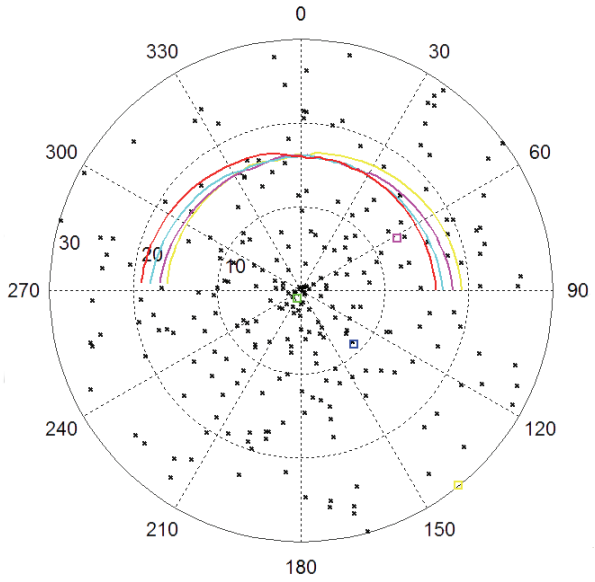


Fig. 2. Simulation of MTT tracking in homogeny field

Fig. 2. presents simulation of MTT tracking four maneuvering targets in homogeny filed. Solid lines present target tracks, rectangles new sources, and x-es present false alarms that are product of MTT clutter. Each scan results with true targets observations and false alarm observation. Large number of false alarms is consequence of large number of scans.

Initial targets positions are on 90 of azimuth and targets are moving in CCW direction. Targets are located far from the radar 15-20 km, and less than 3 m between each other. All tracking range could be divided to three areas: beginning, middle, and ending area. In the beginning area of tracking, targets are enough away from each other and allow quality tracking. The same goes for the ending area. In the middle tracking area, targets are so close and quality of tracking is not on the maximal level. It is possible to say that MTT is functioning on threshold of detection in the middle area.

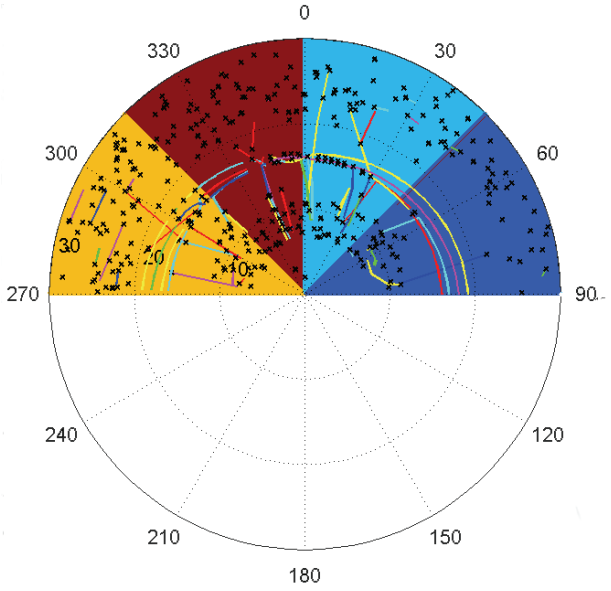


Fig. 3. Simulation of MTT tracking in inhomogeneous relief structure

Fig. 3. presents simulation of MTT tracking four maneuvering targets in inhomogeneous relief structure. Tracking area from Fig. 2. is distributed to four area with different new source density. Initially, the new source density has lowest value and slightly increasing to the end, one by one area. As it was expected, tracking is best at the first area. Number of false alarm that initiated new tracks in this area is 6. In the second area, tracking is significant disturbed and some targets could not be tracked. Reasons are increasing of new source density and closeness of targets. Number of false alarm that initiated new tracks in this area is 9. In the third area, new source density is more increased and that reflected with short losing of tracking. Number of false alarm that initiated new tracks in this area is 10. In the last area, where new source density is even more increased, short losing of tracking is again present. Number of false alarm that initiated new tracks in this area is 13.

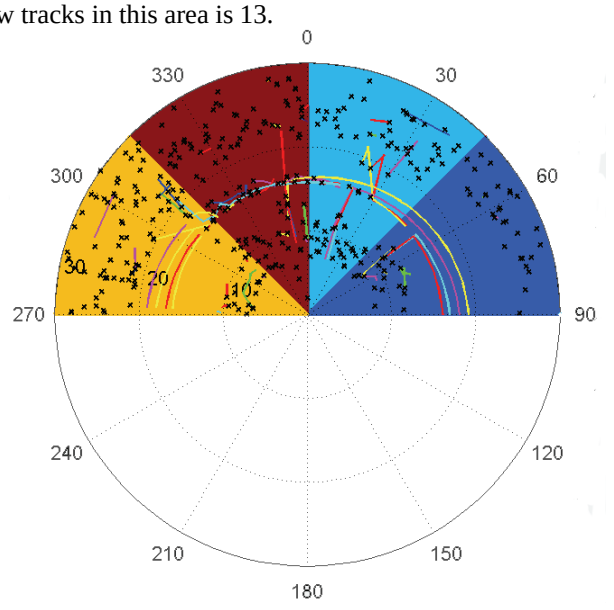


Fig. 4. Simulation of MTT tracking in inhomogeneous relief structure with estimation of new source density

Fig. 4. presents simulation of MTT tracking four maneuvering targets in inhomogeneous relief structure with estimation of new source density. Results of the new source density estimation, improves quality of true targets tracking. Number of false alarm that initiated new tracks is reduced from 6 to 4 at the first segment, from 9 to 7 at the second, from 10 to 9, and from 13 to 9 at the last. It is important to note that many of false initiated tracks are in the relief segment transient areas. With further filtering of transient areas, number of false alarm that initiated new tracks is reduced from 6 to 3 at the first segment, from 9 to 5 at the second, from 10 to 8, and from 13 to 8 at the last.

IV. Conclusion

General conclusion is that toward adaptive initialization of new tracks, achieved by estimation of new source density, could partial depress false targets caused by relief topology of sensor environment and improves the efficiency MTT systems. Obtained results may be applied on radar or sonar systems or any other sensors network that performs multiple target tracking.

Acknowledgement

This work was supported in part by the Ministry of Science of Serbia within the Project “Research and development solutions to improve the performance of wireless communication systems in the microwave and millimeter frequency range”, project N0. TR-32052.

References

[1] S. S. Blackman, *Multi-Target Tracking with Radar Applications*, Dedham, Artech House, 1986.

[2] J. A. Fuemmeler, V. V. Veeravalli, “Energy Efficient Multi-Object Tracking in Sensor Networks in Sensor Networks”, IEEE Trans. on Signal Processing, vol. 57, no. 7, pp. 3742-3750, 2010.

[3] D. Clark, B. Ristic, B. N. Vo, B. T. Vo, “Bayesian Multi-Object Filtering With Amplitude Feature Likelihood for Unknown Object SNR”, IEEE Trans. on Signal Processing, vol. 58, no. 1, pp. 26-37, 2010.

[4] B. T. Vo, B. N. Vo, A. Cantoni, “The Cardinality Balanced Multi-Target Multi-Bernoulli Filter and Its Implementations”, IEEE Trans. on Signal Processing, vol. 57, no. 2, pp. 409-423, 2009.

[5] S. Sen, A. Nehorai, “Sparsity-Based Multi-Target Tracking Using OFDM Radar”, IEEE Trans. on Signal Processing, vol. 54, no. 4, pp. 1902-1906, 2011.

[6] S. Oh, S. Russell, S. Sastry, “Markov Chain Monte Carlo Data Association for Multi-Target Tracking”, IEEE Trans. on Automatic Control, vol. 54, no. 3, pp. 481-497, 2009.

[7] L. Trailovic, , L. Y. Pao, “Computing Budget Allocation for Efficient Ranking and Selection of Variances With Application to Target Tracking Algorithms”, IEEE Trans. on Automatic Control, vol. 49, no. 1, pp. 58-67, 2004.

[8] L. Hong, N. Z. Cui, “An Interacting Multipattern Probabilistic Data Association (IMP-PDA) Algorithm for Target Tracking”, IEEE Trans. on Automatic Control, vol. 46, no. 8, pp. 1223-1236, 2001.

Experimental Studies of Broadband Transmission Line Transformers

Boyan Karapenev[1]

***Abstract* – In this paper are given experimental results obtained from studies of broadband transmission line transformers in the frequency response of their used ferrite toroidal cores 0,5÷30 MHz with coefficients of resistant ratio 1:1, 4:1 and 1:4. Some transformers are practically implemented with 3 and 5 twists per 1 cm of pair copper enamelled wires with ferrite toroidal cores manufactured by the firm Amidon FT82-77 and FT114-77. The obtained results – characteristics and qualitative parameters are generalized and presented in graphical and tabular form.**

***Keywords* – Experimental studies, Broadband transmission line transformers, Qualitative parameters, 3 and 5 twists per 1 cm.**

I. INTRODUCTION

Broadband co-ordination and transformation of input and output resistance of a high-frequency amplifier, between two adjacent amplifier stages as well as broadband power aggregation and division can be carried out by transmission line transformers employing an electromagnetic connection between the primary and secondary windings. To provide the necessary transformation coefficient of resistances and minimum deco-ordination is an important prerequisite for achieving a wide operating frequency band. These transformers have high efficiency and reliability and through them can be made: galvanic dissociation between nodes and units of the equipment, transition from asymmetric to symmetric I/O and vice versa, they have small sizes, etc.

II. IMPLEMENTATION AND EXPERIMANTAL STUDIES OF BROADBAND TRANSMISSION LINE TRANSFORMERS

A. Implementation of broadband transmission line transformers

Broadband Transmission Line Transformers (BTLT) are constructed employing appropriately interconnected transmission lines, positioned on a ferromagnetic core which is mostly of toroidal shape [3]. The input signal excites electromagnetic waves whose linear combinations depending on the type of line connection, determine the output signal voltage. Since the resistance transformation of BTLT is associated with changes in the coefficient of voltage transmission A_U ($nz=n_U^2$) when nz=4:1, n_U=2:1 and A_U=0,5. For nz=1:4, n_U=1:2 and A_U=2. The operating principle of broadband transmission line transformers is illustrated in Fig. 1.

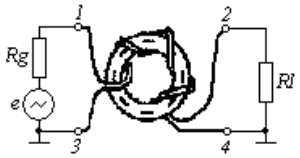


Fig. 1. Transmission Line Transformer

The main component in transmission line transformers is the ferrite toroidal core used. Some of the most important catalog parameters of ferrite toroidal cores manufactured by the firm Amidon FT82-77 and FT114-77 are presented in Table 1 [1].

TABLE I
CATALOG PARAMETERS OF AMIDON
FERRITE TOROIDAL CORES FT82-77 AND FT114-77

Parameter	Ferrite core grade, Amidon	
	FT82-77	FT114-77
D_{dim}, mm	21,000	29,000
d_{dim}, mm	13,100	19,000
h_{dim}, mm	6,350	7,000
l_e, cm	5,260	7,420
A_e, cm^2	0,246	0,375
V_e, cm^3	1,290	2,790
Δf, MHz	0,5÷30	0,5÷30
A_l, mH/1000w	1170	1270
μ_r	2000	2000
ρ, Ω.cm	1.10^2	1.10^2

The transmission lines implemented with pair copper enamelled wires were most widely used in the frequency range 0,1÷50 MHz and to ensure the characteristic impedances from 15÷20 to 100÷150 Ω using BTLT [3]. Analytical determination of the geometry of these lines for necessary characteristic impedance Z_0 is too complicated therefore it is appropriate the construction parameters to be selected experimentally. It is recommended that the wires should be wound with a large number of twists per unit of length, not to change the characteristic impedance along the line [6].

The connection method and the values of input and output resistance of the co-ordinating (nz=1:1) and transforming (nz=4:1 and nz=1:4) the resistance broadband transmission line transformers are shown in Fig. 2.

[1]Boyan Karapenev is with the Department of Communications Technology and Engineering of Technical University, 4 Hadji Dimitar Str., 5300 Gabrovo, Bulgaria, E-mail: bkarapenev@tugab.bg

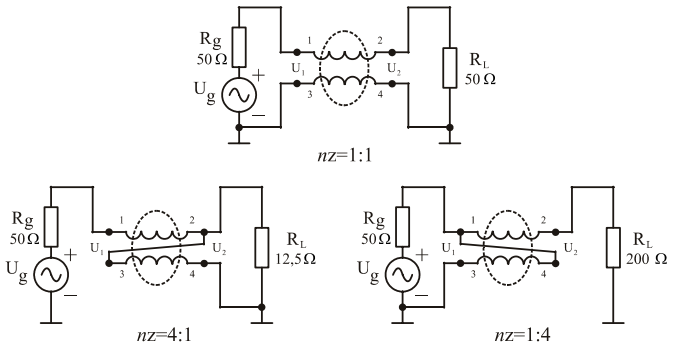


Fig. 2. Connection of BTLT with *nz*=1:1, *nz*=4:1 and *nz*=1:4

Broadband transmission line transformers with ferrite toroidal cores by the firm Amidon FT82-77 and FT114-77 with 7 windings of twisted pair copper enamelled wires with diameter *d*=0,62 mm when the number of twists is equal to 3 and 5 per 1 cm are implemented practically. The same design implementation of the broadband transmission line transformers with a small and large ferrite core is used, and the corresponding coefficient of transformation is provided with the necessary transmission lines connection to the source and load, and their corresponding values (Fig. 2). The experimental studies were carried out with values of the input voltage U_1=48,76 mV and U_1=21,2 mV.

B. Experimental results of BTLT with toroidal ferrite core Amidon FT82-77 with nz=1:1, nz=4:1 and nz=1:4

In Fig. 3 and Fig. 4 are presented the obtained amplitude-frequency responses (AFR) in graphic form for BTLT with ferrite core Amidon FT82-77 with 3 twists per 1 cm respectively when U_1=48,76 mV and U_1=21,2 mV.

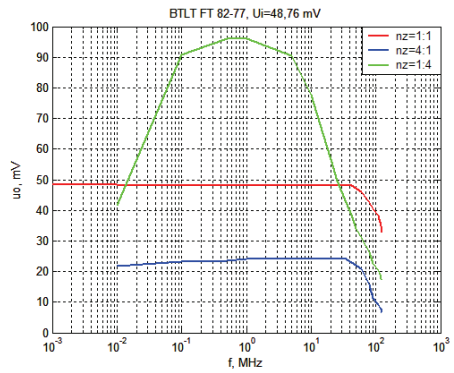


Fig. 3. AFR of BTLT with ferrite core FT82-77, with 3 twists per 1 cm when U_1=48,76 mV

Based on the results presented in Fig. 3 and Fig. 4 it is established that:

- when the coefficient of resistance transformation is *nz*=1:4, a drop of amplitude-frequency response at low and high frequencies is observed, from where there is a clearly formed frequency band of transmission;
- the amplitude-frequency response slopes at high frequencies are steeper when reducing the value of input voltage; there are sections with a different drop in individual frequency bands; in this case the frequency band is wider.

Conclusion: The transmitted frequency band expands when the value of input voltage is reduced.

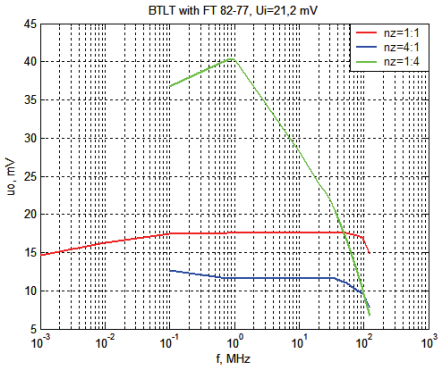


Fig. 4. AFR of BTLT with ferrite core FT82-77, with 3 twists per 1 cm when U_1=21,2 Mv

In Fig. 5 and Fig. 6 are presented graphically AFR of BTLT with ferrite core Amidon FT82-77, respectively when U_1=48,76 mV and U_1=21,2 mV with 5 twists per 1 cm.

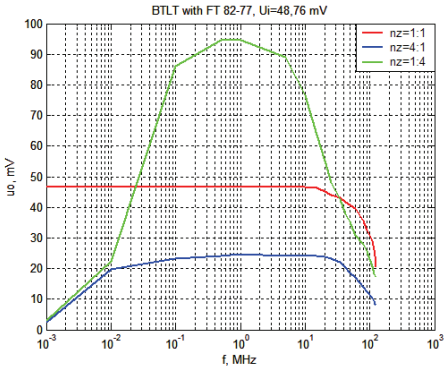


Fig. 5. AFR of BTLT with ferrite core FT82-77, with 5 twists per 1 cm when U_1=48,76 mV

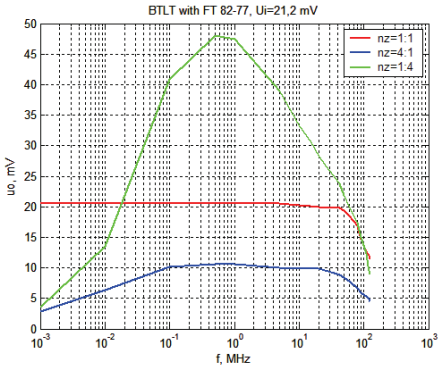


Fig. 6. AFR of BTLT with ferrite core FT82-77, with 5 twists per 1 cm when U_1=21,2 mV

From the presented AFR of BTLT with ferrite core Amidon FT82-77 with 5 twists per 1 cm, it has been found that:

- when transmission coefficients *nz*=1:4 and *nz*=4:1 with 5 twists there is a drop of AFR in the range of low frequencies.

In both cases the permeability of ferrite does not exert significant influence on the low-frequency range, because when *nz*=1:4 there is also Au=1:2, and when *nz*=4:1 – Au=2:1 what reduces the value of the created magnetic flux in the ferrite core;

- when *nz*=1:4 the shape of AFR resembles that of the bandpass filter and the drops at low and high frequencies are more pronounced (steeper). In this case the section with a uniform transmission coefficient in the transmitted band is very small which determines the very narrow transmitted frequency band.

Conclusion: The transmitted frequency band narrows with increasing the number of twists.

C. Experimental results of BTLT with toroidal ferrite core Amidon FT114-77 with nz=1:1, nz=4:1 and nz=1:4

Amplitude-frequency responses of BTLT with ferrite core Amidon FT114-77 are presented in graphical form when input voltages are U_1=48,76 and U_1=21,2 mV with 3 twists per 1 cm respectively in Fig. 7 and Fig. 8.

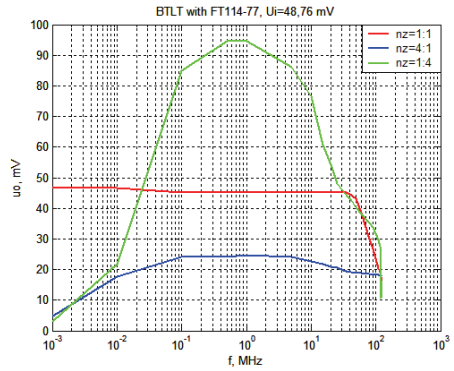


Fig. 7. AFR of BTLT with ferrite core FT114-77, with 3 twists per 1 cm when U_1=48,76 mV

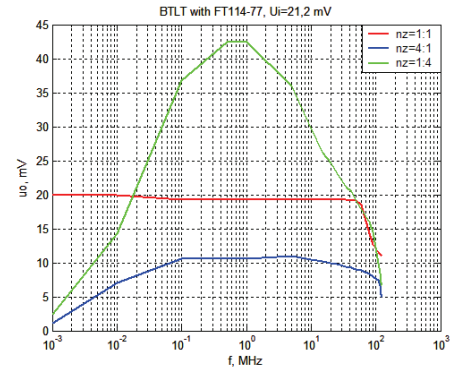


Fig. 8. AFR of BTLT with ferrite core FT114-77, with 3 twists per 1 cm when U_1=21,2 mV

From the experimental studies carried out and their results obtained, it is established that when *nz*=1:1 in the frequency range 1÷30 kHz, the transmission coefficient is a bit higher compared to that in the transmitted band of ferrite core. This is due to the larger size of the used ferrite core and the greater length of transmission lines.

Conclusion: The type of amplitude-frequency response is determined by the method of connecting transmission lines to the source and load, i.e. by *nz*, and when *nz*≠1:1, are established slopes in the low-frequency and high-frequency field. When *nz*=1:1 a wider frequency band is ensured due to the extended low-frequency and high-frequency range.

Amplitude-frequency responses of broadband transmission line transformers with ferrite core Amidon FT114-77 with 5 twists per 1 cm are shown in Fig. 9 and Fig. 10 respectively when input voltages U_1=48,76 mV and U_1=21,2 mV.

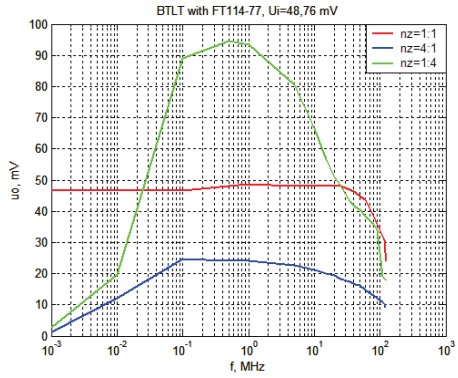


Fig. 9. AFR of BTLT with ferrite core FT114-77, with 5 twists per 1 cm when U_1=48,76 mV

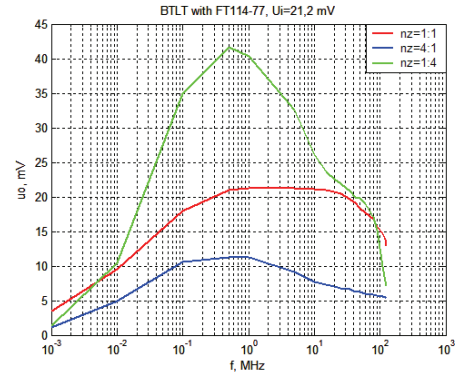


Fig. 10. AFR of BTLT with ferrite core FT114-77, with 5 twists per 1 cm when U_1=21,2 mV

On the basis of obtained amplitude-frequency responses in Fig. 9 and Fig. 10 it has been found that:

- when *nz*=1:1, the transmission coefficient in the frequency response 1÷20 MHz is slightly increased;

- when *nz*=1:4, there is no part in which the transmission coefficient is even. The slopes of growth and drop are more pronounced when the input voltage is lower;

- There are slopes with different drop of amplitude-frequency responses at lower value of operating voltage. For cases presented in Fig. 10 the steepness of these slopes decreases in both cutoff-frequency ranges.

The qualitative parameters of BTLT with ferrite toroidal cores Amidon FT82-77 and FT114-77 with 3 and 5 twists per 1 cm when U_1=48,76 mV and U_1=21,2 mV are presented in the generalized Table II.

TABLE II
PARAMETERS OF IMPLEMENTED BROADBAND TRANSMISSION LINE TRANSFORMERS

Number of turns	U_1, mV	*nz*	BTLT with Amidon FT82-77			BTLT with Amidon FT114-77		
			fb, kHz	fh, MHz	Δf, MHz	fb, kHz	fh, MHz	Δf, MHz
3 twists per 1 cm	48,76	1:1	1	120	119,99	1	73	72,99
		4:1	1	78	77,99	10	125	124,99
		1:4	50	14	13,95	80	13	12,92
	21,20	1:1	1	125	124,99	1	78	77,99
		4:1	1	110	109,99	40	100	99,96
		1:4	1	10	9,99	80	8	7,92
5 twists per 1 cm	48,76	1:1	1	90	89,99	1	100	99,99
		4:1	60	6	5,94	30	35	34,97
		1:4	70	13	12,93	70	10	9,93
	21,20	1:1	1	90	89,99	80	100	99,92
		4:1	60	55	54,94	80	3	2,92
		1:4	90	9	8,91	70	7	6,93

From the obtained comparative assessment of AFR and parameters of BTLT with 3 and 5 twists per 1 cm it is established that:

- when *nz*=1:4 and the ferrite core (FT114-77) is larger, narrowing of AFR is observed as the section with a uniform transmission coefficient remains the same;
- when *nz*=1:1 there is a difference in the transmission coefficient of the order of 6 %;
- when *nz*=4:1 for BTLT with ferrite core FT82-77, a larger section of the transmitted band with the same value of transmission coefficient is established, compared with BTLT with ferrite core FT114-77;
- when *nz*=4:1 for BTLT with ferrite core FT114-77 and 5 twists, the transmitted frequency band lightly shifts to the range of the low frequencies - Fig. 4.10.

Conclusion: A significant difference between AFR of BTLT with ferrite cores FT88-77 and FT114-77 in individual *nz* has not been found. The existing differences when there are 5 twists per 1 cm are smaller than those when the twists are 3. The minimum differences in AFR of BTLT with ferrite cores FT88-77 and FT114-77 with 3 and 5 twists per 1 cm are attributable to manufacturing tolerances of the parameters of the used ferrites - μ_r и A_L.

In Fig. 11 is presented a comparative assessment of the obtained amplitude performances (AP) for different implementations of broadband transmission line transformers with ferrite core FT82-77 with 3 and 5 twists per 1 cm when f=1 MHz.

From the obtained AP it is established that:

- the nature of AP in all cases is linear;
- the difference in AP between BTLT with ferrite core FT82-77 with 3 and 5 twists per 1 cm is minimal.

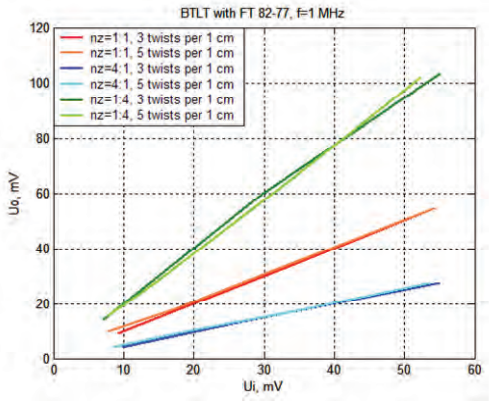


Fig. 11. Comparative assessment of AP of BTLT with ferrite core FT88-77 with 3 and 5 twists per 1 cm when f=1 MHz

III. CONCLUSION

This paper presents a comparative assessment (in graphical and tabular form and analysis) of the experimentally obtained AFR and qualitative parameters of implemented BTLT with ferrite toroidal cores manufactured by the firm Amidon FT82-77 and FT114-77 with 3 and 5 twists per 1 cm for various values of the submitted input voltage and different *nz* – 1:1, 4:1 and 1:4.

The obtained results can be used for optimization of the values of model parameters and qualitative parameters of simulation models of broadband transmission line transformers, and in practice.

REFERENCES

[1] Amidon Associates, „Iron-powder and Ferrite Coil Forms", *Catalogue*, 2008.

[2] G. A. Breed, "Transmission Line Transformer Basics," *Applied Microwave & Wireless*, Vol. 10, No. 4, May 1998.

[3] D. Dobrev, L. Jordanova, *Radiocommunications*, Part One, Sofia, Siela, 2001.

[4] J. Sevick, *Transmission Line Transformers*, New Jersey, Noble Publishing Corp., 4th Edition, 2001.

[5] Ch. Tihchev, *Radiotransmitter Devices*, Tehnika, Sofia, 1992.

[6] C, Trask, *Transmission Line transformers: Theory, Design and Application*, Part I and II, 2006.

High Efficiency RF Amplifier Design for Maximum PAE

Iliya Nedelchev[1]

Abstract **– In most cases during analyzing and designing efficient RF amplifiers ideal switch model is used. This model does not take in the account the power dissipation in the driving loop and its increasing with the frequency. Finally the amplifier gain cannot be determined correctly and it is obvious contradiction between a significant gain and high drain efficiency. To express the input, output and DC power simplified model of MOSFET was used allowing to design high efficiency class E amplifier with maximum PAE(power-added-efficiency). Presented are results of designing of high efficiency RF amplifiers with output power Po=100W and operating frequency 1 and 3 MHz.**

Keywords **– Power amplifier class E, High efficiency, Power-added-efficiency (PAE).**

I. Introduction

Class E resonant power amplifier becomes more attractive in recent development in wireless communications as high efficient transmitter. Lack of new theoretical investigations can be partially explained with the fact that there is no easy implemented standardize techniques for analyzing the schematic. There are many authors who are treating the active device as an ideal switch [3, 7, 8]. The optimal operating conditions of the high efficiency amplifier can be found in [3]. Analytical method for calculating the components of the output loop of the class E amplifier using ideal switch is presented in [8].

In a HF band the parasitic components and related decrease in conductivity are strongly effecting the operation. When using the ideal switch as a model usually the driving power of the real switch is neglected. Even more the model does not include the active conductance of the transistor reflecting the drain efficiency in HF region.

There is a contradiction between high gain and high drain efficiency. By that reason new systematically build method is presented where the effect of the parasitic component of the transistor and drain efficiency are taken in the account. For the mathematical equations for input, output and DC power simplified model has been used. That allows optimization of class E power amplifier with maximum enhanced efficiency PAE. The followed design is limited to FETs only.

[1]Iliya Nedelchev is with the Technical University - Gabrovo, H. Dimitar 4, 5300 Gabrovo, Bulgaria, E-mail:lined@tugab.bg.

II. Systematically Design Method

A. Simplified MOSFET model

Simplified model of MOSFET including the parasitic components is presented on fig. 1. The ideal switch is replaces with active conductance component. The source resistance R_S was neglected in the course of the analysis from GS and DS

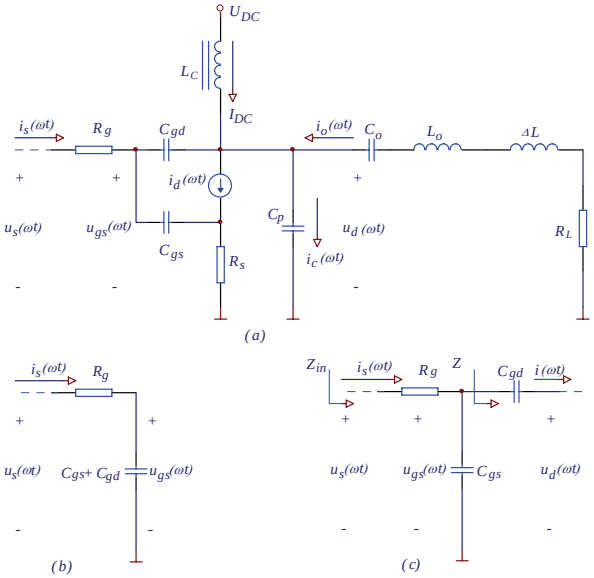


Fig.1 MOSFET model of amplifier class E (a). Equivalent circuits for positive (b), and negative (c) half-cycles.

capacitances as very low value component. Results comparisons for the input power, output power PAE etc.

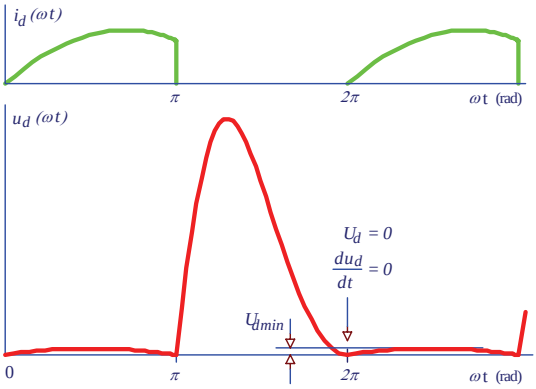


Fig. 2.Drain current and voltage waveforms for class E amplifier

confirm that connecting the capacitances to the ground does not affects the model accuracy.

Drain current and voltage as per optimal mode class E with real MOSFET are shown on fig. 2.

Mathematical expressions for drain current and voltage for class E amplifier for both state of the transistor (OFF and ON) are given in [7]. In the expressions there are additional sub-expressions added to represent the effect of the voltage drop of the sores resistor and minimum drain voltage in the ON state of the transistor ($U_{d\,min}$ –fig. 2).

The max voltage on the drain can be calculated from [5].

$$U_{d\max} = 1.13\frac{I_{DC}}{\omega C_d} + U_{d\min} + 2I_{DC}R_s\,. \qquad (1)$$

In the investigation the transistor is represented as a current source driven by voltage, where the driving voltages are gate-source and drain-source (fig. 3). This allows the DC current through the loop I_{DC} and the drain-source voltage in the ON time ($U_{d\,min}$) to be calculated [1]:

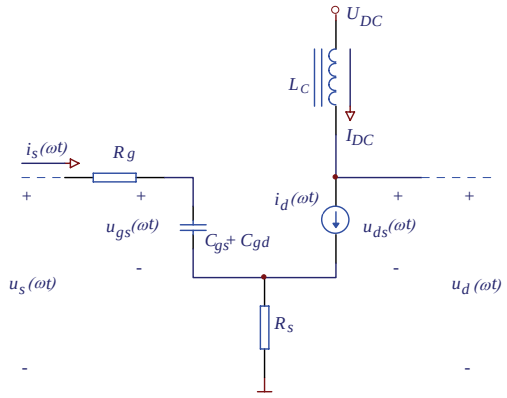


Fig. 3. MOSFET equivalent circuit.

$$I_{DC} = \frac{\pi\omega C_p\left(U_{DC} - U_{d\min}\right)}{1 + 2\pi\omega R_s C_p}, \qquad (2)$$

$$\begin{aligned} U_{d\min} &= -\frac{b + \sqrt{b^2 - 4ac}}{2a} \\ a &= 1 - 2\pi\omega R_s C_p \\ b &= 2\left\{2\pi\omega R_s C_p\left[U_{DC} - u(\pi)\right]\right. \\ &\quad \left. - 2\pi\omega C_p\frac{L}{KW} - u_{gs}(\pi)\right\} \\ c &= 4\pi\omega C_p\frac{L}{KW}U_{DC} \end{aligned}, \qquad (3)$$

where W, L and K are the width, length and internal conductivity coefficient of the MOSFET channel.

B. Maximum PAE Design

The algorithm consists of series of steps for designing of high efficiency RF amplifier with maximum PAE and is presented on fig. 4. The data used for the design are output HF power in the load P_o, power supply DC voltage U_{DC} and operating frequency f_o.

For calculating of the components values based on design data (P_o, U_{DC}, f_o) is required R_{on} but as well C_p and $\omega_o C_p R_{on}$ to be establish. Also the parasitic parameters of the MOSFET are used.

From the design data initial calculations are used for the load resistance R_{init} and the capacitance of C_p, max current of the drain $I_{d\ max}$, and max drain voltage $U_{d\ max}$. For high efficiency operation the max MOSFET resistance in ON time is determine $R_{on\ max}$. Depending of the results MOSFET is chosen [4].

To check the right MOSFET choice the simplified model can be used. For that purpose the power gain is calculated from [1]:

$$G_p = \frac{6.935 I_{DC}^2 R_L\left[1 + \left(\frac{\omega_o R_s C_p}{2}\right)^2\right]\left[1 + \left(\omega_o R_g C_g\right)^2\right]}{U_s\omega_0^2 R_g C_g\left(2U_s C_g + U_{d\max} C_{gd}\right)}, \qquad (4)$$

where $C_g = C_{gs} + C_{gd}$.

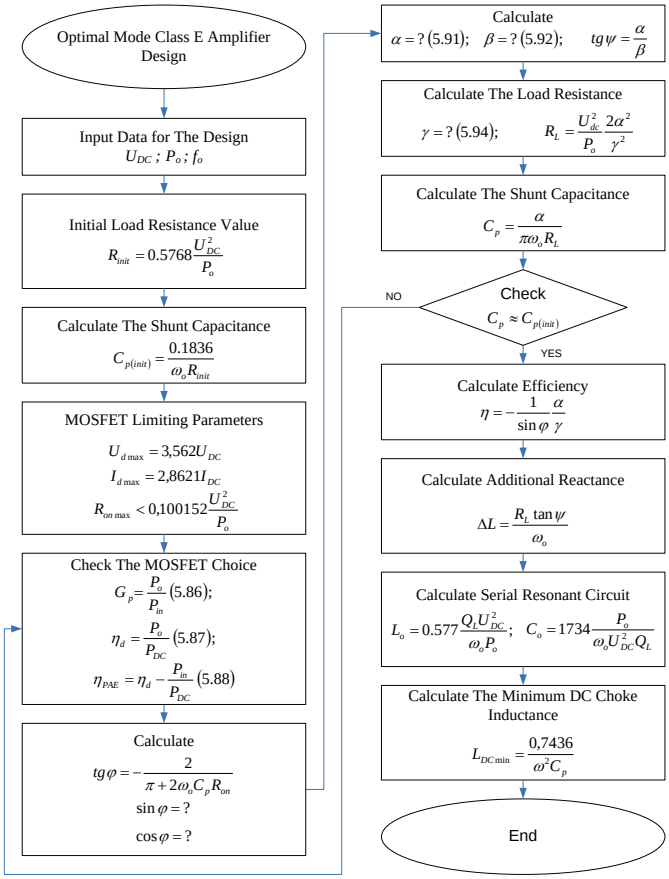


Fig. 4.Design Algorithm

The electron efficiency η_d and PAE are calculated [5].

$$\eta_d = \frac{P_o}{P_{DC}} = 1.734 \frac{I_{DC} R_L \left[1 + \left(\frac{\omega_o R_s C_p}{2}\right)^2\right]}{U_{DC}}, \quad (5)$$

$$\eta_{PAE} = \frac{P_o - P_{in}}{P_{DC}} = \eta_d - \frac{P_{in}}{P_{DC}} \quad (6)$$

$$= 1.734 \frac{I_{DC} R_L \left[1 + \left(\frac{\omega_o R_s C_p}{2}\right)^2\right]}{U_{DC}} - \frac{U_s \omega^2 R_g C_g \left(2U_s C_g + U_{d\max} C_{gd}\right)}{4 I_{DC} U_{DC} [1 + (\omega_o R_g C_g)^2]}.$$

MOSFET has been chosen in order to achieve max PAE (η_{PAE}). Limiting Conditions in this case are the max drain current $I_{d\ max}$, max drain voltage $U_{d\ max}$, and max resistance of the MOSFET in the ON state $R_{on\ max}$. After the choice of the MOSFET the coefficients are calculated [2].

$$\alpha = 2\sin^2\varphi - 2\cos^2\varphi - \pi\sin\varphi\cos\varphi - 2\omega_o C_p R_{on}\sin\varphi\cos\varphi - \frac{\pi}{2}\omega_o C_p R_{on} \quad (7)$$

$$\beta = \pi\sin^2\varphi + 4\sin\varphi\cos\varphi + \frac{\pi}{2} + 2\omega_o C_p R_{on}\sin^2\varphi \quad (8)$$

$$\gamma = 2\omega_o C_p R_{on}\cos\varphi - 3\omega_o C_p R_{on}\pi\sin\varphi - \frac{\pi^2}{2}\sin\varphi - 2\sin\varphi - \pi\cos\varphi. \quad (9)$$

The load resistance R_L and the capacitance of C_p are calculated again in comparison with the initial calculations. If the values are similar next steps of the design process can be carried out (fig. 4). If not the case, new transistor has to be selected and the design process repeats.

III. Simulation and Experimental Results

The propose design method is applied when design class E power amplifier with design data: P_o = 100 W, U_{DC}= 96 V, f_o= 1 MHz и f_o = 3,5 MHz. The component values calculated are compared with other classical methods [3], [7] results, barely different from each other and are presented in table I and II. Selected was MOSFET transistor IRF 730, V_{DSS} = 400 V, $R_{DS(on)}$=1,0 Ω, I_D = 5,5 A, C_{iss} – Input Capacitance – 700 pF, C_{oss} – Output Capacitance – 170 pF, C_{rss} – Reverse Transfer Capacitance – 64 pF.

TABLE I
AMPLIFIER CLASS E – 3.5 MHZ

Parameter	Suggested method	Classic method	Difference [%]
$\Delta L\ [\mu H]$	6,796	9.167	34,89
$C_p\ [pF]$	577	584	1,12
$L_o\ [\mu H]$	42,32	30,62	-27,65
$C_o\ [pF]$	599	827	38,06
$L_{DC}\ [\mu H]$	180	180	0
η_d [%]	97,4	90,2	-7,39
η_{PAE} [%]	96,1	88,9	-7,49
$G_p\ [dB]$	46,9	32,6	-30,49

TABLE II
AMPLIFIER CLASS E – 1 MHZ

Parameter	Suggested method	Classic method	Difference [%]
$\Delta L\ [\mu H]$	1,934	2,62	35,47
$C_p\ [pF]$	164	180	9,76
$L_o\ [\mu H]$	12,09	8,749	-24,51
$C_o\ [pF]$	178	236	32,58
$L_{DC}\ [\mu H]$	50	50	0
η_d [%]	96,3	87,1	-9,55
η_{PAE} [%]	94,7	85,7	-9,50
$G_p\ [dB]$	44,7	30,5	-30,77

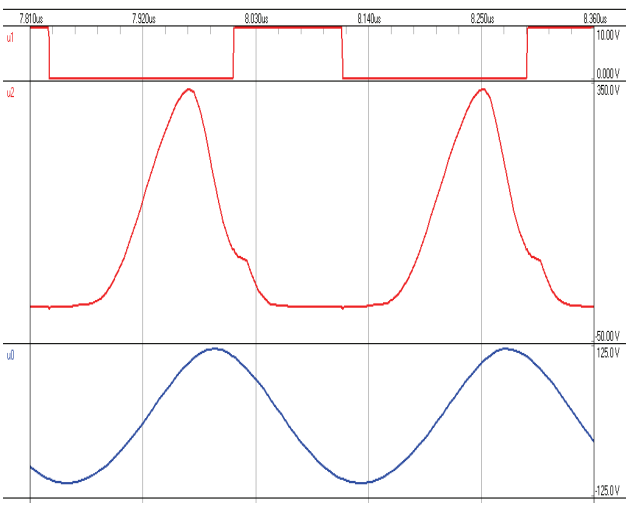


Fig. 5. Amplifier class E voltage waveforms.

The results were checked with computer simulation with Protel 99 SE at operating frequency 3.5MHz. The gate voltage *u1*, drain voltage *u2* and load voltage *u0* presented are on fig. 4. The shape of the waveforms confirms that all optimal conditions for class E operation are met.

Output voltage by default is sine wave but contains about 6% second harmonic. It distort the waveform THD 6.7% (fig. 6). If special output voltage waveform requirements higher order filter can be connected after the series resonant loop in order to suppress the high level harmonics and match the output impedance of the amplifier to the load. As sample application, it can be used as a quadrature modulator in coaxial cable distributed television networks. [6].

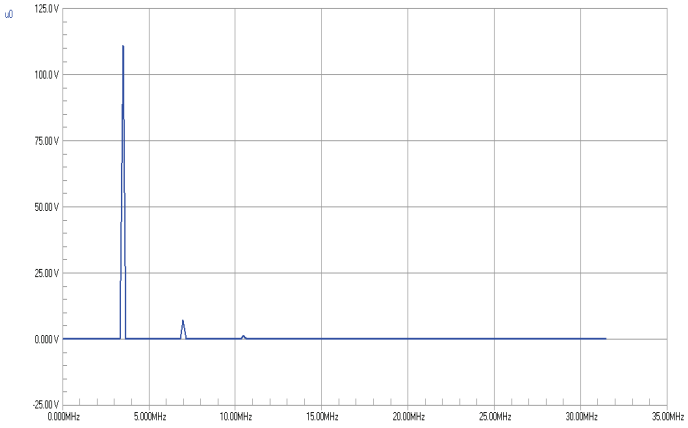


Fig. 6. Load voltage spectrum.

Based on the design prototype unit was developed. Pictured are the drain and load waveform of power amplifier class E operating at 1MHz (fig. 7).

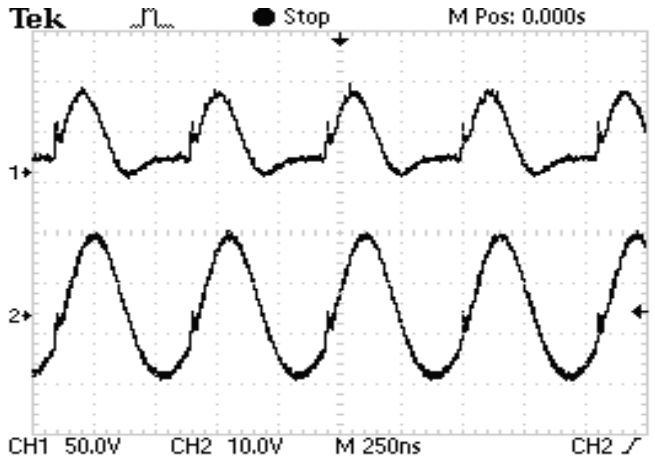


Fig. 7. Waveform of class E:
1 - Drain voltage, 2 - Load voltage.

As can be seen the amplifier class E is in stable operation and all theoretical conditions for optimal mode and minimal losses are met.

IV. Conclusion

The method presented reflects the transistor parameters. That allows by the choice of the MOSFET to compromise between bigger efficiency of the drain (η_D) i.e. minimum ON resistance R_{on} or bigger power gain G_p i.e. small transistor input capacitor C_{iss}. That decreases the driving power, allowing to improve the current and temperature stress of the driving stage and increase the power – added efficiency (η_{PAE}).

When calculating the shunt capacitor value C_p the transistor parameters are taking into a account as well, allowing minimizing the losses and improve the efficiency and maximum operating frequency where the described transistor can be used.

References

[1] D. Choi, S. I. Long, "A Physically Based Analiryc Model of FET Class-E Power Amplifiers-Designing for Maximum PAE", *IEEE Transactions on Microwave Theory and Techniques*, vol. 47 No. 9, pp. 1712-1720, September 1999.

[2] D. Milosevic, J. van der Tang, A. van Roermund, "Analysis of Losses in Non-ideal Passive Components in The Class E Power Amplifier", *ProRISC Workshop on Circuits, Systems and Signal Processing*, 2003

[3] F. Raab, "Idealized Operation of Exact Analiysis of Class E tuned power amplifier", *IEEE Trans., Circuits and Systems*,vol. CAS-24, pp. 725-735, Dec. 1977.

[4] I. Nedelchev, "Analysis of Class E Power Amplifier Using Real Components" *UNITECH'09 International Scientific Conference*, p p.I-295-I-301 Gabrovo, Bulgaria, 2009.

[5] I. Nedelchev, "Model of MOSFET for Class E Power Amplifier", *UNITECH'09 International Scientific Conference*, pp.I-288-I-294, Gabrovo, Bulgaria, 2009.

[6] K. Koitchev, K. Angelov, S. Sadinov, "Boundary Conditions When Using Quadrature Modulations in Coaxial Cable Televisions Networks", *ICEST 2007*, vol.1, pp. 421-424, Ohrid, Macedonia, June 2007.

[7] M. Kazimierczuk, K. Puczko, "Exact Analiysis of Class E tuned power amplifier at Any *Q* and Switching duty cycle", *IEEE Trans., Circuits and Systems*, vol. CA34, pp. 149-159, Feb. 1987.

[8] N. Sokal, A. Sokal, "Class E – A New Class of High Efficiency Tuned Single-Ended Switching Power Amplifiers", *IEEE J. Solid-State Circuits*, vol. SC-10, pp. 168-176, June 1975.

Session RMA II:

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS II

Methods for Generation of Compact Lumped Element Model for Passive Microwave Circuits

N. Dončov[1], F. Mukhtar[2], J. Russer[2], B. Stošić[1], B. Milovanović[1], and P. Russer[2]

***Abstract* – The representation of distributed passive electromagnetic structures by lumped element circuits is considered in this paper. Network models are established by a combination of system identification and circuit synthesis methods and their subsequent application to data obtained by TLM numerical simulation. Two different methods for synthesis of compact lumped element models for linear lossy reciprocal multiports are discussed in the paper. Accuracy and efficiency of developed compact models are verified on the example of a low-pass microstrip filter.**

Keywords **– Distributed passive electromagnetic structures, circuit synthesis methods, compact model.**

I. INTRODUCTION

As modern analog and digital electronics are operating at microwave and millimeter wave frequencies and at gigabit rates, full-wave electromagnetic (EM) tools for the design and modeling of distributed passive circuit structures are required [1]. However, a different approach to overall system analysis is possible, based on an extraction of an equivalent circuit model from a full-wave electromagnetic simulation of the structure under consideration. Such compact lumped element models can be embedded into conventional circuit simulators and treated by methods of network theory [2], [3]. Compared with field oriented simulation, the application of network oriented design methods yields considerable lower computational effort and time.

In general, the representation of distributed circuits by lumped element network models requires that the transfer function of a distributed circuit is realized by an equivalent circuit with an infinite number of lumped circuit elements. This equivalent network should give the same response for a required excitation, as the considered distributed structure (so-called synthesis problem [4]). As this description needs to be valid within a certain frequency range only and within a certain accuracy margin, one can find an equivalent circuit with a limited number of circuit elements. This lumped element model provides a compact description of the distributed circuit. Such representation of distributed circuit sections by a lumped element models can be very useful especially in the case of modeling complex circuit containing also nonlinear and active lumped elements.

[1]Authors are with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mails: [nebojsa.doncov, biljana.stosic, bratislav.milovanovic]@ elfak.ni.ac.rs

[2]Authors are with the Institute for Nanoelectronics, Technical University Munich, Arcisstrasse 21, Munich, Germany, E-mails: [mukhtar, jrusser, russer]@tum.de

In order to generate compact model, the combination of system identification (SI) of microwave structures and subsequent lumped element model synthesis has to be performed by full-wave simulation or measuring of the input and output signals of the device in the time or in the frequency domain [5-8]. To establish the network model of a distributed circuit a three-step procedure is performed:

1. Determination of the transfer functions by numerical EM full-wave analysis or by measurement,
2. Determination of the rational functions representing the transfer functions by system identification using e.g. vector fitting (VF) method,
3. Synthesis of a lumped element equivalent circuit realizing the transfer function.

In this paper, two different methods are discussed in order to perform step 3 - synthesis of compact lumped element models for linear lossy reciprocal multiports. The first method is based on Foster expansion for lossless circuit and its extension to account for lossy structures as suggested in [9]. The second method is Brune's synthesis procedure that ultimately provides network synthesis by positive lumped elements [10-13] with minimum number of elements. Accuracy and efficiency of compact models, developed by these two methods, are verified on the example of a low-pass microstrip filter. The first two steps in the above mentioned procedure to create the network model of a distributed circuit are subsequently performed by the TLM (Transmission-Line Matrix), electromagnetically based, numerical method in the time-domain [14] and VF method originally introduced in [15-17].

II. CIRCUIT SYNTHESIS METHODS

In this section, two systematic synthesis procedures for the generation of lumped element equivalent circuit models for passive microwave structures are described. Initial data to develop a compact model for the structure under consideration can be generated either by numerical full wave analysis or by measurement. Then, for the impedance data, obtained in this way, an approximation by a rational function is performed. The rational fit is computed at discrete frequencies over the bandwidth of interest. This yields the following closed form expression:

$$Z(s) = E \cdot s + D + \sum_{k=1}^{K} \frac{B_k}{s - s_k} . \qquad (1)$$

A. Extended Foster Equivalent Circuit Synthesis

The Foster realization approach starts with the characterization of the given impedance parameters z_{ij} by partial fraction expansion:

$$z_{ij}(s) = \frac{k_{ij}^{(0)}}{s} + \sum_m \frac{2k_{ij}^{(m)} s}{s^2 + \omega_m^2} + \ldots + k_{ij}^{(\infty)} s \,. \qquad (2)$$

For example, for a two-port device, the Foster impedance realization can be obtained if all *Z* parameters are known:

$$\begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} = \left(\begin{bmatrix} k_{11}^{(0)} & k_{12}^{(0)} \\ k_{21}^{(0)} & k_{22}^{(0)} \end{bmatrix} + \sum_m \begin{bmatrix} k_{11}^{(m)} & k_{12}^{(m)} \\ k_{21}^{(m)} & k_{22}^{(m)} \end{bmatrix} + \begin{bmatrix} k_{11}^{(\infty)} & k_{12}^{(\infty)} \\ k_{21}^{(\infty} & k_{22}^{(\infty)} \end{bmatrix} \right) \cdot f^{(m)} \quad (3)$$

where:

$$f^{(m)} = \begin{cases} \dfrac{1}{s}, & \text{for } m = 0 \\ \dfrac{2s}{s^2 + \omega_m^2}, & \text{for } m = 1,2,\ldots,\mathrm{M} \\ s, & \text{for } m = \infty \end{cases} \qquad (4)$$

and s is complex frequency; $k_{ij}^{(0)}$ and $k_{ij}^{(\infty)}$ are residues of the poles at zero and infinity, respectively; $k_{ij}^{(m)}$ is the residue at the intermediate pole of the complex frequency ω_m and M is a total number of intermediate poles.

Once the Foster expansions have been obtained for each impedance parameter, the poles and residues are used to determine the equivalent circuit component values. For each column in the Foster expansion, there will be a corresponding sub T-network [9]. For the equivalent circuit to be realizable, the transformer turns ratio, $a^{(m)}$, for an arbitrary pole *m* must comply within the following equations

$$\frac{\left|k_{12}^{(m)}\right|}{k_{11}^{(m)}} \le a^{(m)} \le \frac{k_{22}^{(m)}}{\left|k_{12}^{(m)}\right|}, \text{ and } \frac{k_{12}^{m)}}{a^{(m)}} \ge 0 \,. \qquad (5)$$

The equations outlined above are for lossless networks. Therefore, all residues k_{ij} are real and poles are approximated to lie along the imaginary axis. For a lossy structure, the Foster expansion term for an intermediate pole ω_m for all z_{ij} has to be modified as suggested in [9]:

$$\frac{2k_{ij}^{(m)} s}{s^2 + \omega_m^2 + js / B^{(m)}}, \qquad (6)$$

where $B^{(m)}$ in the linear term added to the denominator is obtained directly from the real part of impedance calculated either from EM simulation or measurement.

B. Brune's Equivalent Circuit Synthesis

Equivalent lumped element circuits for general lossy or lossless two-ports, such as considered in this paper, can be obtained from Brune's circuit synthesis procedure [10-13]. Cauer or Foster representations of lossless circuits, explained in subsection *A*, can be extended to lossy circuits but negative elements even for passive circuits would result. Brune synthesis yields the realization of passive circuits with minimum number of elements.

A positive real (PR) character is required for the impedance function to be synthesized. The impedance (or admittance) function is of the form:

$$W = \frac{a_n \cdot s^n + a_{n-1} \cdot s^{n-1} + a_{n-2} \cdot s^{n-2} + \cdots + a_0}{b_n \cdot s^n + b_{n-1} \cdot s^{n-1} + b_{n-2} \cdot s^{n-2} + \cdots + b_0} = \frac{P(s)}{Q(s)} \,. (7)$$

For a PR function, all poles and zeros are located in the left half of the complex frequency plane, or on the imaginary axis respectively. Poles or zeros lying on the imaginary axis can be separated from the rational function without disturbing the function's PR character. In Brune's equivalent network synthesis procedure, the impedance function is analyzed and poles and zeros on the imaginary axis can be separated from the impedance (or admittance) function (7), and can be realized in a subcircuit in a straightforward manner. However, if all poles and zeros are strictly in the left half plane a special so called Brune's process is applied. The global minimum of the real part of the function on the imaginary axis is determined. The value of this global minimum is subtracted from the function. Depending on which frequency this minimum value is found at, we have to extract the real part of the impedance function (for $s = 0$ and $s = \infty$) or, if the minimum occurs at a finite frequency, we have to extract real and imaginary part. Possible subcircuit extractions for this Brune's process are shown in Fig.1.

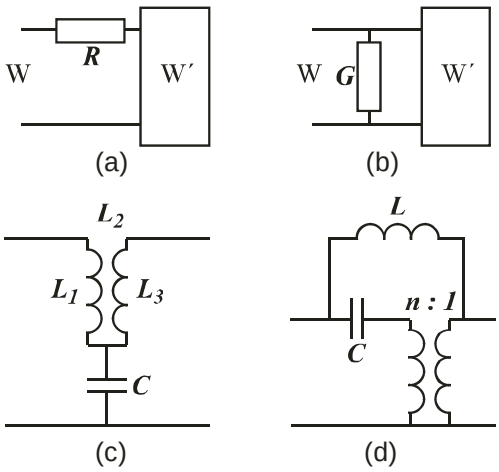


Fig.1. Extracted circuits from the Brune's process

III. NUMERICAL ANALYSIS

For the numerical study, we consider a low-pass microstrip filter, shown in Fig.2, to demonstrate the two methods for synthesis of compact lumped element models for linear lossy reciprocal two-port devices described above.

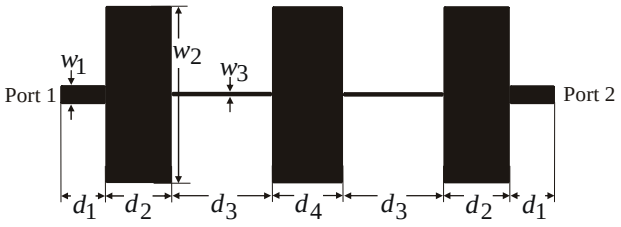


Fig.2. Layout of low-pass microstrip filter

The physical dimensions of symmetrical low-pass filter in millimeters are: line widths w_1 = 0.217054, w_2 = 2.31921 and w_3 = 0.0248336; line lengths d_1 = 0.566318, d_2 = 0.84057, d_3 =1.29201 and d_4 = 0.901333. The substrate height is h = 0.2 mm and its relative permittivity is ε_r =12.9.

The full wave EM analysis results are obtained from TLM simulations. In order to generate the equivalent circuit, impedance parameters of the full-wave analysis have to be de-embedded. Compact lumped element models obtained by Foster and Brune's synthesis procedure are shown in Figs.3 and 4, respectively. Before applying Brune's method, symmetric two-port device from Fig.2 is first transformed into a connection of one-ports using Bartlett's theorem. These one-ports are then synthesized by Brune's process which yields a minimum number of elements.

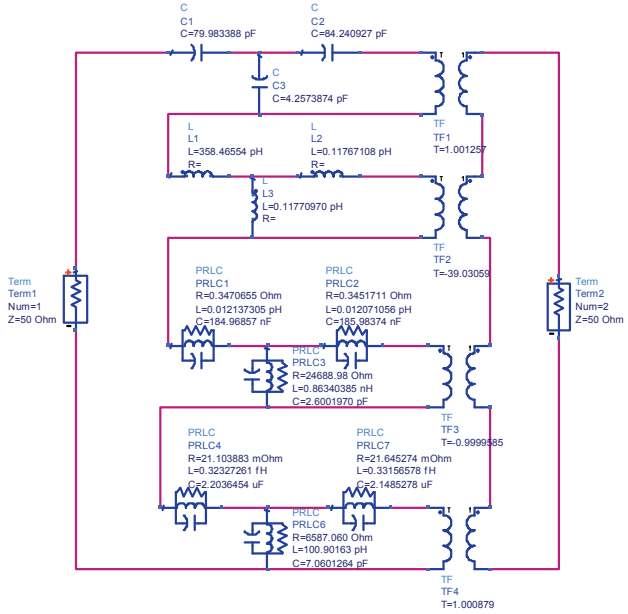
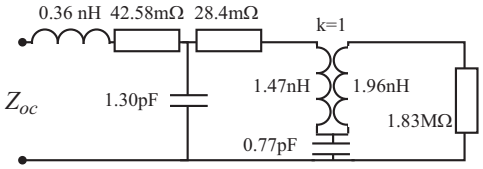
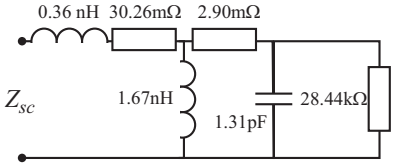


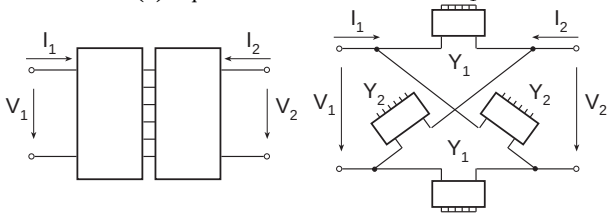
Fig.3. Equivalent Foster circuit



(a) Equivalent Brune's circuit of Y_2



(b) Equivalent Brune's circuit of Y_1



(c) Bartlett's transformation of symmetric two-port network

Fig.4. Equivalent Brune's circuit

Figs.5-8 compare the synthesized equivalent Foster and Brune's circuit scattering matrix results to the results obtained directly from EM simulation.

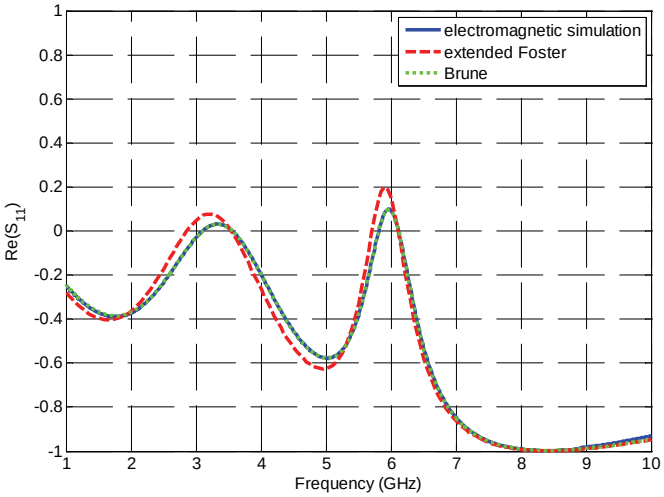


Fig.5. Real part of S_{11} parameter

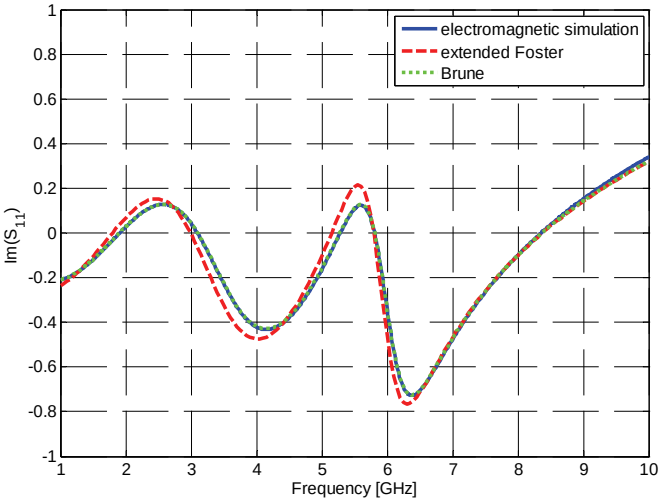


Fig.6. Imaginary part of S_{11} parameter

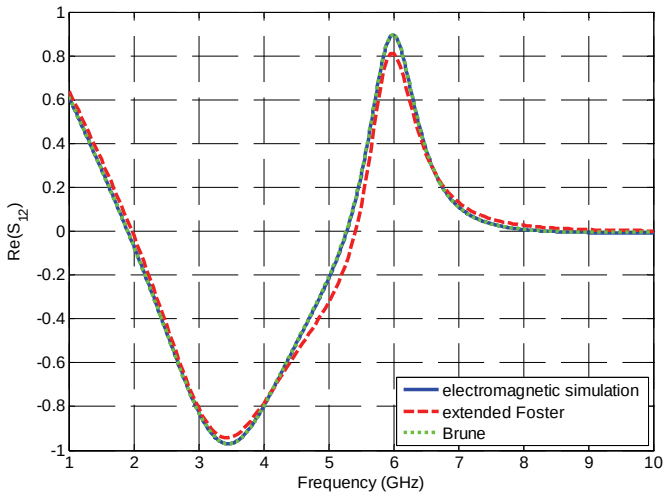


Fig.7. Real part of S_{21} parameter

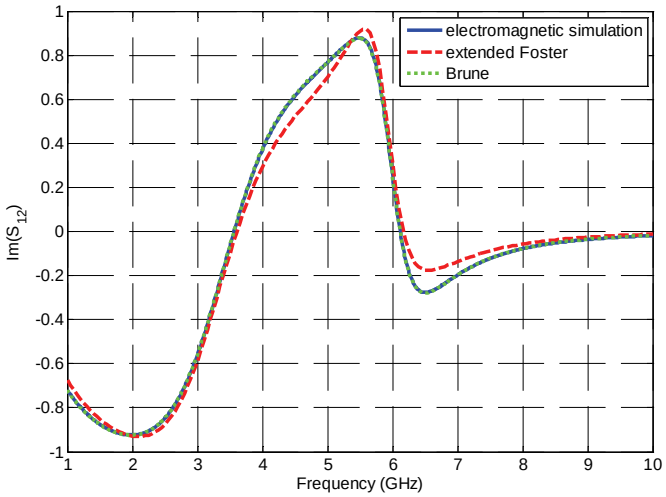


Fig.8. Imaginary part of S_{21} parameter

IV. Conclusion

Two methods for synthesis of compact lumped element models for linear lossy reciprocal multiports are discussed in the paper. Brune's method has provided a better agreement approximation of the EM simulation data, it is more suitable for the lossy structures as it provides only positive lumped elements in the equivalent circuit. However, it is applicable at the moment to the one-port device; symmetrical two-port devices can be transformed into one-ports by using Bartlett's theory. In the extended Foster approach, the agreement of the approximation to the EM data was slightly reduced, and this approach could result in negative elements in the case of lossy structures, but it can be apllied easily to multiport devices.

Acknowledgement

The authors thank the Deutschem Akademischem Austauschdienst (DAAD) and the Serbian Ministry for Education and Science for their support of this work. Also, this work has been partially supported by the Serbian Ministry for Education and Science under the project TR-32052.

References

[1] P. Russer, *Electromagnetics, Microwave Circuit and Antenna Design for Communications Engineering*, 2nd edition, Boston: Artech House, 2006.

[2] L.B. Felsen, M. Mongiardo, and P. Russer, *Electromagnetic Field Computational by Network Methods*, Berlin, Germany: Springer, 2009.

[3] T. Mangold, and P. Russer, "Full-wave Modeling and Automatic Equivalent-circuit Generation of Millimeter-wave Planar and Multilayer Structures", IEEE Transactions on Microwave Theory and Techniques, vol. 47, no. 6, pp. 851–858, 1999.

[4] E.A. Guillemin, *Synthesis of Passive Networks*, New York: Wiley, 1957.

[5] L. Ljung, *System Identification, Theory for the User*, Upper Saddle River, NJ: Prentice Hall PTR, 1999.

[6] P. Russer, M. Righi, C. Eswarappa, and W. Hoefer, "Lumped Element Equivalent Circuit Parameter Extraction of Distributed Microwave Circuits via TLM Simulation", IEEE International Microwave Symposium Digest, vol. 2, pp. 887-890, 1994.

[7] M. Righi, C. Eswarappa, W. Hoefer, and P. Russer, "An Alternative Way of Computing S-parameters via Impulsive TLM Analysis without using Absorbing Boundary Conditions", International Microwave Symposium Digest, Orlando, pp. 1203-1206, May 1995.

[8] C. Eswarappa, and W.J.R. Hoefer, "Fast s-parameter Computation of a Microstrip Interdigital Filter using TLM, Prony's and Digital Filtering Techniques", International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, vol. 9, no. 3, pp. 237-248, 1996.

[9] I. Timmins, Ke-Li Wu, "An Efficient Systematic Approach to Model Extraction for Passive Microwave Circuits", IEEE Transactions on Microwave Theory and Techniques, vol. 48, no. 9, 2000.

[10] O. Brune, "Synthesis of a Finite Two Terminal Network whose Driving Impedance is a Prescribed Function of Frequency", J. Math. And Phys. vol. 10, no. 3, pp. 191-236, 1931.

[11] Y. Kuznetsov, A. Baev, T. Shevgunov. U. Siart, H. Yordanov, and P. Russer, "Generation of Network Models for Planar Microwave Circuits by System Identification Methods", in Proceeedings of International Conference on Electromagnetics in Advanced Applications – ICEAA'09, pp. 966-969.

[12] V. Belevitch, "On the Brune Process for n-ports", IRE Transactions on Circuit Theory, vol. 7, no. 3, pp. 280-296, 1960.

[13] F. Mukhtar, Y. Kuznetsov, C. Hoffmann, P. Russer, "Brune's synthesis of linear lossy distributed one-port and symmetric two-port microwave circuits," German Microwave Conference (GeMIC), 2011, pp.1-4, 14-16 March 2011.

[14] C. Christopoulos, *The Transmission-Line Modelling (TLM) Method in Electromagnetics*, Publication in the Morgan & Claypool Publishers series, 2006.

[15] B. Gustavsen and A. Semlyen, "Rational Approximation of Frequency Domain Responses by Vector Fitting", IEEE Transactions on Power Delivery, vol. 14, no. 3, pp. 1052-1061, 1999.

[16] B. Gustavsen, "Improving The Pole Relocating Properties of Vector Fitting", IEEE Transactions on Power Delivery, vol. 21, no. 3, pp. 1587-1592, July 2006.

[17] D. Deschrijver, M. Mrozowski, T. Dhaene, and D. De Zutter, "Macromodeling of Multiport Systems using a Fast Implementation of the Vector Fitting Method", IEEE Microwave and Wireless Components Letters, vol. 18, no. 6, pp. 383-385, June 2008.

Synthesis of Microwave Filters by Coupling Matrix Optimization

Marin V. Nedelchev, Ilia G.Iliev

Abstract - **This paper presents optimization method for synthesis of generalized microwave filters with arbitrary topology. The method utilizes Nelder-Mead local optimizer for coupling matrix determination. The synthesis procedure converges very fast as for a initial point is used a vector based on the Chebyshev all pole filter for the same degree of the filter. To validate the proposed synthesis method two numerical examples for resonant filters are computed. The frequency responses from the synthesis procedure and the theoretical responses show excellent agreement.**

Keywords - **microwave filter, Chebyshev filter, Nelder-Mead optimization, coupling matrix.**

I. Introduction

Microwave coupled resonator filters play important role in the modern communication systems. The constraint RF/microwave spectrum requires high attenuation in the stop band and low insertion loss in the passband of the filters. These requirements can be met only by cross-coupled microwave filters, realizing attenuation poles on finite frequencies. More over cross coupled filters can exhibit flat group delay, when realizing complex conjugate transmission zeroes. Cross-coupled resonator filters allow using various topologies with variety of frequency responses.

The microwave filter modelling is very important for the fast and accurate design.

In the early 1970's started the development of the theory of cross-coupled resonator filters by Atia and Williams in their basic paper [1]. Cameron extended the theory to general cross-coupled Chebyshev filtering functions synthesis in the papers [2, 3]. The synthesis procedure continues with deriving the transversal coupling matrix from the Chebyshev polynomials. Key point in the obtaining of the coupling matrix corresponding to the practical filter topology is to convert transversal form to folded form using matrix rotations. Folded form of the coupling matrix is starting point for matrix rotation sequences to derive the final coupling matrix. Most of the matrix rotation sequences are given in [4]. It is noticed that this method for synthesis suffers from generality, because the matrix rotations cannot be derived for every one practical filter topology. Some of the matrix rotation sequences cannot converge in order to find the coupling matrix. Some of the disadvantages in this method are solved if arrow form of the coupling matrix is used [5] or Pfitzenmeir method is used [6].

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-mails: mnedelchev@tu-sofia.bg, igiliev@tu-sofia.bg

In many practical cases, it is necessary to define the filter topology in order to satisfy some manufacturing or space requirements. In this case, the exact solution is hard to be found with the conventional synthesis methods.

Many commercially available design packages offer direct optimization of the physical dimensions of the filter topology. This method for synthesis cannot converge in the general case, because there is no "general" optimization method that is suitable for the optimization problem. If a local minimum is found, the sensitivity of this point is unknown. Starting from different points, the optimization process can lead us to different local minimum of the cost function. One of the solutions may be a global minimum, but there is no a priori guarantee for finding it. Some of the solutions may have very big sensitivities toward the manufacturing tolerances, temperature, or could not be realized in the practice. Therefore, the direct synthesis over the geometrical dimensions of the filter is not a good decision for general design method.

One possible general solution to the filter design for arbitrary topology is to apply direct local optimization over the coupling matrix with successive starting point. In the basic papers proposed optimization method for coupling matrix synthesis [7, 8], the starting vector is set to arbitrary values. This makes the local optimization very unstable method for cost function minimization. Another method is to use global optimization method for finding the coupling matrix for certain filter topology. They perform robust optimization, no matter about the starting point. Unfortunately the global optimizers such as genetic or stochastic have very slow convergence to the cost function minimum.

This paper presents optimization method for synthesis of microwave filters with arbitrary topology. The method uses Nelder-Mead local optimizer for coupling matrix determination. The synthesis procedure converges very fast as for a initial point is used a vector based on the Chebyshev all pole filter for the same degree of the filter. The cost function is based on amplitude of the transmission and reflection coefficient zeros and their values at the cut-off frequencies. To validate the proposed synthesis method two resonant filters are designed with asymmetrical responses. The frequency responses from the synthesis procedure and the theoretical responses show excellent agreement.

II. Resonator Filter Characteristics

The synthesis procedure starts with the low-pass prototype with normalized angular frequency of passband $\omega = 1$. The transfer and reflection coefficients may be expressed as a ratio of two N-th degree polynomials as follows:

$$S_{21} = \frac{P_N(\omega)}{E_N(\omega)},$$

$$S_{11} = \frac{D_N(\omega)}{\varepsilon E_N(\omega)}, \quad (1)$$

where ω is real angular frequency and $\varepsilon = \left(1/\sqrt{10^{RL/10}-1}\right).\left(D_N(\omega)/P_N(\omega)\right)\Big|_{\omega=1}$, RL is the prescribed value of the return loss in dB, in the passband of the filter. It is assumed that all polynomials are normalized to their highest degree coefficient. The reflection and the transfer coefficients must satisfy the unitary conditions of the scattering matrix.

$$S_{11}S_{11}^* + S_{21}S_{21}^* = 1. \quad (2)$$

It can be easily found that the transfer coefficient may be expressed in the following way:

$$S_{21}^2(\omega) = \frac{1}{1+\varepsilon^2 C_N^2(\omega)} \quad (3)$$

where $C_N(\omega) = P_N(\omega)/D_N(\omega)$ is the filtering function. For general Chebyshev characteristics, the filtering function is in the form:

$$C_N(\omega) = \cosh\left(\sum_{n=1}^{N} a\cosh(x_n)\right), \quad (4)$$

where $x_n = \frac{\omega - 1/\omega_n}{1-\omega/\omega_n}$, where ω_n is the angular frequency of the prescribed transmission zero.

In order to obtain the coupling matrix, it is necessary to consider the equivalent circuit of general coupled resonator filter shown on Fig.1.

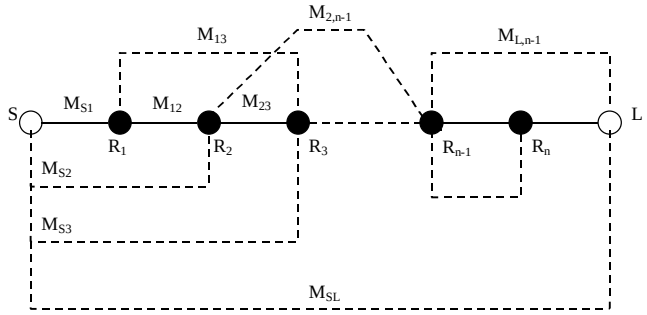


Fig.1. General coupled resonator filter

The equivalent circuit consists of N series coupled resonators with frequency independent couplings M_{ij} ($i \neq j$), between the *i-th* and *j-th* resonators. The circuit is driven by voltage source E with internal normalized resistance $R_1 = 1$ and loaded to normalized impedance $R_2 = 1$. There exists coupling between the Source and Load to perform all N finite frequency transmission zeroes. The resonant frequency of each resonator f_{0i} is represented by the self-coupling coefficient M_{ii} and the center frequency of the filter. They must satisfy the equation:

$$f_{0i}^2 - M_{ii} f_0 f_{0i} - f_0^2 = 0. \quad (5)$$

The transmission and reflection coefficients of a lossless filter of N-th order depend only of the coupling matrix $|M|$ (7):

$$S_{21} = -2j[A]^{-1}_{N+2,1},$$

$$S_{11} = 1 + 2j[A]^{-1}_{11}, \quad (6)$$

where $[A] = -j[R] + \omega[W] + [M]$, a $[R]$ e $(N+2)\times(N+2)$ matrix, which elements are zeroes except $R_{11} = R_{N+2,N+2} = 1$. $[W]$ is a $(N+2)\times(N+2)$ matrix, where the main diagonal elements are unity except $W_{11} = W_{N+2,N+2} = 0$. All remaining elements of $[W]$ are zeroes. $[M]$ is the coupling matrix, symmetrical around the main diagonal.

III. Synthesis of Microwave Filter with Coupling Matrix Optimization

The cost function used in the optimization process is based on the zeroes and poles of the filtering function C_N, assuming that the number of poles is P and zeroes N [8]:

$$Cost = \sum_{i=1}^{N}\left|S_{11}(\omega_{zi})\right|^2 + \sum_{i=1}^{P}\left|S_{21}(\omega_{pi})\right|^2 + \left(\left|S_{11}(\omega=-1)\right| - \frac{\varepsilon}{\sqrt{\varepsilon^2+1}}\right)^2 + \left(\left|S_{11}(\omega=1)\right| - \frac{\varepsilon}{\sqrt{\varepsilon^2+1}}\right)^2 \quad (7).$$

This cost function requires less computational efforts than using the theoretically derived S_{21} and S_{11} for pattern search.

In this way it is possible to formulate the local optimization problem for obtaining the coupling matrix.

The advantages of this method for synthesis are:

1. Design of filter with prescribed transmission zeros with symmetric or asymmetric response.
2. Design of filter with arbitrary topology, even or odd order.
3. Possibility of constraints for the magnitude and sign of the coupling coefficients if a given realization is intended.
4. Elimination of the similarity transformations for the coupling matrix and the extraction technique. There is no possibility for calculation errors or round off errors.

The main disadvantages for the optimization method are:

1. Exact solution is not guaranteed, especially for great number of variables.
2. The filter topology must be able to realize the desired filter response. Then a local minimum is reached by the optimization process.
3. If the initial guess is arbitrary, the global minimum cannot be reached in every filter design.
4. Some of the elements of the coupling matrix, derived in the optimization process may be impossible to realize.

The initial guess for the coupling matrix is very important for the reaching of the global minimum of the cost function (7). Having on mind that a local optimizer is used, the starting vector should be close to the target value in order to assure a fast convergence of the method. One of the possible starting coupling matrices is to set all self-coupling

couplings to zero ($M_{ii} = 0$) and all direct couplings to 1. The cross-coupling coefficients are all set to zero.

The second possible starting coupling matrix is to use classical Chebyshev filter from the same order. All self- and cross-couplings are set to zero. In order to find out which starting point is more computational efficient some numerical designs are investigated.

III. NUMERICAL RESULTS

For verification of the optimization method presented in this paper, it is applied to an asymmetric resonator filters.

A. Asymmetric Three Resonator Passband Filter

This filter is of Chebyshev type and it has return loss more than 20dB in the passband and rejection in the higher stopband better than -15dB. The transmission coefficient zero is placed on normalized frequency $\omega_p = 1.6$. The coupling scheme of the filter is shown on Fig.2.

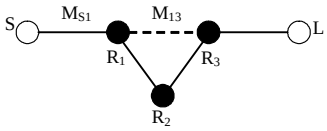


Fig.2. Coupling scheme of an asymmetric three pole filter

The reflection and transmission zeroes calculated and summarized in Table 1.

Table 1. Poles and zeros of asymmetric three resonator filter

№	Reflection zeros	Transmission zeros	Poles
1	*-j0.8061*	*j1.56*	*-0.9542-1.4447i*
2	*j0.9257*	*infinite*	*-1.1781+ 0.5923i*
3	*j0.2430*	*infinite*	*-0.2239 + 1.2150i*

The initial point for the coupling matrix elements for the optimization procedure is to set the values of the all pole three resonator Chebyshev filter $M_{S1} = 1.0825$, $M_{12} = M_{23} = 1.0303$.

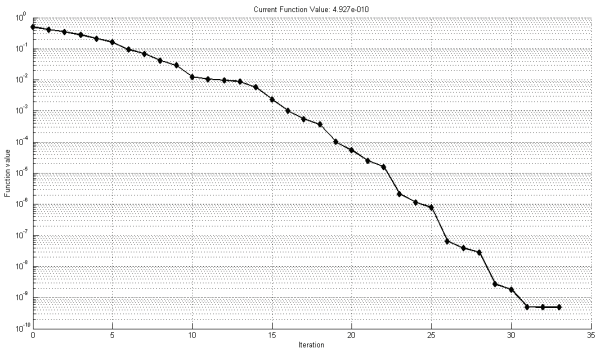


Fig.3 Cost function value for asymmetric three resonator filter

All self coupling and cross coupling coefficients are set to zero. The number of the independent values of the coupling matrix is 7.

After 33 iterations for the optimization coefficient, the procedure converges. The values of the cost function vs the number of iterations is shown on Fig.3. The initial value of the cost function is 0.335 and the end value is $4.927.10^{-10}$. The optimization process stopped because of reaching local minimum of the cost function (7). The final coupling matrix is:

$$M = \begin{bmatrix} 0 & 1.0866 & 0 & 0 & 0 \\ 1.0866 & 0.1741 & 0.8076 & 0.7679 & 0 \\ 0 & 0.8076 & -0.7108 & 0.8076 & 0 \\ 0 & 0.7679 & 0.8076 & 0.1741 & 1.0866 \\ 0 & 0 & 0 & 1.0866 & 0 \end{bmatrix}. \quad (8)$$

The frequency response of the designed filter is shown on Fig.4. It is calculated by the derived in the optimization process coupling matrix (8) and (6). It is clearly seen that the normalized cut off frequency is $\omega_c = \pm 1$, while the transmission zero frequency is $\omega_p = -1.56$. The maximum value of the return loss is with the prescribed value of *-20dB.*

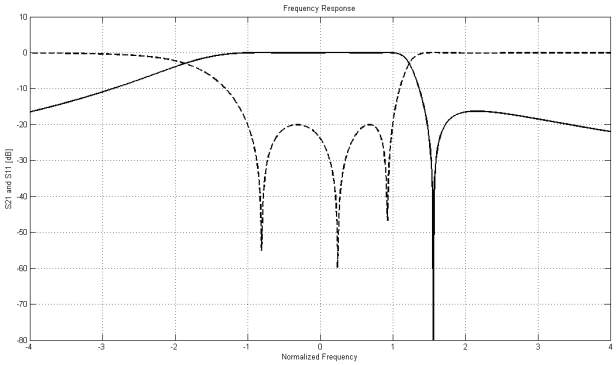


Fig.4 Frequency response of three resonator filter with asymmetric response. Solid line-S_{21}, dashed line- S_{11}

B. Asymmetric Five Resonator Passband Filter

The five resonator filter is formed by two cascaded trisections, each one performing one transmission zero. The filter is Chebyshev type and it has maximum return loss of -20dB. The transmission zeroes are placed on -2.3 and 1.6.

The coupling scheme of the filter is shown on Fig.5.

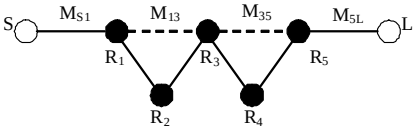


Fig.5. Coupling scheme of an asymmetric five pole filter

Both trisections share a common resonator in order to reduce the filter order. The roots of the polynomials in the numerator and denominator in (1) are shown in Table2.

Table 2. Poles and zeros of asymmetric five resonator filter

№	Reflection zeros	Transmission zeros	Poles
1	*-j0.9522*	*j1.6*	*-0.1838 - 1.1266j*
2	*-j0.5808*	*-j2.3*	*-0.5407 - 0.7163j*
3	*j0.0481*	*infinite*	*-0.6933 + 0.0655j*
4	*j0.6452*	*infinite*	*-0.1377 + 1.1036j*
5	*j0.9619*	*infinite*	*-0.4742 + 0.7961j*

The starting point for the optimization process is based on the Chebyshev coupling matrix elements $M_{S1} = 1.0137$,

$M_{12} = M_{45} = 0.8653$, $M_{23} = M_{34} = 0.6357$. The number of the independent values of the coupling matrix is 12. The optimization process converges very fast in 77 iterations of the optimizer with end cost function value $2.5953.10^{-12}$. The trials with setting of all main couplings to ones and all cross and self-couplings to zero as an initial point lead the optimizer to a local minimum with cost function value of order 1.10^{-3}. Fig. 5 shows the cost function value with respect to the iterations.

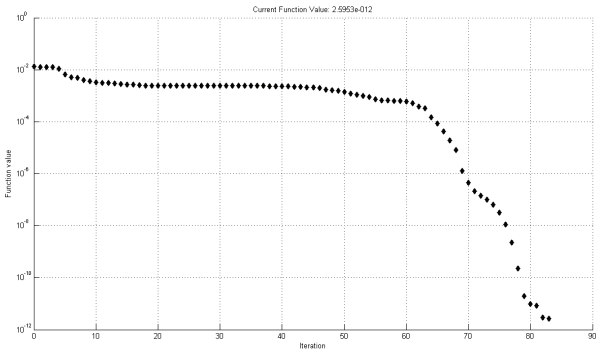


Fig.5 Cost function value for asymmetric five resonator filter

The coupling matrix derived in the optimization process is:

$$M = \begin{bmatrix} 0 & 1.0081 & 0 & 0 & 0 & 0 & 0 \\ 1.0081 & 0.0097 & 0.7590 & 0.3904 & 0 & 0 & 0 \\ 0 & 0.7590 & -0.5480 & 0.5411 & 0 & 0 & 0 \\ 0 & 0.3904 & 0.5411 & 0.0264 & 0.5952 & -0.2527 & 0 \\ 0 & 0 & 0 & 0.5952 & 0.3801 & 0.8152 & 0 \\ 0 & 0 & 0 & -0.2527 & 0.8152 & 0.0097 & 1.0081 \\ 0 & 0 & 0 & 0 & 0 & 1.0081 & 0 \end{bmatrix}$$

The corresponding frequency response calculated by the coupling matrix and Eq.(6) is shown on Fig.6.

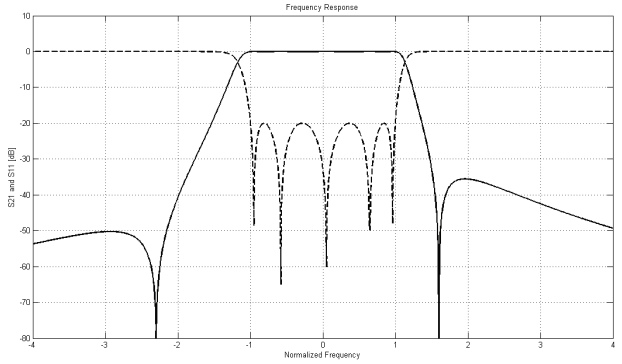


Fig.6 Frequency response of five resonator filter with asymmetric response. Solid line-S_{21}, dashed line- S_{11}

As it is clearly seen from Fig.6, the transmission zeros are placed on the prescribed values of -2.3 and 1.6. The maximum value of the reflection coefficient is -20dB.

Both presented examples show fast convergence of the cost function to a local minimum. In both cases this local minimum is found to be a global minimum corresponding to general Chebyshev filter. In both cases the starting point for the optimization process was the coupling matrix of classic Chebyshev filter. All the cross couplings and self-couplings were set to zero. During the tests of the optimization function, was experienced with another set of starting points. All the main couplings were set to ones and the rest of the couplings were set to zero. In this case the optimization process went to a minimum that was found to be local, but not a global one.

IV. CONCLUSION

This paper presents optimization method for synthesis of microwave filters with arbitrary topology. The method uses Nelder-Mead local optimizer for coupling matrix determination. The synthesis procedure converges very fast as for an initial point is used a vector based on the Chebyshev all pole filter for the same degree of the filter. To validate the proposed synthesis method two resonant filters are designed with asymmetrical responses. Both presented examples show fast convergence of the cost function to a local minimum. In both cases this local minimum is found to be a global minimum corresponding to general Chebyshev filter. The frequency responses from the synthesis procedure are within the expectations and found to be consistent with the theoretical responses and given filter specifications.

REFERENCES

[1] A.E. Atia and A.E. Williams. "Narrow-Bandpass Waveguide Filters." 1972 Trans. on Microwave Theory and Techniques 20.4 (Apr. 1972 [T-MTT]): 258-265.

[2] Cameron, R., Advanced Coupling Matrix Synthesis Techniques for Microwave Filters, IEEE Trans on MTT-50, Jan.2003, pp.1-10.

[3] Cameron, R.J., General Coupling Matrix Synthesis Methods for Chebyshev Filtering Functions, IEEE Trans. on MTT, April 1999, pp.433-442.

[4] Rhodes, J.D., The Design and Synthesis of a Class of Microwave Bandpass Linear Phase Filters, IEEE Trans. on MTT, 1969 pp.189-204.

[5] Macchiarella, G, An Analytical Technique for the Synthesis of Cascaded N-Tuplets Cross-Coupled Resonators Microwave Filters Using Matrix Rotations, IEEE Trans. on MTT, May 2005, pp.1693-1698.

[6] G. Pfitzenmaier, "Synthesis and Realization of Narrow-band Canonical Microwave Bandpass Filters Exhibiting Linear Phase and Transmission Zeros," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-30, pp.1300–1311, Sep. 1982.

[7] Atia W.A., K.A. Zaki and A.E. Atia. "Synthesis of general topology multiple coupled resonator filters by optimization." 1998 MTT-S International Microwave Symposium Digest 98.2 (1998Vol. II MWSYM]): 821-824.

[8] Amari, S., Synthesis of Cross-Coupled Resonator Filters Using an Analytical Gradient-Based Optimization Technique, IEEE Trans on MTT Sept. 2000, pp.1559-1564.

Synthesis of Microstrip Filters Using Miniaturized Pentagonal Resonators

Marin V. Nedelchev[1]

Abstract: This paper presents a study of miniaturized pentagonal microstrip resonators and their application in filter design. Due to their shape, the pentagonal resonators have many coupling topologies in order to achieve couplings of different nature- electric, magnetic, mixed. Each coupling topology is analyzed in full wave electromagnetic (EM) simulator in order to estimate the resonance peaks in the frequency response. Based on the simulation results, coupling coefficient graphs are presented. These design graphs are used in microstrip filter design. Third order microstrip filter for application in GSM900 system is synthesized, simulated, manufactured and measured. There is a very good agreement between the simulated and measured results.

Keywords:, miniaturized pentagonal resonator, coupling coefficient, GSM900.

I. INTRODUCTION

The fast development of the mobile communication systems stimulates the research of microwave filters with symmetrical response [1]. Microstrip filters are preferred for these systems, because of their compact size, low weight, easy integration in integrated circuits, fine adjustment.

Most of the microwave filters are of Chebyshev type. They are equiripple in the passband and maximally flat in the stopband. Such filters can be realized by cascading resonators in series. High filter selectivity requires higher filter order and more resonators. Because of the low unloaded Q factor of the microstrip resonators, the passband loss increases. Alternative way is to use cross-coupled filters with coupling between non-adjacent resonators. Non-adjacent couplings cause transmission zeroes in the stopband or equalization in the group delay.

Among the variety of filter topologies, the classic half-wavelength and hairpin resonator filters are commonly used. Miniaturization is an important requirement for the used resonators. In order to reduce the size of the half wavelength resonator the authors of [2] fold back the ends of the resonator into a "U" shape. The further miniaturization of half wavelength is achieved by the square open loop filters [2]. The resonator is bent in square form. In order to achieve more flexible filter design, the authors of [3] proposed hexagonal structure of the resonator. It inherits the features of the halfwavelength resonator-position of the spur frequency, input impedance and slope parameter.

The authors of [4] propose to use pentagonal open loop resonators in order to achieve compact cascaded quadruplet filter with two symmetrical transmission zeros. The geometrical form of pentagonal resonators allows coupling indifferent coupling schemes, which exhibits various coupling types.

[1]Marin V. Nedelchev – Assistant, PhD in Dept. of Radiocommunication and Videotechnology in Faculty of Telecommunication in TU –Sofia E-mail mnedelchev@tu-sofia.bg

The proposition of cross coupled filters with symmetrical pair of transmission zeros come from the classic papers of Kurzrok [5]. The synthesis of cross coupled filters is based on the early works from Atia and Williams [1], Cameron and Rhodes [6,7] considering waveguide cavity filter design. It is based on the deriving the coupling matrix from the transfer function and its reduction to the corresponding topology form. This technique is found to be useful in the design of microstrip cross-coupled filters. Hong and Lancaster proposed in several papers [2-4], numerical method for cross coupled filter design based on approximation of the low pass filter prototype elements.

This paper proposes miniaturized pentagonal microstrip resonators and a design procedure of all pole Chebyshev type filter. A full wave EM simulator is used for obtaining the coupling coefficients between the resonators. Numerical results for the values of the coupling coefficients are presented. A design example for GSM900 system is manufactured and measured. The frequency responses from full wave EM simulator and measurements are presented and show very good agreement.

II. MINIATURIZED PENTAGONAL RESONATOR COUPLING STRUCTURES

The miniaturized pentagonal resonator is based on the miniaturized resonator proposed in [2,4]. The form of the resonator is shown on Fig.1. Its form helps using the resonator in bigger variety of coupling schemes than using rectangular or square miniaturized resonator. The resonant characteristics of the pentagonal resonator are the same as the characteristics of the rectangular miniaturized resonator.

Fig.1. Topology of pentagonal miniaturized microstrip resonator

The coupling mechanism is based on the fringe fields of closely situated resonators. The nature of coupling depends on the resonator configuration. The coupling coefficient for synchronously tuned resonators can be calculated easily by the resonance frequencies of even and odd mode [8]:

$$k = \frac{f_e^2 - f_o^2}{f_e^2 + f_o^2} \quad (1)$$

The necessary condition for observing these resonance peaks is to set the resonator structure in overcoupled mode. In this case the coupling coefficient is larger than the critical coupling value of $1/Q$, where Q is the quality factor of the resonators [6]. A full wave EM simulator based on the Method of the Moments (MoM) is used to identify the

resonance frequencies. Most of the coupling structures are simulated using the symmetry in their topology for electrical and magnetic wall introducing in-between. When symmetry does not present, the whole structure is simulated. This does not constrain in any way the obtained results. The simulation are carried out for standard FR-4 substrate with $\varepsilon_r = 4.4$, $h = 1.5mm$, $tg\delta = 0.02$.

A. Magnetic Coupling

Both resonators are arranged according to Fig.2. In this case magnetic coupling exists. There presents a line of symmetry and the point **A** has zero potential for the resonant frequency. In this point the electrical field has minimum, but the magnetic field has maximum. This determines the magnetic character of the coupling. The mutual inductance between both coupled lines characterizes the magnetic coupling. The mutual capacitance of coupled lines is negligible because of the minimum of the amplitude of electric field.

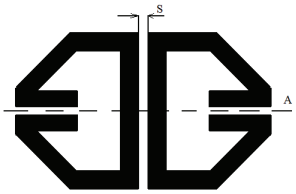


Fig.2. Topology of magnetic coupled pentagonal miniaturized resonators

The coupling coefficient when magnetic coupling presents is with positive sign, because the even mode frequency is higher than the odd mode resonant frequency.

Fig.3. shows the results derived from the full wave EM simulations for the magnetic coupling coefficient dependence with respect to the distance between the coupled lines –***s***.

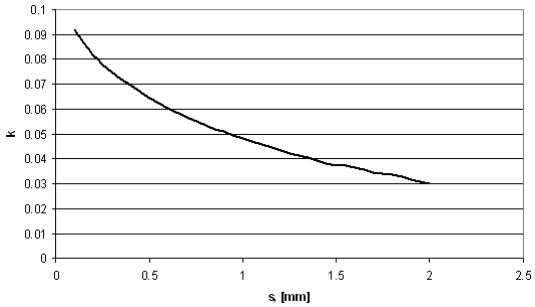


Fig.3 Magnetic coupling coefficient dependence with the distance ***s*** between the coupled lines.

The geometric parameters of the pentagonal resonator are as follows- arm length $l = 13mm$, width of the main transmission line $w = 2.8mm$, width of the symmetrical coupled lines $w_1 = 3.1mm$, distance between the coupled lines $s = 0.3mm$.

When analyse overcoupled resonators, two resonance peaks for even and odd mode are observed in the response. Both frequencies are read and the coupling coefficient is calculated according to Eq.1. The dependence of the coupling coefficient in respect to the distance between the coupled lines is shown on Fig.3.

The topology of magnetic coupled resonators is applicable in realization of positive coupling coefficients in classical or cross-coupled filters.

B. Electric Coupling

The topology of electric coupling between pentagonal microstrip resonators is shown on Fig.4

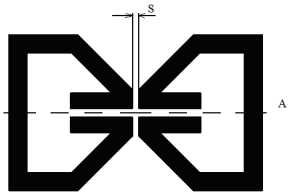


Fig.4. Topology of magnetic coupled pentagonal miniaturized resonators

The nature of coupling is assumed as electrical because the maximum of the electrical field is in the middle of the resonator. In this point the electrical field is predominant over the magnetic field for the resonant frequency. The coupling is defined by the mutual capacitance. The mutual inductance is negligible because of the minimum of the magnetic field. Fullwave EM simulations are carried out over the electrically coupled resonators and the coupling coefficient is computed according to Eq.1. The dependence of the coupling coefficient with respect to the distance between the coupled lines is shown on Fig.5.

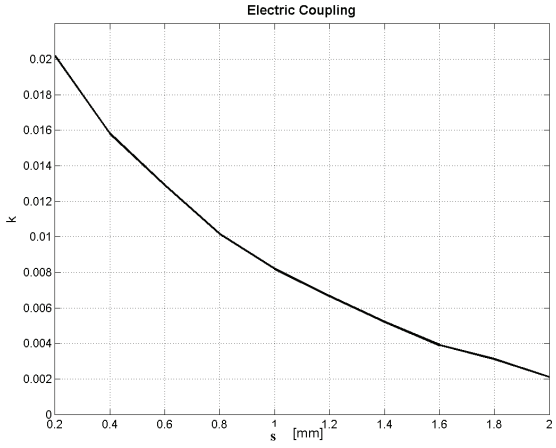


Fig.5 Electrical coupling coefficient dependence with the distance ***s*** between the coupled lines.

The electric coupling coefficient is with negative sign. The coupling between the non-adjacent resonators should be out-of-phase the other couplings. In the case of pentagonal coupled resonators, the coupling is not strong enough. This leads to transmission zeros in the frequency response away from the passband.

The application of the electrically coupled resonators is constrained to realization of negative coupling coefficients for cross-coupled filters.

C. Mixed Coupling

The mixed coupling presents in the coupling structure shown on Fig.6. The currents in the coupled lines are equal in amplitudes, and in-phase.

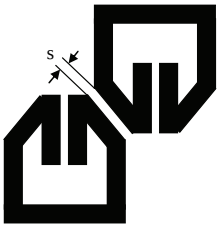


Fig.6. Topology of mixed coupled pentagonal miniaturized resonators

This fact supposes high value of the coupling coefficient. It cannot be estimated which component of the field- electric or magnetic is predominant in this coupling structure. However, the electrical filed decays more rapidly with the distance from the open ends, than the magnetic field. The value of the coupling coefficient is with positive sign.

The coupling structure is simulated in full wave EM simulator according to the above described method and the resonance peaks' frequencies are obtained. The graphical results are presented on Fig.7.

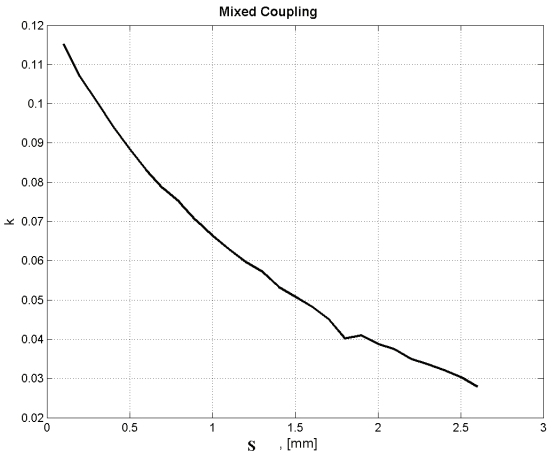


Fig.7 Mixed coupling for the structure shown on Fig.5.

III. Simulation and Measurement Results

In order to prove the applicability of the proposed pentagonal miniaturized microstrip resonators, a third order classical Chebyshev filter is designed, manufactured and measured.

The filter is realized on the FR-4 substrate with the following parameters:

- Relative dielectric permeability $\varepsilon_r = 4.4$;
- Substrate height: $h = 1.5mm$
- Copper foil thickness $t = 17.5\mu m$;
- Dielectric loss tangent: $tg\delta = 0.02$.

The center frequency of the filter is $f_0 = 902.5MHz$, the bandwith is $\Delta f = 75MHz$ with maximum return loss in the passband $RL = -20dB$.

Fig.8 shows the structure schematic of the filter consists of three cascaded resonators.

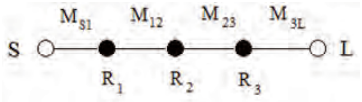


Fig.8. Structure schematic of cascaded three resonator filter. R_i are the resonators, *S*- Source, *L*-load, M_{ij} – the coupling coefficients

For the realization of the parameters of the filter, miniature pentagonal microstrip resonators are utilized. Their topology is very convenient for realizing the form of the filter shown on Fig.9. The geometric dimensions of the resonators are: length of the arm $l = 13mm$, width of the main line $w = 2.8mm$, which corresponds to characteristic impedance $Z_c = 50\Omega$, width of the symmetrical coupled lines $w_1 = 3.1mm$, distance between the coupled lines $d = 0.3mm$.

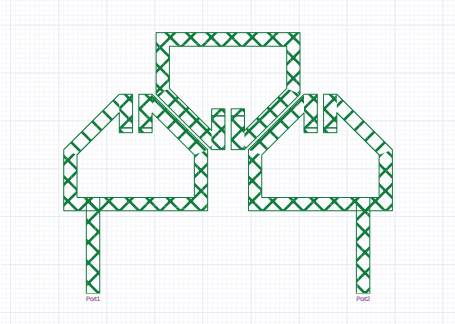


Fig.9. Topology of three resonator microstrip filter with pentagonal resonators.

The topology of the filter requires mixed coupling between the resonators. The coupling coefficient is $M_{12} = M_{23} = 0.081$.The distance between the coupled lines is found to be $s = 0.75mm$, according to the results shown on Fig.6.The results of the fullwave EM simulation of the designed filter are shown on Fig.10.

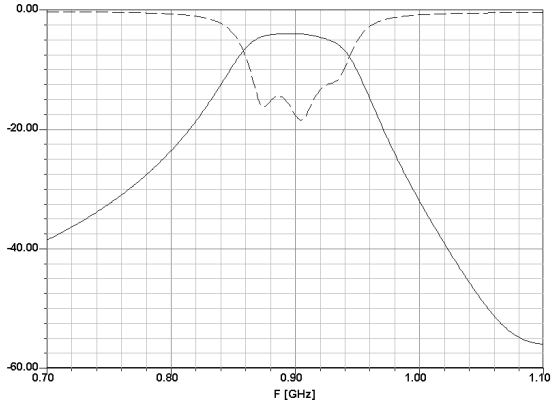
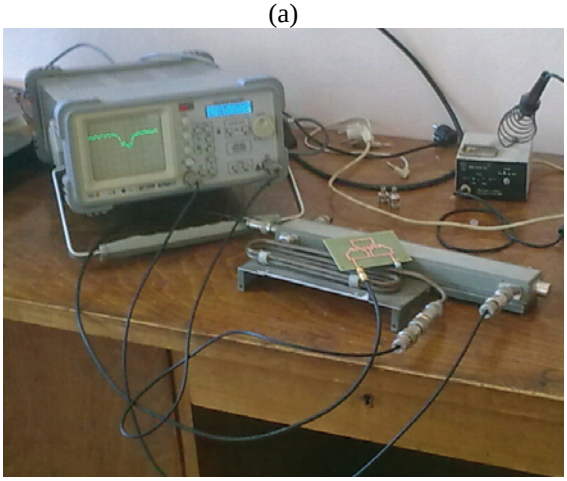


Fig.10 Frequency response from the EM simulation. Solid- transmission coefficient, dashed- reflection coeficient

From Fig.10 is clearly seen that the filter response is of Chebyshev type with equiripple response in the passband. The bandwith is 72MHz and the center frequency is 895MHz. The insertion loss in the passband is less than 4dB. The reflection coefficient is lower than -14dB.

The synthesized filter is manufactured on FR-4 substrate with standard etching technique. The measurement equipment for transmission and reflection coefficients is shown on Fig.11.

(a)



(b)

Fig.11 Measurement equipment for (a) transmission coefficient S_{21} and (b) reflection coefficient S_{11}

The measurements are carried out on spectrum analyser ATTEN AT6011with tracking generator. For the reflection coefficient measurement is used directional coupler Hewllet-Packard. The measurement results are shown on Fig.12.

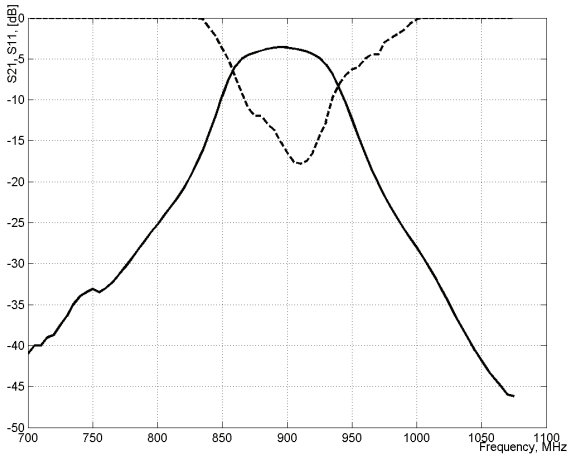


Fig.12 Measured frequency response. Solid- transmission coefficient, dashed- reflection coeficient

The measured results show very good agreement between the simulated and measured results. The measured center frequency is 895MHz, and the bandwidth is 65MHz measured on -3dB. The minimum insertion loss in the pass band is -3.5dB and is due to the very high dielectric loss of the substrate FR-4. The conductor loss is relatively low in comparison to the dielectric loss. The reflection coefficient is less than -12dB with minimum value of -17.5dB. The filter size is 69x55mm and it is 30% less than the same filter using pentagonal open-loop resonators

V. Conclusion

The paper proposes miniaturized pentagonal microstrip resonators and their usage in filters for mobile communication systems like GSM 900. The topology of the resonator yields more flexibility in the possible couplings between the resonators in the structure of the filter. Three coupling topologies are investigated and the coupling mechanism is described. Based on fullwave EM simulations, the coupling coefficient is computed and graphical results are presented. Based on the research of topologies of coupled pentagonal resonators, it is synthesized third order Chebysev filter. The designed filer is manufactured and its frequency responses are measured. There is very good agreement between the theoretical, simulated and measured results.

References

[1] A.E. Atia and A.E. Williams. "Narrow-Bandpass Waveguide Filters", *IEEE Trans. on Microwave Theory and Techniques*, vol. 20, no.4, pp. 258-265, Apr. 1972.

[2] M. Sagawa, K. Takahashi and M. Makimoto. "Miniaturized Hairpin Resonator Filters and Their Application to Receiver Front-End MIC's ", *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 12, pp. 1991-1997, Dec. 1989.

[3] K.F. Chang, K.W. Tam, W.W. Choi, R.P. Martins, "Novel Quasi-Elliptic Microstrip Filter Configuration Using Hexagonal Open-Loop Resonators", *IEEE MTT-S Digest*, pp.863-866, Feb.2002.

[4] S. Romani, et al., "Design of compact microstrip bandpass filters using pentagonal open-loop", *Proceeding CSS '07 Proceedings of the Fifth IASTED International Conference on Circuits, Signals and Systems*, 2007.

[5] R. M. Kurzrok, "General Three-Resonator Filters in Waveguide (Correspondence)", *IEEE Transactions on Microwave Theory and Techniques*, vol. 14, no.1, pp. 46-47, Jan. 1966.

[6] R. Cameron, "Asymmetric Realization for Dual-Mode Bandpass Filters", *IEEE Trans on MTT* , pp.51-58, Jan. 1981.

[7] R. Cameron, "Advanced Coupling Matrix Synthesis Techniques for Microwave Filters", *IEEE Trans on MTT*, vol. 50, no. 1, pp. 1-10, Jan.2003.

[8] J.-S Hong, M.J. Lancaster, "Couplings of Microstrip Square Open-Loop Resonators for Cross-Coupled Planar Microwave Filters", *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, no. 11, pp. 2099-2109, Nov. 1996.

Synthesis of Transfer Wave Matrix Polynomials for Digital Structure of Microstrip Ultra-Wideband Filter utilizing Short-circuited Stubs

Biljana P. Stošić

Abstract – **An efficient method based on transfer wave matrix polynomials is proposed to obtain the scattering parameters in *z*-domain of planar microstrip structure utilizing short-circuited stubs. A wave digital network (WDN) which represents a digital model of a microstrip stub-line structure is observed here. WDN is composed of cascaded unit elements and three-port adaptors. An ultra-wideband filter realized in microstrip technique, proving the response accuracy of the new technique, is given.**

Keywords – **Wave digital approach, wave digital networks, microwave circuits, short-circuited stubs, ultra-wideband filter.**

I. Introduction

The wave digital concept has been introduced in order to obtain digital filter structures that, due to their inherent passivity, possess many advantageous properties such as stability. Wave digital filters (WDFs) represent a class of digital filters with a particular interest. A detailed discussion of WDF theory is given in references [1-5]. Well known theory of WDFs is used for modeling of the planar structures by wave digital elements [5-14].

The basic idea of the 1D wave digital approach is to treat the complex structure as a typical connection of several uniform segments. The delays of uniform segments vary from one another, and because of this each segment has to be represented as cascade of several unit elements (UEs).

The wave digital model of a short-circuited transmission line, which represents the background to the modeling strategies used here, is given in [14].

This paper is devoted to the synthesis of polynomials of transfer wave matrix elements in *z*-domain using wave digital approach. Microstrip structures, such as ultra-wideband filter utilizing short-circuited stubs, can be modeled by use of the wave digital networks (WDNs) in MATLAB environment [15]. WDN response can be calculated in the frequency or in the time domain directly from known network function in *z*-domain that is going to be presented in this paper.

The paper is organized as follows. The WDN of the observed structure is given in Section II. Moreover, the calculating of the scattering parameters of the known WDNs is described.

Finally, to show the validity of the proposed modeling and analysis approaches of a microstrip ultra-wideband filter utilizing short-circuited stubs, the simulated results of wave digital approach are presented and discussed in Section III.

Biljana P. Stošić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, Niš, Serbia, E-mail: biljana.stosic@elfak.ni.ac.rs

II. Network Transfer Wave Matrix

The calculating of the wave transfer matrix polynomilas of the known WDNs is discussed here.

A planar microstrip stub-line structure can be represented as M uniform segments connected in a typical way. The ladder wave digital network of the planar microstrip stub-line structure is composed of several building blocks: UEs, multipliers, and two-port and three-port adaptors.

A WDN with n_t UEs and $M-1$ three-port parallel adaptors, pictured in Fig. 1, is analyzed here. This WDN is a two-port circuit having at each port an input and an output wave variable. Each uniform segment of the planar structure – so-called UTL segment (uniform transmission lines and stubs) - is modeled by n_k cascaded sections, $k=1,2,...,M$ (assigned as $n_k \times T$ blocks in the Fig. 1). Each UE is associated with its delay T, and port resistances R_k at either port. The port impedances of the UEs in those blocks are equal, which means that they can be directly cascade connected (coefficients of two-port adaptors are zeros). The simulation of connections between the three models of UTL segments (one of them is a short-circuited stub connected at dependent port) is achieved by three-port parallel adaptor. The incident wave A_0 is equal to voltage U_S of the source, and reflected wave B_m is equal to voltage $2U_L$ on the load. The first and the last two-port series adaptors are used for matching source and load resistances to the rest of the WDN.

The transfer wave matrix for a single UE is

$$\boldsymbol{T}_{UE} = \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix} = \frac{1}{z^{-1}} \cdot \begin{bmatrix} z^{-1} & 0 \\ 0 & 1 \end{bmatrix}. \quad (1)$$

Finally, the transfer wave matrix for one uniform segment which is modeled by n_k cascaded UEs is

$$\boldsymbol{T}_{UE}^{n_k} = \underbrace{\boldsymbol{T}_{UE} \times \boldsymbol{T}_{UE} \times \cdots \times \boldsymbol{T}_{UE}}_{n_k} = \frac{1}{z^{-n_k}} \cdot \begin{bmatrix} z^{-n_k} & 0 \\ 0 & 1 \end{bmatrix} \quad (2)$$

The k^{th} three-port parallel adaptor with port 2 chosen as dependent port (Fig. 1) is described by set of equations

$$B_{k-1} = B_k + A_k - A_{k-1}, \quad (3a)$$

$$B_{k+1} = B_k + A_k - A_{k+1}, \quad (3b)$$

$$B_k = A_k + \alpha_{k-1} \cdot (A_{k-1} - A_k) + \alpha_k \cdot (A_{k+1} - A_k), \quad (3c)$$

where the multiplier coefficients are

$$\alpha_{k-1} = \frac{2G_{k-1}}{G_{k-1} + G_k + G_{k+1}}, \quad \alpha_k = \frac{2G_{k+1}}{G_{k-1} + G_k + G_{k+1}}, \quad (4)$$

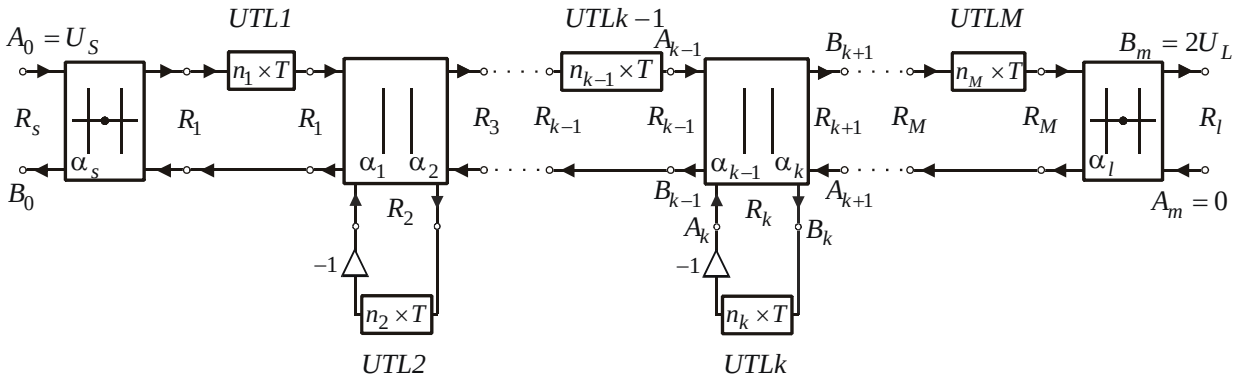


Fig. 1. Wave digital network of a planar microstrip structure with short-circuited stubs

and G_{k-1}, G_k and G_{k+1} are the port conductances. A_{k-1}, A_k and A_{k+1} are incident, and B_{k-1}, B_k and B_{k+1} are reflected waves at ports (k-1)-(k-1)', k-k' and (k+1)-(k+1)', respectively.

A short-circuited stub is connected to the port 2 being dependent and for this port the wave variables can be written

$$A_k = -z^{-n_k} \cdot B_k . \tag{5}$$

If dependent port 2 of three-port adaptor is terminated by a model of a short-circuited stub, its network is reduced to a two-port network, and for wave transfer matrix can be written

$$\begin{bmatrix} B_{k-1} \\ A_{k-1} \end{bmatrix} = \boldsymbol{T}^{n_k}_{\alpha_{k-1}\alpha_k} \cdot \begin{bmatrix} A_{k+1} \\ B_{k+1} \end{bmatrix} = \begin{bmatrix} T^{n_k}_{11} & T^{n_k}_{12} \\ T^{n_k}_{21} & T^{n_k}_{22} \end{bmatrix} \cdot \begin{bmatrix} A_{k+1} \\ B_{k+1} \end{bmatrix} \tag{6}$$

The wave transfer matrix for the k^{th} three-port adaptor, with a short-circuited stub connected on the dependent port 2, $\boldsymbol{T}^{n_k}_{\alpha_{k-1}\alpha_k}$ is obtained by use of relations (2)-(5) and its elements are

$$T^{n_k}_{11} = \frac{-\alpha - z^{-n_k}}{\alpha_{k-1} \cdot (1 - z^{-n_k})}, \tag{7a}$$

$$T^{n_k}_{12} = \frac{(\alpha_{k-1} - 1) + (\alpha_k - 1) \cdot z^{-n_k}}{\alpha_{k-1} \cdot (1 - z^{-n_k})}, \tag{7b}$$

$$T^{n_k}_{21} = \frac{(1 - \alpha_k) + (1 - \alpha_{k-1}) \cdot z^{-n_k}}{\alpha_{k-1} \cdot (1 - z^{-n_k})} \tag{7c}$$

and

$$T^{n_k}_{22} = \frac{1 + \alpha \cdot z^{-n_k}}{\alpha_{k-1} \cdot (1 - z^{-n_k})}, \tag{7d}$$

where

$$\alpha = 1 - \alpha_{k-1} - \alpha_k . \tag{7e}$$

For WDN depicted in Fig. 1, the elements of the wave transfer matrix satisfy relations

$$B_0 = T_{11} \cdot A_m + T_{12} \cdot B_m , \tag{8}$$

$$A_0 = T_{21} \cdot A_m + T_{22} \cdot B_m . \tag{9}$$

The complete transfer wave matrix $\boldsymbol{T}$ corresponding to the analyzed WDN is a product of the wave matrices of network building blocks as

$$\boldsymbol{T} = \boldsymbol{T}_{\alpha_s} \times \boldsymbol{T}^{n_1}_{UE} \times \boldsymbol{T}^{n_2}_{\alpha_1\alpha_2} \times \boldsymbol{T}^{n_3}_{UE} \times \ldots \times \boldsymbol{T}^{n_{k-1}}_{UE} \times \boldsymbol{T}^{n_k}_{\alpha_{k-1}\alpha_k} \times \times \boldsymbol{T}^{n_{k+1}}_{UE} \times \ldots \times \boldsymbol{T}_{\alpha_l} . \tag{10}$$

Consider now the matrices of two-port series adaptors as follows

$$\boldsymbol{T}_{\alpha_s} = \frac{1}{1 - \alpha_s} \cdot \begin{bmatrix} -1 & \alpha_s \\ \alpha_s & -1 \end{bmatrix} = \frac{1}{1 - \alpha_s} \cdot \boldsymbol{Q}_s \tag{11}$$

and

$$\boldsymbol{T}_{\alpha_l} = \frac{1}{1 - \alpha_l} \cdot \begin{bmatrix} -1 & \alpha_l \\ \alpha_l & -1 \end{bmatrix} = \frac{1}{1 - \alpha_l} \cdot \boldsymbol{Q}_l , \tag{12}$$

where adaptors' coefficients are

$$\alpha_s = \frac{R_s - R_1}{R_s + R_1}, \text{ and } \alpha_l = \frac{R_M - R_l}{R_M + R_l}, \tag{13}$$

with the port resistances R_s, R_1, R_M, and R_l assigned as shown in Fig. 1.

The matrix of one uniform segment modeled with n_{2j-1} UEs can be written in the form

$$\boldsymbol{T}^{n_{2j-1}}_{UE} = \frac{1}{z^{-n_{2j-1}}} \cdot \begin{bmatrix} z^{-n_{2j-1}} & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{z^{-n_{2j-1}}} \cdot \boldsymbol{Q}_{n_{2j-1}} , \tag{14}$$

where is $j = 1,2,...,N_1$.

The matrix of three-port parallel adaptor with stub on dependent port 2 which is modeled by n_{2j} UEs is given in the form

$$\boldsymbol{T}^{n_{2j}}_{\alpha_{2j-1}\alpha_{2j}} = \frac{1}{\alpha_{2j-1} \cdot (1 - z^{-n_{2j}})} \cdot \boldsymbol{Q}_{n_{2j}} \tag{15a}$$

where the matrix $\boldsymbol{Q}_{n_{2j}}$ elements are

$$Q^{n_{2j}}_{11} = -\alpha - z^{-n_{2j}} , \tag{15b}$$

$$Q^{n_{2j}}_{12} = (\alpha_{2j-1} - 1) + (\alpha_{2j} - 1) \cdot z^{-n_{2j}} , \tag{15c}$$

$$Q^{n_{2j}}_{21} = (1 - \alpha_{2j}) + (1 - \alpha_{2j-1}) \cdot z^{-n_{2j}} , \tag{15d}$$

$$Q^{n_{2j}}_{22} = 1 + \alpha \cdot z^{-n_{2j}} , \tag{15e}$$

and $j = 1,2,...,N_1$. The analyzed planar microstrip structures can have different number of UTL segments. The number N_1 depends on the number of segments in structure as

$$N_1 = \begin{cases} M/2 & \text{for } M \text{ being even,} \\ (M-1)/2 & \text{for } M \text{ being odd.} \end{cases} \quad (16)$$

According to the relations (10)-(16), the polynomials can be written in the form

$$W_e(z) = (1-\alpha_s)\cdot(1-\alpha_l)\cdot\prod_{j=1}^{N_1}\left(\alpha_{2j-1}\cdot(1-z^{-n_{2j}})\cdot z^{-n_{2j-1}}\right)$$

for even M, (17)

or $W_o(z) = W_e(z)\cdot z^{-n_M}$, for odd M. (18)

The complete matrix can be represented in the form

$$\boldsymbol{Q}_e(z) = \boldsymbol{Q}_S \times \prod_{j=1}^{N_1}\left(\boldsymbol{Q}_{n_{2j-1}} \times \boldsymbol{Q}_{n_{2j}}\right) \times \boldsymbol{Q}_l, \text{ for even } M, \quad (19)$$

or

$$\boldsymbol{Q}_o(z) = \boldsymbol{Q}_S \times \prod_{j=1}^{N_1}\left(\boldsymbol{Q}_{n_{2j-1}} \times \boldsymbol{Q}_{n_{2j}}\right) \times \boldsymbol{Q}_{n_M} \times \boldsymbol{Q}_l, \text{ for odd } M \quad (20)$$

where the matrix $\boldsymbol{Q}_{n_M}$ corresponds to the last segment in the series branch.

Finally, the complete wave transfer matrix $\boldsymbol{T}$ due to the number of segment in the structure can be written in one of two forms

$$\boldsymbol{T} = \frac{1}{W_e(z)}\cdot\boldsymbol{Q}_e(z), \text{ for even } M \quad (21)$$

or $$\boldsymbol{T} = \frac{1}{W_o(z)}\cdot\boldsymbol{Q}_o(z), \text{ for odd } M. \quad (22)$$

The wave matrix elements are the rational polynomial functions of z^{-1}. In other words, the complete wave transfer matrix can be written in the form of polynomials

$$\boldsymbol{T} = \begin{bmatrix} T_{11}(z) & T_{12}(z) \\ T_{21}(z) & T_{22}(z) \end{bmatrix} = \frac{1}{W_{e/o}(z)}\cdot\begin{bmatrix} Q_{e/o11}(z) & Q_{e/o12}(z) \\ Q_{e/o21}(z) & Q_{e/o22}(z) \end{bmatrix} \quad (23)$$

where index e/o corresponds to even (e) or odd (o) M. Only two elements, $T_{12}(z)$, and $T_{22}(z)$, have to be calculated. If necessary, two other elements, $T_{11}(z)$, and $T_{21}(z)$, can be derived from the previously ones.

Providing $A_m = 0$, the output response (forward voltage transmission coefficient) is

$$S_{21} = \left.\frac{B_m}{A_0}\right|_{A_m=0} = \frac{W_{e/o}(z)}{Q_{e/o22}(z)}, \quad (24)$$

and the input response (input reflection coefficient) is

$$\Gamma_0 = S_{11} = \left.\frac{B_0}{A_0}\right|_{A_m=0} = \frac{Q_{e/o12}(z)}{Q_{e/o22}(z)}. \quad (25)$$

III. RESULTS

The objective of this section is to prove the accuracy of the proposed modeling and analyzing approaches. To demonstrate the main idea and approach, a microstrip ultra wideband filter with a central frequency of 2.1 *GHz* [16] is depicted. The layout is shown in Fig. 2. It is analyzed on FR-4 substrate with dielectric constant $\varepsilon_r = 4.6$, and the board thickness $h = 0.6\,mm$. Metalisation is cooper and the metal thickness is $t = 17.5\,\mu m$. Its wave-based model is implemented in MATLAB/Simulink in paper [14]. The filter is symmetrical and approximated by connection of 19 uniform segments with parameters given in the Table I.

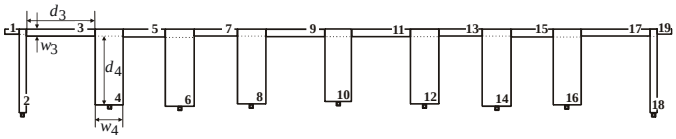


Fig. 2. Layout of ultra-wideband filter with short-circuited stubs

TABLE I. PARAMETERS OF UNIFORM SEGMENTS

nv	d [mm]	w [mm]	Zc [Ohm]	Tv [ps]
1,19	3.1000	1.1000	49.5082	19.2338
2,18	18.0000	1.5000	40.9808	113.1258
3,17	14.8000	1.5000	40.9808	93.0145
4,16	16.3000	6.0000	14.3797	108.8778
5,15	9.0000	1.7000	37.7671	56.8691
6,14	16.6000	6.2000	13.9839	111.0299
7,13	9.3000	1.5000	40.9808	58.4483
8,12	16.1000	6.1000	14.1790	107.6145
9,11	12.6000	1.6000	39.3056	79.4078
10	15.6000	5.7000	15.0181	103.9822

For given error of $n\_er = 0.01\%$, a total minimal number of sections in *WDN* is $n_t = \sum_{k=1}^{19} n_k = 582$. The numbers of sections in individual segments $n_k = round\left[q\cdot T_k / T_{\min}\right]$ are 7, 41, 34, 40, 21, 40, 21, 39, 29, 38, 29, 39, 21, 40, 21, 40, 34, 41, and 7, respectively. A total delay for the digital model of the structure is $T_t = n_t \cdot T_{\min} / q = 1599.1561\,ps$ where a multiple factor is $q = 7$ and a minimum delay is $T_{\min} = \min\{T_1, T_2, ..., T_{19}\} = 19.2338\,ps$. A total real delay of the structure is $T_\Sigma = \sum_{k=1}^{19} T_k = 1599.2257\,ps$. A sampling frequency of the digital model of the planar structure for the chosen minimal number of sections is $F_s = n_t / T_t = 363.9420\,GHz$. In this case, a relative error of delay is $er = \frac{T_\Sigma - T_t}{T_\Sigma}\cdot 100\,\% = 0.004352\,\%$. According to the relation (4), the three-port adaptor coefficients are $\alpha_1 = \alpha_{18} = 0.5855$, $\alpha_2 = \alpha_{17} = 0.7073$, $\alpha_3 = \alpha_{16} = 0.4053$, $\alpha_4 = \alpha_{15} = 0.4398$, $\alpha_5 = \alpha_{14} = 0.4327$, $\alpha_6 = \alpha_{13} = 0.3988$ $\alpha_7 = \alpha_{12} = 0.4054$, $\alpha_8 = \alpha_{11} = 0.4227$, $\alpha_9 = \alpha_{10} = 0.4332$ The two-port adaptor coefficients are $\alpha_S = -\alpha_L = 0.004942$.

Fig. 3 shows responses both simulated in MATLAB by a new proposed approach and obtained in ADS simulator [17]. Microstrip circuits inevitably incorporate transmission line discontinuities, such as T-junction. A complete understanding of microstrip circuits requires characterization of various discontinuities included in the circuit. Only in that case, WDN curve is going to have better agreement with other results. That will be considered in the future.

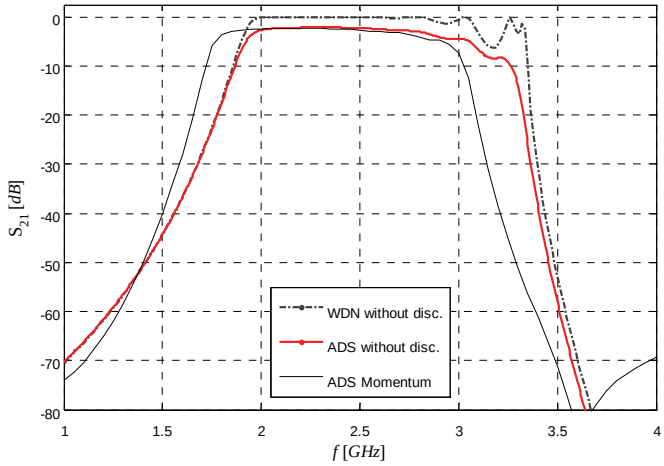


Fig. 3. Frequency response

IV. Conclusion

There are several main conclusions that can be drawn here:

1) A synthesis method is presented for evaluating scattering parameters of microwave circuits formed by connecting several multiport networks together. The synthesis is quite general and can be employed to analyze structure with any number of segments.

2) The analysis of wave digital structures is efficiently automated, which is inevitable when structures with larger numbers of building blocks are to be dealt with.

3) As has been told previously, response can be calculated in the frequency or in the time domain directly from known network function in *z*-domain. Known network functions in *z*-domain can be used as input data in some other simulations.

4) A great advantage of this method is its computational efficiency. The proposed approach is implemented on a processor Intel® Pentium® Dual CPU E2220 @ 2.4 *GHz* . A time for a response calculation in the frequency domain directly is 0.0753 *s* . This approach provides the fast structure simulation versus complex and time consuming 3D models.

5) In order to prove the accuracy of the proposed modeling and analyzing approaches, the computer simulated results obtained by WDN are compared to those of linear and momentum simulations obtained in ADS. One can observe that results obtained by described approach have good agreement with ADS data in whole frequency band.

6) Implementation of WDS in analysis of microwave structures can be used by microwave engineers because of the associated simplicity and accuracy.

Acknowledgment

This paper is supported by Ministry of Education and Science, Serbia, project number TR32052.

References

[1] A. Fettweis, “Digital Circuits and Systems”, I*EEE Transactions on Circuits and Systems*, vol. CAS-31, no. 1, pp. 31-48, January, 1984.

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

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

[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] 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, pp. 2150-2153, 2007.

[7] B.P. Stošić, and M.V. Gmitrović, “Direct Analysis of Wave Digital Network of Microstrip Structure with Step Discontinuities”, *Proceedings of the 7th WSEAS International Conference on System Science and Simulation in Engineering – ICOSSSE’08*, pp. 25-29, Venice, Italy, 2008.

[8] B.P. Stošić, and M.V. Gmitrović, “Wave Digital Approach - Different Procedures for Modeling of Microstrip Step Discontinuities”, *International Journal of Circuits, Systems and Signal Processing*, issue 3, vol. 2, pp. 209-218, 2008.

[9] B.P. Stošić, and M.V. Gmitrović, “A Wave Digital Approach in Obtaining *z*-domain Functions for Microstrip Stub-line Structures”, *Proceedings of the 9th Intern. Conference on Telecom. in Modern Cable, Satellite and Broadcasting Services - TELSIKS 2009*, vol. 1, pp. 193-197, Niš, Serbia, 2009.

[10] B.P. Stošić, and M.V. Gmitrović, “Wave-based Modeling and Analysis of Microstrip Stub-line Structures”, *Proc. 17th Telecom. forum TELFOR 2009*, pp. 533-539, Belgrade, Serbia, 2009.

[11] B.P. Stošić, “Using z-variable Functions for the Analysis of Wave-based Model of Microstrip Stub-line Structure”, *Microwave Review*, vol. 15, no. 2, pp. 6-11, 2009.

[12] B.P. Stošić, “A T-matrix of Digital Model of Structure with Cascaded Transmission Line and Open Stub in Parallel Branch”, *Proc. 54th ETRAN Conference*, pp. MT1.1-1-4, Donji Milanovac, Serbia, 2010 (in Serbian).

[13] B.P. Stošić, “Application of the Equations for Elliptic Structure Digital Model to T-resonator Model”, *Proc. 18th Telecom. forum TELFOR 2010*, pp. 630-633, Belgrade, Serbia, 2010 (in Serbian).

[14] B.P. Stošić, and J.J. Joković, ”Generation of Wave-based Model of Microstrip Ultra-wideband Filter utilizing Short-circuited Stubs and its MATLAB/Simulink Implementation”, *Proc. 55th ETRAN Conference*, pp. MT1.1-1-4, Republic of Srpska, Banja Vrućica (Teslić), June 6-9, 2011.

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

[16] N. Timko, M. Kovač, and M. Potrebić, „Miniaturization of Microwave Ultra-wideband Filter utilizing Short-circuited Stubs“, *Proc. 54th ETRAN Conference*, pp. MT1.5-1-4, Donji Milanovac, Serbia, June 7-11, 2010.

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

Low Power IR-UWB Pulse Generator in 0.13um CMOS Technology

Jelena Radic[1], Alena Djugova[2] and Mirjana Videnovic-Misic[3]

***Abstract* – A low-power ultra-wideband (UWB) pulse generator is investigated in this paper. It consists of a glitch generator, a three-stage ring oscillator, an amplifier and a pulse shaping filter. The pulse oscillator is switched on by the glitch generator only for sub-nanosecond duration to produce a UWB signal. As impulse radio-based generator operates in burst mode with low duty cycle it has low power consumption with no static power dissipation. The circuit presented was designed in a 0.13µm UMC CMOS technology with 1.2V supply voltage. Simulation results showed spectrum that fully complies with the corresponding FCC spectral mask. The signal generator has been evaluated for the best performance supporting on-off keying modulation and 100 MHz pulse repetition frequency (PRF).**

***Keywords* – CMOS technology, impulse radio, low power, on-off keying (OOK), pulse generator, ultra-wideband (UWB).**

I. INTRODUCTION

During last decade ultra-wideband was becoming one of the most promising radio technologies in both academic and industrial circles within very high data rate short-range communication, and low data rate communication related to localization, targeting both low cost and low power consumption. Recently, UWB standard is highly used in short range wireless sensor networks, local and personal area networks, ground penetration radars, the position detection and inventory tracing systems, medical imaging system (remote cardiopulmonary monitoring) and many other medical applications [1-3].

The American Federal Communications Commission (FCC) defines a signal as ultra-wideband if it occupies more than 500 MHz of radio frequency spectrum or exhibits a fractional bandwidth of at least 25% [4]. Nowadays, two different approaches are used for the implementation of UWB communication system: the impulse radio (IR-UWB) technique where data are transmitted as a series of extremely short pulses modulated in time, polarity or amplitude [5], and the multi-band approach consisting in modulating several carriers by applying Orthogonal Frequency Division Multiplexing (MB OFDM). The main advantage of IR-UWB systems is their implementation that can lead to low complexity and low power architectures with increased battery lifetime. Additionally, IR-UWB technology offers wide bandwidth up to several GHz and high fading margin for communication systems in multipath environments [3].

Since the FCC allocated frequency spectrum for UWB technology is 3.1−10.6 GHz, the power level from the UWB transmitter should be small enough not to interfere with the already existing communication systems such as WiMax, Bluetooth and GSM. This requirement limits output power level of UWB TXs at −41.3 dBm/MHz [4]. In spite of these regulations, there have been many reports of interferences with wireless local area network (WLAN) systems operated in the 5-6 GHz band. Therefore, for practical reasons UWB spectrum is split in two sub-bands: 3-5 GHz (lower band) and 6-10.6 GHz (higher band).

There are many papers published about low-band UWB pulse generator due to its easier implementation. Although higher-band operation offers advantages like wider bandwidth, higher data rate, and immunity from interference with WiMax, this research is still at an early stage because of design difficulties (higher operation frequency and shorter pulse duration). In this paper pulse generator, with low power consumption and low complexity, designed for higher band UWB applications is proposed. The presented topology is analyzed in Section III. Simulated FOMs are given in Section IV followed by discussion of advantages and comparison with the other pulse generator designs found in literature. The Section V concludes the paper.

II. IR-UWB PULSE GENERATOR ARCHITECTURE

The pulse generator represents a key block in impulse UWB communications. As pulse shape determines the spectrum characteristic of the UWB signal and effectively dictates specific system requirements, its generation is one of the essential considerations in the UWB design. Fig. 1 shows the

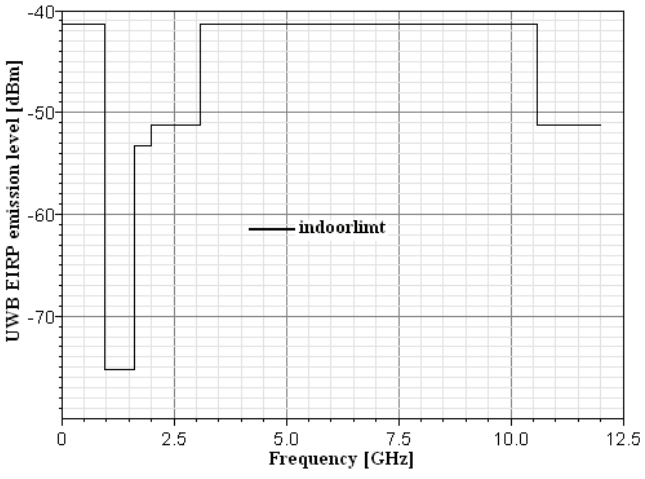


Fig. 1. FCC Indoor mask for UWB.

[1] Jelena B. Radic is with the Faculty of Technical Sciences, Trg D. Obradovica 6, 21000 Novi Sad, Serbia, E-mail: jelenar_@uns.ac.rs.

[2] Alena M. Djugova is with the Faculty of Technical Sciences, Trg D. Obradovica 6, 21000 Novi Sad, Serbia, E-mail: alenad@uns.ac.rs.

[3]Mirjana S. Videnovic-Misic is with the Faculty of Technical Sciences, Trg D. Obradovica 6, 21000 Novi Sad, Serbia, E-mail: mirjam@uns.ac.rs.

FCC mask for indoor UWB communications [4] that pulse spectrum has to comply.

The FCC rules only define the frequency bands and radiated power spectral density but there are no requirements on the time-domain shapes. Pulse shapes usually used in impulse radio technology are based on the Gaussian pulse and its derivatives. For indoor systems, the 5th or higher order derivative of the Gaussian pulse should be used to comply with allocated indoor spectrum mask [6].

Implementation of the pulse generator in standard CMOS process with low power and small area constraints is very challenging. Furthermore, assuming a single band operation, the sub-nanosecond pulse length is required to use the band efficiently. The requirements are even more demanding when higher band is used. The all-digital pulse generator architectures usually require power amplifier at the output to provide sufficient signal strength [7]. Although digital circuits have minimal power dissipation, the total power consumption of the mostly digital UWB transmitter is significantly increased due to power amplifier dissipation. Moreover, IR-UWB low system complexity is further degraded introducing additional PA design constraints. The approach, where the UWB pulse is generated using LC resonant circuits, requires considerable die area, making these architectures less suitable for area constrained applications. The slow transient response of the LC pulsed oscillators restricts the bandwidth and the pulse amplitude, as the oscillation is not able to settle sufficiently in a short duration. To use the high sub-band efficiently a pulse generating oscillator with a fast transient response is needed. One such approach is ring oscillator-based pulse generator whereby the central frequency of transmitted spectrum is defined by the oscillator frequency, and signal bandwidth is determined by the gate pulse duration. This architecture is analyzed in this paper.

III. Proposed UWB Pulse Generator

The ring oscillator-based pulse generator is shown in Fig.2. It consists of a glitch generator, a pulsed oscillator and a pulse shaping filter.

The proposed pulsed oscillator is composed of a three-stage ring oscillator, (M_1-M_3), an oscillation-enabling switch (M_4), and a buffer (M_5). The ring oscillator has been chosen for its simplicity and short start-up time. It has small resistance at each feedback nod which allows fast transient response.

The buffer isolates the ring oscillator from pulse shaping filter, and simultaneously provides the current driving capability of the pulsed oscillator.

The oscillation-enabling switch, as its name says, controls the oscillation process. When the control switch turns off, each transistor goes into a sub-threshold region and no signal is generated at the output of the oscillator. At the same time, the output signals of the inverters M_1-M_3 go close to V_{dd} effectively shutting down the buffer. When the oscillation-enabling switch turns on, the inverters M_1-M_3 outputs have voltage values determined by the size ratio of the corresponding pMOS and nMOS transistors. Due to the small inverter reactance, the oscillation can start immediately. Also, during oscillation period voltage values at the output of the ring oscillator are 'low' enough to keep the buffer turned on.

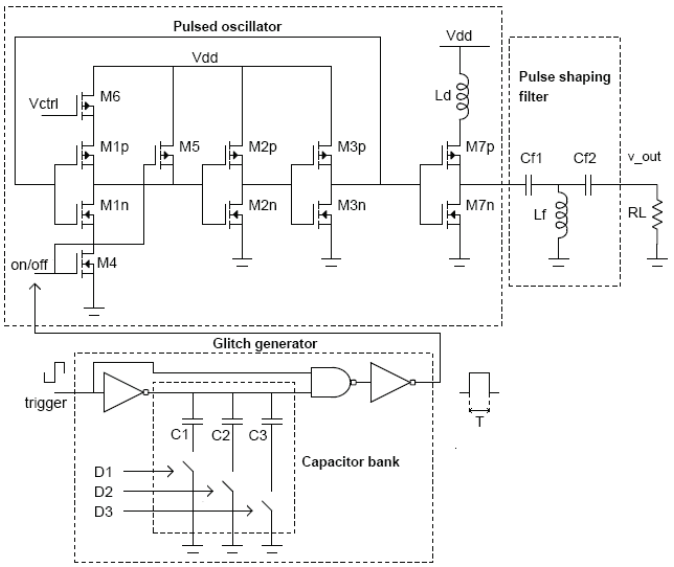


Fig. 2. Proposed UWB pulse generator.

Again, when the control switch turns off, the oscillation can stop immediately due to small reactance at the output node of each inverter. The duration of the control signal *T* (at the input of oscillation-enabling switch) determines the length of the signal generated at the ring oscillator output and thus its bandwidth. If the control signal is shorter, the output signal bandwidth gets wider.

It is already noted that IR-UWB technology requires sub-nanosecond signal duration to generate wide bandwidth. As a single microcontroller can not give such a short signal, the control pulse is produced by the glitch generator. Signal is generated in accordance with input trigger signal which should be controlled by microcontroller. The glitch generator is composed of an inverter, a NAND gate, a digitally controlled capacitor bank and a buffer/inverter at the output of the circuit. In this work, the digitally controlled capacitor bank was added at the inverter output to allow adjustment of the control signal duration and thus the bandwidth of the output signal. Digital signals D_1, D_2 and D_3 switch off/on adequate capacitors determining the total capacitance of the capacitor bank which is directly proportional to the duration of the glitch generator output signal. In this way, the control signal duration and indirectly the output signal and its bandwidth

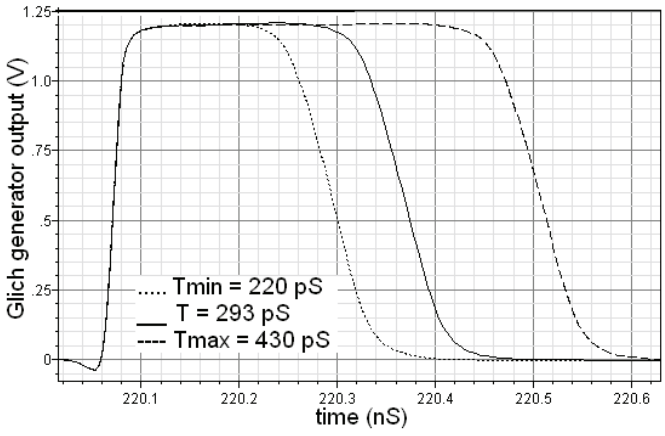


Fig. 3. The control signal at the glitch generator output.

could be digitally tuned. The smallest bank capacitance (D_1=1, D_2=D_3=0) gives the shortest control signal duration of T_{min}=220 pS, shown in Fig. 3. For the highest bank capacitance (D_1=D_2=D_3=1), the longest duration of the control signal (T_{max}=430 pS) has been achieved, depicted in Fig. 3. In this work, the control signal of T=293 pS (obtained for the digital signals combination of D_1=1, D_2= 1, D_3=0) turns on the pulsed oscillator. Due to its fast transient response, sub-nanosecond pulse with very wide bandwidth is generated at the output of the circuit. The buffer/inverter was inserted at the output of NAND gate to isolate the glitch generator from the pulse generator and provide the necessary logic.

The pulse shaping filter is high pass filter composed of inductor L_f and two capacitors C_{f1} and C_{f2}. Its main role is to shape the signal at the output of the pulsed oscillator so its spectrum complies with the FCC spectral mask. To enable the side-lobe suppression, the ring oscillator has to be designed in such a way that its oscillation frequency f_0 fall into the pass band of the pulse shaping filter. This maximizes the total power efficiency and the output amplitude as the filtered-out components are only the side-lobes that have small portion of the total generated spectral components.

IV. The Simulations Results and Discussions

The presented IR-UWB pulse generator has been designed and simulated in standard UMC 0.13 μm CMOS eight-metal technology. Supply voltage of this technology is 1.2V. Simulation has been performed using SpectreRF Simulator from Cadence Design System. Initially, the circuit was designed to generate pulse with central frequency in the middle of the higher UWB band (around 8 GHz) and optimized for frequency range from 6 GHz to 10.6 GHz, while driving a 50-Ω load and supporting on-off keying modulation. Even though the UWB antenna impedance is generally not restricted to 50 Ω, the value was chosen because it is used for most of models based on miniature commercial antennas. This type of modulation is adopted due to its simplicity. Next, the circuit was fine-tuned so that the output bandwidth could be pushed to the limits (the lower frequency limit) without degrading the other key circuit performance parameters.

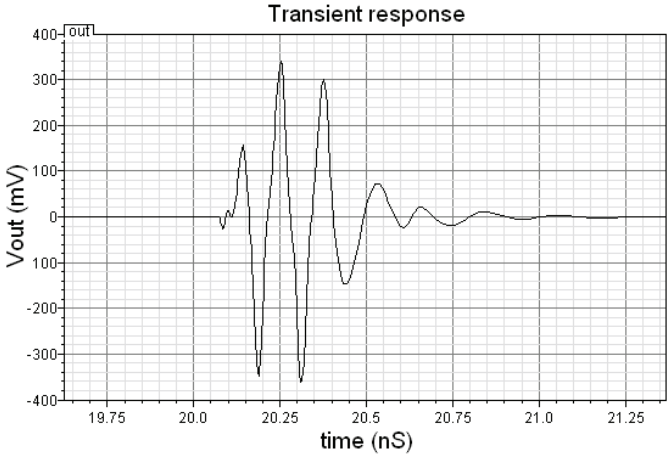


Fig. 4. Time response of the ring-oscillator based pulse generator.

The pulse generator operates in burst mode with low duty cycle and pulse repetition frequency (PRF) of 100 MHz, and thus has very low power dissipation. The IR-UWB topology was optimized with the main aim to meet FCC spectrum demands and minimize power consumption while still keeping acceptable values for remaining Figures of Merits (FOMs).

The simulation results for the generated UWB pulse and its PSD (Power Spectral Density) are shown in Figs. 4 and 5. The center frequency was observed to be around 7.66 GHz and the -10 dB bandwidth was 5.6 GHz, from 5 GHz to 10.6 GHz. In fact, the upper cut-off frequency is slightly greater than 10.6 GHz, but the mentioned point represents the upper UWB limit.

As it could be noticed, the pulse duration is about 0.75 ns and spectrum fully complies with FCC spectrum mask with more than 20 dB of the side-lobe rejection. The peak-to-peak amplitude on a 50 Ω output load is around 700 mV. The average power consumption including the glitch generator, the buffer stage and the filter is 478 μW for PRF of 100 MHz. This corresponds to 4.78 pJ energy consumption per pulse.

Simulation results show that the presented design is very suitable for high-band IR-UWB pulse generation with extremely low power consumption. The possible application could be a radar system which receives its own signal reflected from targets or a communication system with non-coherent detection.

The comparisons of the obtained Figures of Merit with pulse generators performances found in literature [8-10] are summarized in Table I. Unfortunately, the simulation results for the same pulse repetition frequency (100 MHZ) and higher UWB sub-band in 0.13μm CMOS technology could not be found in the literature. Therefore, the pulse generators optimized for the whole UWB band in the same technology (except in Ref [9]) have been used for the performance comparison.

As it was already noted that the proposed design was initially optimized for the higher band and then fine-tuned at the expense of the lower frequency limit, bandwidth of the presented pulse generator is slightly narrower compared to the observed results. Furthermore, it can be seen that design proposed in this work has by far the lowest power consumption and significantly higher amplitude compared to values reported in published papers. Other Figures of Merits are comparable to the results given by other authors.

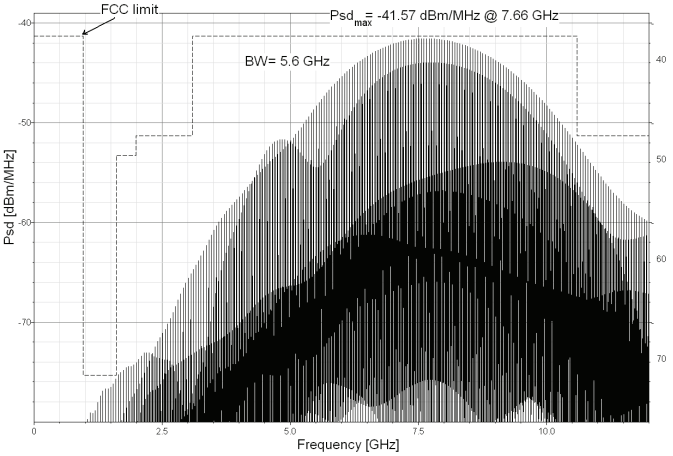


Fig. 5. Simulated spectrum of the output signal.

TABLE I
SIMULATION RESULTS SUMARY AND COMPARISON WITH REPORTED PULSE GENERATORS

Reference	This work	Ref [8]	Ref [9]*	Ref [10]
Ener. cons. [pJ/pulse]	4.78	21	380	5.6
PRF [MHz]	100	200	100	100
Amplitude (V_{pp}) [mV]	700	230	127	190
Pulse length [nS]	0.75	1	0.31	0.75
Band [GHz]	5–10.6	3.5–7.5	3.6–10.2	3.9–10.4
Technology	UMC 0.13μm CMOS	IBM 0.13μm CMOS	TSMC 0.18μm CMOS	0.13μm CMOS

*Simulation results are obtained using Advance Design System (ADS)

V. CONCLUSION

In this paper IR-UWB pulse generator for high-band applications is presented. The results demonstrated that proposed ring oscillator-based architecture has significantly lower power consumption and considerably higher peak to peak amplitude compared to the previously reported UWB pulse generators using CMOS technology. Moreover, it was shown that design has high signal bandwidth and very good side-lobe rejection performance. In addition, it should be noted that the bandwidth could be adjusted by tuning capacitance of the digitally controlled capacitor bank. The pulse generator is feasible for applications in a radar system which receives its own signal reflected from targets or in non-coherent detection communication systems.

ACKNOWLEDGEMENT

This research was supported by the Serbian Ministry of Science and Technological Development (contract number III-43008).

REFERENCES

[1] M. Ghavami, L. B. Michael, and R. Kohno, *Ultra Wideband Signals and Systems in Communications Engineering*, John Wiley&Sons Ltd, 2004.

[2] K. Siwiak and D. McKeown, *Ultra-Wideband Radio Technology*, John Wiley&Sons Ltd, 2004.

[3] J. R. Fernandes and D. Wentzloff, "Recent Andvances in IR-UWB Transceivers: An Overview ," *IEEE Int. Conf. on Circuits and Systems*, pp. 3284-3287, 2010.

[4] First Report and Order: *Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems Federal Communications Commission (FCC)*, ET Docket 98-153, Adopted February 14, 2002, Released Apr. 22, 2002.

[5] M. Z. Win and R. A. Scholtz, "Impulse radio: how it works," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 36–38, 1998.

[6] H. Sheng, P. Orlik, A. M. Haimovich, L. J. Cimini, and J. J. Zhang, "On the spectral and power requirements for ultra-wideband transmission," *IEEE Int. Conf. on Communications*, vol. 1, pp. 738–742, 2003.

[7] O. Novak and C. Charles, "Low-Power UWB Pulse Generators for Biomedical Implants," *IEEE Int. Conf. on Ultra-Wideband*, pp. 778-782, 2009.

[8] Y. Shamsa and W. A. Serdijn, "A 21pJ/pulse FCC Compliant UWB Pulse Generator," *IEEE Int. Conf. on Circuits and Systems*, pp. 497-500, 2010.

[9] J. Ch. Li, S. Jung, M. Lu, P. Gui, and Y. Joo, "A Current-Steering DAC-Based CMOS Ultra-Wideband Impulse Generator," *IEEE Int. Conf. on Communications and Information Technology*, pp. 971-975, 2009.

[10] X. Wang et al., "FCC-EIRP-Aware UWB Pulse Generator Design Approach (Invited)," *IEEE Int. Conf. on Ultra-Wideband*, pp. 592-596, 2009.

A 6–9 GHz Resistive Feedback Low Noise Amplifier Designed in 0.18µm CMOS Technology

Alena Djugova[1], Jelena Radic[2] and Mirjana Videnovic-Misic[3]

***Abstract* – In this paper two-stage low-noise amplifier (LNA) designed for operation over the common frequency range, 6–9 GHz, allocated for ultra-wideband (UWB) communications both in USA, Europe, and Japan is presented. Implemented in 0.18µm CMOS technology, it features a maximum gain (S_{21}) of 17.89 dB dissipating only 6.5 mA from the 1.8 V supply. In addition, LNA shows less than –10 dB input return loss (S_{11}) and less than –14.33 dB output return loss (S_{22}). The noise figure (NF) ranges from 3.29 dB to 3.68 dB over the band of interest. The high circuit stability parameters K_f > 37.24 and B_{1f} > 0.98 are obtained.**

***Keywords* – Ultra-wideband (UWB), Low noise amplifier (LNA), Resistive shunt-feedback topology.**

I. INTRODUCTION

Ultra-wideband (UWB) presents low power technology applicable for short-range, high-speed wireless communications. UWB differs from other communication techniques due to employing extremely short-time pulses (sub-nanosecond) to communicate between transmitters and receivers. The short duration of UWB pulses generates very wide bandwidth (GHz) in frequency domain. According to the rules set defined by the US Federal Communications Commission's (FCC) in 2002, allocated band for the unlicensed use of UWB devices is between 3.1 GHz and 10.6 GHz frequency range [1]. In order to protect the existing radio services from UWB interference maximum allowed power spectral density is –41.3dBm/MHz. Compared to the FCC in United States, the regulatory bodies in Europe decided on more stringent emission masks and specific restrictions. As frequency band from 3–5 GHz is allocated to the fixed services/WiMax systems, unrestricted frequency range in Europe is between 6 GHz and 8.5 GHz. Meanwhile in Japan and Korea, only band from 7–9 GHz is released without restriction. In addition, standard for the UWB spectrum in China, with frequency band from 6–9 GHz, will be allocated soon. Thus, frequency range of 6–9 GHz is the most interesting for UWB systems with worldwide usage.

One of the demanding tasks in up-to-date receiver design is design of low noise amplifier (LNA). As first circuit in the receiver's chain it needs to fulfill several stringent requirements. Sufficient gain is obligatory to amplify received weak signal and to overcome the noise of the subsequent stages, adequate input and output matching, to minimize reflection, and low noise figure (NF), to improve sensitivity. All this have to be obtained within defined bandwidth and with low power consumption.

In Section II basic circuit design techniques for obtaining wide frequency operational bandwidth, are described, while chosen broadband LNA architecture, two-stage common-source resistive shunt-feedback based topology, is analyzed in Section III. Simulated figures of merit (FOMs) are given in Section IV followed by discussion and comparison with other UWB LNA designs found in literature. The Section V concludes the paper.

II. BROADBAND LNA DESIGN TECHNIQUES

Four major UWB LNA's design approaches have been reported in literature: distributed amplifier, common-gate circuit, inductively degenerated common-source amplifier with additional multistage input filters and common-source amplifier with resistive shunt-feedback technique.

A. Distributed Amplifier (DA)

The renewed interest in distributed circuits arises mainly due to the capability of designing on-chip transmission lines (realized using either coplanar waveguides or cascaded LC circuits) and high-Q inductors. The general block diagram of a DA comprises transmission lines and gain stages distributed along them. Whit this architecture impedance matching over a wide bandwidth as well as a broadband gain, equal to the sum of the scalar gains from each stage, can be achieved. Consequently, this approach is power hungry and area consuming. Moreover, NF optimization for this technique is still an issue [2].

B. Common-gate Amplifier

The input common-gate stage provides the wideband input impedance matching with less design complexity and small area occupancy. However, the main disadvantage of this type of amplifier is high noise figure (noise from the load network and channel thermal noise are referred to the input in their entirety limiting the minimum possible noise figure). As input matching is coupled with noise matching, noise canceling methods need to be used. Additionally, their relatively low transconductance value can not provide low noise and high gain in a whole frequency range. This type of amplifier is

[1]Alena Djugova is with the Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia, E-mail: alenad@uns.ac.rs.

[2]Jelena Radic is with the Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia, E-mail: jelenar_@uns.ac.rs.

[3]Mirjana Videnovic-Misic is with the Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia, E-mail: mirjam@uns.ac.rs.

usually combined with additional amplifying stage which provides high-frequency gain and enhances the amplification bandwidth [3].

C. Inductive Degenerated Common-Source Amplifier Incorporated with Passive Input Matching Networks

Input multistage reactive network of the amplifying device is embedded in a multi section filter structure so that the overall input reactance is resonated over the whole band from 3.1 to 10.6 GHz. Inductive source degeneration is used for proper input matching and the size of the transistor is no longer coupled to the impedance matching requirement. Therefore, gain flatness and low power consumption are achieved. The main drawbacks of this approach are the large silicon area occupied by the integrated inductors of the passive networks and the noise contribution introduced by the passive elements at the input of the LNA. The NF rises rapidly at high frequencies, consequently decreasing the receiver sensitivity in the higher UWB bands [4]. In addition, the high-quality-factor integrated inductors require high accuracy of the electromagnetic design, and the traditional mismatch reduces the design reliability.

III. Resistive Feedback Based LNA Circuit Design

The proposed wideband LNA is shown in Fig. 1. Proper input matching over the whole band is achieved with resistive shunt-feedback technique in both stages. This technique provides wideband input matching and flat gain [5]. Additionally the inductive series-peaking technique is adopted in the input terminal of the output stage. The source follower output buffer is used as a low output impedance (i.e., 50Ω) interface.

All bias circuits are composed of resistors R_{biasn} and R_{refn}, n = 1, 2, 3, and transistor M_n, n = 5, 6, 7, where transistor forms DC current mirror with corresponding transistor (M_1, M_2, M_4). To decrease the overall power consumption (P_D), bias transistor width, should be small fraction of amplifying transistor width. The R_{biasn} value is chosen large enough to provide high impedance path to the RF signal while making a small contribution to the circuit noise.

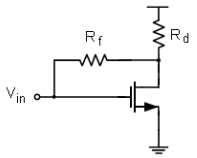


Fig. 2. Common source LNA with resistive feedback

Capacitors C_g and C_b are input and output DC blocking capacitors. Their values are chosen large so they do not influence the resonant frequency of the input and output circuit, respectively.

A. Input Stage

The first block of the proposed LNA topology can be roughly seen as basic common source amplifier with resistive shunt-feedback, shown in Fig.2. R_{f1} forms resistive shunt-feedback and drain impedance is equal to: $Z_D = j\omega L_{d1}$, hence input resistance is given by:

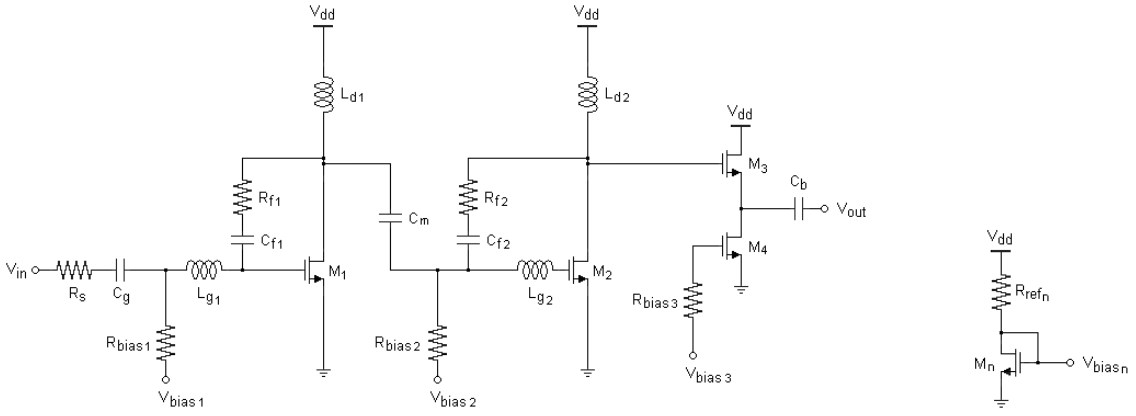
$$R_{in} = \frac{R_{f1} + Z_D}{1 + g_m \cdot Z_D}. \tag{1}$$

Increase in value of feedback resistor, R_{f1}, results in R_{in} increase. Therefore, to obtain 50Ω input match, feedback resistor tends to be a few hundred ohms, which leads to significant NF degradation.

Therefore, additional degree of freedom is provided by the gate inductance L_{g1}. Miller equivalent input resistance of R_{f1} is approximately given by [6]:

$$R_{feq} = \frac{R_{f1}}{(1 - A_v)} \tag{2}$$

where A_v is the open-loop voltage gain of the first LNA stage. With this transformation Q-factor of the input circuit shown in Fig. 1 is given by:



(a) (b)

Fig. 1. The designed 6−9 GHz CMOS LNA: (a) amplifying circuit, and (b) biasing circuit

$$Q = \frac{\omega_0 \cdot L_{g1}}{R_{feqs}} \tag{3}$$

where $R_{feqs} = \dfrac{R_{feq}}{(Q_p^{\ 2} + 1)}$ and $Q_p = R_{feq} \cdot \omega_0 \cdot C_{gs}$.

Decreasing value of inductor L_{g1} Q-factor is lower and higher bandwidth occurs.

Simultaneously impedance and noise matching over the band of interest is achieved by appropriately selected values of L_{g1}, R_{f1}, and the size and bias of the input transistor M_1, i.e. C_{gs1} and g_{m1}.

Voltage gain of resistive feedback based LNA is equal to:

$$A_v \approx -\frac{g_m \cdot R_{f1} \cdot Z_D}{R_{f1} + g_m \cdot R_s \cdot Z_D} = -\frac{1}{\dfrac{1}{g_m \cdot Z_D} + \dfrac{R_s}{R_{f1}}}. \tag{4}$$

Due to strong dependence of voltage gain on the amplifying transistor transconductance, the amplifier with resistive feedback requires a large amount of current to achieve high gain. Also, it can be seen from (4), that parameter R_{f1} increase results in voltage increase.

Combined with parasitic capacitances, purely resistive loads would result in limited high frequency performance. Thus, inductive load, L_{d1}, is used to increase bandwidth and improve gain flatness. In addition value of L_{d1} determines the gain of the first stage.

Coupling capacitor C_m is used to couple RF signal from transistor M_1 to transistor M_2.

B. Second Stage

The second stage is another common-source resistive feedback based stage. A peaking inductor L_{g2} is added to the gate-terminal of M_3 for LNA gain and NF improvement at high frequencies. Inductor L_{d2} resonate with the total parasitic capacitance at the drain of M_2, and determines the gain of the second stage and flat gain.

IV. Simulation Results

The designed circuit has been simulated in UMC 0.18µm CMOS six-metal technology using Spectre Simulator from Cadence Design System. BSIM3V3 models for all circuit components were introduced.

LNA topology was optimized with the main aim to obtain proper input matching, high flat gain, while minimizing P_D, *NF* and keeping acceptable values for remaining FOMs. Simulation results for two-stage resistive feedback LNA topology are given in Figs. 3–5 (S_{11} and S_{22}, S_{21}, *NF* and NF_{min}).

Parameter S_{11} is below –10 dB in the band of interest (–13.36 dB at 6 GHz and –10.41 dB at 9 GHz). This demonstrates the effectiveness of wideband input matching obtained by resistive feedback topology. Decrease in R_{f1} value

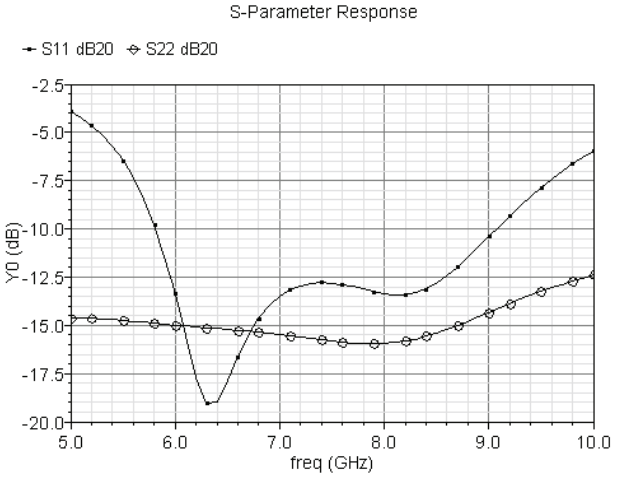


Fig. 3. LNA input (S_{11}) and output (S_{22}) return loss

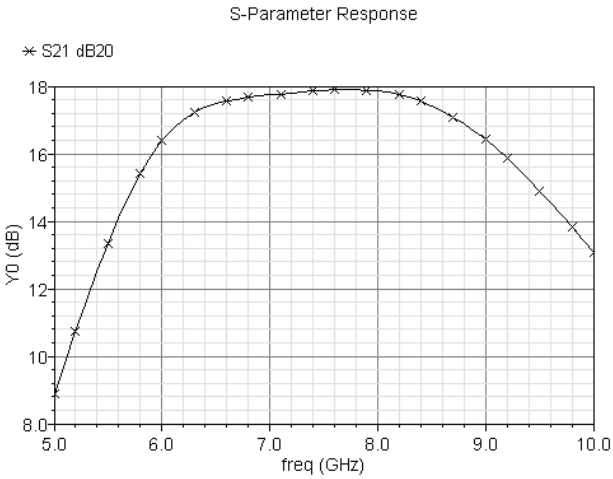


Fig. 4. LNA gain (S_{21})

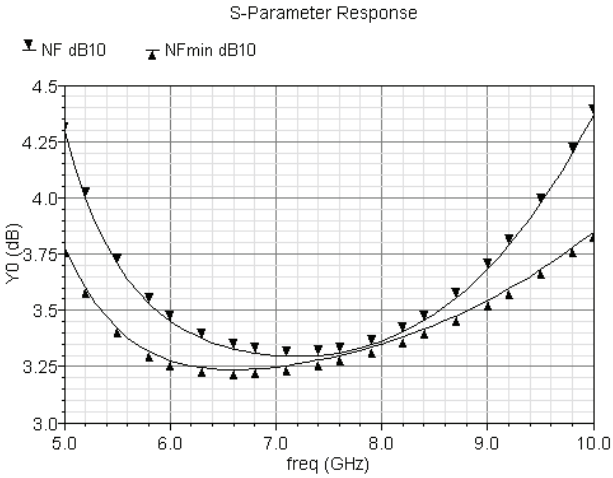


Fig. 5. LNA noise figure (*NF*) and minimum noise figure (NF_{min})

results in lower S_{11} value and higher bandwidth, while decrease in L_{d1} inductor value shifts S_{11} to higher frequencies and reduces minimum S_{11} value. L_{g1} increase narrow the S_{11} characteristic, therefore trade-off between its value and wanted bandwidth has to be achieved. Parameter S_{22} is less than –14.33 dB for the whole simulated range. This is achieved by enhancement LNA topology with output buffer.

Further excellent reverse isolation (S_{12}) below –55.45 dB is obtained. Regarding the gain, maximum S_{21} parameter value of 17.89 dB is achieved at 7.7 GHz (16.38 dB at 6 GHz and 16.42 dB at 9 GHz). Variation in S_{21} value in band of interests is 1.51 dB and the 3-dB bandwidth covers frequency range from 5.72 GHz to 9.5 GHz. Widths of amplifying transistors M_1 and M_2 determines the LNA gain while higher values of resistors R_{f1} and R_{f2} improve it. Transistor values are chosen to satisfy the gain requirement, while keeping at the same time P_D at reasonable level.

NF exhibits small variation (up to 0.39 dB) across entire simulated band (3.45 dB at 6 GHz and 3.68 dB at 9 GHz). Minimum NF value of 3.29 dB is obtained at 7.2 GHz. NF values are close to NF_{min}, the minimum possible NF values (3.28 dB at 6 GHz and 3.54 dB at 9 GHz).

For voltage supply of 1.8 V, LNA core consumes 6.5 mA. The rest is consumed by the output buffer (2.84 mA) and biasing circuits (3.98 mA).

Unconditional stability requirements for the whole simulated range are satisfied. Minimum values for stability factors K_f (Rollet stability factor) and B_{1f} (alternate stability factor) are equal to 37.24 and 0.98, respectively, which are values much higher than 1 and 0 [7], [8].

The optimized FOMs of the designed LNA and FOMs values of similar LNA topologies found in literature are summarized in Table I. In comparison with the amplifier design presented in [9], proposed LNA circuit in this paper dissipates something higher amount of power, gives larger gain and better input matching. Furthermore, LNA in [9] achieves the wideband input matching by a three-section band-pass Chebyshev filter configuration, requiring usage of large number of reactive elements that causes higher NF and larger chip area. A CMOS dual-wideband low-noise amplifier (LNA) in [10] shows lower S_{21} and higher S_{11}. Result for higher *NF* and power consumption is due to more complex topology which consists of a wideband input impedance matching network, two stage cascode amplifiers with shunt-peaked load, a notch filter and an output buffer.

TABLE I
LNA PERFORMANCE COMPARISON

	This work	[9]	[10]	[11]
BW [GHz]	6–9	6–10	6–10.3	4.5–11
S_{11} [dB]	<–10.41	<–9	<–8.31	<–8.2
S_{22} [dB]	<–14.33	<–12	N/A	N/A
S_{21} [dB]	17.89	11.6	10.34	14.2
NF [dB]	3.68	5.3	3.8*	5.1
Power [mW]	11.7	11.6	24.07	20.4
Technology	0.18µm	0.18µm	0.18µm	90nm

* Minimum noise figure.

The same drawbacks occur in [11] where differential topology with integrated balun is used. Despite usage of technology with lower DC supply voltage (1.2 V) differential design makes this approach power hungry.

V. CONCLUSION

A 6–9 GHz two stage LNA designed in 0.18µm UMC CMOS technology for frequency band from 6 GHz to 9 GHz is demonstrated in this paper. Resistive feedback technique for ultra-wideband design is used. This technique gives adequate input match without introducing large number of additionally components (inductors and capacitors) and improves LNA gain. Obtained simulation results show high gain of 17.89 dB with consumption of only 11.7 mW while maintaining low noise figure.

ACKNOWLEDGEMENT

This research was supported by the Serbian Ministry of Science and Technological Development (contract number TR-32016).

REFERENCES

[1] G. R. Aiello and G. D. Rogerson, "Ultra-Wideband Wireless Systems", *IEEE Microwave Magazine*, vol. 4, pp. 36–47, 2003.

[2] P. Heydari and D. Lin, "A Performance Optimized CMOS Distributed LNA for UWB Receivers", *CICC'05, Conference Proceedings*, pp. 337–340, San Jose, California, USA, 2005.

[3] K. Chen, J. Lu, B. Chen and S. Liu, "An Ultra-Wide-Band 0.4–10-GHz LNA in 0.18-µm CMOS," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 3, pp. 217-221, 2007.

[4] A. Bevilacqua and A. M. Niknejad, "An Ultrawideband CMOS Low Noise Amplifier for 3.1–10.6-GHz Wireless Receivers", *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 2259-2268, 2004.

[5] C. W. Kim, M. S. Kang, P. T. Anh, H. T. Kim and S. G. Lee, "An Ultra-Wideband CMOS Low Noise Amplifier for 3–5-GHz UWB System," I*EEE Journal of Solid-State Circuits*, vol. 40, pp. 544-547, 2005.

[6] B. Razavi, *Design of Analog CMOS Integrated Circuits*, New York, McGraw-Hill, 2001.

[7] SpectreRF simulator, Cadence Design System, Available: http://www.cadence.com

[8] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, Prentice Hall, 1984.

[9] Y.-Ch. Chen and Ch.-N. Kuo, "A 6~10-GHz Ultra-Wideband Tunable LNA", ISCAS 2005, Conference Proceedings, vol. 5, pp. 5099–5102, Kobe, Japan, 2005.

[10] Z.-Y.Huang and Ch.-Ch.Huang, "CMOS Dual-Wideband Low-Noise Amplifier with Notch Filter for 3.1GHz–10.6GHz Ultra-Wideband Wireless Receiver", A*SICON'07, Conference Proceedings*, pp. 415–418, Guilin, China, 2007.

[11] A. Bevilacqua, C. Sandner, M. Tiebout, A. Gerosa and A. Neviani, "A 6–9-GHz Programmable Gain LNA with Integrated Balun in 90-nm CMOS", *ICUWB2008, Conference Proceedings*, vol.1, pp. 25–28, Hannover, Germany, 2008.

PKI ANNs in Noise Wave Modelling of Microwave Transistors

Zlatica Marinković, Olivera Pronić - Rančić, Vera Marković

Abstract- **A new noise wave model of microwave field-effect transistors valid for various bias conditions is presented in this paper. The proposed model consists of an empirical noise wave model and a prior knowledge input artificial neural network trained to predict accurately transistor noise parameters. The inputs of the network are frequency, bias current and the noise parameters obtained by the noise wave model, which represent additional knowledge to the ANN. The validity of the presented model is exemplified by modelling of a specific MESFET device in packaged form.**

Keywords – **prior knowledge input artificial neural network, microwave transistor, noise model, bias dependence**

I. INTRODUCTION

Reliable and accurate small signal and noise models of microwave FETs (MESFETs, HEMTs) are necessary for the design of active circuits in modern communication systems. During the last few decades an extensive work has been carried out in the field of signal and noise modelling of these devices. Their physical models are complex and require many input technological parameters [1]-[2]. Therefore the empirical models, mostly based on equivalent circuits are often used [3]–[5].

The complete characterization of microwave transistors includes knowledge about device signal and noise parameters (scattering parameters and noise parameters: minimum noise figure, optimum source reflection coefficient and normalized noise resistance) which are frequency-, temperature- and bias-dependent. The measurements of these parameters, especially of the noise parameters, are complex and time-consuming procedures. Therefore, for accurate and efficient circuit design, models of active devices should include bias and/or temperature dependence. However, most of the existing transistor noise models are valid only for a single operating point.

The noise model [5] is considered to be very suitable for implementation into the standard commercial microwave circuit simulators. It is based on *H* representation of MESFET / HEMT intrinsic circuit with two uncorrelated noise sources. However, at microwave frequencies a treatment of noise in terms of waves seems to be more appropriate [6]-[7] because it allows the use of scattering matrices for the noise computations, leading to advantages in CAD of microwave networks. The wave approach is useful not only for noise modelling but also for measurement of microwave FETs [7]. The noise wave modelling procedures of MESFETs / HEMTs and dual-gate MESFETs based on *T* representation of transistor intrinsic circuit are proposed in [8] and [9]. A new more accurate noise wave model of microwave FETs, including the error correction functions into the basic noise wave model is proposed in [10].

In order to include bias dependence in noise wave modelling of MESFETs / HEMTs, artificial neural networks (ANNs) have been used and the results are given in this paper. Once developed, an ANN provides fast response for various input values (even for those not seen by the ANN during model development). ANNs are very convenient to be used as a modelling tool and are especially useful for problems not fully mathematically described. There are many publications referring results of applications of ANNs in the microwave area, [11 – 16]. ANNs have been applied in modelling of either active or passive devices, in microwave circuit analysis and design, etc.

The most frequently used ANN is a standard multilayer perceptron (MLP) ANN, consisting of basic units (neurons) grouped into layers: an input layer, an output layer, as well as several hidden layers [11]. Each neuron is connected to all neurons from the adjacent layers, whereas there are no connections among neurons belonging to the same layer. Each neuron is characterized by a transfer function and each connection is weighted. Information flows forward from the input layer to the output layer.

An ANN learns relationship among sets of input-output data (training sets) by adjusting network connection weights and thresholds of activation functions. There is a number of algorithms for training of ANNs. The most frequently used are backpropagation algorithm and its modifications, e.g. Levenberg Marquard algorithm [11].

The model presented in this paper is a new bias dependant hybrid empirical-neural noise model of microwave FETs. The model consists of the empirical noise wave model [10] and an additional artificial neural network aimed to introduce bias dependence in the model. The used ANN belongs to the Prior Knowledge Input (PKI) ANNs [16] which have additional information about the problem being modelled at their inputs. The model and its development are described in the paper. An example of modelling the specific MESFET device in packaged form is provided as well.

II. PROPOSED PKI ANN NOISE MODEL

The noise wave model of microwave MESFETs / HEMTs described in [10], similar to other empirical noise models, is valid only for one specific bias point. In order to include bias dependence in the noise modelling, a new hybrid model, which is shown in Fig.1, is proposed in this paper. The model

Authors are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18 000 Niš, Serbia
e-mail: zlatica.marinkovic@elfak.ni.ac.rs, olivera.pronic@elfak.ni.ac.rs, vera.markovic@elfak.ni.ac.rs

consists of the empirical noise wave model [10] developed for one representative (reference) bias point and an ANN through which dependence on bias conditions is included in the model.

Since we considered the case where dc drain-to-source voltage is constant, bias conditions are defined by dc drain-to-source current. The inputs of the neural network are: bias current (I_{ds}) and frequency (f) as well as the prior knowledge - noise parameters for the reference bias point obtained from the noise wave model for that value of frequency (F'_{min}, $\mathrm{Mag}(\Gamma'_{opt})$, $\mathrm{Ang}(\Gamma'_{opt})$, r'_n). The ANN's outputs are device noise parameters (F_{min}, $\mathrm{Mag}(\Gamma_{opt})$, $\mathrm{Ang}(\Gamma_{opt})$, r_n).

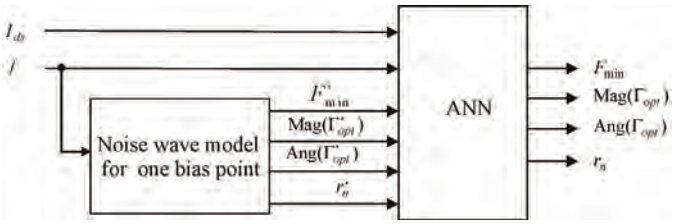


Fig. 1. Proposed PKI ANN model

The model is developed as follows: first, a reference bias point from the device bias operating range is chosen. Then, the equivalent circuit parameters for the noise wave model are extracted from the scattering (S-) parameters and noise parameters measured at several frequencies. For ANN training, it is necessary to acquire the measured values of noise parameters for certain number of bias points over the considered frequency range, and for each considered frequency to perform noise simulations using the developed noise equivalent circuit in order to obtain the noise parameters represent the ANN prior knowledge. To resume, the training set consists of samples, each containing the following: values of bias current, frequency, noise parameters obtained from noise wave model for that value of frequency and measured values of noise parameters as the network target values. After the training and evaluation of ANN, the mathematical expressions describing the ANN are further added on the mentioned device noise wave model schematic by means of a block dealing with variables and expressions (VAR). Inputs of the VAR block are the same as the inputs of ANN. The noise parameters calculated from these expressions are bias dependent values that have to be assigned to a device. Such schematics with added ANN can be used as a library element with the bias current as inputs.

After extraction of the elements of an equivalent circuit and training of the neural network, measured values are not requested for later determination of the noise parameters.

III. Numerical Example

Numerical results related to noise modelling of GaAs FET packaged microwave transistor, type ATF21186 by Agilent (HP), are presented in this paper. Measured values of *S* and noise parameters for biases (2V, 10mA), (2V, 15mA) and (2V, 20mA), in the frequency range (0.5 – 8) GHz, taken from the device datasheet, were used for the model development. All simulations were performed using microwave circuit simulator ADS, [17].

The small-signal equivalent circuit of the considered transistor is shown in Fig. 2. Transistor intrinsic circuit is denoted by the broken line and it is embedded in a network representing device parasitics. The noise of the device is described by means of the noise wave sources that emanate from the input port of the intrinsic circuit. The equivalent noise wave temperatures are assigned to the noise wave sources [10].

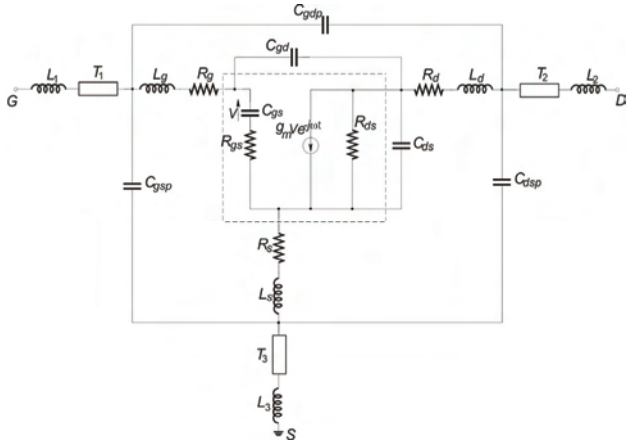


Fig. 2 Equivalent circuit of MESFET / HEMT package

At the beginning, the equivalent circuit parameters (equivalent circuit elements and noise wave temperatures) were extracted from the measured values of the device scattering and noise parameters for a bias point (2V, 15mA) and noise parameters are simulated. In order to eliminate deviations that exist between measured and modelled values of noise parameters, the error correction functions were determined and included in the transistor noise wave model. Namely, each intrinsic circuit noise parameter obtained by the basic noise wave approach is multiplied by the corresponding error correction function, [10]. In order to determine the error correction functions, for each of four noise parameters the ratio of the measured and simulated transistor noise parameters is calculated over the entire frequency range. Then, curve-fitting procedure is applied on these sets of data, in order to obtain suitable frequency dependences. In this way, corresponding mathematical functions are determined for all four noise parameters. In this case, the error correction functions were determined for the minimum noise figure, magnitude of optimum reflection coefficient and noise resistance. All functions have the polynomial form. The simulated values for angle of optimum reflection coefficient are in excellent agreement with measured ones and there was no need to include any correction for that parameter in the noise model.

The extracted equivalent circuit parameters were used for simulation of the noise parameters used as the prior knowledge for the ANN training. After building of the training set in a way described in the previous section, the training of the ANN was performed. Since number of hidden layers and hidden neurons cannot be a priori set, in order to find the ANN with the best performance, ANNs with different

number of hidden neurons were trained. After training and evaluation of the ANNs, a network with one hidden layer consisting of 15 neurons gave the best results. The mathematical expressions describing the ANN were then added on the equivalent circuit schematics and noise simulations have been performed. The most illustrative results are given below.

Frequency dependences of the minimum noise figure and magnitude of optimum reflection coefficient for bias point (2V, 15mA) obtained by the noise wave approach (dashed line), as well as by the proposed PKI ANN model (solid line) are compared with measured data (symbols) and shown in Fig. 3. The proposed PKI ANN model has better accuracy than noise wave model. It is obvious that almost perfect match between measured noise parameters and those obtained by the suggested PKI ANN model is achieved.

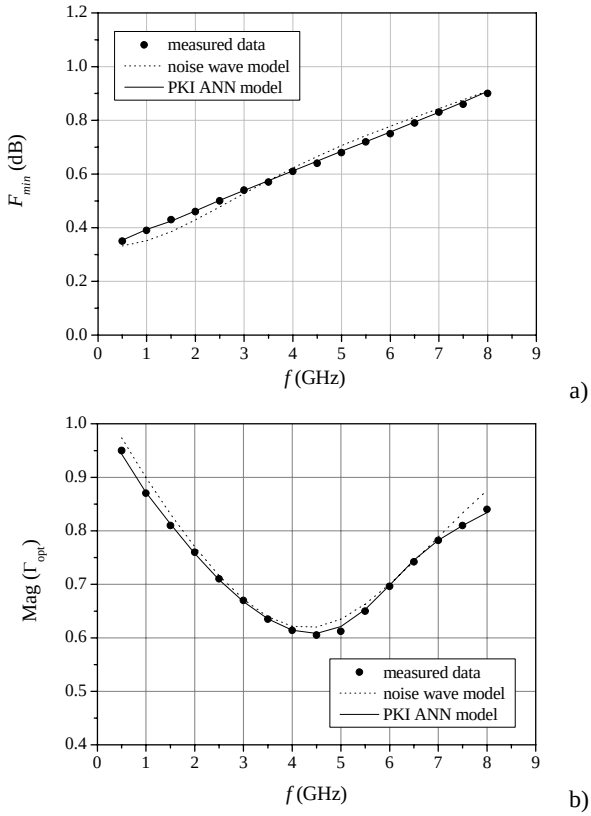


Fig. 3 Accuracy improvement in noise parameters modelling: a) minimum noise figure, b) magnitude of optimum reflection coefficient

In Fig. 4 and Fig. 5 there are plots of the noise parameters simulated by the proposed PKI ANN model and compared with the measured values. Comparison of measured and simulated values for the minimum noise figure, optimum source reflection coefficient and normalized noise resistance versus frequency is shown in Fig. 4. Circles denote the measured values and solid lines – the simulated ones obtained by the proposed PKI ANN model. It can be seen that the predicted noise parameter values match very well with the measured ones.

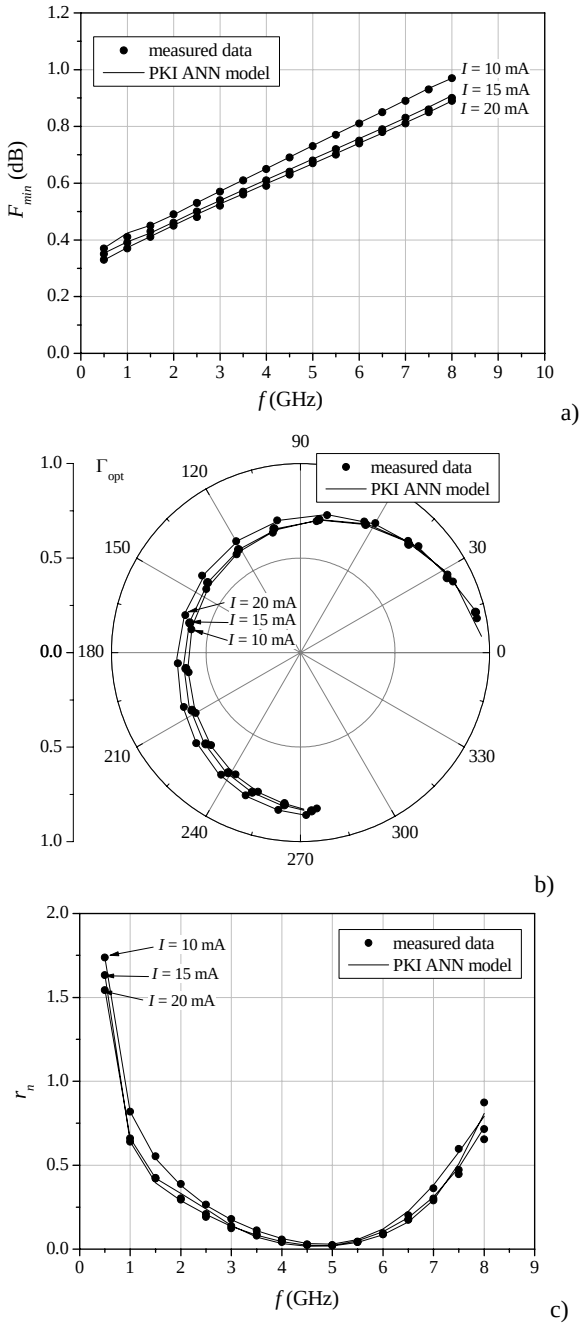


Fig. 4 Noise parameters versus frequency a) minimum noise figure, b) optimum reflection coefficient, c) normalised noise resistance

As a further illustration, the dependence of the minimum noise figure and the magnitude of optimum reflection coefficient on bias current is shown in Fig. 5. The presented data refer to frequencies (0.5, 2.5, 5 and 8) GHz. It can be observed that the simulated values match well the measured ones.

IV. Conclusion

Bias dependant noise parameters modelling of microwave FETs is proposed in the paper. The presented model consists of the empirical noise wave model, which does not directly

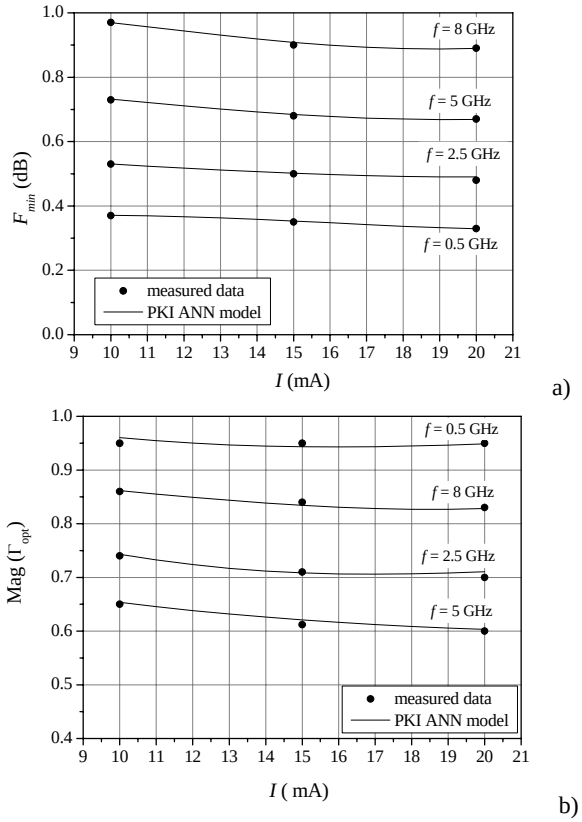


Fig. 5 Noise parameters versus bias current: a) minimum noise figure, b) magnitude of optimum reflection coefficient

incorporate the bias dependence, and ANN through which bias dependence is incorporated in the model. Besides the bias conditions and frequency, the ANN has as inputs the noise parameters simulated by the empirical model which represent additional knowledge about the modelled parameters. Therefore, this ANN is a PKI ANN.

Once developed, this model does not require any change in its structure or in values of its parameters. Extraction of the equivalent circuit parameters is performed only once, omitting in this way measurements and extraction procedures for each bias point. This makes bias-dependent noise modelling of microwave transistors less time consuming and more efficient.

The validity of the proposed bias dependant noise model is exemplified by modelling a specific MESFET device. The presented model has better accuracy than empirical noise wave model for the bias point used for the noise wave model development. Further, the presented model provides results that agree well with the measured characteristics in the considered bias and frequency operating range.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science and Technology Development of Serbia within the project No. TR-32052: 'Research and development of solutions for performance improvements of wireless communication systems in microwave and millimeter frequency bands'.

REFERENCES

[1] R.A. Pucel, H.A. Haus, H. Statz, *Signal and noise properties of gallium arsenide microwave field-effect transistors,* Advances in Electronics and Electron Physics, New York: Academic Press, pp. 195-265, 1975.

[2] Cappy, A., Vanoverschelde, A., Schortgen, A., Versnaeyen, C., Salmer, G., *Noise modelling in submicrometer-gate two-dimensional electron-gas field-effect transistors*, *IEEE Trans. Electron Devices*, Vol. 32, pp. 2787-2795, 1985.

[3] H. Fukui, *Design of microwave GaAs MESFET's for broad-band low-noise amplifiers*, *IEEE Trans. Microwave Theory Tech*, vol. 27, pp. 643-650, 1979.

[4] Gupta, M.S., Pitzalis, O., Rosenbaum, S.E., Greiling, P.T., *Microwave noise characterization of GaAs MESFETs: Evaluation by on-wafer low-frequency output noise current measurement*, *IEEE Trans. Microwave Theory Tech*, vol. 35, pp. 1208-1218, 1987.

[5] M. W. Pospieszalski, Modelling of noise parameters of MESFET's and MODFET's and their frequency and temperature dependence, *IEEE Trans. Microwave Theory Tech.*, vol. 37, pp. 1340-1350, 1989.

[6] R. P. Hecken, "Analysis of liner noisy two-ports using scattering waves", IEEE Trans. Microwave Theory Tech., vol. 29, pp. 997-1004, 1981.

[7] S.W. Wedge, D. B. Rutledge, "Wave techniques for noise modelling and measurement, *IEEE Trans. Microwave Theory Tech.*, vol. 40, pp. 2004-2012, 1992.

[8] O. Pronić, V. Marković and N. Maleš-Ilić: "MESFET Noise Modelling Based on Noise Wave Temperatures", in *Proc. TELSIKS'99*, Niš, Yugoslavia, 1999, pp. 407-410.

[9] O. Pronić and V. Marković, "A wave approach to signal and noise modelling of dual-gate MESFET", *AEÜ- Archiv für Elektronik und Ubertragungstechnik (International Journal of Electronics and Communications)*, vol. 56, no. 1, pp. 61-64, 2002.

[10] O. Pronić- Rančić, Z. Marinković, V. Marković, "An accurate procedure for noise wave modelling of microwave FETs versus temperature", *Microwave Review*, vol, 16, no. 2, pp. 37-41, 2010.

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

[12] F. Giannini, G. Leuzzi, G. Orengo, and M. Albertini, "Small-signal and large-signal modelling of active devices using CAD-optimized neural networks", *Int J RF Microw C E*, vol. 12, pp. 71-78, 2002.

[13] J. E. Rayas-Sanchez, "EM-based optimization of microwave circuits using artificial neural networks: The state-of-the-art," *IEEE Trans. Microw.Theory Tech.*, vol. 52, no. 1, pp. 420–435, Jan. 2004.

[14] Z. D. Marinković, O. R. Pronić, V. V. Marković: "Bias-dependent scalable modelling of microwave FETs based on artificial neural networks", *Microwave and Optical Technology Letters*, Vol. 48, No. 10, pp. 1932-1936, 2006

[15] Z. D. Marinković, O. R. Pronić-Rančić, V. V. Marković, "Bias-dependent models of microwave transistors based on PKI artificial neural networks," *Proceedings of the European Microwave Association*, vol. 4, no. 1, pp. 82-87, 2008.

[16] H. Kabir, Lei Zhang, Ming Yu, P. Aaen, J. Wood, and Q. J. Zhang "Smart modelling of microwave devices", *IEEE Microw Mag*, vol. 11, pp.105–108, 2010.

[17] Advanced Design Systems (ADS), Version 1.5, Agilent EEsof EDA, 2000.

Strong FEM Calculation of the Influence of the Conductor's Position on Quasi-Static Parameters of the Shielded Stripline With Anisotropic Dielectric

Žaklina J.Mančić[1] Vladimir V. Petrović[2]

***Abstract* – Strong formulation of the Finite Element Method is applied for calculation of effective permittivity and characteristic impedance of asymmetric shielded stripline filled with anisotropic dielectric. The influence of the position of the strip conductor is studied for two typical anisotropic substrates — sapphire and boron nitride. Rectangular elements of the third degree are applied.**

***Keywords* – Anisotropic dielectric, asymmetric stripline, shielded stripline, strong FEM formulation.**

I. Introduction

TEM transmission lines with anisotropic dielectric cannot be easily analyzed by the Method of Moments, even when dielectric is homogeneous and linear. In these cases Finite Element Method (FEM) [1, 2] is a very good choice. It was shown that strong FEM formulation of the Galerkin type [3,4] can be successfully applied to 2-D closed electrostatic problems for both isotropic [4, 5] and anisotropic [6, 7] dielectrics.

In this paper strong FEM formulation is applied for calculation of quasi-static parameters of the shielded asymmetric stripline with homogeneous anisotropic dielectric.

II. Basic FEM Methodology for Anisotropic Media

Consider a 2-D calculation domain, as shown in Fig.1. Starting for differential equation for electrostatic potential,

$$\mathrm{div}_S(\bar{\bar{\varepsilon}}\,\mathrm{grad}_S V)=0, \qquad (1)$$

where div_S is a surface divergence, grad_S is a surface gradient and $\bar{\bar{\varepsilon}}$ is a diagonal 2×2 permittivity tensor, $\bar{\bar{\varepsilon}}=\mathrm{diag}[\varepsilon_{xx}\ \varepsilon_{yy}]$. Calculation domain is divided into M subdomains (elements) and approximate solution for $V(x,y)$ presented as a linear combination of basis functions, $V\approx f=\sum_{j=1}^{N} a_j f_j$, where a_j are unknown coefficients. In this analysis rectangular elements are applied. Every basis function is nonzero on the surface of one element (for singlet functions) or on several neighboring elements (for doublet and quadruplet functions) [4, 6, 7]. According to Galerkin's procedure [1, 2], a system of linear algebraic equations is formed,

$$[K_{ij}][a_j]=[G_i], \quad i,j=1,\ldots,N, \qquad (2)$$

$$K_{ij}=\int_S(\mathrm{grad}\,f_i)(\bar{\bar{\varepsilon}}\,\mathrm{grad}\,f_j)\,\mathrm{d}S, \quad G_i=\int_{C_2} f_i D_{n0}\,\mathrm{d}l, \qquad (3)$$

where D_{n0} is the given normal component of vector **D** on contour C_2, i and j are global indices of basis functions, S is the surface of the calculation domain, $S=\bigcup_{e=1}^{M} S^e$, and S^e is the surface of the e-th element. By solving (2), unknown coefficients a_j are obtained, and, thus, the approximate distribution of potential V.

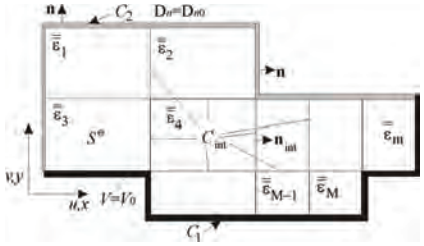


Fig. 1. Two-dimensional calculation domain divided into elements.

III. Strong FEM Basis Functions

In [3] strong FEM formulation for 1-D electromagnetic (EM) problems is introduced by the use of special, C^1 continuous, basis functions (strong basis functions). In corresponding formulation for 2-D electrostatic problems [4, 5] strong 2-D basis functions automatically satisfy continuity of potential V (C^0 continuity) and continuity of D_n (generalized C^1 continuity) on interelement boundaries (C_{int} in Fig.1). Complete set of strong basis functions for 2-D problems in homogeneous (isotropic or anisotropic) media consists of four types of quadruplets (Fig.2), two types of doublets (Fig.3) and various singlets (Fig.4).

[1]Žaklina Mančić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: zaklina.mancic@elfak.ni.ac.rs.

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

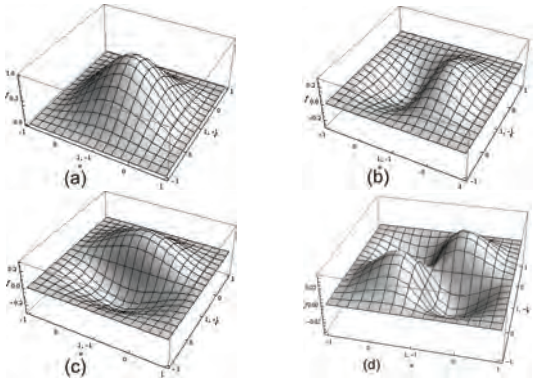


Fig. 2. Quadruplets for the strong FEM formulation. They provide continuity of (a) V, (b) D_{nx}, (c) D_{ny}, and (d) both D_{nx} and D_{ny}.

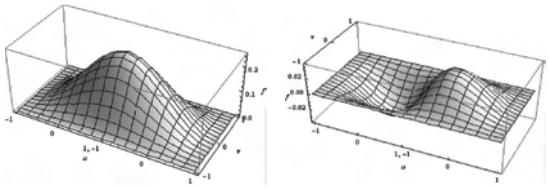


Fig. 3. Doublets for the strong FEM formulation.

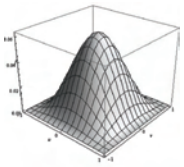


Fig. 4. Typical singlet for the strong FEM formulation.

IV. NUMERICAL RESULTS

Two geometries, depicted in Fig.5 and Fig.11, are considered. Although formally the same, the two geometries differ in orientation (horizontal and vertical) of the strip. Two typical anisotropic substrates for planar transmission lines are considered. The first is sapphire ($\varepsilon_{rxx}=9.4, \varepsilon_{ryy}=11.6$), and the second is boron nitride, ($\varepsilon_{rxx}=5.12, \varepsilon_{ryy}=3.4$). Effective relative permittivity, ε_{re}, and characteristic impedance, Z_c, were calculated by standard formulas for non-magnetic dielectrics, $\varepsilon_{re}=C'/C_0'$ and $Z_c=\left(c_0\sqrt{C'C_0'}\right)^{-1}$, where C' is the capacitance per unit length of the analyzed line, C_0' is its capacitance per unit length when dielectric is replaced by vacuum, and c_0 is the speed of light in free-space. Capacitances per unit length were calculated by the use of energy formula,

$$W'=\frac{1}{2}C'U^2=\sum_e\int_{S^e} w_e\mathrm{d}S=\frac{1}{2}\sum_e\int_{S^e}\mathbf{E}\cdot\mathbf{D}\mathrm{d}S,\quad \mathbf{E}=-\mathrm{grad}V$$

We used $N=512$ elements of the third order in both *x*- and *y*-directions ($n_x=n_y=3$). The third-order elements represent the lowest-order approximation for the strong FEM formulation [4].

IV.A. Example 1

For the stripline whose cross-section is given in Fig.5, the strip is parallel to the wider side of the rectangular shield. Relative dimensions of the line are chosen as: $a/b=3$, $w/b=1$, $t/b=0.1$, while h/b and s/b are variable.

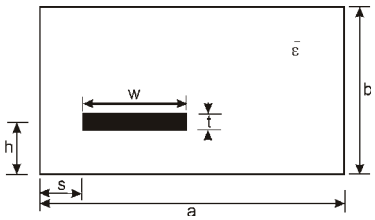


Fig. 5. Geometry of the first analyzed structure.

In Figs.6 and 7 relative effective permittivity, ε_{re}, and characteristic impedance, Z_c, are presented for sapphire, and in Figs. 8 and 9, the same parameters are presented for boron nitride. In Fig.10, field lines for sapphire, $h/b=0.45$, $s/b=0.5$ are depicted (by the use of the commercial software [8]).

From those four diagrams it can be seen that ε_{re} and Z_c are changing rapidly when the strip is close to the shield. When the strip is relatively far from the shield, these parameters are changing slowly.

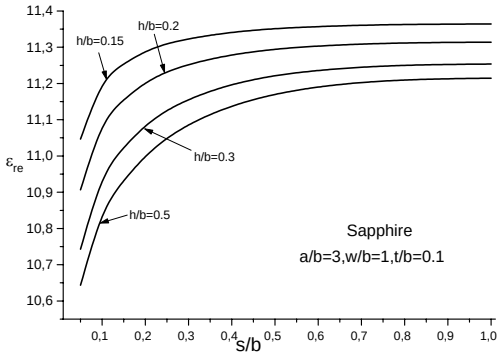


Fig. 6. Effective relative permittivity of the shielded stripline filled with sapphire, as a function of s/b.

From Figs.6 and 8 can be seen that when the strip approaches the wall of the shield, ε_{re} tends to values closer to ε_{rxx}. This can be explained by the edge effect in which the *x*-component of the field is dominant. When the strip moves away from the wall towards the center of the structure, the *y*-component of the field relatively increases (compared to the *x*-component), and ε_{re} tends to values closer to ε_{ryy}. When the strip is closer to the bottom of the structure, ε_{re} tends to values closer to ε_{ryy}, as the *y*-component of the field then

increases even more. The sketch of electric field lines, shown in Fig.10, can help for visual illustrations of these effects.

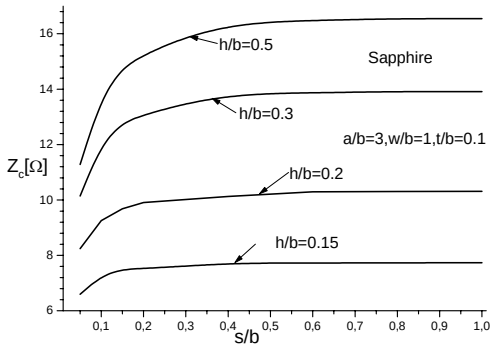


Fig. 7. Characteristic impedance of the shielded stripline filled with sapphire, as a function of s/b.

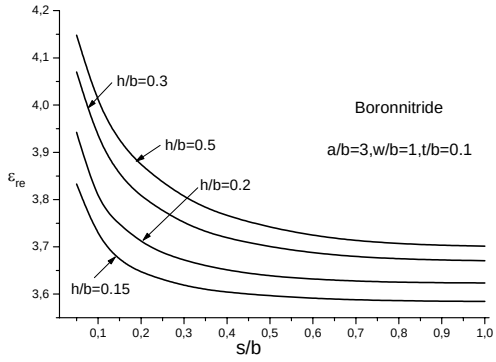


Fig. 8. Effective relative permittivity of the shielded stripline filled with boron nitride, as a function of s/b.

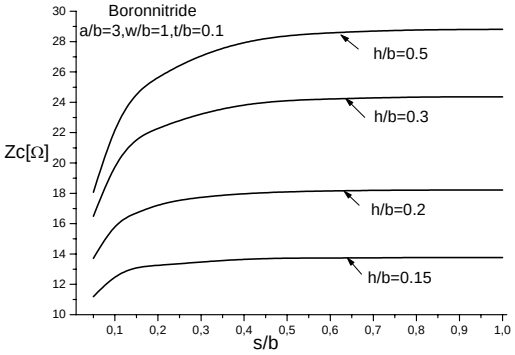


Fig. 9. Characteristic impedance of the shielded stripline filled with boron nitride, as a function of s/b.

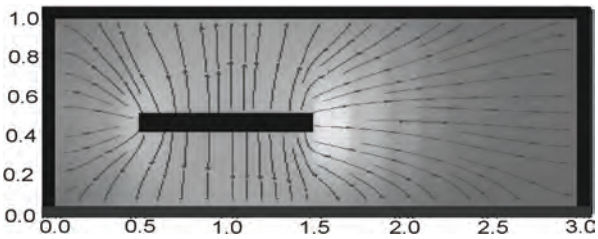


Fig. 10 Electric field lines for sapphire ($a/b=3,\ w/b=1,\ t/b=0.1,\ h/b=0.45, s/b=0.5$).

As for the characteristic impedance, in all cases it increases with moving the strip away from the shield towards the center of the structure. This is due to decrease of the capacitance per unit length and is not directly associated with anisotropic characteristics of the substrate.

IV.B. Example 2

For the stripline whose cross-section is given in Fig.11, the strip is parallel to the narrower side of the rectangular shield. Relative dimensions of the line are chosen as: $a/b=3$, $w/b=0.5$, $t/b=0.1$, while h/b and s/b are variable.

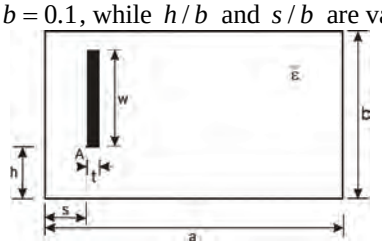


Fig. 11. Geometry of the second analyzed structure.

In Figs.12 and 13 relative effective permittivity, ε_{re}, and characteristic impedance, Z_c, are presented for sapphire, and in Figs. 14 and 15, the same parameters are presented for boron nitride. In Fig.16, field lines for sapphire, $h/b=0.45$, $s/b=0.5$ are depicted (by the use of the commercial software [8]).

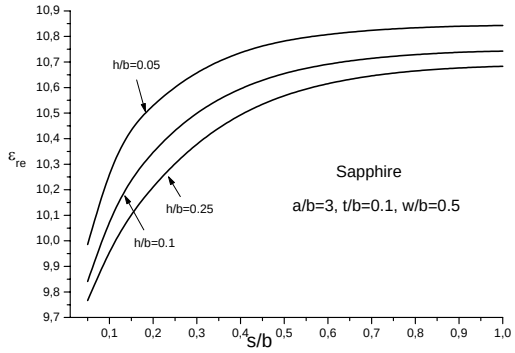


Fig. 12. Effective relative permittivity of the shielded stripline filled with sapphire, as a function of s/b.

Those diagrams practically illustrate the same effects as in example 1.

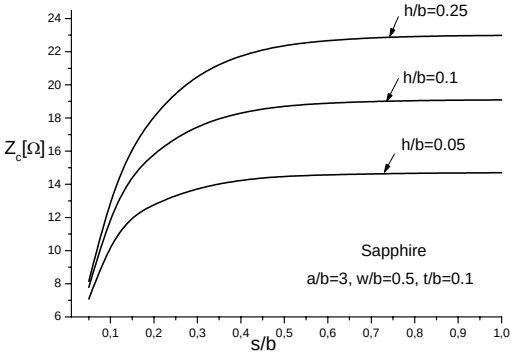


Fig. 13. Characteristic impedance of the shielded stripline filled with sapphire, as a function of s/b.

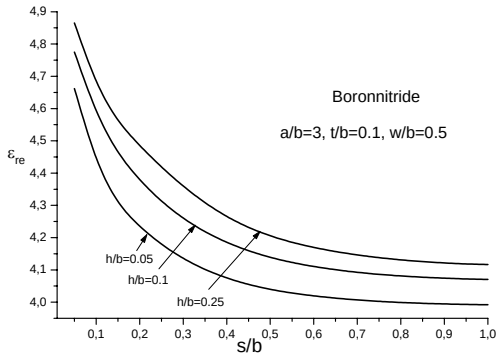


Fig. 14. Effective relative permittivity of the shielded stripline filled with boron nitride, as a function of s/b.

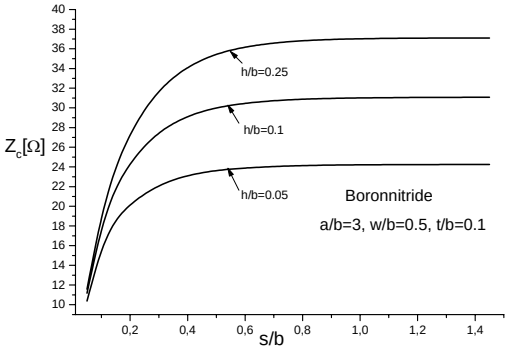


Fig. 15. Characteristic impedance of the shielded stripline filled with boron nitride, as a function of s/b.

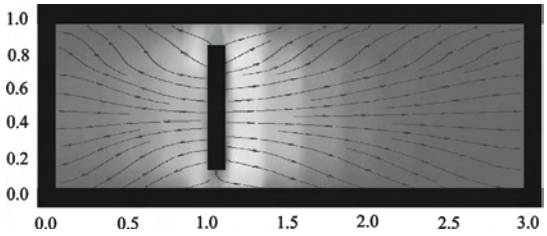


Fig. 16. Electric field lines for sapphire ($h/b = 0.1,\ s/b = 1, a/b = 3,\ t/b = 0.1,\ w/b = 0.8$).

V. CONCLUSION

Quasi-static parameters of asymmetric shielded stripline, are obtained by the strong FEM formulation for various positions of the strip. It was shown that effective relative permittivity of the stripline tends to values closer to ε_{rxx} when the strip moves closer to the side wall, and to values closer to ε_{ryy} when the strip moves closer to the bottom wall. This behavior can be explained by the edge effect.

Characteristic impedance increases with moving the strip away from the shield towards the center of the structure, which is due to general decrease of the capacitance per unit length and is not directly associated with the anisotropy of the substrate.

ACKNOWLEDGEMENT

This research was partly funded by the Serbian Ministry of science and technological development, under grant TR-32052.

REFERENCES

[1] J. Jin, *The Finite Element Method in Electromagnetics*, Wiley:NewYork, 1993.

[2] P. P. Silvester, *Finite Elements for Electrical Engineers*, 3rd Ed., Cambridge University Press, 1996.

[3] V. V. Petrović, B. D. Popović, "Optimal FEM solutions of one-dimensional EM problems," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, Vol. 14, No. 1, pp. 49–68, Jan-Feb 2001.

[4] Ž. J. Mančić and V. V. Petrovic, "Strong and Weak FEM Formulations of Higher Order for Quasi-Static Analysis of Shielded Planar Transmission Lines," *Microwave and Optical Technology Letters*, Vol. 53, No. 5, pp. 1114–1119, May 2011.

[5] Ž. J. Mančić, S. R. Aleksić, Vladimir V. Petrović, "Comparison of FEM, MoM and EEM in Solving a Benchmark 2D Electrostatic Problem," Proc. 8th International Conference On Applied Electromagnetics – PES 2007, Niš, Serbia, September 3–5, 2007.

[6] Z. J. Mancic, V. V. Petrovic, "Analiza kvadratnog koaksijalnog voda sa anizotropnim dielektrikom metodom konačnih elemenata," Proc. XVIII Telekom. forum TELFOR, Belgrade, Serbia, November 2010, pp. 850–853 (in Serbian).

[7] V. V. Petrovic, Z. J. Mancic, "Strong FEM Formulation for Quasi-Static Analysis of Shielded Planar Transmission Lines in Anisotropic media," 5th European Conference on Circuits and Systems for Communications (ECCSC'10), Belgrade, Serbia, November 2010, pp. 253–255.

[8] *Mathematica*, Ver.7, Wolfram Research, Inc., USA.

System of Square-shaped Electrodes as a Pillar Grounding System

Nenad N. Cvetković[1]

***Abstract* – The pillar grounding system formed of two horizontal square-shaped wire electrodes and an iron armature connected to the main grounding connector is analyzed in the paper. Recently proposed procedures for modeling the influence of a concrete cylinder on a grounding system, and for approximating a square-shaped electrode with a ring electrode, are used.**

***Keywords* – Ground inhomogeneity, grounding systems, method of moments, pillar foundation, quasi-stationary EM field.**

I. INTRODUCTION

Pillars used as a part of the overhead power, telecommunications, or lightning protection system necessarily include corresponding grounding systems. A basic electrode of such systems is usually formed of a basic star, rectangular or circularly shaped electrode [1-2] connected to an iron armature of the pillar foundation, which can be treated as a second part of the grounding system.

Official publications, as [1], usually neglect influence of concrete foundation. This influence can be modelled as a hemispherical, [3-4], inhomogeneity in a shape of a parallelepiped [5-7]. A simple procedure for approximate modelling of this influence was proposed in [5] and it provides simplification of the analyzed problem to a problem of a grounding sys tem in the homogeneous ground. The level of foundation's influence depends on electrical parameters of the concrete and the ground, which can have different values depending on the ground structure, humidity, etc.

In [7], a grounding system having one rectangular basic electrode is analyzed in the quasi-stationary regime. The same approach is applied in this paper to the analysis of the grounding system having main electrodes' system formed of two rectangular basic electrodes, Fig. 1. Leakage currents` distributions are assumed as constant while the earthing conductor's influence is neglected. The surrounding ground is modelled as a homogeneous semi-conducting media of known electrical parameters, while the feeding current is modelled by an ideal LF current generator. Calculations are carried out for various values of concrete's specific resistivity and different embedding depths, whose values correspond to those of grounding systems realized in practice [1-2, 8-11].

As in [9], the obtained results are also compared with the results for ring basic electrodes [2, 4], since such problem is simpler from the numerical point of view, and for some approximations it is possible to analyze circular instead of rectangular electrode making error which is acceptably small**.**

[1]Nenad N. Cvetković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73, 18000 Niš, Serbia, E-mail: nenad.cvetkovic@elfak.ni.ac.rs.

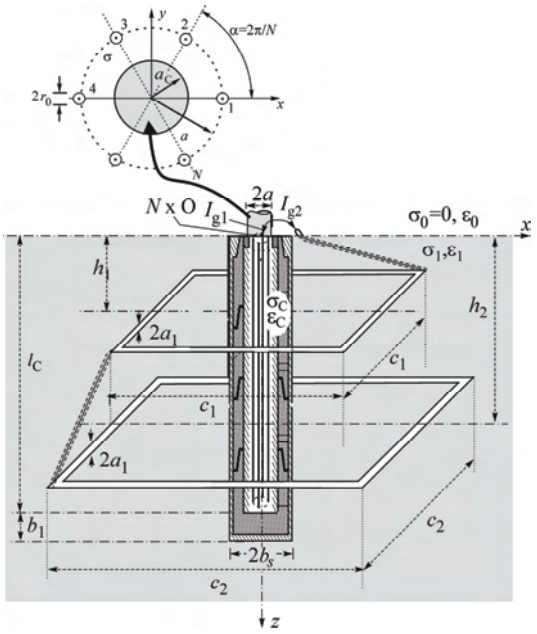


Fig. 1. Pillar grounding system.

II. THEORETICAL BACKGROUND

A. Description of the Grounding System

The pillar foundation grounding system formed of two square-shaped wire electrodes of sides c_1 and c_2, and embedded at depths h_1 and h_2, is observed (Fig. 1). The electrode is made of a strip conductor (usually FeZn strip) assumed to be a wire conductor having circular cross-section of an equivalent radius a_1, [4]. The structure of the observed grounding system also includes a single vertical electrode, which actually models the wire armature cage formed of N parallel conductors of length l_C, having a circular cross-section of radius r_0, (upper part of Fig. 1). The equivalent vertical wire electrode is of the same length l, and the circular cross-section of equivalent radius, $a_C = a \sqrt[N]{N\, r_0 / a}$, [5]. Armature (i.e. equivalent vertical wire electrode) and square-shaped ground electrodes' system are connected at the main ground point.

Foundation is parallelepiped concrete domain of a square cross-section having side length b_s, while concrete is of specific conductivity σ_C and permittivity ε_C. The surrounding ground is assumed as linear, isotropic and homogeneous semi-conducting media of specific conductivity σ_1 and permittivity ε_1. For the assumed quasi-stationary regime, complex con-

ductivities are $\underline{\sigma}_k = \sigma_k + \mathrm{j}\omega\varepsilon_k \approx \sigma_k$, $k = 0,1,\mathrm{C}$, (for air, surrounding ground and concrete, respectively), while ω is angular frequency. Also, since for air $\sigma_0 = 0$, the value of the complex reflection coefficient can be considered: $R_{i0} = (\underline{\sigma}_i - \underline{\sigma}_0)/(\underline{\sigma}_i + \underline{\sigma}_0) \approx 1$, $i = 1,\mathrm{C}$.

Applying the procedure described in [5] the vertical electrode system and the concrete foundation can be replaced by a vertical electrode of equivalent length $l_1 = \mathrm{K_e}\, l_\mathrm{C}$ and cross-section radus $a_\mathrm{e} = \mathrm{K_e}\, a_\mathrm{C}$, placed in a homogeneous ground of specific conductivity σ_1, Fig. 2. Parameter $\mathrm{K_e}$ is determined from the expression:

$$\mathrm{K_e^{-1}} = \sigma_1/\sigma_\mathrm{C} + (1-\sigma_1/\sigma_\mathrm{C})\ln(1+l_\mathrm{C}/b)/\ln(1+l_\mathrm{C}/a_\mathrm{C}), \qquad (1)$$

where is $b = b_s(1+\sqrt{2})/4$, [5].

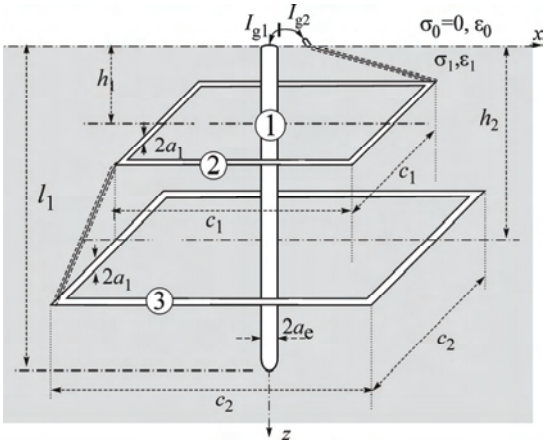


Fig.2. Equivalent grounding system in homogeneous ground.

This way, the problem is reduced to analysis of the equivalent grounding system in the homogeneous ground shown in Fig. 2. Equivalent vertical electrode (labeled by 1) and square-shaped wire electrodes' system (electrodes labeled by 2 and 3) are fed by very low frequency currents I_{g1} and I_{g2}, respectively. For practical values it is reasonable to assume that $a_\mathrm{e} << l_1$, $a_1 << l_2, l_3$ and $a_1, a_\mathrm{e} << \lambda_1$ ($l_2 = 4c_1$ and $l_3 = 4c_2$ are length of electrodes 2 and 3, while λ_1 is wavelength in the surrounding ground). Unknown longitudinal current distributions along the vertical electrode (1) and square wire electrodes (2 and 3) are labeled by $I_k(s'_k)$ $s'_k \in [0, l_k]$, $k = 1,2,3$. Generally, for the purpose of obtaining the grounding resistance, the current distribution in the grounding network is not critical for the calculation of an accurate resistance value. Therefore, leakage currents` distributions are assumed as constant,

$$I_{\mathrm{leak}\,k}(s'_k) = -\partial I_k(s'_k)/\partial s'_k = -I'_k(s'_k) = I_{\mathrm{L}\,k}/l_k, \; k = 1,2,3. \qquad (2)$$

where $I_{\mathrm{L}k}$, $k = 1,2,3$ are total leakage currents from the corresponding conductors. Since earthing conductors influence has not been taken into consideration, and $I_1(l_1) = 0$, the following conditions are satisfied

$$I_{\mathrm{L}1} = I_{\mathrm{g}1}, \; I_{\mathrm{L}2} + I_{\mathrm{L}3} = I_{\mathrm{g}2}. \qquad (3)$$

B. Electrical Scalar Potential and Determining of Unknown Current Distributions and "$\underline{Z}$" Parameters

Taking into consideration all presumptions pointed out in the previous text, electric scalar potential at the points defined by the field vector $\vec{r}$ in the vicinity of the equivalent grounding system shown in Fig. 2, can be determined using the expression

$$\varphi(\vec{r}) = \sum_{k=1}^{3} \frac{I_{\mathrm{L}\,k}}{4\pi\sigma_1 l_k} \int_{l_k} \left(\frac{1}{r_k} + \frac{1}{r_{k\,\mathrm{i}}} \right) \mathrm{d}\, s'_k, \qquad (4)$$

where r_k and $r_{k\,\mathrm{i}}$ denote distances between the current element, i.e. its image, and the field point, respectively. Since the quasi-stationary approach is applied, it is assumed that the surface of the electrode 1 is equipotential, and that the same is valid for square-shaped electrodes 2 and 3 that are of the same potential value. Applying the Method of Moments [12] and matching potential value on the n-th ($n = 1,2,3$) conductor`s surface defined by the field vector $\vec{r}_1 = (l_1/2)\,\hat{z}$ (vertical conductor), $\vec{r}_2 = 0.5c_1\,\hat{x}$ and $\vec{r}_3 = 0.5c_2\,\hat{x}$ (points on the square-shaped electrodes), the following system of integral equations is formed:

$$\varphi(\vec{r} = \vec{r}_1) \cong U_1, \; \varphi(\vec{r} = \vec{r}_k) \cong U_2, \; k = 2,3. \qquad (5)$$

In order to solve the described system, it is needed to adopt potential values U_1 and U_2. The system is solved in two regimes, so-called symmetric ($U_1 = U_2 = 1\,\mathrm{V}$) and anti-symmetric ($U_1 = -U_2 = 1\,\mathrm{V}$) feeding regimes. Based on these solutions, the solution for "Z" parameters of the electrode system are obtained [4-5]. They are formulated as

$$U_1 = \underline{Z}_{11} I_{\mathrm{g}1} + \underline{Z}_{12} I_{\mathrm{g}2}, \; U_2 = \underline{Z}_{21} I_{\mathrm{g}1} + \underline{Z}_{22} I_{\mathrm{g}2}, \qquad (6)$$

where $\underline{Z}_{11}$ and $\underline{Z}_{22}$ are self-impedances, while $\underline{Z}_{12}$ and $\underline{Z}_{21}$ are mutual-impedances of the wire electrode 1 and the square electrodes' system (electrodes 2 and 3). "Z" parameters represent integral grounding system characteristics and indicate level of mutual-influence between the two electrodes. For linear systems $\underline{Z}_{12} = \underline{Z}_{21}$. If electrodes are connected, i.e. they form a unique grounding system, substituting $U_1 = U_2 = 1\,\mathrm{V}$ in (6), grounding impedance can be determined as

$$\underline{Z}_\mathrm{T} = 1/(I_{\mathrm{g}1} + I_{\mathrm{g}2}) = \frac{\underline{Z}_{11}\underline{Z}_{22} - \underline{Z}_{12}\underline{Z}_{21}}{\underline{Z}_{11} + \underline{Z}_{22} - \underline{Z}_{12} - \underline{Z}_{21}}. \qquad (7)$$

C. Approximation of Square-shaped Wire Electrode by an Equivalent Ring Electrode

As it has been already emphasized, the pillar grounding system with circular (instead of square) basic electrode is more simple for numerical solving. Because of that, it could be of interest to find some appropriate relationship between the square side and the radius of the equivalent circular wire electrode that would approximately model the square electrode. Such analysis has been already carried out for the case of a single square-shaped electrode as a main ground electrode

[7]. A case of the square electrodes being approximated by rings of the same length and same surface, as well as the case of ring electrodes having equivalent radii determined by estimation method [4], is analyzed in this paper. The mathematical model has the same form, and in that case l_2 and l_3 in (4) label contours of the ring electrode, instead of square wire electrodes. Matching points are chosen on ring surfaces, and presumption of constant leakage current also stays valid.

Related to the square side *c*, which in this case takes values c_1 or c_2, equivalent ring radius *R* is now, [9],

$$R = 0.25c\left(1+\sqrt{2}\right) \text{ (estimation method)}, \quad (8a)$$

$$2R\pi = 4c \Rightarrow R = 2c/\pi \text{ (same length)}, \quad (8b)$$

$$R^2\pi = c^2 \Rightarrow R = c/\sqrt{\pi} \text{ (same surface)}. \quad (8c)$$

III. Numerical Results

Based on the presented model, the corresponding program packages are developed and applied to approximate solving of the pillar foundation grounding system formed of a vertical conductor in a concrete cylinder and two square-shaped electrodes, Fig. 1. Values of geometry parameters and concrete specific conductivity are selected according to the values from [1, 8-11]. Used values of all parameters in Figs. 1 and 2 are: $\rho_1 = 1/\sigma_1 = 100\,\Omega\text{m}$, $a = 0.25\,\text{m}$, $b_s = 0.4\,\text{m}$, $b_1 = 0.2\,\text{m}$, $l_C = 2\,\text{m}$, $c_1 = 1.5\,\text{m}$, $c_2 = 2.5\,\text{m}$, $h_1 = 0.5\,\text{m}$, $a_1 = 9.7\,\text{mm}$, $r_0 = 0.007\,\text{m}$ (radius of armature conductors` cross-section), $N = 10$, $b = 0.483\,\text{m}$, and $a_C = 0.18\,\text{m}$.

Mutual-resistance $R_{12} = \text{Re}\{\underline{Z}_{12}\}$ of the system from Fig. 1 versus embedding depth of the lower square electrode h_2, having concrete specific resistivity $\rho_C = 1/\sigma_C$ taken as parameter, is shown in Fig. 3. It is noticeable that the position of the lower square electrode related to vertical electrode's system influences the mutual resistance, and this influence differs for different specific resistivity concrete/ground ratio (ρ_C/ρ_1).

Mutual resistance R_{12} of the system from Fig. 1, versus the embedding depth h_2 when $\rho_1 = \rho_C$, for the square electrodes and different radii of equivalent ring electrodes (8a-c) is presented in Fig. 5. It is obvious that best approximations of the square electrodes are the ring electrodes of radii (8c).

In Fig. 6, graphs of the self-resistance $R_{22} = \text{Re}\{\underline{Z}_{22}\}$ versus the embedding depth h_2 when $\rho_1 = \rho_C$, for the square electrodes and different radii of equivalent ring electrodes (8a-c), are presented. In this case, rings having radii given by the expression (8a) are the best approximation of the square electrodes.

Finally, in Fig. 7, the total resistance R_T of the grounding system from Fig. 1 versus the embedding depth h_2 when $\rho_1 = \rho_C$, for square electrodes and different radii of the equivalent ring electrode (8a-c) is shown. The best approximation of the square electrode is using ring electrodes of radii (8c) and (8a), depending on the embedding depth h_2.

Total resistance $R_T = \text{Re}\{\underline{Z}_T\}$ of the system from Fig. 1, versus embedding depth h_2, and concrete specific resistivity $\rho_C = 1/\sigma_C$ taken as parameter, is shown in Fig. 4. It is expected that the total resistance will decrease while increasing depth h_2, and also, that the curves corresponding to different values of specific resistivity of concrete are mutually shifted.

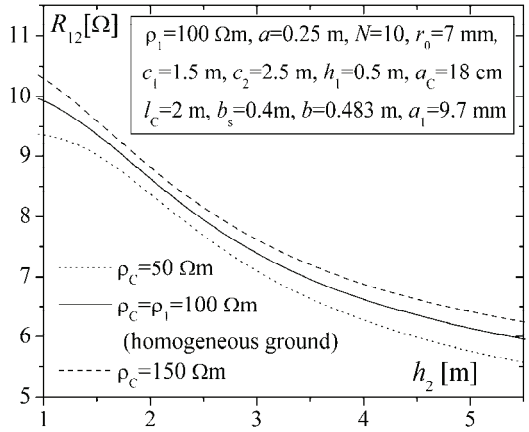


Fig.3. Mutual resistance of the grounding system shown in Fig. 1 versus embedding depth h_2, when ρ_C is taken as a parameter.

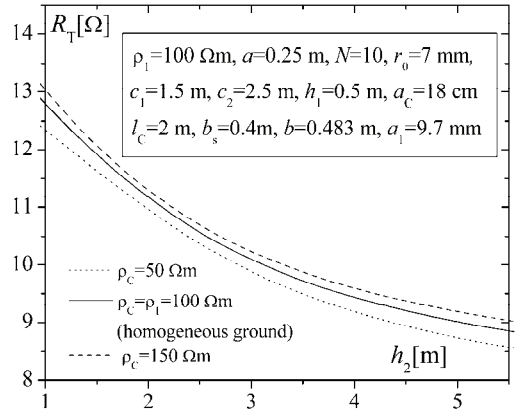


Fig.4. Total resistance of the grounding system shown in Fig. 1 versus embedding depth h_2, when ρ_C is taken as a parameter.

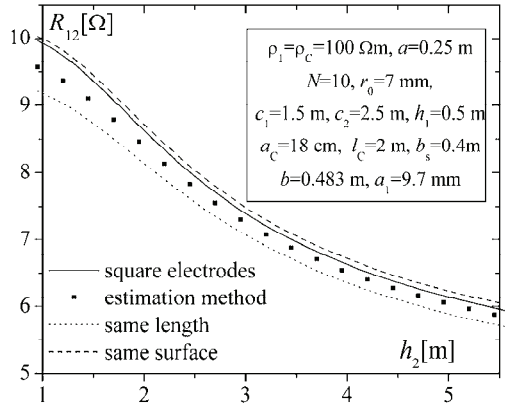


Fig.5. Mutual-resistance of the grounding system shown in Fig. 1 versus embedding depth h_2 for ρ_1=ρ_C and different dimensions of the equivalent ring electrodes.

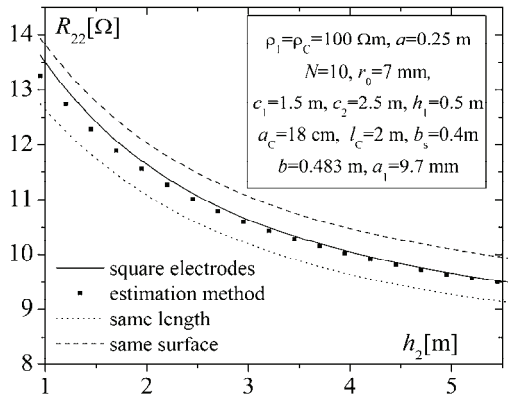


Fig.6. Self- resistance R_{22} corresponding to the basic electrode of the grounding system shown in Fig. 1 versus embedding depth h_2 for $\rho_1=\rho_c$ and different dimensions of the equivalent ring electrodes.

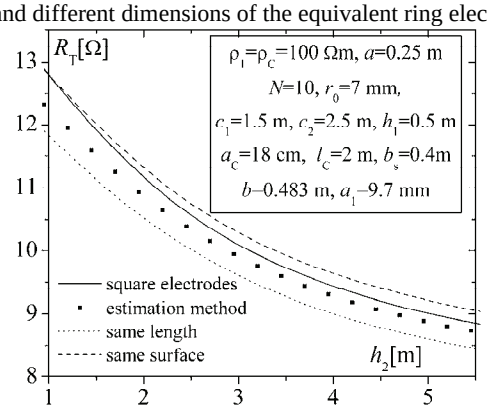


Fig.7. Total resistance of the grounding system shown in Fig. 1 versus embedding depth h_2 for $\rho_1=\rho_c$ and different dimensions of the equivalent ring electrodes.

IV. Conclusions

The analysis of the influence of the pillar foundation on characteristics of the pillar grounding system is carried out in this paper and the results are obtained in the quasi-stationary regime assuming the surrounding ground as a linear homogeneous semi-conducting media of known electrical parameters. A grounding system realization applied in practice, with a rectangular wire conductors' system as a basic electrode, is analyzed. The foundation influence is modeled using a recently proposed procedure [7], based on approximating the pillar foundation by a concrete domain having parallelepiped shape of square cross-section.

Based on the obtained results one can conclude that the foundation`s influence on grounding system`s impedance can be significant for real values of concrete`s specific resistivity and it should be taken into consideration during the design of such systems. Since from the numerical point of view, it is easier to solve an analogue grounding system with a ring basic electrodes' system, different approaches for approximating a square electrode with an equivalent ring electrode are also presented in the paper.

Presented results indicate a possibility to approximate well the square electrode using an equivalent ring electrode, but the most appropriate approximations differ depending on the embedding depth of the basic electrode. Another collateral conclusion that follows the presented analysis is the significance of the influence of the shape of the electrodes, which form basic electrodes' system, on the self-, mutual- and consequently, the total impedance. The obtained results show that selection of the approximating method given by Eqs. (8) depends on the number of the electrodes in the basic electrodes' system, since the results presented in the paper differ a little bit from those ones given in [7].

Acknowledgements

This paper is partly supported by the project grant III44004 (2011-2014) financed by the Ministry of Education and Science of Republic of Serbia.

The results presented in the paper were obtained by the joint research of the paper author with prof. Predrag D. Rančić (1947-2011) from the University of Niš, Faculty of Electronic Engineering. Author is grateful to prof. Rančić for his support and help during the years of successful collaboration.

References

[1] *Technical Recommendation of Electric Power Industry of Serbia: Grounding of the Power Network Pillars, TR No .9*, Belgrade, Serbia, 2000. [in Serbian]. Available at: http://www.eps.rs/publikacije/interni_standardi/preporuke/tp9.zip.

[2] J. Nahman, *Neutral Grounding of Distribution Networks*, Naučna knjiga, Belgrade, Serbia, pp. 281-288, 1980. [in Serbian]

[3] N. N. Cvetković and P. D. Rančić, "A Simple Model for a Numerical Determination of Electrical Characteristics of a Pillar Foundation Grounding System", *Engineering Analysis with Boundary Elements*, Elsevier, vol 33, pp. 555-560, 2009.

[4] N. N. Cvetković and P. D. Rančić, "Influence of Foundation on Pillar Grounding System's Characteristics", *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, Emerald Group Publishing Limited, vol. 28, no 2, pp. 471-492, 2009.

[5] N. N. Cvetković and P. D. Rančić, "Model for Analyzing a Pillar Grounding System with Cylindrically-Shaped Concrete Foundation", *Electromagnetics*, vol. 29, no. 2, pp. 151–164, 2009.

[6] P. D. Rančić and N. N. Cvetković, "Pillar Star-shaped Grounding System", *Przeglad Elektrotechniczny (Electrical Review)*, R. 86 NR 5/2010, pp.100-104, 2010.

[7] P. D. Rančić and N. N. Cvetković, "Square shaped Electrode as Pillar Grounding System", *TELSIKS`09, Conference Proceedings*, pp. 351-354, Niš, Serbia, 2009.

[8] T. Radojičić, *Concrete Constructions*, Prosveta, Niš, Serbia, 1981. [in Serbian]

[9] "Ferro Concrete Montage Foundation EBB-MT1", *Web presentation of ELEKTROIZGRADNJA Bajina Bašta Company*, 2009. [In Serbian]. Available at: http://www.ebb.rs.

[10] M. Chouteau and S. Beaulieu, "An Iinvestigation on Application of the Electrical Resistivity Tomography Method to Concrete Structures", *Geophysics*, Los Angeles, CA, USA, 2002. Available at: http://www.dot.ca.gov/hq/esc/geotech/gg/geophysics2002/050chouteau_resistivity_tomography.pdf.

[11] J. Ribič and J. Podlipnik, "Grounding and Negative Effects of Using Zinc Ground Electrodes in Objects Foundation", *Web presentation of Hermi Company*, 2008. [In Serbian]. Available at: http://www.hermi.si/sr/Projektanti/Uzemljenje/tabid/523/Default.aspx

[12] R. F. Harrington, *Field computation by Moment Methods*, New York: The Macmillan Company, 1969.

Session MRS:

METROLOGY AND REMOTE SENSING

Detecting the Direction of the Shaft Rotation by Using Incremental and Virtual Absolute Encoders

Dragan Denić[1], Jelena Lukić[2], Aleksandar Jocić[3], Miroljub Pešić[4] and Dejan Prolović[5]

***Abstract* – The position encoders are widely used sensors for detection of angular and linear motion in many systems such as machine tools, industrial robots, a variety of instruments, computer peripherals. The classification and the main properties of position encoders are presented in this paper. Special attention is given to the optical rotary encoders i.e. incremental, absolute and virtual absolute encoders. In the second part of the paper are presented solutions for detecting the movement direction of the shaft by using incremental and virtual absolute encoders. The solutions are implemented and tested in LabVIEW environment.**

***Keywords* – Optical rotary encoders, Rotation direction, Virtual instrumentation, LabVIEW**

I. INTRODUCTION

Linear and angular position measurement of a movable system (MS) is essential for the functioning of many systems, such as machines and robots in industry, computer peripherals, generators in power plants, radars, telescopes, etc. These measurements are usually performed by using the position encoders [1]. The easiness with which position encoders are interfaced to digital systems made them very popular. Due to increased complexity of the modern automated industrial systems, which increase demands for position measurement in sense of resolution, reliability, accuracy and etc., the different types of encoders were developed over time. The Section 2 of this paper will be devoted to the classification of different types of position encoders. After the Section 3, where the key features of incremental, absolute and virtual absolute encoders will be mentioned, two different solutions for determination of shaft rotation direction will be presented, and they can be applied with incremental and virtual absolute encoders.

[1]Dragan Denić is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.denic@elfak.ni.ac.rs.

[2]Jelena Lukić is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: jelenalukicpk@gmail.com.

[3]Aleksandar Jocić is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: aleksandar.jocic@elfak.ni.ac.rs.

[4]Miroljub Pešić is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: miroljub.pesic@elfak.ni.ac.rs.

[5]Dejan Prolović is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: prolovic@gmail.com.

II. CLASSIFICATION OF POSITION ENCODERS

By observing the classification shown in Fig. 1, the first division of encoders to non-contact and contact encoder types is done depending on whether there is the physical contact during code reading. The group of non-contact encoders is consisted of magnetic, inductive and optical encoders, where the division is made according to the operation principle of the sensing element of encoder. The most common contact encoder type is electrical. Each of the previously mentioned encoder types can be further divided, by the shape of coding scale, into rotary and linear types. In this paper will be analyzed the properties of rotary optical encoders, so there will not be given further classification and analysis of other types of encoders. By the position finding method, optical rotary encoders can be divided into absolute, incremental and virtual absolute encoders [1].

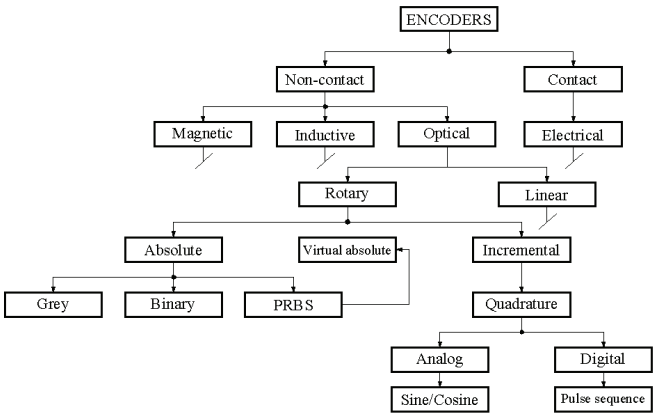


Fig. 1. The classification of position encoders

Absolute encoders give the position information in unique digital form, while the incremental encoders determine position by counting pulses. Virtual absolute encoders are new type of absolute encoders, and also the result of effort to avoid the application of a large number of coding tracks used in the absolute encoders for high resolution measurements. This was achieved by using pseudorandom binary sequences (PRBS). Except pseudorandom binary sequences, widely used with absolute encoders are two more position coding methods - binary (so-called natural) coding and Gray coding. The final classification is based on the type of coding method in case of absolute, and on the type of output signal in case of incremental encoders. In the case of incremental encoders at the output are generated quadrature signals, phase shifted by quarter of signal period, i.e. 90°. Quadrature signals can have analog format (sine and cosine signals) and digital format (sequence of rectangular pulses).

III. Characteristics of Optical Rotary Encoders

Optical rotary encoders have perforated code disc made of glass, Mylar film or metal, over which surface the adequate sequences of transparent and non-transparent parts are placed, along one or more concentric code tracks (see Fig. 2). The code disc is placed between the light source and photodetector. Environment to which the equipment will be subjected (temperature, humidity, shock, cleanliness) should be considered when choosing code disc material.

a) b) c)

Fig. 2. Code disc examples for: a) Incremental, b) Absolute and c) Virtual absolute encoder

Incremental encoders are the cost-effective and simple devices for measuring the position, but they do not give information about absolute position. In the simplest variant one pulse sequence is used, whereat the pulses are counted in order to obtain information about the position and in this way information about direction of rotation can not be provided. In that case, quadrature incremental encoder which generates two pulse sequences, mutually phase shifted by 90°, are used to determine the direction of rotation. Often, incremental encoders have a third, so called, index channel, which generates one pulse per revolution, which is used to define the reference "zero" position, and to record whole revolutions. The advantages of incremental encoders are lower cost, simpler structure and smaller dimensions, while the disadvantages are low reliability and error accumulation. The fact that incremental encoders lose the information about previous position with the loss of power, or signal transmission error, is also disadvantage. Another important parameter of position encoders is resolution, and its values are taken from GPI (Gurley Precision Instruments) encoders manufacturer catalogue, so this value for incremental encoder goes up to 7,2 million counts/revolution (this is not native resolution, it is the number of counts per revolution with electronic divide-by-four and interpolation) [4].

In the case of absolute encoders the number of code tracks on a code disc defines the number of bits, and so resolution of position determination. Absolute encoders contain *n* optical sensors that in parallel generate a set of *n* digital outputs which together describe the position of the rotating shaft. Set of *n* digital outputs represents *n* bit long digital word. Single and multiturn absolute encoders are available. The advantages of absolute encoders are: they give absolute position at the output, the information about previous position is remembered even after loss of power, and high reliability, while the disadvantages are: they have complex structure, they are expensive and have larger dimensions. Their resolution goes up to 25 bits, which also depends on code disc diameter [4].

Virtual absolute encoder [2, 3] is a pseudorandom absolute encoder with serial code reading. The virtual absolute encoder code disc is similar to the incremental encoder code disc. The difference is that the index track of incremental encoder is replaced with pseudorandom code track. The advantages of virtual absolute encoders are: they give absolute position at the output, they are cheaper than absolute encoders, because of smaller number of code tracks they have simpler structure than absolute encoders, they have high reliability and possibility of automatic adjustment of zero position when mounting. The disadvantages are: large dimensions, necessity of initial moving after the first powering on and the problem of code reading synchronization. The resolution value for virtual absolute encoders goes up to 24 bits [4].

IV. Determining of Shaft Rotation Direction

In this section, solutions for determination of shaft rotation direction by using the incremental and virtual absolute encoders, will be worked out [5]. In order to ensure the reliability of the information, the impact of the oscillation of rotating shaft in the direction of motion will be considered. The oscillations of rotating shaft at the border of two states can disturb determining of shaft rotation direction. That "hysteresis" is good in the light of reduction of error which could occur during direction determination.

A. Determining the rotation direction without "hysteresis"

The signal obtained from one sensor head of the incremental encoder does not carry information about the rotation direction. To obtain information about the rotation direction at least two sensors should be used. Distance between them is determined by the required phase shift of output signals. The required phase shift is usually 90°.

The time characteristic of output signals, in the form of logic signals A and B, is shown in Fig. 3b).

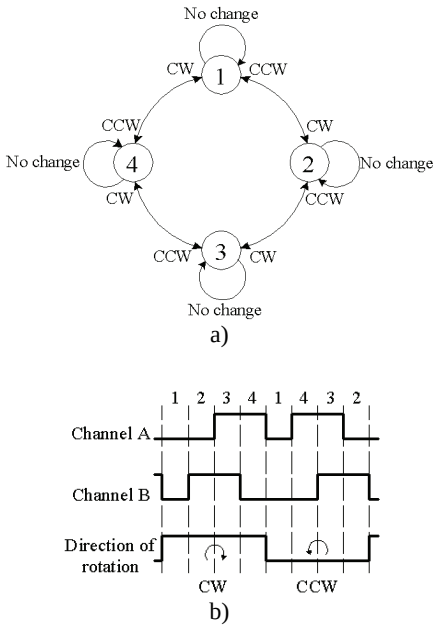


Fig. 3. a) State diagram, b) Output signals

If the individual levels of the output signals A and B are determined by logic values 0 and 1, then a pair of logic signals AB can appear in one of four states: 00,01,11,10 [6]. Those states are respectively marked with 1, 2, 3 and 4. States are cyclic linked, and transitions between states 1-3 and 2-4 are not allowed. The principle of transition between two adjacent states is shown in Fig. 3a), with a simple state diagram. CW and CCW represent outputs of the circuit for determining the rotation direction (CW-clockwise, CCW-counter-clockwise). The state diagram points out the fact that reliable recognition of the direction assumes knowledge of the initial state and the current state. Values of the logic signals A and B should be taken at regular time intervals, which is achieved by using clock pulses. This is how, in certain time instants, synchronization values As, Bs are determined, while the values Ao, Bo are stored from the previous time instant. The time period of clock pulses, at the maximum angular speed of rotation, must be shorter than one-half of the width of pulses A and B.

The group of signals As, Bs, Ao, Bo is a group of four logic variables which provide sixteen different combinations. Each individual combination, except the forbidden, defines values of logic functions CW and CCW, see Table I. One software implementation of this solution, based on application of a virtual instrumentation concept, is presented below.

TABLE I
STATES SEQUENCE AND CORRESPONDING VALUES OF LOGIC FUNCTIONS CW AND CCW WITHOUT "HYSTERESIS"

A_S	B_S	A_0	B_0	CW	CCW
0	0	0	0	0	0
0	0	0	1	0	1
0	1	0	0	1	0
0	1	0	1	0	0
0	0	1	0	1	0
0	0	1	1	-	-
0	1	1	0	-	-
0	1	1	1	0	1
1	0	0	0	0	1
1	0	0	1	-	-
1	1	0	0	-	-
1	1	0	1	1	0
1	0	1	0	0	0
1	0	1	1	1	0
1	1	1	0	0	1
1	1	1	1	0	0

B. Determining the rotation direction with "hysteresis"

In order to avoid errors in determining direction during the oscillations of rotating shaft at the border of two states, the knowledge about states in two successive discrete time intervals is not enough. The knowledge about the previous state is required, or in other words, the knowledge about state prior to the initial state. A sequence of previous, initial and current state would now define the values of logic functions CW and CCW, Table II. Sequences of states that occur due to interruption of rotation or oscillations at the border of two states should be marked with passive values of logic functions CW and CCW. In this paper is suggested that the forbidden sequences of states (413, 124, etc.) should also be taken into consideration, because these sequences can occur due to contamination of the incremental track. Adoption of the fact that the forbidden sequence of states does not lead to the change in information about the shaft rotation direction, is also proposed. This is a realistic assumption, because it is unlikely that the transition which just has been detected, in the next moment will not be detected. However, to avoid increase of the maximum error caused by the forbidden sequence of states, retaining of the passive values of logic functions CW and CCW is suggested. Also, this solution with “hysteresis” is implemented and tested in LabVIEW environment, which is presented below.

TABLE II
STATES SEQUENCE AND CORRESPONDING VALUES OF LOGIC FUNCTIONS CW AND CCW WITH "HYSTERESIS"

SEQUENCE OF STATES			CW	CCW
4	1	4	0	0
4	1	1	0	0
4	1	2	1	0
4	1	3	0	0
1	2	1	0	0
1	2	2	0	0
1	2	3	1	0
1	2	4	0	0
2	3	2	0	0
2	3	3	0	0
2	3	4	1	0
2	3	1	0	0
3	4	3	0	0
3	4	4	0	0
3	4	1	1	0
3	4	2	0	0
1	4	3	0	1
1	4	4	0	0
1	4	1	0	0
1	4	2	0	0
2	1	4	0	1
2	1	1	0	0
2	1	2	0	0
2	1	3	0	0
3	2	1	0	1
3	2	2	0	0
3	2	3	0	0
3	2	4	0	0
4	3	2	0	1
4	3	3	0	0
4	3	4	0	0
4	3	1	0	0

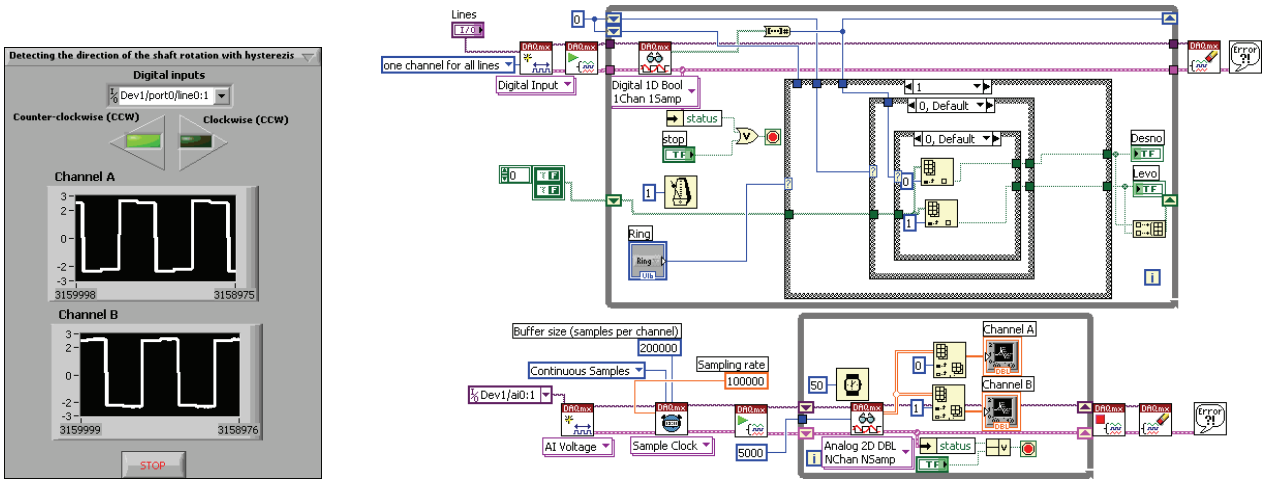


Fig. 4. Front panel and block diagram of realized virtual instrument for direction detection of shaft rotation

C. Implementation and testing of proposed solutions in LabVIEW environment

Virtual instrumentation concept is widely applied for purpose of measurement and testing during research work [7]. The virtual instrumentation in this paper is firstly used for simulation of output signals of incremental encoder and for generating of different test sequences. Simulator of incremental quadrature encoder shown on Fig. 5, provides the choice of number of revolutions per minute, number of pulses per revolution and direction of rotation.

Either of proposed solutions (with and without "hysteresis") for rotation direction determination is implemented in one common virtual instrument by using LAbVIEW 8.0 software and acquisition boards PCI 6251 and USB 6008, see Fig. 4. The output signals from incremental encoder simulator are brought on two digital inputs of USB 6008 acquisition board, and any type of acquisition board which has digital inputs, can be used for this purpose [8].

The output information about shaft rotation direction is obtained by applying of Tables I and II using shift register and case structure functions in the block diagram. The input signals can be seen on the front panel and the choice of used method for rotation direction determination is realized through ring control. This virtual instrument can be implemented within solution for angular position or velocity determination.

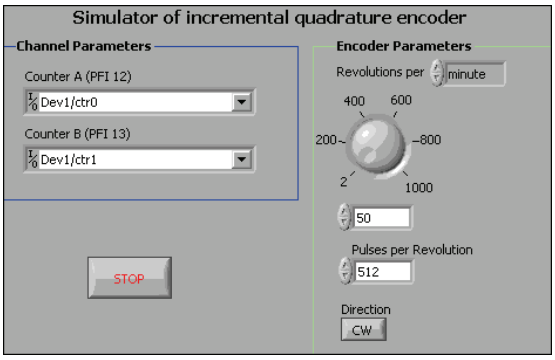


Fig. 5. Simulation of incremental encoder

V. Conclusion

To make the position or angular velocity information of the rotating object complete, among other parameters, the direction of rotation must be known. In this paper are proposed solutions for determining the rotation direction with and without the presence of oscillations of the rotating shaft on the border of two adjacent states, and each of them was implemented in LabVIEW software environment. These realizations can be used as an SubVI into instrument structure which measure displacement or angular velocity. The aforementioned solutions can have grater importance as part of complex systems for remote monitoring of movable objects in the industry, than in the case of individual application.

Acknowledgement

Research activities presented in this paper, are supported by funds of the Ministry of Science and Technological Development, having the reference project number TR32045.

References

[1] J. G. Webster, *The measurement, instrumentation and sensors handbook*, New York, CRC Press, 1999.
[2] D. Denić and G. Miljković, "High Resolution Virtual Absolute Encoders", Conf. Proc. ICEST 2003, Sofia, Bulgaria, pp. 390-393, October 16-18, 2003.
[3] D. Denić and G. Miljković, "Code reading synchronization method for pseudorandom position encoders", Sensors and Actuators A: Physical 150, pp. 188-191, 2009.
[4] http://www.kompauto.se/files/encoderline09.pdf
[5] P. Thajchayapong and M. Chinakarn, "Further improvement in a counting and direction-sensing circuit", International journal of electronics, vol. 66, no. 6, pp. 935-938, 1989.
[6] M. Milly, "Recognition of direction of incremental sensor shaft rotation", Int. J. Electronics, vol. 68, no. 3, pp. 471-481, 1990.
[7] DAQ M Series User Manual, National Instruments, 2008.
[8] N. Patrascoiu, A. Poanta, A. Tomus and B. Sochirca, "Detection of motion direction implemented by virtual instrumentation", ICAI'10, pp. 72-75, Iasi, Romania, 2010.

Virtual Instrumentation used for Adaptive Angular Velocity Measurements

Goran S. Miljković[1], Miodrag Z. Arsić[2], Dragan B. Živanović[3] and Milan M. Simić[4]

Abstract **– In many applications where some kind of motion is performed, for example in robotics, it is of high importance the angular velocity measurement. Concept of virtual instrumentation is more and more incorporated in modern industrial system, and therefore the capabilities of virtual instrumentation for angular velocity measurement are tested and presented in this paper. It is presented the realized virtual instrument for angular velocity measurement, which applies some known classical and one adaptive method for measurement. The adaptive method for angular velocity measurement had the main goal maintaining good measurement accuracy and resolution in wide velocity range.**

Keywords **– Angular velocity measurement, virtual instrumentation, LabVIEW.**

I. Introduction

Nowadays, as computers and microprocessors are more often implemented in modern industrial measurement and control systems, the digital techniques for angular velocity measurement becomes more and more dominate in relation to analog techniques. Digital techniques have some clear advantages such as error reduction, easier processing, easier presentation, easier data transfer, etc. The digital tachometer is the most frequently used digital device for angular velocity measurement [1]. High-quality digital tachometers are usually incorporated in every application which requires high-accuracy motion control. These applications can be in servo, mechatronic, robotic and precision production systems. To improve stability and smoothness of digital motor control, the velocity feedback is added along with position feedback.

An incremental digital encoder mounted on the shaft of some machine provides pulses for both position and velocity estimation. Different methods of encoder pulses processing for velocity estimation are developed, and implemented using different hardware. The first discovered, so-called classical measurement methods which are used in velocity estimation are direct counting method (or M method) and the indirect counting method (or T method).

The M method [2] is based on counted pulses from the optical encoder in fixed-time intervals. In this method, one counter is needed for counting of encoder pulses and one timer for determination of fixed time interval. Disadvantage of this method is poor accuracy in low speed range, but it is used in some applications due to its simplicity, low cost and minimal required hardware. For improving accuracy of the method, the averaging of counter values in more than two successive time intervals is applied, but in this case measurement time is not constant.

The T method [3] is especially applicable in a low speed range because it measures duration between two successive encoder pulses. The internal clock signal is used for measurement of the time interval, and counter determines number of clock pulses between two successive encoder pulses. In contrast to previous method this method has poor accuracy in high speed range.

In order to obtain a method which would be applicable in wide speed range, more complex method is designed as a combination of previous two methods (M/T method) [4]. So, a constant elapsed time (CET) method is developed [5], which is based on counting encoder pulses in a prescribed time interval. The buffered method [6] is based on both pulse counting and measurement of the fractional pulse period during the fixed sampling period. To reduce sensitivity to sensor nonideality, the new PEM-CSDT (parallel edge measurement, constant sample-time digital tachometer) structure is proposed, implemented by employing an FPGA [7]. For four possible transitions from two encoder channels, it uses four independent sets of counters. Lygouras in his paper [8] describes a new approach to velocity estimation based on adaptive sampling period according to the instant rotational velocity, and in this way better response times are achieved at medium or high speed ranges.

So far, the microprocessors are the most often used for implementation of various velocity measurement methods. FPGA circuits have become popular in the last decade because of its reasonable cost and very good performance. Virtual instrumentation, based on PC, powerful software and modular hardware, has reached mainstream acceptance and is used in thousands of applications in the industry. Flexibility, relatively easy learning of graphical programming, easy modifications of the realized methods, powerful data presentation are some of advantages. In this paper, the initial idea was to explore capabilities of virtual instrumentation in angular velocity measurement [9]. The multifunctional acquisition boards are used, concretely National Instruments PCI 6251 board [10], which has two general purpose 32-bit counters on-board. So, the first step was the realization of classical methods for angular velocity measurement such as M method, and T method. Then, one adaptive method with improved accuracy in wide speed range is developed as virtual instrument. The experimental results of the realized methods are presented, compared and analyzed.

[1]Goran S. Miljković is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E -mail: goran.miljkovic@elfak.ni.ac.rs

[2]Miodrag Z. Arsić is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: miodrag.arsic@elfak.ni.ac.rs

[3]Dragan B. Živanović is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.zivanovic@elfak.ni.ac.rs

[4]Milan M. Simić is with University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milan.simic@elfak.ni.ac.rs

II. Classical Methods for Angular Velocity Measurement

The concept of virtual instrumentation delivers solutions with faster development time, lower costs, and greater flexibility. In order to explore the capabilities of virtual instrumentation for the angular velocity measurement, we started with realization of simple methods such as M and T method. PCI multifunctional acquisition board NI 6251, PC with installed LabVIEW 8.0 software and pulse/function generator HP 8116A for encoder pulses simulation are used for realizations and testing of this methods. Universal counter HP 5316B was used for the accurate measurement of the function generator output frequency. The acquisition board has two general purpose counters, which are used in realization of various methods for angular velocity measurement. Also, it has internal 80 MHz clock.

Firstly, it was realized virtual instrument for implementation of previously mentioned methods M and T method. M method or direct counting method is simple and widely used method, and it is based on counting of pulses in fixed time interval. C_P denotes the number of counted pulses in fixed sampling period T_S. The measured velocity ω is expressed as follows:

$$\omega = \frac{60C_P}{PT_S} \text{ [rpm]}, \tag{1}$$

where P denotes the encoder pulse number per rotation, and it can be changed in the front panel of the realized virtual instrument, which depends on the applied encoder type. Measurement time is equal to the sampling period T_S, and it is constant. One counter is used for determination of sampling period and the other counter is used for counting the encoder pulses (Fig. 1). This method has good accuracy at high speed.

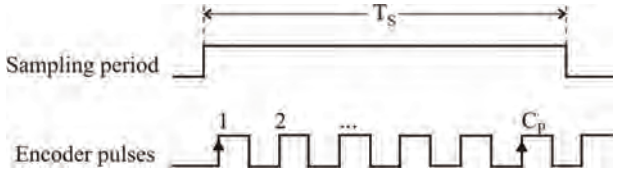


Fig. 1. Principle of the M method

Measurement of velocity with the T method or indirect counting method is based on the time interval measurement of the encoder pulse period. Accuracy of the method depends on the resolution of the time interval measurement, and in this case the internal 80 MHz clock (f_c) is used for that purpose (Fig. 2). Velocity can be evaluated based on relation,

$$\omega = \frac{60 f_c}{PC_t} \text{ [rpm]}, \tag{2}$$

where C_t denotes the number of the counted clock pulses.

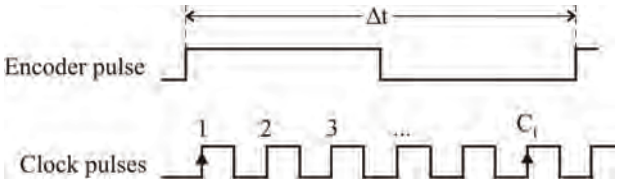


Fig. 2. Principle of the T method

This method has good accuracy at low speed. Disadvantages of this method are the variable measurement time which depends on the velocity, and low accuracy at high speed.

The errors which appear in angular velocity measurement by using virtual instrumentation are generally classified to: the errors introduced from quartz oscillator on acquisition board δ_{f_c}, the triggering errors due to noise in signal $\delta_{trigger}$ and the quantization error δ_q:

$$\delta_\omega = \frac{\Delta\omega}{\omega} = \delta_{f_c} + \delta_{trigger} + \delta_q, \tag{3}$$

Error values introduced from quartz oscillator on acquisition board can be determined from manufacturer's specifications. Oscillator accuracy of PCI 6251 acquisition board is about $5\text{x}10^{-5}$, but this error can be reduced by calibrating the designed virtual instrument with more accurate external instrument.

The triggering error due to noise in signal depends on the signal quality at the acquisition board input, and can be reduced with the appropriate conditioning circuits. This error does not dominate if M method is performed, or if measurement with averaging of several input signal periods is performed.

The quantization error is defined by relation:

$$\delta_q = \left| \frac{\omega(C+1) - \omega(C)}{\omega(C)} \right|, \tag{4}$$

where C is counter value.

Regardless of the method used to determine the angular velocity, either by M method or by T method, it can be derived that the quantization error directly depends on the counter value [1],

$$\delta_q \approx \frac{1}{C}. \tag{5}$$

III. Adaptive Angular Velocity Measurement

The virtual instrumentation is then utilized for implementation of one adaptive method for angular velocity measurement. In basic T method it is measured time interval of one encoder pulse period, and here is proposed solution which is based on time interval measurement of k input signal periods. This variable value k will help in maintaining quantization error within specified boundaries. This value is very important for measurement because if it's too big it can cause counter overflow or too long measurement. On the other hand, if it's too small, it can not be sufficient to increase measurement resolution, i.e. decrease quantization error.

Block diagram of realized virtual instrument determines itself the value of multiplication factor k according to previously defined maximal quantization error. In that way, the minimal measurement time for the defined quantization error is also achieved.

Principle of using counters on-board during realization of the suggested adaptive method and signal waveforms are shown in Fig. 3. One counter is used for the adaptive

multiplication of the encoder output signal period, and second counter for clock period counting contained in obtained sampling period. According to the proposed method, the angular velocity is calculated based on relation

$$\omega = \frac{60 f_c k}{P C_t} \text{ [rpm]}, \quad (6)$$

where f_c is the clock frequency, and C_t is the number of clock periods counted by second counter.

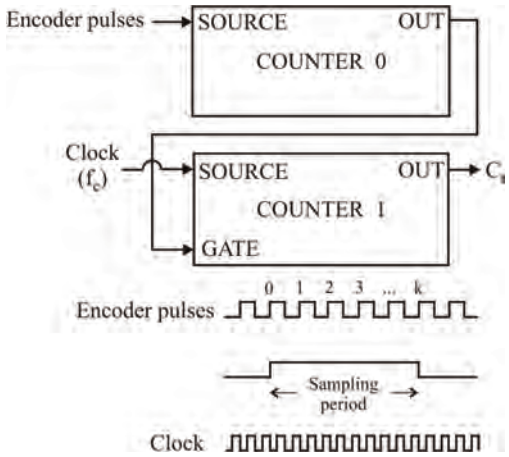


Fig. 3. Principle of adaptive method

Two-step measurement method which is based on using the adaptive multiplication factor k is explained in detail in algorithm shown in Fig. 4. Measurement using the method for multiperiod measurement for value k=4 is performed in the first step. Changing the value of k is then performed depending on the obtained counter content C_t. Then, the next measurement with new value of k is performed. It can be seen in the algorithm that the value of multiplication factor k in second step is within range of 4 to 200000, depending on the instant angular velocity. In that way, the final counter content of approximately 200000 is achieved in a wide range of measurement, producing also constant value of the quantization error. The method is developed and tested for angular velocities of up to approximately 300000 rpm. Value of 200000, which is used in algorithm to determine the adaptive multiplication factor, is chosen to achieve compromise between the smallest possible measurement time and the smallest possible quantization error. However, the quantization error must not be unrealistically small in relation to the total error and the on-board quartz oscillator instability. Value of k is determined over and over again in each sampling period.

The front panel and block diagram of the realized virtual instrument for angular velocity measurements are shown in Fig. 5 and 6, respectively. The number of encoder pulses per rotation can be entered in the front panel. Also, the angular velocity measurement results can be seen on the numerical indicator during measurement, and number of counted pulses. Input signal can be seen on waveform chart. In block diagram is used flat sequence structure for two-step measurement.

The experimental results of velocity measurement which are obtained by using these three methods are presented in Table I. One can see testing frequency, measurement this frequency with external counter, calculated angular velocity, measurement results of three methods, and obtained relative errors. Last row presents values of multiplication factor k for each particular measurement. The relative errors obtained by proposed adaptive method are in wide measurement range equalized and somewhere less than the values obtained with classical methods, as it can be seen from Table I.

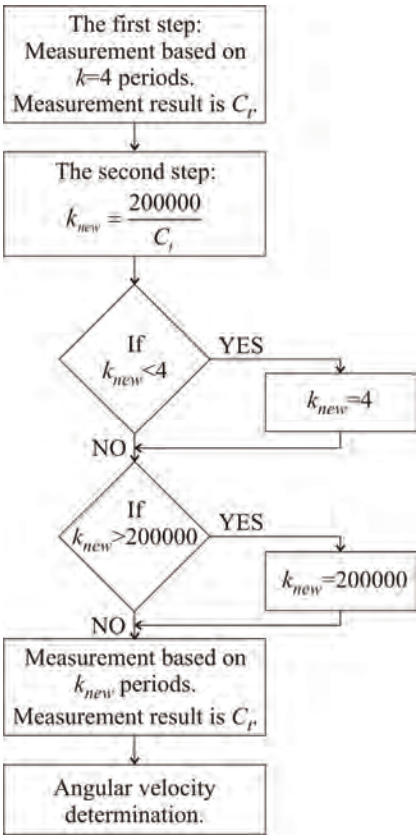


Fig. 4. Algorithm of adaptive angular velocity measurement

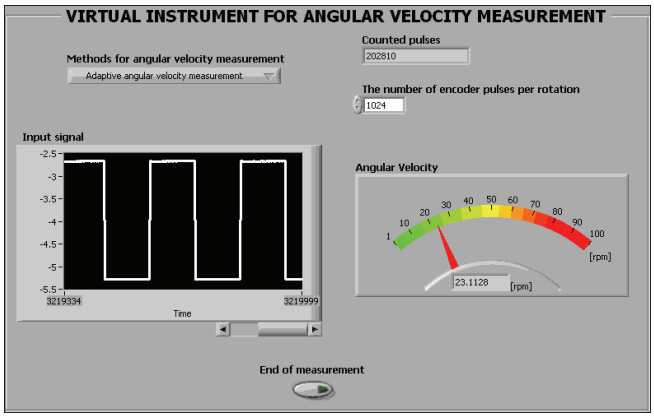


Fig. 5. Front panel of realized virtual instrument for angular velocity measurements

The digital tachometer, realized in the described way, uses only counters on the acquisition board. Therefore, additional resources of the board, such as analog and digital inputs and outputs, can be used for other purposes.

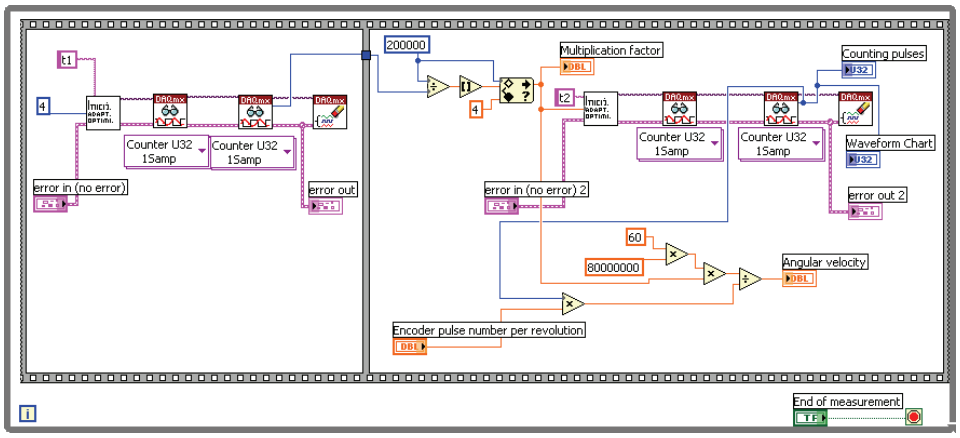


Fig. 6. Block diagram for adaptive method

TABLE I
EXPERIMENTAL RESULTS OF ANGULAR VELOCITY MEASUREMENTS

Output of function generator		10 Hz	100 Hz	1 kHz	10 kHz	100 kHz	1 MHz	2 MHz	5 MHz
Frequency counter reading		10.09 Hz	99.39 Hz	1.0071 kHz	10.031 kHz	103.054 kHz	1.0389 MHz	2.0678 MHz	5.126 MHz
Equivalent angular velocity (for P = 1024) [rpm]		0.5912	5.8236	59.0097	587.754	6038.32	60873.0	121160	300351
M method	**Velocity [rpm]**	0.538	5.55	58.85	587.95	6037.7	60868.1	121155	300349
	Relative error [%]	8.96	4.698	0.27	0.033	0.01	0.008	0.004	0.0006
T method	**Velocity [rpm]**	0.591198	5.8238	59.0085	587.748	6036.5	60325.1	114980.8	278875.9
	Relative error [%]	0.0003	0.0034	0.002	0.001	0.03	0.9	5.1	7.15
Adaptive method	**Velocity [rpm]**	0.591199	5.82375	59.008	587.77	6038.2	60873.8	121158	300370
	Relative error [%]	0.00016	0.0026	0.0029	0.003	0.002	0.001	0.002	0.006
	k	4	4	4	6	62	625	1250	3125

IV. CONCLUSION

Capabilities of virtual instrumentation for angular velocity measurement are examined in the paper. The experimental results can be easy memorized or analyzed in LabVIEW.

Proposed adaptation method of the angular velocity measurement is realized using standard acquisition board which has two counters. The achieved relative errors are in wide measurement range equalized. This method is not suitable for fast changing angular velocities, and the measurement time is not constant which can be drawback in some applications.

ACKNOWLEDGEMENT

Research activities presented in this paper, are supported by funds of the Ministry of Science and Technological Development, having the reference project number TR32045.

REFERENCES

[1] D. Denic, G. Miljkovic, D. Zivanovic, “Microcomputer based wide range digital tachometer”, Electronics and Electrical Engineering, vol. 3, no. 67, pp. 31-36, 2006.

[2] N. Kirianaki, S. Yurish, N. Shpak, and V. Deynega, “Data acquisition and signal processing for smart sensors”, John Wiley & Sons Ltd, ISBN: 0-470-84317-9, 2002.

[3] C.D. Cenzo, B. Szabados and N. K. Sinha, “Digital measurement of angular velocity for instrumentation and control”, IEEE Trans. Ind. Electron. Contr. Instrum., vol. IECI-23, pp 83-86, 1976.

[4] T. Ohmae, T. Matsuda, K. Kamiyama, M. Tachikawa, “A microprocessor-controlled high-accuracy wide-range speed regulator for motor-drives”, IEEE Transactions on Industrial Electronics, vol. 29, no. 3, pp. 207–211, 1982.

[5] R. Bonert, “Design of a high performance digital tachometer with a microcontroller”, IEEE Transactions on Instrumentation and Measurement, vol. 38, no. 6, pp. 1104–1108, 1989.

[6] M. Prokin, “Double buffered wide-range frequency measurement method for digital tachometers”, IEEE Trans. on Instrumentation and Measurement, vol. 40, no. 3, pp. 606-610, 1991.

[7] R.C. Kavanagh, “Improved digital tachometer with reduced sensitivity to sensor nonideality”, IEEE Transactions on Industrial Electronics, vol. 47, no. 4, pp. 890–897, 2000.

[8] J.N. Lygouras, “Accurate velocity evaluation using adaptive sampling interval”, Microprocessors and Microsystems, vol. 24, no. 5, pp. 269–275, 2000.

[9] N. Patrascoiu, A. Poanta, A. Tomus and B. Sochirca, “Virtual instrumentation used for displacement and angular speed measurements”, International Journal of Circuits, Systems and Signal Processing, issue 2, vol. 5, pp. 168-175, 2010.

[10] DAQ M Series User Manual, National Instruments, 2008.

System for Testing of the Current Measuring Transformer Basic Parameters supported by LabVIEW Software

Milan Simić[1], Dragan Denić[2], Dragan Živanović[3] and Goran Miljković[4]

***Abstract* – Procedure for verification of the current measuring transformer basic parameters, which is functionally supported by virtual instrumentation software LabVIEW, is presented in this paper. Hardware configuration of developed measurement and acquisition system includes regulation source of AC voltages, power transformer 150A/0.4V, standard computer configuration with data acquisition card NI 6008 and virtual instrumentation software support. Virtual instruments developed in LabVIEW software environment perform signal presentation, measurement and software processing of measurement results, regarding to the basic parameters of input and output transformer waveforms. Statistical processing of measurement results includes recordings of transformer transfer characteristics, graphical presentation of diagrams for measured signal parameter values and calculations of measurement errors and standard uncertainty components.**

***Keywords* – Current transformers, Virtual instrumentation software, Standard measurement uncertainty.**

I. INTRODUCTION

Development and further improvement of the measurement methods, instruments and additional equipment applicable for metrological verification of the voltage and current measuring transformer basic parameters, are very important and complex activities, which are significantly determined by development of the transformers itself. Special attention to these problems must be dedicated in the process of measurement and analysis of the relevant power quality parameters. Primary purpose of these measurement procedures is to provide preconditions for increasing of total efficiency level concerning electrical power production, distribution and consumption process. Important metrological parameters of the voltage and current measuring transformers applied in the mentioned processes significantly affect to the total accuracy level of complete measurement procedure. From previously mentioned facts can be concluded that metrological verification procedure applied to the current measuring transformer basic characteristics, is very significant activity and therefore must be performed on regular basis [1]. Measurement instruments and additional equipment applied to these specific purposes must be designed to provide nominal measurement accuracy level significantly better than specified nominal accuracy value of the current measuring transformer under test. Besides initial testing of the measuring transformer nominal characteristics immediately after production process inside factory testing facility, there is necessity for performing of periodic checking of the measuring transformer parameters, at precisely determined time intervals. Such testing procedure can be performed inside specialized metrological laboratories or alternatively directly on the remote locations of individual measurement stations. Measurement and control instruments must be adjustable for work at various operational conditions, that can be significantly different from laboratory conditions. Among the rest this involves instrument capability to hold on various possible transportation problems, including resistance to tremors and noise. These measurement equipment primarily includes reference instrumentation for generation of reference input signals having the nominal values of basic parameters according to the relevant standards and documents. A second necessary component of this system is control instrumentation which performs measurement of basic parameters regarding to signals obtained from measuring transformer outputs [2,3].

Solution of the laboratory measurement system described in this paper, developed for metrological verification of current measuring transformer basic parameters, is functionally based on the data acquisition USB card NI 6008 [4], supported by software application in LabVIEW progrramming package [5]. As reference system component which provides generation of the reference input current signals for measuring transformer inputs, here is used power transformer with specified nominal output current of 150A maximum value. Power transformer is supplied by means of the AC voltage regulation source having maximum RMS voltage values of 245V and signal frequency of 50Hz Programming application of virtual instrumentation software in PC environment controls execution of the testing procedure algorithm. This involves continuous measurement, graphical presentation, recording and statistical processing of measurement results concerning basic parameters of input and output signals for current measuring transformer under test.

II. DESCRIPTION OF THE MEASUREMENT SYSTEM HARDWARE CONFIGURATION

Simplified hardware block configuration of the laboratory measurement system, which is developed for the purpose of current measuring transformer basic parameter verification, is illustrated on the Fig 1. Presented solution of the measurement laboratory system, developed on Department of Measurement

[1]Milan Simić is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: milan.simic@elfak.ni.ac.rs.

[2]Dragan Denić is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dragan.denic@elfak.ni.ac.rs.

[3]Dragan Živanović is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dragan.zivanovic@elfak.ni.ac.rs.

[4]Goran Miljković is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: goran.miljkovic@elfak.ni.ac.rs.

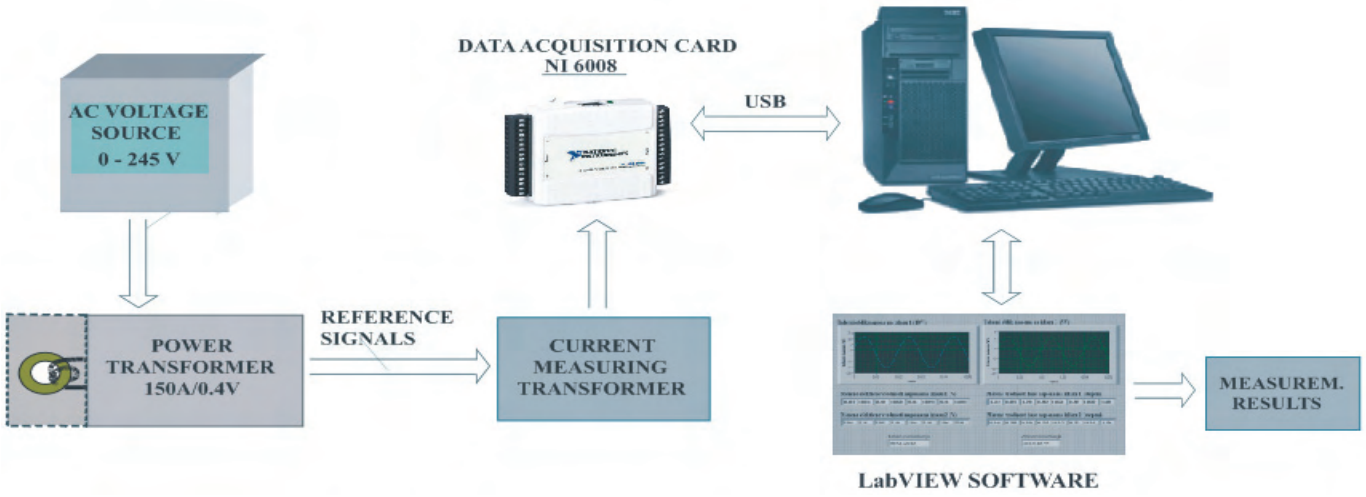


Fig. 1. Hardware configuration of the laboratory measurement system for current measuring transformer testing procedure

at the Faculty of Electronic Engineering in Niš, is consisting of the AC voltage regulation source, 150A power transformer used as generator of the reference current input signals, USB data acquisition card NI 6008 and standard PC configuration supported with virtual instrumentation programming package LabVIEW. Depending on the actual requirements, AC voltage regulation source provides continuous adjustment of the basic parameters for reference waveforms on the current measuring transformer inputs. Using this specific regulation AC voltage source and power transformer, reference input waveforms can be generated with maximum RMS current value of 150A, but upper limit value of the reference current signals used for this laboratory application was significantly smaller from nominal maximum value. Recording and graphical presentation of the measuring transformer current and voltage characteristics is provided by measurement and recording of the basic current and voltage signal parameters on the measuring transformer outputs, in dependence on the continuous variations of input reference current values. In order to perform signal acquisition waveforms generated from the current measuring transformer outputs must be sent directly to the inputs of analog to digital converter within data acquisition card, developed with 12-bit resolution [4]. Data acquisition 12-channel USB card NI 6008 developed with possibility for analog to digital and digital to analog data conversion, from manufacturer company National Instruments Corporaton, for this specific application purpose uses a several analog input channels, which receive measuring acquisition signals from tested current transformer outputs.

Direct two-way signal communication and data interchange between data acquisition card and PC computer is provided by means of standard USB communication interface. Defining of the basic parameters used for acquisition of the measurement signals is performed according to the specific user demands by means of the specialized control software application DAQ Assistant. Front panel of this control application for defining of the signal acquisition parameters is presented on the Fig 2.

Previously presented and described procedure applied to the metrological verification of the current measuring transformer basic parameters and characteristics, including measurement and statistical processing of the obtained measurement results, is software controlled by programming application developed in LabVIEW software environment. Some solutions of virtual measurement instruments, developed as metrological support of presented measurement and data acquisition system, will be illustrated and analyzed in the following segment of the paper.

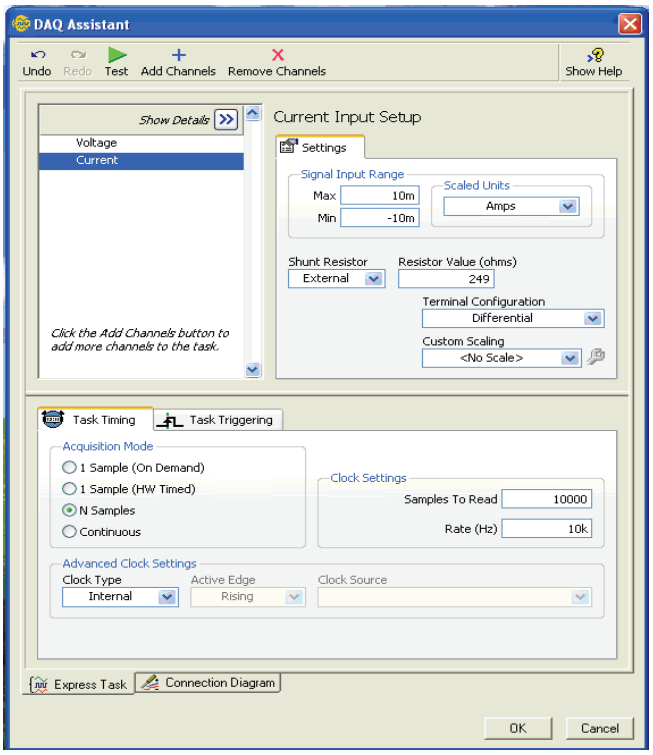


Fig. 2. Front panel of the software application for defining of the basic parameters for measurement signal acquisition

III. Programming Support of the Procedure

Virtual instrumentation presents software oriented method for realization of the measurement instruments with traditional functions, which is based on using of the standard computer configurations, hardware components for measurement signal acquisition and graphical programming packages capable for measurements, software processing, recordings and graphical

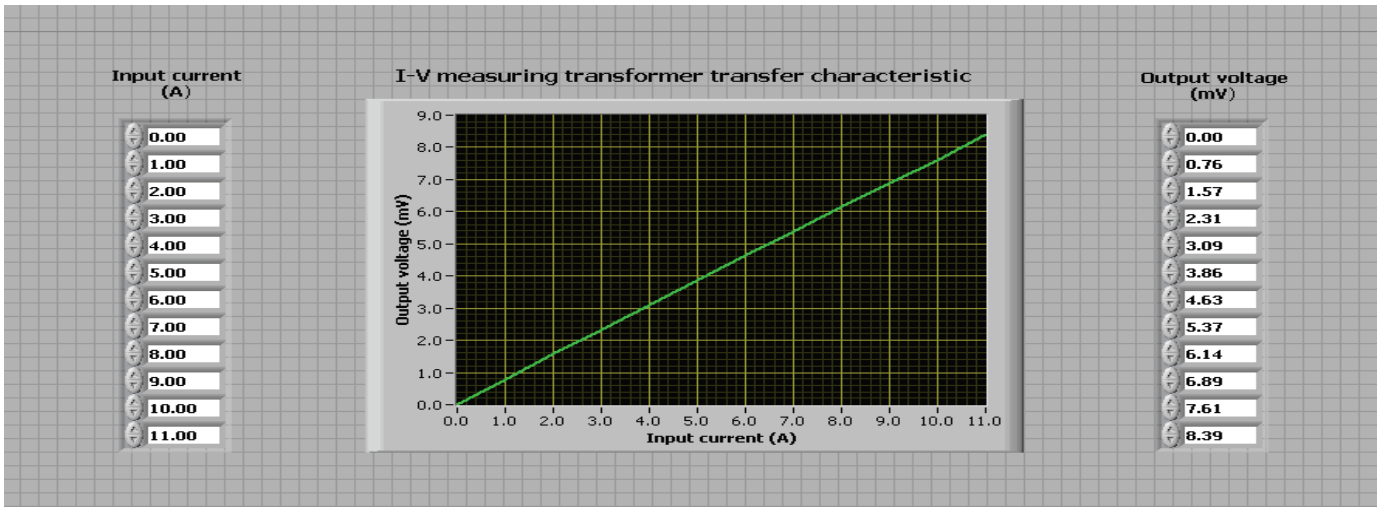


Fig. 3. Input and output transfer characteristic of the current measuring transformer recorded in LabVIEW environment

presentation of obtained measurement results. Functionality of developed virtual instruments can be easily changed by simple corections of the software algoritm sequence which controls execution of entire measurement and acquisition process [6].

Software application designed using virtual instrumentation programming package LabVIEW, which performs automated recording and graphical presentation of measuring transformer transfer characteristic, is presented by virtual instrument front panel on the Fig 3. Illustrated transfer characteristic function indicates dependence of the RMS voltage values measured on transformer outputs from variations of the RMS input current values. Besides function diagram of the recorded current and voltage transfer characteristic, on presented virtual instrument front panel are indicated measurement results regarding to the RMS values of the measuring transformer input current and output voltage. These illustrated measured current and voltage parameters are obtained from twelve successively performed measurement cycles, which are corresponding to variations of transformer input current values within a range from 0 to 11A.

Front panel of the LabVIEW based virtual measurement instrument, which simultaneously performs presentation of the recorded waveforms and measurement of the basic parameters for input and output measuring transformer signal waveforms, is illustrated on the Fig 4. Besides graphical presentation of voltage signal waveforms recorded at measuring transformer inputs, on presented virtual instrument front panel are shown measurement results regarding to the RMS values and phases of the voltage waveform from measuring transformer outputs. In this specific case from totally one hundred measured values for each of the transformer signal parameters, obtained during measurement procedure, in order to provide better visibility options on the presented virtual instrument are indicated only sixteen repeatedly measured values of the signal parameters.

Besides indicated measured values of output voltage basic parameters, virtual instrument previously shown on the Fig 4. indicates some additional information regarding to the precise date and time for recording of signal waveforms and measured RMS and phase values of the output voltage waveforms. Also,

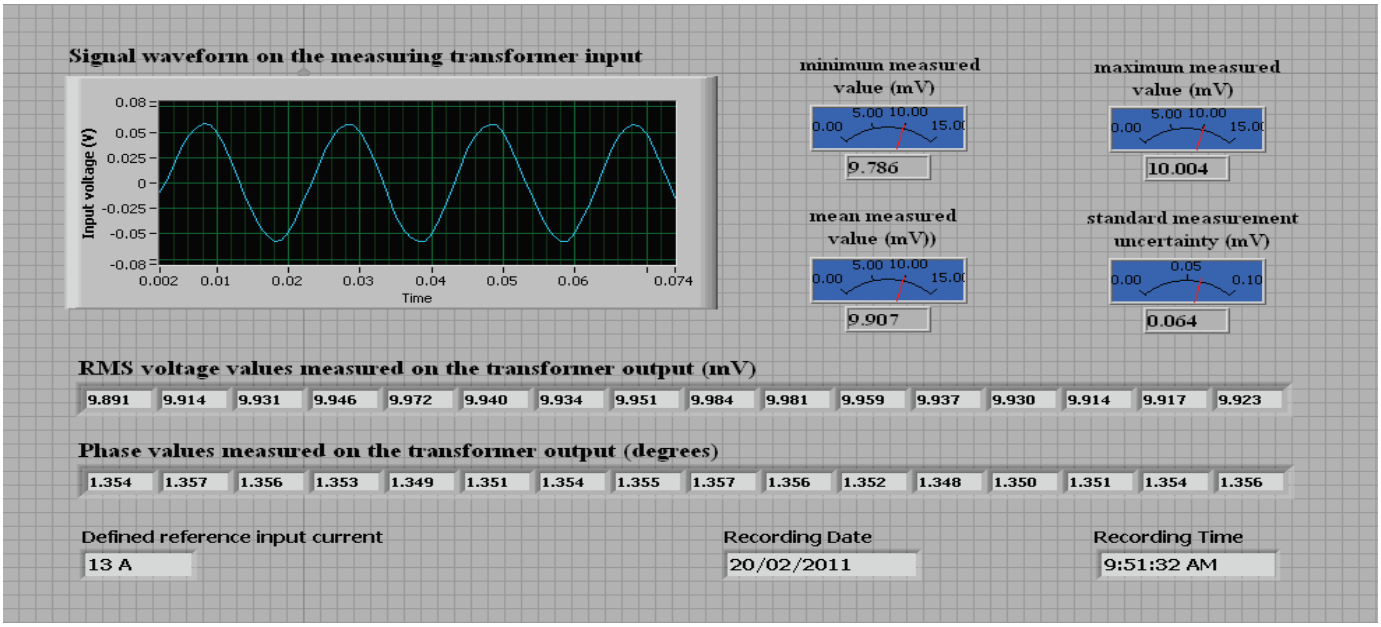


Fig. 4. LabVIEW virtual instrument developed for measurement and software processing of the measured signal parameters

here need to be mentioned that presented signal waveform and measurement results of basic voltage parameters are recorded for the case of previously defined RMS value of the reference measuring transformer input current of 13A. This predefined specific value of reference input current can be easily changed according to actual user requirements and demands, by means of the previously described reference system consisting of the regulation AC voltage source and power current transformer.

Control application of the virtual instrumentation software performs measurement, presentation, recording and statistical processing of the obtained measurement values. Some results concerning statistical processing of the measured RMS output transformer voltage values are presented by virtual instrument from Fig 4. Statistical processing of the measurement results generally gives possibilities for recordings and presentation of time diagrams and corresponding histograms of the measured signal parameter values. A large number of the measurement results from successively performed measurement procedures demands additional information about mean measured values and standard measurement uncertainties, that can be estimated by statistical procedures applied to the obtained measurement results. According to these mentioned demands, front panel of the virtual instrument given on the Fig 4. performs calculation and indication of the minimum and maximum measured RMS transformer output voltage values, obtained during previously described measurement procedure. Besides that, this software analysis of the measurement results involves calculations and indication of mean measured output voltage values, including calculations of the standard measurement uncertainty values.

Calculation of the standard measurement uncertainty values is performed according to recommendations of the document: Guide to the Expression of Uncertainty in Measurement [7], prescribed by International Standard Organization, ISO. Mean value of the measured output voltage parameters is calculated as arithmetic mean value of the obtained measurement results. Standard measurement uncertainty value is estimated from the statistical based procedures applied to the measurement results software analysis. Specifically measurement uncertainty value corresponds to the standard deviation of the measured values.

Finally, need to be mentioned that laboratory measurement results presented in this paper are only part of the beginning segment in development of the measurement and acquisition system for described purposes. Next segment of this process will be focused on upgrade and improvement of the described measurement procedure. Special attention will be oriented on measurement and software processing of measurement results regarding to the phase and high-order harmonic components.

IV. CONCLUSION

Solution of the measurement system developed for testing of the current measuring transformer basic characteristics and parameters, based on virtual instrumentation software package LabVIEW, is described in this paper. Hardware configuration of the presented experimental measurement and acquisition system includes regulation source of the AC voltage signals, power transformer in this specific case applied as generator of reference current input signals, data acquisition card NI USB 6008 and computer configuration with application software. Control programming application developed using LabVIEW software package monitors measurement procedure execution and performs recording, graphical presentation and statistical processing of the measurement results regarding to measuring transformer basic signal parameters. This statistical processing of the measured values includes recording and presentation of input and output transformer transfer characteristics, detection and indication of the minimum and maximum values obtained from measurement process, calculation of the mean measured parameter values and correspondent measurement uncertainty values, according to the valid documents for calculations and presentation of the uncertainty in measurements. By means of presented solution based on computer and cost-effective data acquisition hardware components, can be provided significant level of the software supported automation in the procedures for verification of current measuring transformer parameters.

ACKNOWLEDGEMENT

Research activities and experimental measurement results, presented in this paper, are financially supported by funds of the Joint European Research Project (JERP), under the SEE-ERA.NET Plus Joint Call, with the Reference Project Number ERA 78/01.

REFERENCES

[1] D. Naumovic-Vukovic, S. Skundric, D. Kovacevic and S. Milosavljevic, “Calibration of High Accuracy Class Standard Current Transformers”, Proceedings of the XIX IMEKO World Congress, September 6-11, 2009, Lisbon, Portugal.

[2] T. Chiulan and B. Pantelimon, “Power Transformer Units Condition Assessment Using Virtual Instrumentation“, Journal of Electronics and Electrical Engineer., Telecommunications Engineering, No. 6(86), 2008.

[3] S. Skundric, D. Kovacevic and D. Naumovic, “The Role and Importance of Software Application in Instrument Transformers Accuracy Testing”, Proceedings of the XVIII IMEKO World Congress, Rio de Janeiro, September 17-22, 2006.

[4] NI, *DAQ USB 6008 – User Specifications*, National Instruments Corporation, USA, (http://www.ni.com), 2005.

[5] NI, *LabVIEW - User Manual*, National Instruments Corporation, USA, (http://www.ni.com), 2007.

[6] S. Tumanski, *Principles of Electrical Measurements, Chapter 6. Computer Measuring Systems, Virtual Measuring Systems*, pp.426-456, Taylor & Francis Group, 2006.

[7] ISO, *Guide to the Expression of Uncertainty in Measurement*, International Standard Organization, Geneva, Switzerland, 1993

Software Package for Measuring of Generators Temperatures

Stanko Stankov[1], Zoran Jovanović[2], Miodrag Spasić[3], Nikola Danković[4] and Dejan Mitić[5]

***Abstract* – Description of a software package for the temperature monitoring in the generators and turbines of hydroelectric power plant is given in this paper. The proposed package and hardware configurations have been developed at the Faculty of Electronic Engineering in Niš and applied in HE "Vrla 2" in Surdulica, which is part of "Vlasinske hidroelektrane". The software receives data from the sensors and displays them in tables and diagrams on a PC screen. Exceeding in temperature of a certain measuring point is signalled to the operator in the command centre, by changing the colour chart and sound alarming. This software allows the archiving of data, generating and printing daily and periodic reports.**

***Keywords* – Software, Power plant, Temperature measuring**

I. INTRODUCTION

Temperature measuring and monitoring at key points of the generator is of great importance for the process of producing electrical energy. Normal values of temperature are a sign of the correctness of the generator, which with turbines are vital elements of the hydroelectric power plant. In certain areas of constructive parts of the generators and turbines, measuring elements (Pt 100) are installed and used to measure temperature [1-3]. The resistance of these elements changes with temperature (at 0°C the resistance is 100Ω and when temperature is increasing, the resistance value is also increasing). Software package of the system for temperature monitoring in the generators and turbines is used for data acquisition, archiving of measured data, numerical and graphic representation of data and alarm activation based on measured temperature values. Access to operating functions is allowed by user authorization system.

Software for Measuring of Temperature in Generators (SMTG2006) is developed for Windows XP operating system, [4] with installed MS Office package and InterBase 6.5 database server. New partition on the hard drive is created in order to improve efficiency of database. Data acquisition from measuring elements is enabled by multiplexer which is connected to serial PC computer port. Software package is realized by several modules described below:

- LMSU-SM-SMTG-06-0000 – Data acquisition module provides communication between PC terminal and hardware module for data acquisition via serial port. Module is operating independently and real-time processes are priorities.
- LMSU-SM-SMTG-06-0001 – Alarm control module monitors measured signals values. If these values are higher than allowed, the module sends request for sound and visual alarm activation. The module receives operator request for sound alarm activation. Visual alarm is activated until the cause of alarm activation is eliminated. Real-time process is priority in this case, also.
- LMSU-SM-SMTG-06-0002–Configuration system module manages system configuration database. This module provides hardware module acquisition data, module addresses acquisition, description of measured points, and graphic interface configuration for monitoring and alarm values.
- LMSU-SM-SMTG-06-0003 – User authorisation module manages authorized users database. This module is used for storage of user names, passwords and user access control data. Data coding before storing on hard drive and decoding during authorization are executed in this module.
- LMSU-SM-SMTG-06-0004 – Activity storing module receives systems activities, errors emergence, login and logout data and stores them to hard drive. Afterward review and generating of system reports are enabled.
- LMSU-SM-SMTG-06-0005 – Data archiving module receives measured data and sends them to archive on hard drive, monitors archive capacity and erases old data. Graphic representation of temperature values at measured points is allowed.
- LMSU-SM-SMTG-06-1000 – Graphic interface for monitoring allows representation of measured values, alarms and other system activities. User can intervene in the case the alarm is activated. The interface is connected to data acquisition module, alarm control module, activity storing module and data archiving module.
- LMSU-SM-SMTG-06-1001 – Graphic interface for configuration is graphic environment which provides hardware module configuration for data acquisition, graphic interface configuration for monitoring and setting of alarm values. Interface is connected to configuration system module.
- LMSU-SM-SMTG-06-1002 – Graphic interface for authorization is used for user authorization and database managing. This interface is connected to user authorisation module.

[1]Stanko Stankov is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Nis, Serbia, E-mail: stanko.stankov@elfak.ni.ac.rs.
[2]Zoran Jovanović is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Nis, Serbia, E-mail: zoran.jovanovic@elfak.ni.ac.rs.
[3]Miodrag Spasić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Nis, Serbia, E-mail: miodrag.spasic@elfak.ni.ac.rs.
[4]Nikola Danković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Nis, Serbia, E-mail: nikola.dankovic @elfak.ni.ac.rs.
[5]Dejan Mitić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, P.O. Box 73 18000 Nis, Serbia, E-mail: dejan.mitic@elfak.ni.ac.rs.

• LMSU-SM-SMTG-06-1100 – Report generator creates graphic reports of temperature values in certain period and table report about registered activities. This generator is connected to data archiving and activity registration modules.

II. Operating Functions of SMTG 2006

Function key F9 displays following operating functions (Fig. 1):
1. MONITORING TEMPERATURE (Temperature history)
2. IZVEŠTAJI (Reports)
3. POSTAVLJANJE ALARMA (Alarm settings)
4. AUTORIZACIJA OPERATERA (User autorisation)
5. KONFIGURISANJE (Configuration)

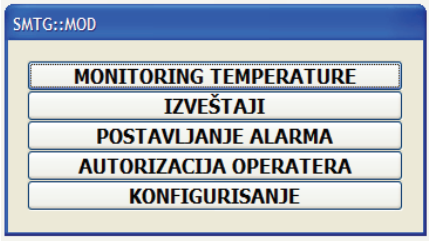


Fig. 1. Welcome screen with operating functions

A. Temperature Monitoring

Primary operating function of SMTG 2006 is temperature monitoring In generators (Fig. 2). Temperature monitoring screen consists of:
1. Screen heading
2. Table with current temperature values on generators
3. Activity listing
4. Status line

Fig. 2. Temperature monitoring screen

1. Screen heading includes the title and software package version, name of authorized user, working unit which the system is installed on and current date and time. If there is not authorized user, user name field is red.

2. Table with current temperature values for each generator is obtained by functional keys from F1 to F8. In the upper part of the table there is the sign “Fn: Generator n”. In the left part of the table below the picture of generator there are light indicators of correctness of communication with activation modules, where each module has its own light indicator. When computer starts communication with acquisition module, light indicator becomes yellow. If module answered according to protocol, light indicator is green. In the case when there is no response, or answer is not adequate to protocol, light indicator is red. Temperature values are shown as digital indicators for each measured point. Colour of the digital indicator corresponds to appropriate temperature range. Blue colour is for allowed range, yellow colour shows entry in risk temperature range and finally, red colour shows temperature values above allowed limit. On the left side of digital indicator there is a short name of appropriate measured point.
3. Activity listing includes a daily list of activities such as authorisation access to operating function, activation, acknowledgement and alarm stopping, errors emergence and other important system activities. Each group of activities has different colour in order to easier notice the group. Especially important activities are shown by red colour with short-timed sound signal.
4. Status line shows data such as: the name of current authorized user, login time, alarm information (the number of active and unacknowledged alarms) and finally the name of software package designer.

B. Temperature history

Composed data are stored in database each second. Precise temperature history graphic representation at selected point is available in this way (Fig. 3).

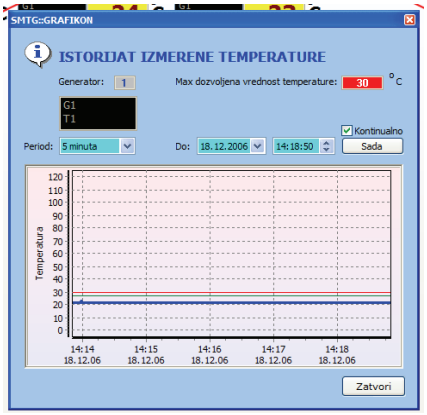


Fig. 3. Screen with temperature history

Temperature on the generator is presented by blue line; low limit of risk range is presented by green; and maximal allowed temperature is presented by red line on the chart.

Desired temperature history chart can be obtained from the list by choosing date and time period. Key “Sada” sets current date and time. If field „Kontinualno“ is checked the newest composed data for chosen period is being permanently presented.

C. Alarm Settings

The most important function of temperature monitoring in generators is alarm activation in the case of some monitored temperature value exceeds the allowed limit. This system incorporates warning subsystem. In the case that temperature enters to the risk area defined by warning temperature and alarming temperature, warning subsystem changes colour of digital indicator from blue to yellow.

If temperature of any measured point exceeds allowed limit, sound-visual alarm signalization will be activated, appropriate digital indicator becomes red and graphic temperature history at the measured point with warning is showed. Alarm activities data are being updated in status line. This activity is registered into activity listing. Computer starts to emit discontinuous sound signal. User separately acknowledges alarms for appropriate measured points by pressing the button „Potvrdi“. After acknowledgement of all alarm activities, alarm is deactivated, sound-visual signalization is turned off and the window with alarm warning and graphic temperature history is closed. Each acknowledgement is being recorded in activity listing.

If temperature value returns into allowed range, alarm will be deactivated for this point, even without user acknowledgement.

There is need to define warning and alarming temperature value to allow the system to operate (Fig. 4).

Fig. 4. Alarm settings

When we click on the button „ALARM“, the form for input of warning and alarming temperature values will be opened (Fig. 5).

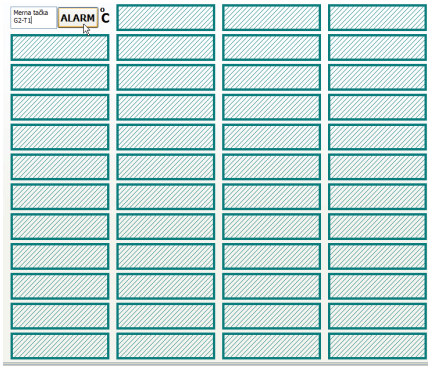


Fig. 5. Alarm values form

D. User Authorisation

Operating function access is determined by access control defined during user authorisation (Fig. 6). Authorisation system provides the services of adding new user deleting some users from database and changing access control data.

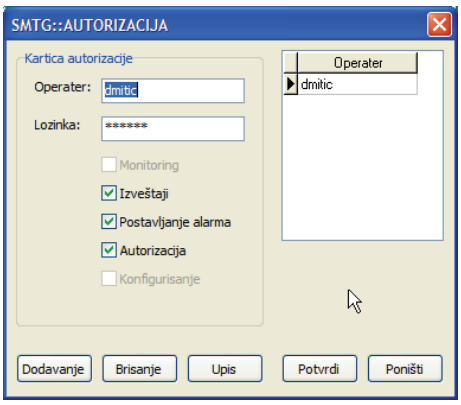


Fig. 6. User authorisation

If we want to create new user account we need to press button „Dodavanje“. In such a way new user card is created in database, then the name and password are entered and access control is being defined. After pressing the button „Upis“, these data are being recorded into database. If we want to access existing card, we need to click the name of user in users list. Button „Upis“ can be used for changing access control for some users and we can delete users by button „Brisanje“. Finally, all changes will be permanently written into database using button „Potvrdi“.

E. Reports

This function provides report generating (Fig. 10) which refers to:

- User login
- Alarm emergence

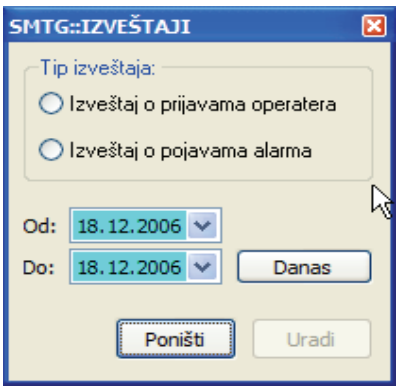


Fig. 7. Report settings

After choosing report type and time period, by pressing the button „Uradi“, report will be obtained as Microsoft Office Word document which can be viewed or printed. It is stored in document archive.

F. Configuration

Only specialized user, with high access level, may configure system. Communication serial port, active measured elements and its addresses are being defined during configuration process. Hardware connections for measured points and display positions could be mutually linked or unlinked. Each generator was given the chart with 48-fielded matrix for digital indicators (Fig. 8). Certain number of measured points is left for turbine. Each measured element has 6 slots with 8 ports each, for hardware connection with measured point. Current unlinked hardware connection is green and free positions have diagonal lines. We create link from hardware connection to free display position using „Drag and Drop“ method. Now, hardware connection is becoming red and two-rowed field for input of measured point description is appearing.

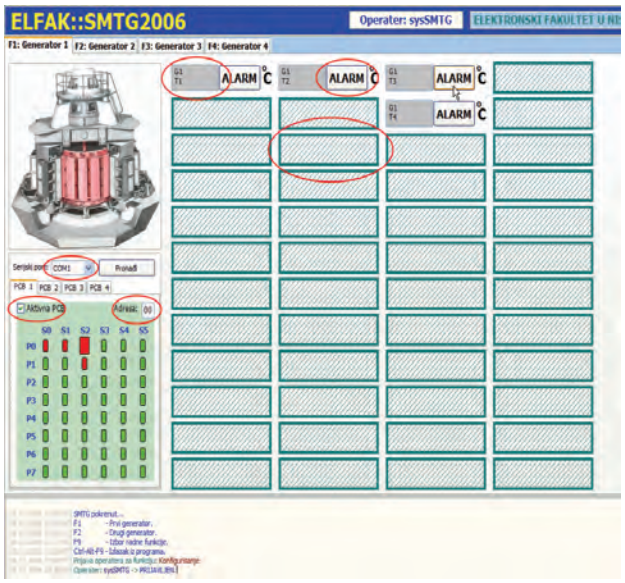


Fig. 8. Configuration system screen

Button „ALARM“ is also appearing. Warning and alarming temperature values for appropriate measured point are defined using this button.

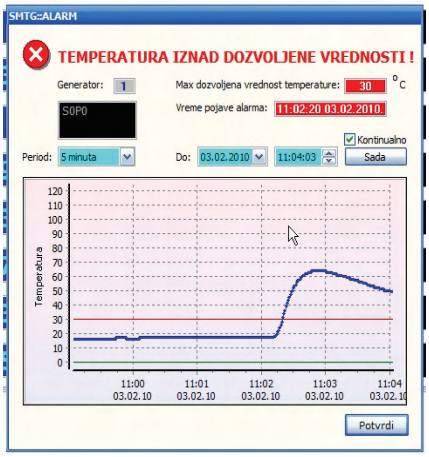


Fig. 9. Chart of current temperature values in generator

III. Conclusion

Described software package is not limited only to temperature measuring. It can be adapted for other measuring systems, for measuring of other physical quantities and parameters at different processes and objects. There is need for an adequate measuring converter for appropriate physical quantity with corresponding proportionality coefficient and creating appropriate sign for this quantity. Graphic and chart representation of measured values, alarm display, data archiving, report generating and printing are provided.

Acknowledgement

The work presented here was supported by the Serbian Ministry of Education and Science (projects III 43007, III 44006 and TR 35005).

References

[1] D. Mitić, S. Stankov, Z. Jovanović, “Projekat merenja temperatura na generatorima” G1, G2, HE “Vrla 2”, u Surdulici, pomoću sistema SMTG, Elektronski fakultet Niš, 2006.

[2] U.Frankenfeld, H.R.Schmidt, S.Popescu, “Temperature monitoring system for the ALICE TPC”, 2004

[3] P.R.N.Childs, “Practical temperature measurement”, Butterworth-Heinemann; First Edition, 2001.

[4] G. Buczek, “Delphy 6 Vodič za programere (+CD)”, Mikro knjiga, Beograd, 2002.

Area Monitor Sensor for Broadband Electromagnetic Environmental Pollution Monitoring

Miodrag Milutinov, Nikola Djuric, Dragisa Miskovic and Dragan Knezevic

Abstract: **The electromagnetic fields are generated by natural phenomena, but also by human activities, mainly through the use of electricity. Development of the man-made electricity and rapid technological progress have multiplied sources of those fields and diversified their characteristics. The present public opinion requires answer if the exposure to electromagnetic fields may increase the risk of some adverse health effects. Therefore, in this paper we present a broadband monitoring system as a solution for automated and constant supervision of electromagnetic fields. The Ministry of Sciences and Technological Developments of the Republic of Serbia recognized the importance of this system within the program of technological development of the Republic Serbia, for the period of 2011-2014.**

Keywords: **EM fields, monitoring, sensors network.**

I. INTRODUCTION

Development of the man-made electricity and rapid technological progress in variety of electrical applications have multiplied sources of the electromagnetic (EM) fields and diversified their characteristics, resulting in the so-called EM environmental pollution.

Exposure to man-made sources of EM fields, ranging from high radio frequencies (mobile phones) through intermediate frequencies (computer screens) to extremely low frequencies (power lines), has increased dramatically in the last few decades. Devices generating non-ionizing EM fields in frequency range from 100 kHz to 300 GHz are in widespread use [1]. The key sources of those fields include mobile phones, cordless phones, local wireless networks and radio and TV broadcasting towers.

Recently, it can be noticed the huge efforts that scientists devote for research, both potential effects on groups of people who have been exposed to electromagnetic fields in their daily lives (epidemiological evidence) and the potential effects observed in laboratory experiments carried out on human volunteers, animals, and cell cultures (experimental evidence). [1]-[3]. Information about the strength of EM fields generated by a given source is readily available and useful in determining compliance with safety limits [4]-[5], but it is also important to consider multi-source exposure and not to focus on single sources, e.g. mobile phone base stations. Therefore, the advanced systems for non-ionizing radiation brad-band monitoring are of vital importance.

In this paper a remote monitoring system, based on wireless sensor information network, for automated, remotely and selectively monitoring of the overall level of EM fields for non-ionizing radiation has been presented [6]-[8]. In addition, the Multi-band Area Monitor Sensor [9]-[10], as a key component of this system, is described.

This system has been recognized by Ministry of Sciences and Technological Developments of the Republic Serbia [11], as valuable support for municipal and provincial agency for non-ionizing radiation protection [12]. Ministry approves development of this monitoring system within the program of the technological development of the Republic of Serbia, for the period of 2011-2014.

The basic concept of the monitoring system is briefly described in Section II, while some details about multi band area monitor sensor in Section III and Section IV conclusion of this paper.

II. BASIC CONCEPT OF THE SYSTEM

The proposed monitoring system has been designed as the advanced solution available to meet the growing demands for systematic and continual, 24-hours, monitoring of cumulative EM fields, from all sources on the observed territory, as it is shown in Fig. 1.

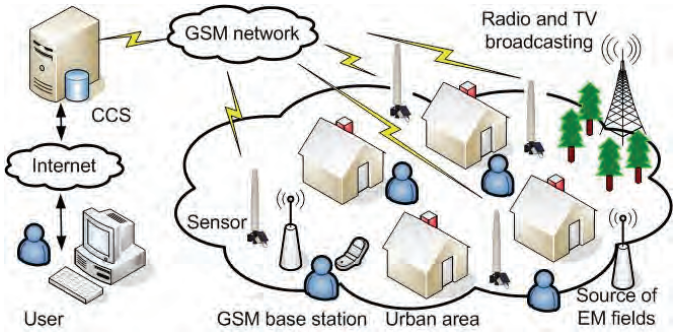


Fig. 1 EM field monitoring system

System performs monitoring both remotely and selectively, avoiding continuous surveillance of sites by technical personnel. This system implements the sophisticated technology of the wireless sensor networks and consists of the

a) spatially distributed autonomous sensors to monitor total level of the EM fields from all sources around the sensors over supervised territory,

1 Miodrag Milutinov is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: mi-odragm@uns.ac.rs

2 Nikola Djuric is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: ndjuric@uns.ac.rs

3 Dragisa Miskovic is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: dragisa@uns.ac.rs

4 Dragan Knezevic is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: tomaja@uns.ac.rs

b) Central Control Station (CCS), that coordinates remote sensor elements, acquire monitoring data, processes and stores them in centralized database,

c) communication network, providing connection and inter-action between remote sensors and CCS, and

d) operating software that manage and supports employment of information network.

The proposed monitoring system is designed so that on a daily basis, investigates the overall level of the EM field and population exposure to the EM field at the particular location, especial over the sites of the special interest.

Compared with the conventional method of investigation, the proposed system does not require presence of technical personnel on the site. As a result, the system provides faster and more efficient monitoring and performs continuous investigation, 24 hours per day during desired period [6]-[9].

III. Multi-Band Area Monitor Sensor

The Multi-Band Area Monitor Sensor (Sensor) is one of the numbers of area sensors which is used for monitoring of electric (E) and/or magnetic (H) fields. The common characteristic of all monitoring equipment is continuous measuring of total electric and magnetic field. The Sensor described in this paper, manufactured by NARDA [10], fulfils investigation of EMFs in sense of public exposure to sources such as radio/TV, GSM, UMTS, or any other wireless system.

The Sensor has ability to measure total electric field over wide frequency range. Frequency range and measuring range depend on used probe. The Sensor has internal circuits for estimating average (AVG or RMS) or maximum (MAX) electric. With internal memory and power supply the Sensor measure and store electric field continuously 24 hours per day. The most important characteristic of this Sensor is ability to estimate contribution of electric field of specific pass bands (GSM, UMTS) to total electric field, as it described in Table I.

TABLE I.
Overall characteristic of Multi-band Area Monitor Sensor

Type of measurement	Total Electric field
Frequency range	100 kHz – 7 GHz
Measurement range	0.03 - 200 V/m
Stored field values	AVG, RMS, MAX
Duration of measurement	24 hours per day
Applications	Wideband, GSM900, GSM1800, UMTS

The Sensor consist of: probe, post-processing data logger with internal memory, GSM modem and RS232 connector and internal rechargeable battery power supply, as shown in Fig. 2.

Beside these, the Sensor has solar panel for recharging internal battery and external power supply during a day, which is highly important for autonomous work. Whit this autonomy Sensor accomplishes measurement in arbitrary long time period as a means of assessing the long-term public exposure to potentially hazardous electromagnetic fields. In case of total darkness internal battery provides full operation of the Sensor of more than 80 days.

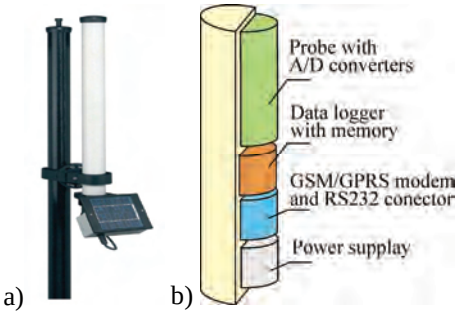


Fig. 2. a) Sensor unit with solar panel, b) Sensor's hardware component.

Functionality of the Sensor can be briefly described in three main stages, as shown in Fig. 3.

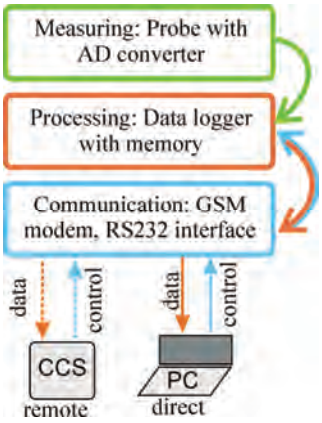


Fig. 3. Functionality of MBAMS

First, signal obtained from field probe and digitalized in A/D convertor continuously coming to the data logger. Second, data logger according the program setup processes and stores the signal in internal memory. Third stage is remote communication between the Sensor and CCS via GSM modem uploading stored data to CCS. Small amount of data like daily report could be sent by SMS to mobile phone. Additionally communication block has ability to receive instruction from CCS to setup data logger. RS232 interface enable direct connection to the Sensor for measurement and collection of stored data on the site.

One of the most appropriate and useful probe regard to high frequency EMF public exposure is a multiband quad electric probe, described in this paper. The main characteristic of the probe are frequency and measurement range described in Table II, frequency discrimination and isotropy.

The probe allows discrimination between the electric fields generated by different sources in frequency bands of GSM 900, GSM 1800 and UMTS 2100, inside the wide band. This

is useful in order to find GSM and UMTS contributions to total electric field.

TABLE II
ELECTRIC FIELD QUAD-BAND PROBE

Application	Frequency range [MHz]	Meas. range [V/m]	Meas. resolution [V/m]
Wideband	0.1–3000	0.2-200	0.01
GSM900	925-960	0.03-30	0.01
GSM1800	1805-1880	0.03-30	0.01
UMTS2100	2110-2170	0.03-30	0.01

The quad-band probe performs simultaneous three-axis measurements, as shown on Fig. 4, which allows obtaining the total field independently of the tri-axial orthogonal arrangement. Isotropic measurement avoids technical personnel on the sites and manual rotation of the sensor. Maximum probe anisotropy is calculated according IEEE Standard for Calibration of Electromagnetic Field Sensors and Probes, IEEE Std. 1309-2005 and it is ±0.8dB calculated at centre frequencies of each band [10].

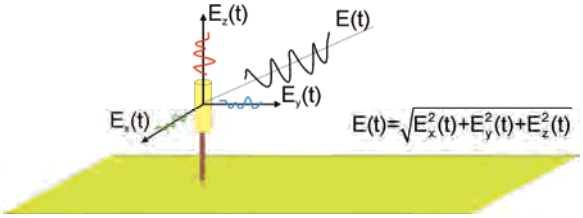


Fig. 4. Isotropic response.

Probe has internal pass band circuits, A/D converters a microcontroller and calibration data memory that produce measurement results in digital form, depicted on Fig. 5. With this filters the sensor has ability to processes data to yield the AVG or RMS and Peak isotropic result which are simultaneously output of all bands.

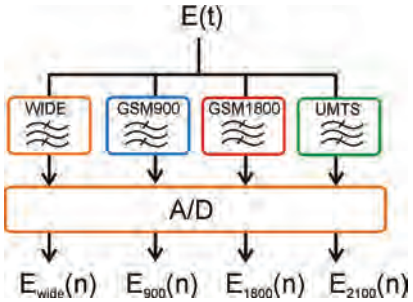


Fig. 5. Probe pass bands circuits

Other probe characteristic relevant for measurement uncertainty such as flatness, out of band attenuations, centre frequency drift, H-field rejection and temperature error are fully described in data sheet obtained from manufacturer [10].

The second part of the Sensor, named data logger, performs sampling, post-processing and storing of data obtained from the probe. The sampling time is 3 seconds, which mean that Sensor capture 20 samples of measurement data per minute.

Investigations of public exposure to electromagnetic fields for frequencies between 100 kHz and 10 GHz according to regulations [4], [5] demands measurement of the appropriate quantities (electric or magnetic field) averaged over any 6 minutes period. The Sensor has ability to perform averaging of numbers of sampled data, which is finding very useful in case of investigation public exposure to the EMFs. In order to meet regulations demands, the Sensor provides averaging over number of collected samples in arbitrary time period (t_{AVG}). The averaging can be arithmetic (AVG), appropriate for frequencies below 10 MHz as described in Eq.(1)

$$E_{AVG} = \frac{\sum_{n=1}^{N} E_n}{N}, \qquad (1)$$

and quadratic (RMS), appropriate for frequencies between 100 kHz and 10 GHz, described in Eq. (2)

$$E_{RMS} = \sqrt{\frac{\sum_{n=1}^{N} E_n^2}{N}}. \qquad (2)$$

Beside this, the Sensor detects maximum (MAX) value over the same time period as described in Eq. (3)

$$E_{max} = \max\{E_1, E_2, \ldots, E_n, \ldots, E_N\}. \qquad (3)$$

The number of samples N, depends on averaging time interval, as described in Eq. (4)

$$N = 20\left[\frac{\text{sample}}{\text{min}}\right] \cdot t_{AVG}\ [\text{min}]. \qquad (4)$$

Due to internal memory, the Sensor can perform storing of processed results at one of selectable time interval (30 sec, 1 min, 2 min, 6 min or 15 min). Fig. 6 shows an example of storing results of electric field measured data at every 1 min with averaging period of 6 min, based on 120 samples collected in the previous 6 minutes. The Sensor simultaneous process and stores E_{AVG}, E_{RMS} and E_{MAX} to the internal memory.

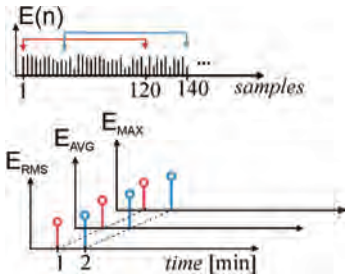


Fig. 6. Example of processing and storing measured values.

With 4MB flash internal memory, the Sensor provides autonomous work and logging from a number of days which depends on storing time. When the memory is full, the new data are overwritten on the oldest to ensure availability of the data for the most recent measurement period. The maximum

numbers of days before overwriting is described in Table III. Sensor provides automatic downloading data from Sensor's memory at selectable time avoiding data overwriting.

TABLE III.
MEMORY CAPACITY OF THE SENSOR.

Storing time	Memory capacity
30 sec	5 days
1 min	10 days
2 min	20 days
6 min	60 days
15 min	169 days

All data stored in Sensor's memory are downloading trough communication block (GSM modem) to the PC, described in Section I. The appropriate software creates separate ASCII files containing downloaded measurement results for each Sensor and for each storing time. Example of several successive measurements averaging in 6 minutes and storing every 30 sec is shown in Fig. 7

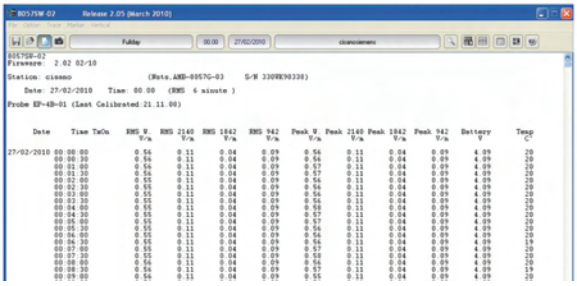


Fig. 7. Example of data table display.

Due to Sensor's small size and narrow shape as well as solid weather-proof unit, it is able to be easily installed indoor or outdoor. It is possible to use appropriate poles for mounting on the any surface. For example, sensor could be placed on the public square or playground in vicinity of base station, on the roof top of surrounding buildings or inside the schools, hospitals, or any other object of interest, as shown on Fig. 8. Moreover, several Sensors connected can be used to build reliable monitoring systems to cover large geographical areas, including nationwide coverage [6]-[8].

IV. CONCLUSION

The proposed project implements a modern and up-to-date technology that automatically and continuously performs investigation and monitoring of EM fields and records the overall level of electromagnetic pollution on environment.

The Sensor manufactured by Narda Sts. with its main characteristics described in this paper is finding very useful in monitoring high frequency electric field from sources such as TV/Radio, GSM, UMTS and WLAN. The Sensor estimate wide band total electric field and partial contribution of GSM and UMTS due to built-in band pass circuits.

System of numbers of Sensors can be placed anywhere and connected to build reliable monitoring systems to cover large geographical areas.

Fig. 8. Sensor located outdoor on the roof.

ACKNOWLEDGEMENT

This paper has been supported by Ministry of Sciences and Technological Development of the Republic of Serbia, through the grant for project of technological development TR 32055.

REFERENCES

[1] International Commission on Non-Ionizing Radiation Protection (ICNIRP) – "Exposure to high frequency electromagnetic fields, biological effects and health consequences (100 kHz-300 GHz)", 2009

[2] EU Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) – "Possible effects of Electromagnetic Fields (EMF) on Human Health", 2007

[3] EU Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) – "Health and electromagnetic fields", 2009,

[4] European Committee for Electrotechnical Standardization (CENELEC) – Council Recommendation on the limitation of exposure of the general public to electromagnetic fields (0 Hz to 300 GHz),

[5] "Regulation on the limits exposure of non-ionizing radiation", the law of Republic of Serbia, no. 104/09,

[6] N. Djuric, "Razvoj informacione mreze za kontinualno i udaljeno isptivanje elektromagnetskih polja," submitted for the Serbian journal Telekomunikacije 2011,

[7] N. Djuric, M. Prsa, K. Kasas-Lazetic, "Information network for continuous electromagnetic fields monitoring," submitted for conference PES 2011, Niš 2011.

[8] N. Djuric, M. Prsa, K. Kasas-Lazetic, "Serbian remote monitor-ing system for electromagnetic environmental pollution," submitted for conference TELSIKS 2011, Niš 2011.

[9] M. Milutinov, N. Djuric, B. Vukobratovic, "Multi-band area monitor sensor in information network for electromagnetic fields monitoring," submitted for conference PES 2011, Niš 2011.

[10] Narda Safety Test Solutions GmbH, AMB-8057 User's Manual, Narda, 2007.

[11] Ministry of Sciences and Technological Developments - http://www.nauka.gov.rs/,

[12] Municipal Agency for the Environmental Protection – City of Novi Sad – http://www.environovisad.org.rs.

Sensor Communication in Wireless Electromagnetic Field Monitoring System

Bogdan Vukobratovic[1], Nikola Djuric[2], Dragisa Miskovic[3] and Dragan Knezevic[4]

Abstract – **The high level of attention about the environmental electromagnetic field pollution stimulates technical solutions that provide even more sophisticated systems for non-ionizing radiation monitoring. In this paper the communication of the sensor in a wireless electromagnetic field monitoring system has been considered. This system is intended for continuous, remote electromagnetic fields supervision and it has been recognized by Ministry of Sciences and Technological Developments of the Republic of Serbia, which grants its development within the program of the technological development of the Republic of Serbia, for the period of 2011–2014.**

Keywords – **EM pollution, monitoring, sensors, wireless.**

I. INTRODUCTION

Electromagnetic (EM) fields occur in nature and have thus always been present on earth. However, during the last few decades, the exposure to the man-made sources of EM fields has steadily increased due to social implementation of various electrical appliances.

Virtually entire population is exposed to EM fields through sources such as high-voltage power lines, household electrical appliances, computers, radio and TV broadcast facilities, mobile telephones and their base stations and a variety of wireless technologies systems. The frequencies of EM fields in the electrical applications vary between 0 Hz to 300 GHz with diverse characteristics [1].

Over the years, scientists made efforts in research about the potential health effects of EM fields and published theirs research results in many scientific papers about this topic [2]-[3]. Respecting the results of research in all these paper it prevails that in order to provide a high level protection of public health the establishment of a legal framework for the protection of workers and citizens is required.

As support for efforts of the Serbian municipal and provincial agency for non-ionizing radiation protection [4], our team has been proposed developments of a wireless sensor information network [5]-[7], for automated, remotely and selectively monitoring of the overall level of EM fields for non-ionizing radiation. This system is conceived as an answer for continual 24-hours supervision of all sources that emit EM fields, over the territory covered by sensor network.

Additionally, the system is capable for daily checking the protection limits set by the national regulations [8]-[10] and it has been recognized by Ministry of Sciences and Technological Developments of the Republic of Serbia, which approves its development within the program of the technological development of the Republic Serbia, for period of 2011–2014.

The crucial element of this monitoring system is multi-band area monitoring sensor [11]-[14], that is responsible for wireless and remote data acquisition in EM field measurements. In this paper the communication techniques implemented in this sensor are presented. In Section I the basic concept of the proposed remote monitoring system is given, while in Section II some details about sensors and central control station (CCS) communication. Section III presents GPRS/GSM protocol and Section IV conclusion of this paper.

II. EM FIELDS MONITORING SYSTEM

The monitoring system is proposed as the most suitable solution for constant, wireless supervision of EM fields and total level of the exposure to these fields. This system is a significant support in efforts to take systematic care of potential effects of non-ionizing EM fields on the health of the population, taking into account their concern about long-term exposure to EM fields.

This system implements the sophisticated technology of the wireless sensor networks and it is deliberate to supervise and control if the cumulative level of the EM fields is below the nationally prescribed limits [8], [10].

The basic concept of proposed monitoring system is shown in Fig. 1.

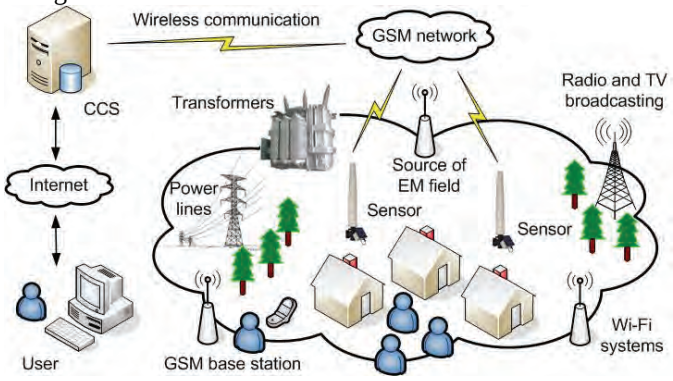


Fig. 1 Remote EM fields monitoring system

This system consists of the

a) autonomous sensor elements, which are spatially distributed to observe cumulative level of EM fields over super-

[1]Bogdan Vukobratovic is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: vukobratovic@uns.ac.rs

[2]Nikola Djuric is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: ndjuric@uns.ac.rs

[3]Dragisa Miskovic is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: dragisa@uns.ac.rs

4Dragan Knezevic is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: tomaja@uns.ac.rs

vised territory and from all sources around the sensors, in particular frequency range,

b) Central Control Station (CCS), that coordinates activity of the remote sensor elements, to collects data obtained by monitoring, processes and stores them in centralized database,
c) communication network, providing connection and interaction between remote sensors and CCS, and
d) operating software that manages and supports employment of information network.

CCS and sensors communication is performed using the existing GSM mobile networks. Both, the CCS and sensors are equipped with a quad-band GSM/GPRS modem for uploading/downloading data and remote control [13].

The monitoring system is able to perform investigation for a number of sources, that can be found in a power system and systems for the power transmission (transmission lines, distribution and substation equipment), through a system of radio and TV broadcasting, to the wireless communication networks (like GSM, UMTS or Wi-Fi systems) [5]-[7].

III. Sensor Communication Approaches

One of the major challenges in developing an acquisition network is collecting data from measurement nodes. Straightforward way of collecting data is to visit periodically remote and scattered measurement nodes and manually collect data. Sometimes, this is not adequate approach and therefore if the solution can be found such that it allows remote and automatic collecting of the measurement data, such acquisition network becomes a very powerful monitoring tool.

In our monitoring system sensors are based on GSM wireless communication. Using GSM/GPRS modem sensors have option of transferring data without intervention of technical personnel, as it is shown in Fig. 2.

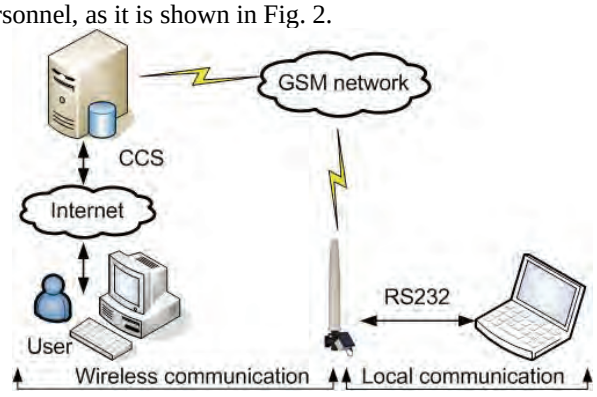


Fig. 2 Communication with sensor

Additionally, communication with sensor can be performed locally via RS232 protocol.

A. *Remote connection through GSM network*

Information network proposed here, takes advantage of GSM network and Internet for it's data transmission. National GSM mobile network covers more than 95% of Serbian territory, thus allowing GSM connection from nearly every spot within. This allows sensor elements to continuously communicate from the point of measurement with CCS, sending real time measured data. Furthermore, mobile operators give option for GSM modems to connect to the Internet through the GSM network that enables placing CCS virtually anywhere.

There are two ways for the sensor to connect to GSM network [13]:

a) using internal GSM/GPRS modem that allows connection or
b) connecting to GSM network via another device (laptop or PC) with GSM modem, as shown in the Fig. 3.

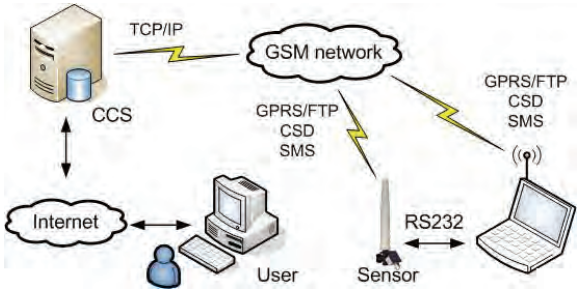


Fig. 3 Sensor-CCS connection over GSM

It should be noticed that laptop has to have the appropriate software, which will enable communication with sensor via RS232 protocol.

It is know that radio communication consumes energy, and the duration of autonomic sensor operation is crucial in wireless sensor networks. Thus, the GSM module is shut down for the majority of time. This means that sensor is unavailable for communication over GSM network but in short preset times of the day. Besides that, sensor switches on the module whenever an alarm condition occurs or when the normal conditions are restored.

Through the modem sensor can make use of three different protocol stacks commonly available in GSM networks:

a) Short Message Service (SMS),
b) Circuit Switched Data (CSD) and
c) General Packet Radio Service / File Transfer Protocol (GPRS/FTP)

In Table 1 the availability of various functions are given in respect of the protocol used.

TABLE I
List of all possible functions of the sensor and their availability through different protocols

Possible functions	SMS	GSM	RS232
Settings	YES	YES	YES
Reading status and alarms	YES	YES	YES
Reading max value	YES	YES	YES
Reading average value	YES	YES	YES
Download data	NO	YES	YES
Reading of battery	YES	YES	YES
Reading of internal temperature	YES	YES	YES
Spontaneous call	NO	YES	NO
Reporting alarms	YES	YES	NO
SMS reporting of daily maximum	YES	NO	NO

The following sections describe the communication with the sensor through various protocols listed above.

B. SMS Communication

SMS is the text communication service used for transfer of short text messages between the devices in network. Alarms, some settings and queries for some small amount of data can be communicated with SMS as shown in Fig. 4.

Fig. 4 SMS communication

The sensor can be set with up to two mobile phone numbers that will receive alarm and daily report messages. Daily report message contains: maximum measured field, minimum battery voltage and status of the sensor, telephone number of the sensor, time and date.

Sensor allows setting up several alarms that can be sent on following circumstances [13]:

a) when the measured property exceeds a threshold setting for attention (warning),
b) when measured property exceeds an alarm threshold,
c) when battery voltage is too low,
d) when one or more field probe parameters is faulty,
e) on exceeding the temperature limit inside the field monitor,
f) when the data memory is almost full and
g) when the external container is opened

Commands that can be sent to the sensor have the following format:

#F7Qcommand(parameters)*

while particular parameters are explained in Table 2

TABLE II
SMS COMMAND FORMAT EXPLANATION

Token	Use
#	opens the command string
F7	string always included
Q	? - for query commands S - for setting commands
command	command string
(parameters)	value of setting parameters (if any)
*	closes the command string

Additionally, commands that the sensor understands include the ones listed in Table 3.

TABLE III
EXAMPLES OF SMS COMMANDS

Command	Description
#F7?IDN psw*	request ID + password authentication
#F7?CLK*	request complete internal time
#F7?BAT*	request battery voltage status
#F7?PRB*	request field probe data
#F7SALR (par)*	set alarm threshold
#F7SAVG (par)*	set time of averaging

Every SMS sent to the sensor must contain as a first command the invitation to authenticate the password, which is the protection mechanism against unauthorized access [13].

C. CSD Communication

CSD is the type of the communication where the direct transmission line is established between two parties, identically as it is done with a voice call. Drawbacks of CSD over the GPRS protocol are: less bandwidth (9.6 kbps), need for a dial-up connection establishing and charging per duration of the connection (as with voice calls). Also the CCS has to have GSM modem to receive the calls and dial-up the sensors in order to communicate.

D. GPRS Communication

GPRS is a packet oriented mobile data service, which offers higher data transfer rates (for an example up to 53.6 kbps offered by Telenor GSM mobile operator in Serbia). Internet protocol TCP/IP can be used with GPRS, that enables the use of many standard internet application layer protocols, such as: FTP, HTTP, IMAP and many other.

The File Transfer Protocol (FTP) is standard protocol used to copy a file from one host to another over a TCP-based network. The sensor is able to use FTP application layer protocol when set to communicate via GPRS with CCS [14]. In order to use FTP communication the CCS must have internet access and has to have static IP internet address.

The advantage of using GPRS/FTP is reduced communication costs, because GPRS data communication is charged on the basis of the very low volume of data transferred. Furthermore, there is no additional cost related to the distance, as the FTP server can be accessed from anywhere that Internet connection is available.

E. Data exchange flow

Remote stations continuously perform field measurements and store results in their internal memory. At programmed time each station uploads the stored measurement data to the CCS.

In order to make the connection parameters, shown in Table 4, have to be set.

TABLE IV
SENSOR PARAMETERS

Parameter	Description
Access Point Name	The GPRS provider name
User Name	User Name required for GPRS
Password	Password required for GPRS
CCS IP address	IP address of FTP server
CCS user name	User name for accessing FTP server
CCS password	Password for accessing FTP server

When connecting to the server, the sensor will check for the existence of any request to modify its settings or to upload the results related to some specific time interval. Precisely what sensor does is as follows:

- first it loads, if any, all the settings which have been written by client,

- then it writes all measurement data that have not yet been uploaded
- finally it erases all previous required settings in order to avoid reloading them at the next connection (data related to field measurements are never deleted by the sensor) and
- after that connection is terminated, as shown in Fig. 5

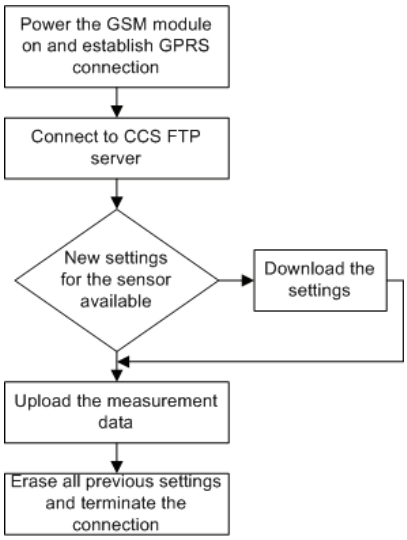


Fig. 5 Sensor – CCS Data exchange flow over FTP

The client software provided by vendor possesses all functions needed for controlling sensor elements. When client sets the parameters for the sensors, the control software will create on CCS one directory for each sensor, if it does not already exist, and place settings files inside. The directories are named after the serial numbers of the sensors, so that each sensor checks it's own folder. In the same folder the sensor places the measurement data.

F. RS232 Communication

RS232 (Recommended Standard 232) is the serial binary single-ended data and control signals standard. The sensor has an option of connecting to a local PC through RS232 for sending commands, downloading data or connecting to CCS over GSM if PC is equipped with GSM module.

The messages which sensor understands when RS232 communication is used are identical to the messages that can be sent over SMS with some exceptions:

- RS232 can be used to upload measurement data and
- RS232 can be used to download a new sensor firmware

The sensor has 4MB of flash storage and can keep the measurement data for up to several months, depending on the settings. So, in applications where real-time tracking of data is not important, client can collect data through RS232 periodically without the need of GSM communication.

G. Power supply

Crucial problem to be solved in designing an autonomic sensor is the power supply. The sensor comprises solar panel and an internal rechargeable Pb-gel battery for unlimited outdoor operation. For indoor applications, the autonomy of the sensor is at least 80 days, depending on the settings used.

IV. Conclusion

Thanks to the communication via widespread GSM mobile network, the proposed monitoring system has a high degree of mobility and can be installed in any part of the area of interest.

The mobility of the sensor elements enable their spatial deployment and integration in the information network for EM field monitoring over the large area of interest, which is virtually impossible using conventional approach.

By implementing the proposed monitoring system, municipal authorities for environmental protection [4], gain the valuable tool for interested users and the general public, in timely informing, prevention and population protection against exposure to the EM fields.

Acknowledgement

This paper has been supported by Ministry of Sciences and Technological Development of the Republic of Serbia, through the grant for project TR 32055.

References

[1] International Commission on Non-Ionizing Radiation Protection (ICNIRP) – "Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz)", 1998, http://www.icnirp.de/documents/emfgdl.pdf

[2] EU Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) – "Possible effects of Electromagnetic Fields (EMF) on Human Health", 2007 http://ec.europa.eu/health/ph_risk/committees/04_scenihr/docs/scenihr_o_007.pdf,

[3] EU Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) – "Health and electromagnetic fields", 2009 http://ec.europa.eu/health/archive/ph_determinants/environment/emf/brochure_en.pdf,

[4] Municipal Agency for the Environmental Protection – City of Novi Sad – http://www.environovisad.org.rs

[5] N. Djuric, "Razvoj informacione mreze za kontinualno i udaljeno isptivanje elektromagnetskih polja," submitted for the Serbian journal Telekomunikacije 2011

[6] N. Djuric, M. Prsa, K. Kasas-Lazetic, "Information network for continuous electromagnetic fields monitoring," submitted for conference PES 2011, Nis, 2011

[7] N. Djuric, M. Prsa, K. Kasas-Lazetic, "Serbian remote monitoring system for electromagnetic environmental pollution," submitted for conference TELSIKS 2011, Nis, 2011

[8] V. Bajovic, N. Djuric, D. Herceg, "Serbian laws and regulations as foundation for electromagnetic field monitoring information network," submitted for conference PES 2011, Nis, 2011

[9] "Law on Non-Ionizing Radiation Protection", the law of Re-public of Serbia, no. 36/09

[10] "Regulation on the limits exposure of non-ionizing radiation", the law of Republic of Serbia, no. 104/09

[11] M. Milutinov, N. Djuric, B. Vukobratovic, "Multi-band area monitor sensor in information network for electromagnetic fields monitoring," submitted for PES 2011

[12] M. Milutinov, N. Djuric, D. Miskovic, D. Knezevic, "Area Monitor Sensor for Broadband Electromagnetic Environmental Pollution Monitoring," submitted for conference PES 2011, Nis, 2011

[13] Narda Safety Test Solutions GmbH, AMB-8057 User's Manual, Narda, 2007

[14] Narda Safety Test Solutions GmbH, AMB-8057 User's Guide to the GPRS/FTP communication, 2010

Wireless Sensor System for Measuring Parameters of UV Radiation

Zoran Petrušić[1], Uglješa Jovanović[2], Igor Jovanović[3], Ljubomir Vračar[4] and Dragan Mančić[5]

***Abstract* – Development of our own mobile wireless sensor networks for tracking the critical parameters of environmental protection (gas emissions, ionizing and non-ionizing radiation, noise and various meteorological parameters) in the specific area has led to the development of different detecting measuring devices. This paper presents a two-channel wireless measuring system for measuring the intensity of UV radiation and UV index. The system can perform on-line tracking of these parameters on maximum distance of 300m from the recepting unit in case of RF transmission, while the distance is not a critical parameter for GSM/GPRS module.**

***Keywords* – UV radiation, UV photodiode, Wireless sensor system, RF transmission.**

I. Introduction

Solar radiation is an important factor which creates the climate on Earth and therefore it formes the whole environment of the biosphere. The spectrum of electromagnetic radiation includes radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation and also *x* and γ radiation. Radiation that reaches sea level, where the most part of the biosphere is, includes part of the spectrum with wavelengths between 290nm and 3000nm. Energy disposition in this range is as follows (see Fig. 1a):

- ultraviolet radiation – UV (λ=290-490nm), makes 5% of radiation spectrum;
- visible radiation – VIS (λ=400-780nm), makes 39% of radiation spectrum;
- infrared radiation – IR (λ=780-3000nm), makes 56% of radiation spectrum.

The ultraviolet part of the solar spectrum has an important role in many processes in the biosphere. However, despite the large number of beneficial effects, some of the effects of ultraviolet radiation can be very harmful if they exceed a certain level of radiation safety. Because of increased energy, a photons of ultraviolet radiation can produce biological activities relevant for human health. If the amount of UV radiation is high enough, the ability of self-protection of certain biological organisms and humans becomes insufficient so they may suffer significant damage. In the human body that damage is primarily related to organs such as skin (causes burns) and eyes (a cataract is formed), and in some cases can cause degradation of DNA molecules [1].

In order to avoid harmful exposure to UV radiation a international unit is defined, so called UV index, which quantitatively characterizes erythemal effects and which should warn people on the level of harmful UV radiation. The most important factors which affect on the intensity of UV radiation are: ozone layer, position of Sun, mean sea level, dispersing in the atmosphere, clouds, atmospheric visibility and reflection on the ground [2].

The division of UV radiation defined by CIE (*Commission Internationale de l'Eclairage, Publication No. 69, 1985*) is shown in Fig. 1b. Definition by the biological effects, the chemical effects, as well as the division from the standpoint of spectroscopy.

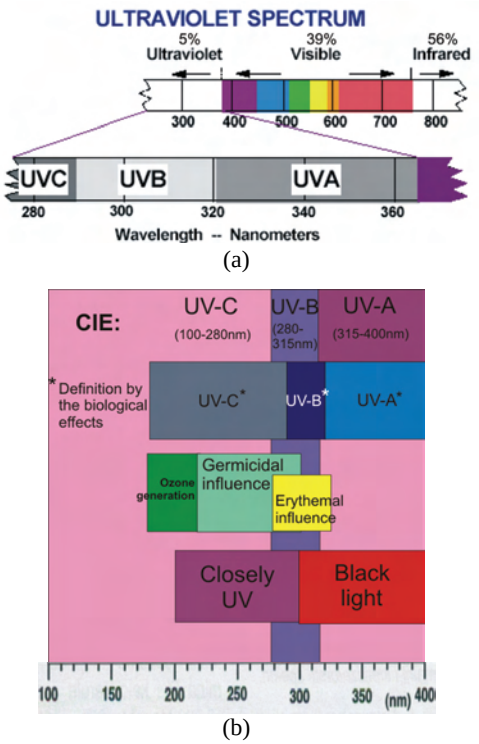


Fig. 1. UV radiation: (a) energy disposition; (b) radiation division

UV index is measured in relation to the horizontal surface at different centers in the world who are either responsible for providing daily weather forecast or do some research work. However, as people expose their bodies to radiation in all directions and because of continuous apparent motion of the Sun in the sky it is necessary to know levels of UV index in different directions in order to reliably reconstruct solar incidence throughout the human body. In cases with particularly high UVI values, such as in the Antarctic due to high reflection, on construction sites where workers are exposed to sun for several hours, or in regions where the

Autors are with the University of Nis, Faculty of Electronic Engineering, A.Medvedeva 14, 18000 Nis, Serbia, E-mail: [1]zoran.petrusic@gmail.com; [2]ugljesha85@yahoo.com; [3]igor_j@elfak.rs; [4] ljubomir.vracar@elfak.ni.ac.rs; [5] dragan.mancic@elfak.ni.ac.rs.

atmosphere filters out UV radiation significantly less due to reduction or absence of the ozone layer, it is necessary to perform several measurements in different directions. This issue has been discussed in several papers from different aspects, paper [3] uses five UV sensors for measuring along different directions.

In this paper, the goal was to design and develop a prototype of sensor node suitable for measuring UVI values which will become the basis for future establishment of a wireless sensor network for measuring UVI in previously mentioned special conditions.

II. UV Sensors and Interfaces

Measuring the intensity of UV radiation in different bands requires the use of modern UV sensors and adequate measuring systems. In principle, depending on to the mechanisms of detection that they use sensors of UV radiation can be classified into two main groups, as a photon or thermal sensors. The basic classification of UV sensors is shown in Fig. 2.

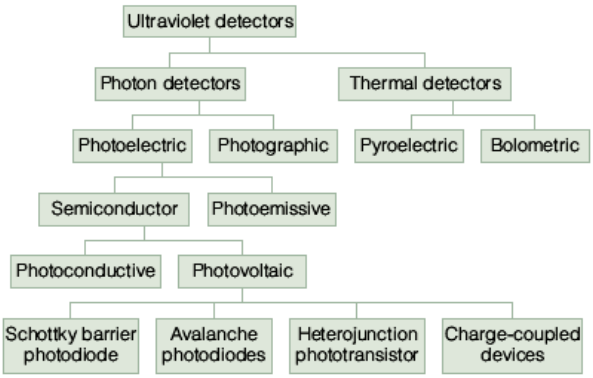


Fig. 2. Classification of UV sensors depending on the mechanisms of detection that they use

The most numerous group of photoelectric sensors are semiconductor sensors. Therefore, the analysis in this paper are focused on the photonic sensors and on the photoelectric semiconductor UV sensors.

Semiconductor UV sensors have a negative feature to the very extreme absorption and high intensity of radiation which induces aging effect in the majority of semiconductor materials. Because of that, at one point of time, development of highly efficient semiconductor quantum UV sensors was a slow. Semiconductor UV sensors can be used in both modes, in photoresistive and in photovoltaic mode. There are several variants of photovoltaic sensors and they depend on the applied effect, so there are: Schottky barriers, PIN diodes, avalanche photodiodes, phototransistor with heterogeneous connection, or devices based on the principle of CCD (charge coupled devices).

Studies in modern semiconductor UV photodetectors are focused on modification of wide semiconductor barriers and on various applications of adequate materials in this structure, such as silicon carbonate (SiC), diamond, gallium and aluminum nitride (AlGaN) alloy, GaAsP, etc.., with characteristics of the materials that make up this family of semiconductors unique for the use in UV sensors. In general, semiconductor UV sensors are combination of photodiode chips based on previously mentioned alloys (SiC, AlGaN, GaAsP) and adequate UV filters. In addition, these sensors must meet the appropriate demands in terms of physical robustness, they must be chemically inert, they must have high corrosion resistance and must not be toxic. On the other hand, modern measuring systems require the use of miniature compact UV sensors with the extra peripheral circuits located in the sensor.

The developed measuring system uses the UV photodiodes developed by *Sglux SolGel Technologies Germany*, with labels *SG01L*-5 and *Eryca* whose spectral characteristics are shown in Fig. 3 [4].

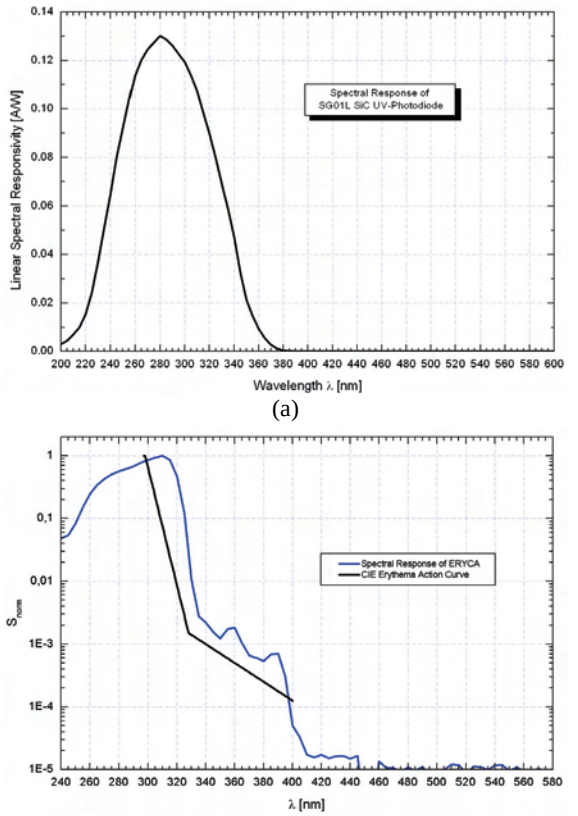


Fig. 3. Characteristics of UV photodiodes: (a) *SG01L-5*; (b) *Eryca*

These UV photodiodes have been made of detector materials based on silicon carbide (SiC), with an approximative spectral band of 210nm to 380nm and are not sensitive to UV radiation outside this band. The application of these diodes depends of spectral characteristics. Thus, the UV photodiode *Eryca* has a spectral response curve matching to erythemal's curve as shown in Fig. 3b, making measurements more accurate with an accuracy of ±0.5 UVI.

These UV photodiodes are ideal detectors for reliable use in tracking of UV spectrum due to their exceptional durability, and due to long-term enduring to UV radiation. Practice has shown that the diode detectors made of materials such as Si, TiO_2, GaN or diamond, by its characteristics can not compete with UV photodiodes based on SiC because of the following benefits:

- internal spectral response is limited to a band of 200nm to 400 nm, and doesn't require additional blocking of unwanted visible and infrared radiation;
- outstanding long-term stability even when exposed to high doses of UV radiation up to the value of 1000W/m^2;
- exceptional temperature stability, temperature coefficient Tk<-0.06%/K provides stable operation on temperatures up to 150°C;
- by integrating with different filters they can be applied in water, flame detection, spectroscopy, food inspection at different UVA, UVB, UVC, UVBC, UVBC2 measurements and in measurements of various biological effects of UV radiation.

Typical levels of electric current generated in the silicon UV photodiode sensors are in range of nA and therefore it is necessery to use different amplifier configurations in order to increase the power of a signal. A standard transimpedanse amplifier which converts electric current into voltage was used in the prototype.

The amplifing stage, with the current into voltage converter, is consisted of three parts, and its block diagram is shown in Fig. 4. The first part converts current flowing through the UV photodiode into the voltage. A fixed level of an amplifier's gain is defined which corresponds to the selected sensor.

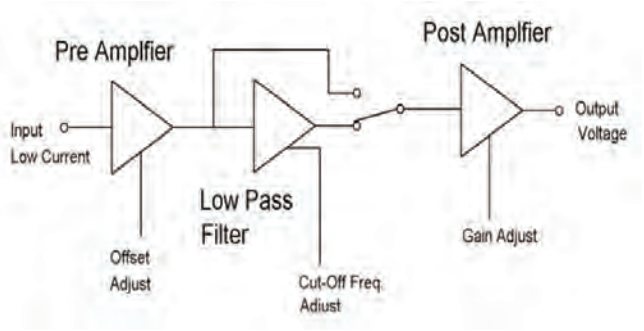


Fig. 4. Block diagram of the amplifing stage

The second part of the amplifing stage is a low-pass filter which attenuates noise signals up to 100Hz. In the third part of the amplifing stage aditional amplification of signal is carried out, so that an output signal corresponds to the level of input signal of A/D converter in PIC16F887 microcontroller.

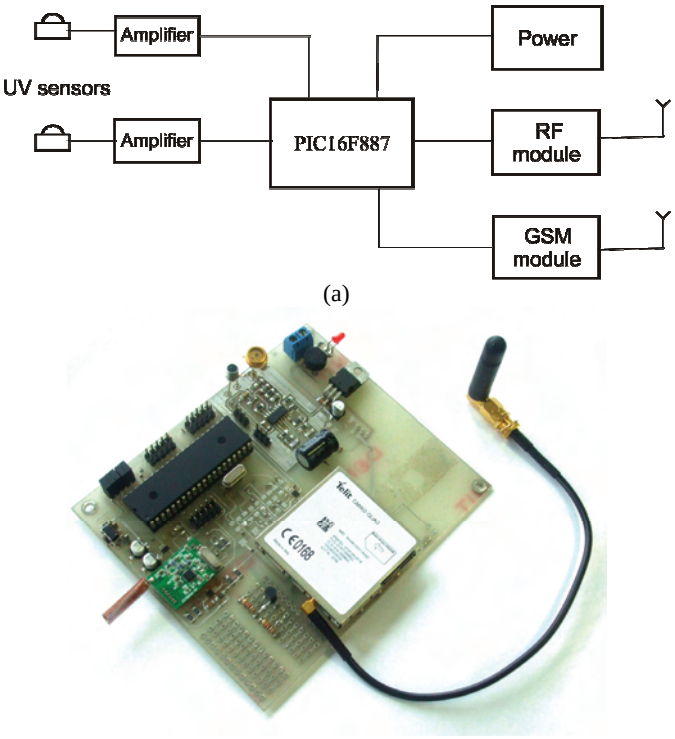
For adjustment and calibration of the measuring system in a specified band of UV radiation it is necessary to use highly specialized sources of radiation, but due to high cost of these sources a variant with currently available UV sources was selected. For calibration of UV sensors in a certain bands a different sources of UV radiation were used (*blacklight blue* fluorescent tube made by *GE Lighting* with power of 6W, UV lamp *Type3* with label *Philips HPA 400W*, a commercial lamp used in mercury lamps, fluorescent tubes with labels *Philips 5M*/*5W*).

III. System Development

A wireless measuring detector has a flexible hardware configuration and with the minimum hardware and software changes it can become the basic unit of an automatic, mobile, measuring system for tracking and measuring of various physical quantities and parameters in environment. Basic features which such a measuring detector must possess are: automatic operation, power autonomy, wireless communication, compatibility with various sensor elements, the ability to memorise a certain number of data, alarm level settings, autonomy in decision making at the first level of degree, reliable transfer of measured data, measuring multiple values, etc.

In this case, the intelligent wireless measuring detector has a two channels, for measuring the intensity of UV radiation and for measuring UV index. Its basic components are:

1. PIC microcontroller (PIC16F887);
2. UV diode *SG01L-5* and UV diode *Eryca*;
3. RF transceiver *ST-TR1100*;
4. GSM/GPRS module *Telit GM862*;
5. Amplifier *LPC660AIM*;
6. Power circuit.



(a)

(b)

Fig. 5. Detecting unit: (a) block diagram; (b) photo of realized unit

Due to many advantages, microcontroller PIC16F887 was selected for development of the detecting unit. Besides the basic functions such as data acquisition, testing of conditions, data processing and sending and SMS sending, microcontroller also has a total of 14 10-bit analog inputs for A/D conversion, allowing connection of multiple sensors. The detecting unit transfers measured values to the receiving unit using RF transceiver. Data transfer can also be done by using GSM/GPRS module directly on a web server via GPRS, or to a mobile phone via SMS. Measured values can be directly displayed on the detecting unit using the optional LCD display which may exclude the use of the receiving unit. Block diagram of the detecting unit is shown in Fig. 5a, and photo of realized detecting unit is shown in Fig. 5b.

Since the secondary focus was on a RF transmission and therefore the receiving unit is consisted only of a microcontroller PIC18F2550 and the RF transceiver. The receiving unit can be connected to the PC via USB or serial port, if it is necessary measurements can also be displayed on the optional LCD. Additional sensors can also be connected on this unit with minimal hardware and software modification.

Developed wireless measuring system is shown in Fig. 6. The presented measuring system measures UV radiation from the spectrum of solar radiation at the projected distance. After each measurement, the system wirelessly sends measured data which are then displayed on a PC. PC application conducts data processing and memorising. Main window of the PC application is shown in Fig. 7.

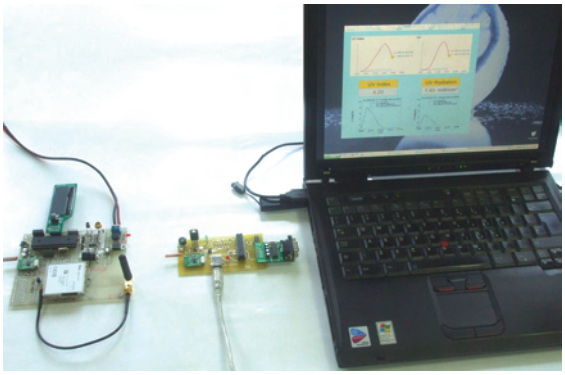


Fig. 6. Developed measuring system

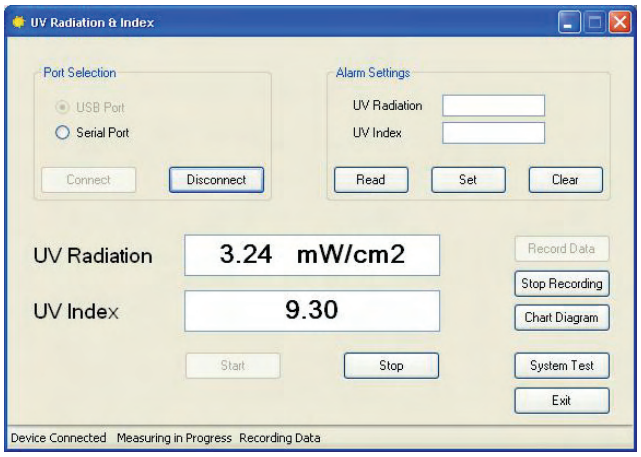


Fig. 7. Main window of the PC application

The PC application supports both USB and serial communication and the selection is made by checking on the appropriate port at the top left corner of application window. After the communication selection it is necessary to connect the PC application with the receiving unit by clicking on the *Connect* button. If the connection was successful a message "Device Connected" will appear in the status bar. Displaying of measured results begins by clicking on the *Start* button, after which a message "Measuring in Progress" is displayed in the status bar. Current measured results are displayed in two text boxes, for the intensity of UV radiation and for the UV index. Displaying of results is stoped by clicking on the *Stop* button.

The application can memorise measuring results into the database by clicking on the *Record Data* button, followed by appearance of "Recording Data" message in the status bar. Data memorising is done in programmable discrete steps. The application can draw a chart diagram of UV radiation and UV index from database by clicking on the *Chart Diagram* button. The application can set the value of alarm levels for intensity of UV radiation and for UV index. Upon the alarm triggering system automatically sends SMS message, while on the PC a new window appears with "The alarm level reached" message. Alarm level settings are stored in the microcontroller's internal EEPROM memory. *Disconnect* button interrupts the communication between the application and the receiving unit followed by appearance of "Device disconnected" message in the status bar, but measuring system continues to operate.

The goal of the development was on the automatic transmision of measured data when the programmed conditions are reached. All conditions can be programmed in the PC application.

IV. Conclusion

This paper presents a developed version of two-channel measuring system for measuring the intensity of UV radiation and UV index. Future research in applying these types of sensor elements will be based on the various modifications of the amplifing stage.

Because of its complexity and mobility selected hardware structure enables development of our own families of various wireless measuring devices. By developing an appropriate PC software these devices would, due to many advantages, quickly became an integral part of most research labs, as well as a standard part of many weather stations.

Acknowledgement

This work has been funded by the Serbian Ministry for Science under the projects TR-33035 and III-43014.

References

[1] *Environmental effects of ozone depletion and its interaction with climate change: 2006. Assessment*, Nairobi, Kenya, UNEP, 2006.

[2] A. F. Mc Kinlay, B. L. Diffey, "A reference action spectrum for ultraviolet induced erythema in human skin", CIE Research Note, CIE-Journal, vol. 6, no.1, pp.17-22, 1987.

[3] J. G. Belmonte, E. Giandoménico, R. M. Martínez, E. D'agostino, R. Corti, R. Piacentini, "Wireless sensor network development for measuring ultraviolet radiation on human health application", U.N.R. Journal**,** vol. 2, pp.613-621, 2010.

[4] *Ultraviolet selective SiC based UV sensor*, Sglux GmbH, Germany, 2011.

Session TST:

TELECOMMUNICATION SYSTEMS AND TECHNOLOGIES

Performance of Quasioptimal Algorithm for Multiuser Detection and M-QAM Modulations

Ilia Georgiev Iliev[1] and Boyan Ruzhkov Kehayov[2]

***Abstract* – The paper presents a research on previously developed quasioptimal algorithm for MUD [1]. There have been many experiments and analysis (with different modulation rates, number of surrounding vectors and number of iterations) to compute the final decision on our MUD algorithm. Program models are realized in MATLAB, in order to research the algorithm performance. Graphic presentations from the measurements are given to prove the exactness of the analytical formulas.**

***Keywords – MUD, CDMA, M-QAM, BER performance*.**

I. INTRODUCTION

CDMA is an efficient method for sharing a mobile radio channel. The noise resistance of the system decreases due to internal system interference – MUI (multi user interference). The correlation receiver is optimal when no MUI exists. The multi user detection (MUD) is used to minimize the influence of the MUI [2].

The MUD receiver, based on the maximum likelihood criteria (ML) [2], gives an optimal solution and checks all possible combinations of transmitted symbols. The number of the calculations grows exponentially with the number of the active users, which is a disadvantage of ML MUD. There are many suggested methods and algorithms for suboptimal receiving, decreasing the needed number of detection calculations. They are compromise between calculation complexity and quality parameters of the receiver. [6,7].

The research is focused on the parameters and performance of the quasioptimal algorithm of MUD, developed in [1]. The algorithm uses discrete consecutive search. The fast-response of the algorithm is due to strong criteria of search discontinuation and selection of start point of the optimization after the single correlation receiving.

A block diagram of Fig. 2 is used for programming of the algorithm. The result of the analysis is the error probability of the algorithm, the dependency of its accuracy from the number of iterations, users and modulation type.

II. BLOCK DIAGRAM OF QUASIOPTIMAL ALGORITHM FOR MUD

The block schematic of the studied system is shown on Fig.1. The processing is in baseband and consists of *K* number of receivers. It is supposed that the signals from different transmitters are statistically independent. *K* users transmit synchronously signals with direct spread spectrum (DSS) and M-QAM modulation.

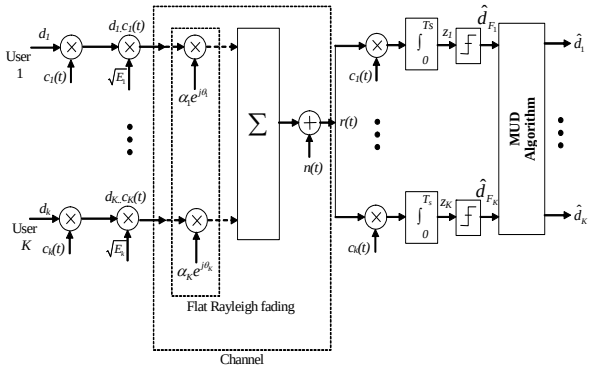


Fig.1

Every branch is a correlation receiver with output the received vector $\boldsymbol{z}$, $\boldsymbol{d} = [d_1, d_2, ..., d_K]^T$ is a vector column that consists the value of the modulated symbol with duration T_s coming from the *k* user. *n(t)* is the realization of complex additive white Gaussian noise (AWGN) in the input of the receiver with independent real and imaginary component and each of them has $\sigma^2 = No/2$ [W/Hz]. $\boldsymbol{H} = diag[\alpha_1 e^{j\theta_1}, \alpha_2 e^{j\theta_2}, ..., \alpha_K e^{j\theta_K}]$ is a diagonal matrix and the elements in the main diagonal are the complex transmission coefficients of the channel for each user in *k*-th channel. The channel is supposed to be with slow flat 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* - $N = T_S / T_c$, T_c is the element duration.

A block diagram of the algorithm is shown on figure 2. The two base parameters are number of iterations and number of vectors that are independent points on the diagram of states.

The objective function of likelihood for the algorithm is expressed in matrix form as:

$$\Omega(\boldsymbol{d}) = \left|(r - H^*.d)\right|^2. \quad (1)$$

The decision for the transmitted symbols is made by:

[1] Ilia Iliev is with the Faculty of Telecommunications, TU-Sofia, Kliment Ohridski 8, Sofia 1000, Bulgaria, E-mail: igiliev@tu-sofia.bg

[2] Boyan Kehayov is with the Faculty of Telecommunications, TU-Sofia, Kliment Ohridski 8, Sofia 1000, Bulgaria, E-mail: bkehayov@tu-sofia.bg

$$\hat{\boldsymbol{d}} = \arg\left\{\min_{\boldsymbol{d}}\left[\Omega(\boldsymbol{d})\right]\right\}. \quad (2)$$

The work of algorithm is described with following five steps:

1. d_{MF} is hard decision vector from demodulator.
2. Generation of K+1 vectors d. They have vector space with Hamming distance H_d=1 with consecutive changes of each element of the vector. d_{K+1} vector is d_{MF} vector. If modulation rate is higher, Hamming distance can be set to 2 or 3. Then the number of vector K will be higher, as is shown on figure 3.
3. Calculate Ω function for every K+1 vectors and find minimum value of Ω and number of corresponding vector in d.

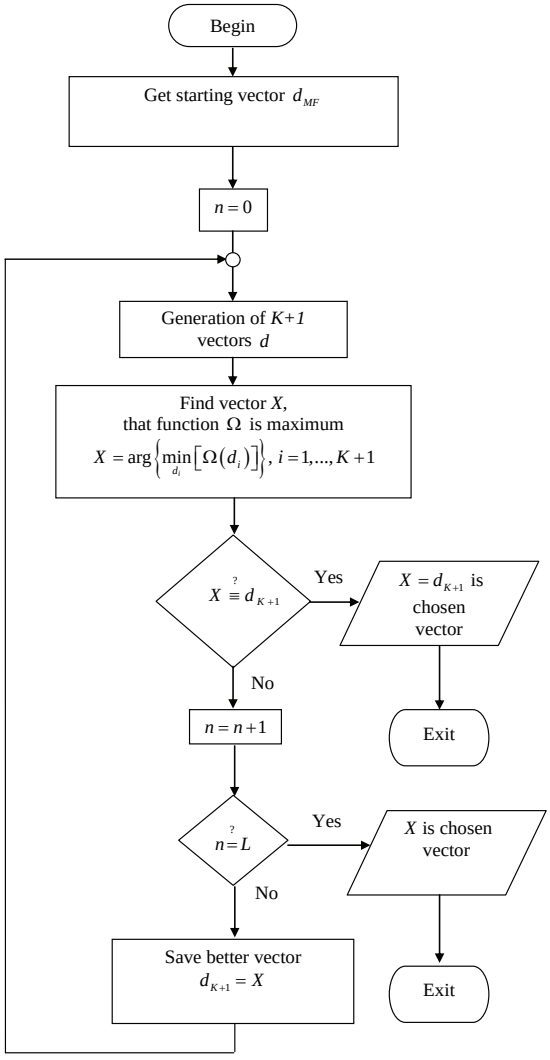


Fig. 2. Block diagram

4. If X is equal to d_{K+1} vector, the initial state d_{MF} is optimum and exit is reached.
5. If $\boldsymbol{X}$ is not equal with d_{K+1} vector, new best vector is estimated and d_{K+1} vector became $\boldsymbol{X}$ vector. This process stop after L iterations.

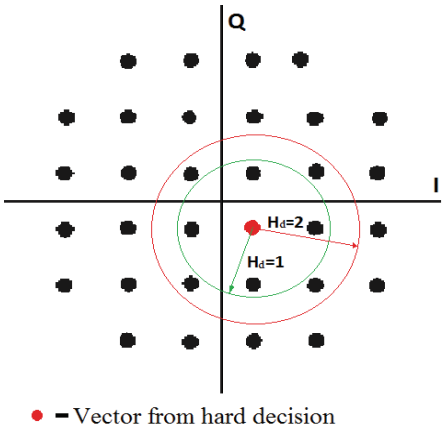


Fig. 3. 32 QAM vector diagram and number of vectors when H_d = 1, 2

With this block diagram different values of Hamming distance can be evaluated. Also different values of number of cycles can be investigated.

III. Results of Simulations

For the purposes of the research a model program in MATLAB, which finds bit error probability (BER) in different cases, is created.

Studies have been made in base band under the following conditions. The CDMA processing gain of the spreading sequences is N = 31, and the sequences are randomly generated. Perfect power control and synchronisation is assumed. The channel is modelled as AWGN with slow Raleigh fading, and channels transmission coefficients are known. Hamming distance for all simulations is H_d=1.

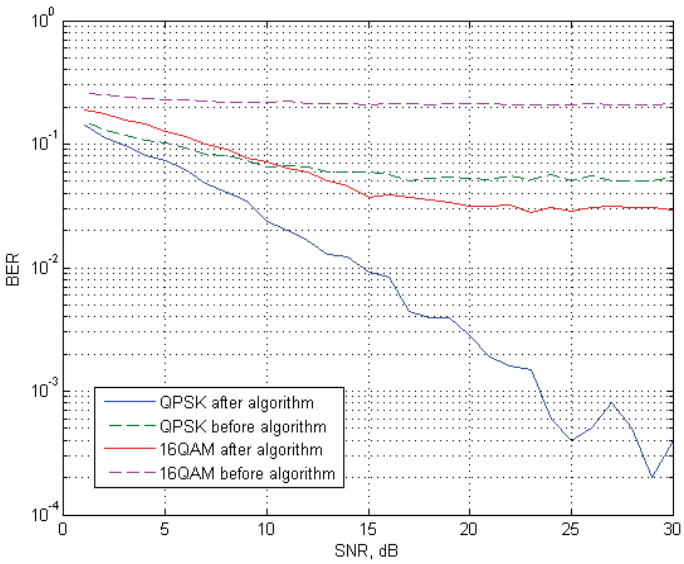


Fig. 4. BER depending on *SNR* and K=5

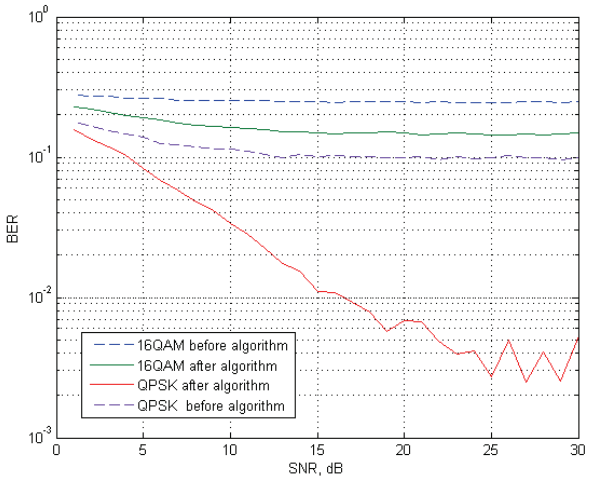


Fig. 5. BER depending on *SNR* and K=10

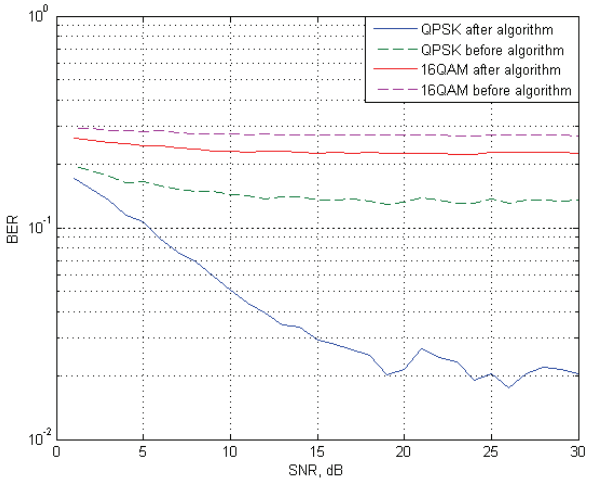


Fig. 6. BER depending on *SNR* and K=15

The results of figures were obtained in five iterations of the algorithm, *L*=5. As described above, experiments were done for 5, 10 and 15 users of the system in two types of modulation QPSK and 16QAM.

On figure 4, 5 and 6 is shown BER depending on SNR and modulation type for five, ten and fifteen active users respectively. Curves from hard decision ML receiver are marked as "before algorithm" and curves after MUD algorithm are marked as "after algorithm". It is seen from figure 4,5 and 6 that the algorithm for MUD reduces error probability in contrast of hard decision ML receiver.

When increasing the modulation rate *M*, values of BER increases and the algorithm needs more iterations to reach the same levels of bit error probability as at lower modulation rate.

Changing the number of iterations of the algorithm has great influence on BER. The results are shown in the next figures – fig.7, fig.8.

Experiments were made with 1, 3 and 5 iterations of the algorithm, two types of modulation QPSK and 16QAM and ten users of the system.

It is important to note that each next algorithm's iteration reduces the BER. This continues until a certain optimum value for the number of cycles *L*. After reaching this value BER decrease stops and increase of *L* becomes meaningless.

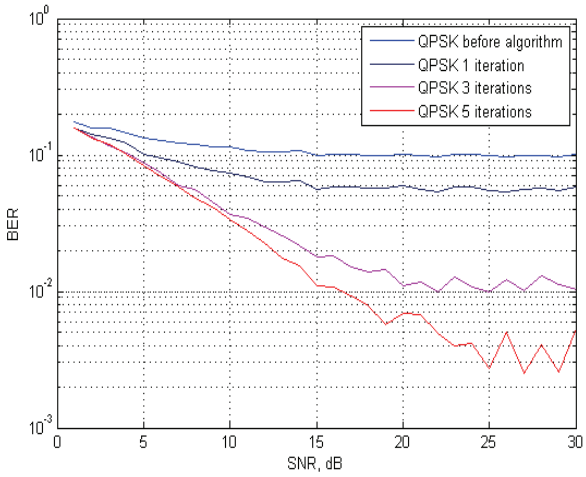


Fig. 7. BER depending on *SNR* and number of iterations *L* for QPSK modulation K=10

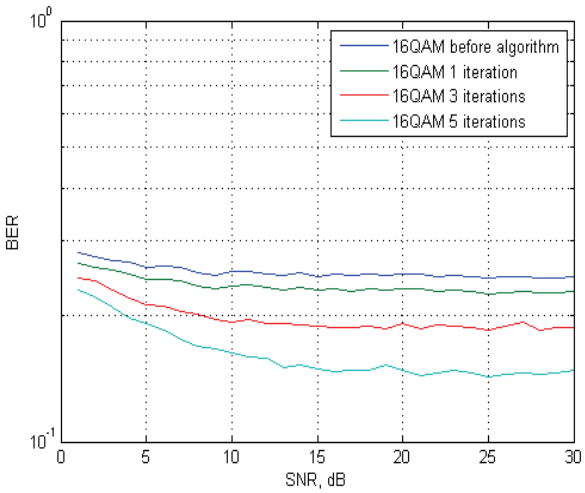


Fig. 8. BER depending on *SNR* and number of iterations *L* for 16QAM modulation K=10

Figure 9 presents BER depending of number of users *K* for hard decision (before algorithm) and after proposed quasioptimal algorithm for MUD. Curves present values for QPSK and 16QAM modulations by SNR=20dB.

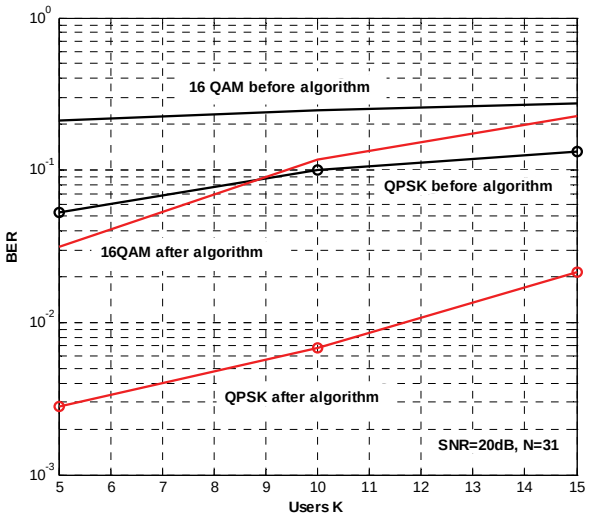


Fig. 9. BER depending on *K* for QQPSK and 16QAM

IV. Conclusion

The paper presents a development of quasioptimal algorithm for multiuser detection. Through it identifies: performance of the algorithm - respectively the lower limit of the bit error probability in a channel with Rayleigh flat fading and dependency on modulation rate and number of iterations. Program model of system is realized in MATLAB, in order to research the algorithm performance and their parameters. Graphic presentations from the measurements are given to prove the exactness of the analytical formulas.

The results show that the algorithm for MUD reduces error probability in contrast of hard decision ML receiver. By increasing the modulation rate *M*, values of BER increases and the algorithm needs more iterations to reach the same levels of bit error probability as at lower modulation rate.

Smaller values of BER after algorithm are expected outcome of the experiments and can be recorded from charts. In the model have not been used any time and phase distortion or synchronization methods. This would lead to lower values of BER under the same parameters. These are plans for future research.

References

[1] Iliev I. G., Nedelchev M. V., Kehayov B. K., Theoretical model of the MUD algorithm based on the diagram of states and Markov's chains, ICEST 2010, Ohrih, Proceedings of Papers, pp.169,172.

[2] S. Verdú, Multiuser Detection. New York: Cambridge Univ. Press,1998.

[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] 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

[6] Iliev I. G., Nedelchev M. V., Performance Analysis of a suboptimal multiuser detection algorithm, ICEST 2007, Macedonia, pp.531 - 534

[7] Iliev I. G., Nedelchev M. V., Antenna Diversity Multhi User Detection Algotithm for Synchronous CDMA System, ICEST 2009, V. Tarnovo, Proceedings of Papers, pp.67-71

Design of Novel Two-Level Quantizer with Extended Huffman Coding for Laplacian Source

Zoran Peric[1], Jelena Nikolic[2] and Lazar Velimirovic[3]

***Abstract* – This paper proposes the novel model of two-level scalar quantizer with extended Huffman coding that is designed such that to achieve as close as possible approaching of the bit rate to the source entropy under the given constrain that the SQNR value does not deviate more than 1 dB from the optimal SQNR Lloyd-Max's quantizer value. Unlike to the Lloyd-Max's quantizer, for the proposed quantizer, the asymmetry of representation levels is assumed to provide an unequal probability of representation levels for the symmetric Laplacian probability density function, that in turn provides the proper basis for the further implementation of a lossless compression techniques. The convergence of the proposed quantizer's bit rate to the source entropy is examined in the case of two and three symbol blocks.**

***Keywords* – Scalar quantizer, Extended Huffman coding.**

I. INTRODUCTION

One of the most important steps in the process of converting analog to digital signal is quantization. Discretization of the signal amplitude is done by a quantizer. When developing new models of quantizers it is usually of great importance to find the manner to increase the quality of the quantized signal for a given bit rate or to provide minimization of the bit rate for the desired level of the quantized signal's quality [1] - [3]. The design of a specific quantizer model is simpler and its realization less complex when the required signal quality measured by the Signal to Quantization Noise Ratio (SQNR), is achieved by quantizing signal samples with fewer bits [1] - [3]. In many modern applications, combination of quantizer and lossless coder is used. Most often, quantizer and lossless coder are designed separately, due to simplicity, but obtained performances are not optimal. However, desired performances can be obtained only with joined design of quantizer and lossless coder, which is done in this paper.

In this paper we propose the novel model of scalar quantizer that has a goal to achieve as close as possible approaching of the bit rate to the source entropy under the given constrain that the SQNR value does not deviate more than 1dB from the optimal SQNR Lloyd-Max's quantizer value. In fact, the Lloyd-Max's quantizer [2], [3] presents a special case of the novel quantizer that is proposed in this paper. Unlike the two-level Lloyd-Max's quantizer having the decision threshold settled in zero, the novel quantizer with the same number of quantization levels proposes that the determination of the variable decision threshold is performed in a way that it has a non-negative value, which is designed depending on which SQNR has to be achieved. The basic idea described in this paper is that, unlike to the Lloyd-Max's quantizer, the asymmetry of representation levels is assumed such that to provide an unequal probability of representation levels for the symmetric Laplacian probability density function (PDF). This in turn provides the proper basis for the further implementation of a lossless compression techniques. Among many lossless compression techniques the most suitable for utilization is the extended Huffman coding technique that achieves the lowest average length of code words [3] - [5]. The performance of four types of quantizers with Huffman coding for small and moderate bit rate are analyzed in [6], where it is shown that the best performance is achieved by the hybrid quantizer, composed of the uniform quantizer and the Lloyd-Max's quantizer when Huffman coding technique is applied. Due to the efficient initialization problem of the Lloyd-Max's quantizer's algorithm and the high design complexity of the Lloyd-Max's quantizer with a large number of quantization levels [7], as well as due to the lack of an effective implementation of the Huffman coding technique on the quantizers with large number of quantization levels [1], [4] - [6], we propose the quantizer having only two representation levels. As with the Lloyd-Max's quantizer these representation levels are determined from the centroid condition. The design procedure of the scalar quantizer having the representation levels also determined in accordance with the centroid condition for the Laplacian and Gaussian source is given in [8] along with the analysis of the absolute and the mean-square error distortion for the low bit rate.

This paper is organized as follows. After brief introduction, the novel quantizer with variable decision threshold is described in Section 2. In Sections 3 and 4 a detailed description of the proposed quantizer's code book determination along with the formulation of the extended Huffman coding is provided. The achieved numerical results for the proposed quantizer with extended Huffman coding and the Laplacian source are the topics addressed in Section 5. Additionally, Section 5 is devoted to the conclusions which summarize the contribution achieved in the paper.

II. DESIGN OF TWO-LEVEL QUANTIZER WITH VARIABLE DECISION THRESHOLD

This section contains a detailed description of the novel quantizer model having variable decision threshold. Quantization is the process of replacing analog samples with

[1]Zoran Peric is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: zoran.peric@elfak.ni.ac.rs .

[2]Jelena Nikolic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: jelena.nikolic@elfak.ni.ac.rs .

[3]Lazar Velimirovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: velimirovic.lazar@gmail.com .

the nearest allowed value from a discrete set of N amplitude values [2], [3]. An N - level scalar quantizer Q is defined by mapping Q: $R \to Y$, where R is the set of real numbers, and

$$Y \equiv \left(y_1, y_2, y_3, \ldots, y_N\right) \subset R \tag{1}$$

is a set of representation levels that makes the code book of size $|Y| = N$ [2], [3]. Associated with every N - level scalar quantizer is partition of the set of real numbers into N cells $R_i = (t_{i-1}, t_i]$, $i = 1, \ldots, N$, where t_i, $i = 0, 1, \ldots, N$ are decision thresholds and it holds $Q(x) = y_i$, $x \in R_i$. The quantizer designed iteratively in accordance with the centroid condition and the condition of the nearest neighbor is the optimal Lloyd-Max's quantizer [2], [3]. In other words, for a given PDF $p(x)$ of the input signal of variance σ^2 and for the considered number of quantization levels N, the minimum value of distortion, i.e. the maximum value of SQNR is achieved by using the Lloyd-Max's quantizer. The quantizer we propose in this paper is defined by the variable decision threshold along with the two representation levels (see Fig. 1) that are, as for the Lloyd-Max's quantizer, determined from the centroid condition. Particularly, this variable decision threshold we determine depending on the quality, measured by SQNR that has to be achieved. Note that in the special case, when the mentioned variable decision threshold has zero value, the proposed quantizer becomes optimal. For the assumed Laplacian PDF of the unit variance [2], [3]

$$p(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right), \tag{2}$$

the representation levels of the proposed quantizer are determined as follows

$$y_1 = \frac{\int_{-\infty}^{t_1} x p(x) dx}{\int_{-\infty}^{t_1} p(x) dx} = \frac{\sqrt{2} + 2t_1}{2 - 4\exp\left(\sqrt{2}t_1\right)}, \tag{3}$$

$$y_2 = \frac{\int_{t_1}^{\infty} x p(x) dx}{\int_{t_1}^{\infty} p(x) dx} = t_1 + \frac{1}{\sqrt{2}}, \tag{4}$$

where the variable decision threshold is denoted by t_1. From the last two equations it is obvious that the representation levels of the proposed quantizer are not symmetrical.

The performances of the quantizer are often determined by SQNR which is defined as follows [2], [3]

$$\mathrm{SQNR} = 10 \log\left(\frac{\sigma^2}{D}\right) \tag{5}$$

and expressed in dB where σ^2 is the variance of the input signal x, while D is the distortion added with quatization. Assuming the unit variance for the given range of SQNR values one can firstly determine the appropriate D values

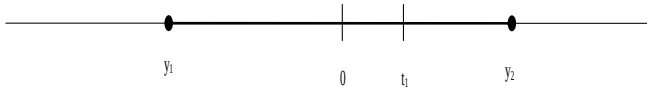


Fig. 1. Model of a two-level quantizer with asymmetric representation levels

$$D = \frac{\sigma^2}{10^{\frac{\mathrm{SQNR}}{10}}} = \frac{1}{10^{\frac{\mathrm{SQNR}}{10}}}. \tag{6}$$

Further defining the distortion of the proposed quantizer

$$D = \int_{-\infty}^{t_1} (x - y_1)^2 p(x) dx + \int_{t_1}^{\infty} (x - y_2)^2 p(x) dx \tag{7}$$

and by combining with Eqs. 3 and 4, a closed form expression for the distortion of the proposed quantizer is derived as a function of the variable decision threshold t_1

$$D = \frac{3 - 4\exp\left(\sqrt{2}t_1\right) + 2\sqrt{2}t_1 + 2t_1^2}{2 - 4\exp\left(\sqrt{2}t_1\right)}. \tag{8}$$

Using the last expression and the distortion values obtained from the Eq. 6 for the considered range of SQNR values, the appropriate value of the threshold t_1 can be determined, and hence, the design of the proposed quantizer is enabled.

III. Extended Huffman Coding for Two Symbol Sources

In this section a basic concept of the very popular lossless compression technique, called extended Huffman coding is presented. The procedure of Huffman coding includes determining the optimal length of code words for a given probability of symbols [1], [3], [5], [9]. Note that it is sometimes beneficial to additionally reduce the bit rate by blocking more than one symbol together. In the mentioned cases, the extended Huffman coding technique is used. Particularly, the extended Huffman coding is the procedure of determining the optimal length of code words for blocks of two or more symbols [3], [5], [9].

Let us denote by p_1 the probability that sample of the input signal has a lower value than the value of decision threshold t_1

$$p_1 = \int_{-\infty}^{t_1} p(x) dx = 1 - \frac{1}{2}\exp(-\sqrt{2}t_1), \tag{9}$$

and by p_2, the probability that sample of the iput signal has a greater value than the value of decision threshold t_1

$$p_2 = \int_{t_1}^{\infty} p(x) dx = \frac{1}{2}\exp(-\sqrt{2}t_1). \tag{10}$$

These probabilities actually refer to the symbol probabilities, i.e. to the probabilities of occurence of representation levels y_1 and y_2. Since we consider two-level quantizer we in fact observe two symbol source. As the extended Huffman coding procedure blocks more than one symbol together, we can now define probabilities of two and three symbol blocks as

$$P_{i,j} = p_i p_j, \quad i = 1,2,\ j = 1,2, \tag{11}$$

$$P_{i,j,k} = p_i p_j p_k, \quad i = 1,2,\ j = 1,2,\ k = 1,2. \tag{12}$$

Note that blocking two symbols together means the symbol alphabet size goes from m to m^2, where m is the size of the initial symbol alphabet. In this paper, we consider two cases, of two and three symbol blocks, such that the size of the extended alphabet is 4 and 8, respectively. For the proposed quantizer with extended Huffman coding we examine the

convergence of the bit rate to the source entropy. Source entropy for two and three symbol blocks are given by the following expressions, respectively [9]

$$H = \sum_{i=1}^{2}\sum_{j=1}^{2} P_{i,j} ld \frac{1}{P_{i,j}}, \quad (13)$$

$$H = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} P_{i,j,k} ld \frac{1}{P_{i,j,k}}. \quad (14)$$

The average bit rate of the observed quantizer in the case of two symbol blocks can be determined as

$$\overline{R} = \sum_{i=1}^{2}\sum_{j=1}^{2} P_{i,j} l_{i,j}, \quad (15)$$

where $l_{i,j}$, $i = 1, 2$, $j = 1, 2$ stands for the length of the code words. Similarly, the average bit rate in the case of three symbol blocks is determined by

$$\overline{R} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} P_{i,j,k} l_{i,j,k}. \quad (16)$$

The procedure of determining the length of the code words using the extended Huffman coding and the code book construction is consisted of the following steps:

Step 1. Determining the symbol block probabilities, further sorting in descending order and finally assigning appropriate probabilities to the initial nodes of the graph.

Step 2. Application of an iterative process, where in each iteration the connection of the two nodes with the smallest probabilities is done and the sum of their probabilities is assigned to a new node. Processing further until the nodes' sum of the probabilities joining in the last step becomes equal to one (see Fig. 2).

Step 3. The construction of code words. Code word for each symbol is determined by beginning from the tree root (node with probability 1) and branches, to which the allocation of zero value is acquired (upper branch) and 1 (lower branch). Assignment process continues to the left until all possible branches are covered. Code word is formed from zeros and ones that are on the path from the root to the node that corresponds to that symbol (see Fig. 2)

An example of extended Huffman coding is illustrated in Fig. 2 for the values of symbol probabilities $p_1 = 0.8958$ and $p_2 = 0.1041$, and for the size of the symbol blocks of $M = 3$. In this case, there are 8 probability symbol blocks $P_{i,j,k}$, $i = 1, 2$, $j = 1, 2$, $k = 1, 2$, having the values given in Table I. The result of the extended Huffman coding procedure is an extended Huffman codebook which for the illustrated example is also given in Table I. From the Table I one can notice that the extended Huffman coding procedure creates variable length codes, where higher probability symbol blocks are coded by shorter codes.

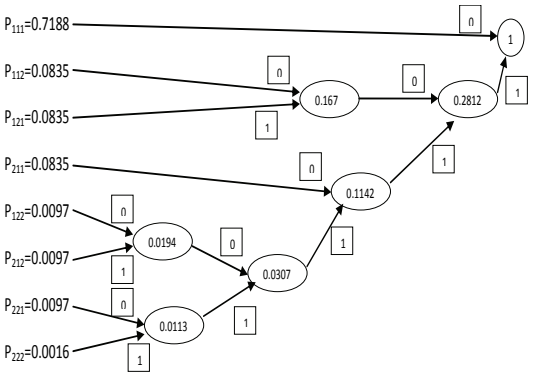


Fig. 2. Example of extended Huffman code construction: forming the tree and assigning the code words

TABLE I
EXTENDED HUFFMAN CODEBOOK EXAMPLE

i,j,k	$P_{i,j,k}$	Extended Huffman code	$l_{i,j,k}$
1,1,1	0.7188	0	1
1,1,2	0.0835	100	3
1,2,1	0.0835	101	3
1,2,2	0.0097	11100	5
2,1,1	0.0835	110	3
2,1,2	0.0097	11101	5
2,2,1	0.0097	11110	5
2,2,2	0.0016	11111	5

IV. ALGORITHM OF DESIGNING NOVEL QUANTIZER WITH VARIABLE DECISION THRESHOLD AND EXTENDED HUFFMAN CODING

The determination of the proposed quantizer performance, for a given signal quality, i.e. the desired level of SQNR, consists of the following steps:

Step 1. The determination of the distortion D by using Eq. 6 for the desired quality of the quantized signal, i.e. for an assumed value of SQNR.

Step 2. The design of the decision threshold t_1 according to Eq. 8 for D calculated in step 1.

Step 3. Determining the representation levels y_1 and y_2 (Eqs. 3 and 4), the probabilities p_1 and p_2 (Eqs. 9 and 10), for the value of the decision threshold t_1 calculated in the previous step.

Step 4. Determining the symbol block probabilities $P_{i,j}$, $i = 1, 2$, $j = 1, 2$ (Eq. 11) and $P_{i,j,k}$, $i = 1, 2$, $j = 1, 2$, $k = 1, 2$ (Eq. 12) using the probabilities p_1 and p_2.

Step 5. The determination of the code word lengths by using the extended Huffman coding, $l_{i,j}$, $i = 1, 2$, $j = 1, 2$ (in the case of two symbol blocks), $l_{i,j,k}$, $i = 1, 2$, $j = 1, 2$, $k = 1, 2$ (in the case of three symbol blocks).

Step 6. Calculating the source entropy H (Eqs. 13 and 14) and the average bit rate $\overline{R}$ (Eqs. 15 and 16).

V. Numerical Results and Conclusion

Numerical results presented in this section for the proposed two-level quantizer with extended Huffman coding are obtained for the cases where the SQNR value does not deviate more than 1 dB from the optimal quantizer SQNR value with the same number of quantization levels. The optimal SQNR value of the Lloyd-Max's quantizer having two quantization levels is 3 dB, which means that the SQNR range in which we consider the performance of the proposed quantizer is [2 dB , 3 dB]. The calculated performance of the proposed quantizer in the case of two and three symbol blocks are shown in Fig. 3. Particularly, Fig. 3 shows the dependence of the average bit rate $\overline{R}$ and the source entropy H on the distortion for the proposed quantizer in the case of the extended Huffman coding applied on two and three symbol blocks. One can notice that the $\overline{R}$ of the proposed quantizer approaches the source entropy H where this convergence is greater in the case of three symbol blocks than in the case of two symbol blocks. By blocking more and more symbols together, the size of the alphabet exponentialy grows, and the extended Huffman coding technique becomes impractical [9]. Accordingly, in this paper our analysis is constrained to the

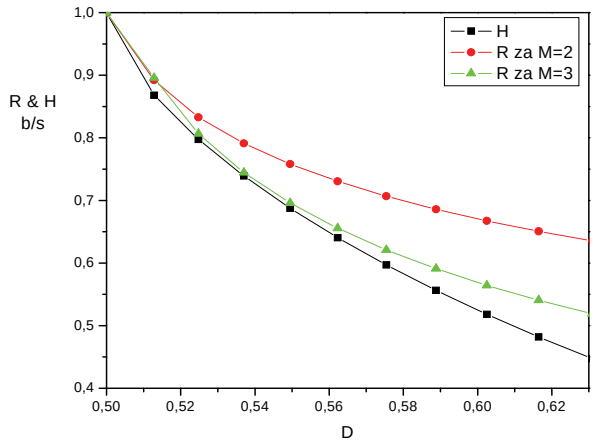


Fig. 3. The dependency of the bit rate and the entropy on the distortion for the proposed quantizer

Table II
Performance of the Proposed Quantizer in the Case of Two and Three Symbol Blocks

SQNR	*D*	*H*	t_1	p_1	p_2	$\overline{R}$ $M = 2$	$\overline{R}$ $M = 3$
2	0.631	0.447	1.188	0.906	0.093	0.635	0.519
2.1	0.617	0.482	1.109	0.895	0.104	0.650	0.540
2.2	0.603	0.518	1.032	0.883	0.116	0.667	0.564
2.3	0.589	0.556	0.955	0.870	0.129	0.685	0.590
2.4	0.575	0.597	0.876	0.855	0.144	0.706	0.620
2.5	0.562	0.641	0.794	0.837	0.162	0.730	0.655
2.6	0.549	0.688	0.709	0.816	0.183	0.758	0.695
2.7	0.537	0.739	0.618	0.791	0.208	0.791	0.744
2.8	0.525	0.797	0.515	0.758	0.241	0.832	0.806
2.9	0.513	0.868	0.387	0.710	0.289	0.892	0.895
3	0.5	1	0	0.5	0.5	1	1

case of three symbol blocks. Table II contains the values of $\overline{R}$ for two and three symbol blocks, distortion, decision thresholds, as well as the values of probabilities p_1 and p_2 that for the considered range of SQNR are achieved by the proposed quantizer. From the results given in Table II and Fig. 3 one can observe that when the SQNR value deviats up to 0.5 dB from the optimal SQNR value, there is a little deviation of $\overline{R}$ from H in the case of three symbol blocks. However, when the mentioned deviation of SQNR is in the range of 0.5 dB to 1 dB, a slightly larger deviation of $\overline{R}$ from H can be perceived. Observe that in both ranges the increasing convergence of $\overline{R}$ to H is achieved in the case of three symbol blocks. It is important to notice that for the proposed quantizer in the case of three symbol blocks with an average bit rate reduction of 0.35 bits, the reduction in SQNR of 0.5 dB is achieved. This is about 0.9 dB smaller SQNR reduction for the same amount of the compression than the one ascertained in the considered range of $\overline{R}$ [9]. Hence, it is obvious that the proposed quantizer represents a very efficient coding solution. Finally, from the last row in Table II one can notice that the optimal Lloyd-Max's quantizer is actually the special case of the proposed quantizer. Particularly, when the decision threshold t_1 of the proposed quantizers is settled to zero, the proposed quantizer is Lloyd-Max's quantizer that has the symmetrical representation levels, i.e. equal probabilities p_1 and p_2. For such values of probabilities, the values of the H and $\overline{R}$ of the proposed quantizer are equal and amount to one.

In this paper a novel class of quantizers having variable decision thresholds with extended Huffman coding is presented. Based on the proposed quantizer analysis, it is shown that by using the extended Huffman coding technique and the set of quantizers with variable decision thresholds, approaching of the average bit rate to the source entropy can be achieved.

References

[1] D. Salomon, *A Concise Introduction to Data Compression*, Springer, 2008.

[2] N. S. Jayant, P. Noll, *Digital Coding of Waveforms*, Prentice Hall, 1984.

[3] L. Hanzo, C. Somerville, J. Woodard, *Voice and Audio Compression for Wireless Communications*, John Wiley & Sons, IEEE Press, October, 2007.

[4] ITU-T, Recommendation T.81, *Information Technology Digital Compression and Coding of Continuous-tone Still Image Requirements and Guidelines*, 1992.

[5] R. L. Rabiner, W. R.Schafer, *Introduction to Digital Speech Processing*, Foundations and Trends in Signal Processing, 2007.

[6] M. Dincic, Z. Peric, "Design of Quantizers with Huffman Coding for Laplacian Source", Electronics and Electrical Engineering, No. 10(106), pp. 129 - 132, 2010.

[7] Z. Peric, J. Nikolic, "An Effective Method for Initialization of Lloyd-Max's Algorithm of Optimal Scalar Quantization for Laplacian Source", Informatica, No. 18(2), pp. 279 - 288, 2007.

[8] D. Marco, D. L. Neuhoff, "Low-Resolution Scalar Quantization for Gaussian and Laplacian Sources with Absolute and Squared Error Distortion Measures", Tech. report, January7, 2006.

[9] K. Sayood, *Introduction to Data Compression*, San Francisco, 2006.

OP Comparison of Dual SC Systems using Desired and SIR Power Algorithm in Presence of Interference

Aleksandra Panajotović[1], Nikola Sekulović[2], Mihajlo Stefanović[3], Dragan Drača[4] and Dušan Stefanović[5]

***Abstract* – An analytical expression for evaluation outage performance of dual selection combining (SC) diversity system employing desired signal power decision algorithm is presented in the paper. The diversity system operates over correlated Rician fading channels in the presence of Rayleigh cochannel interference (CCI). Numerical results are presented to show influence of fading severity and branch correlation on outage probability (OP). Moreover, they are used to compare performance of dual SC systems applying two different decision power algorithms, i.e. desired and signal-to-interference ratio (SIR) power algorithm.**

***Keywords* – Cochannel interference, Outage probability, Selection combining, Rayleigh fading, Rician fading.**

I. INTRODUCTION

In wireless system, the main causes of the performance degradation are fading due to multipath propagation and cochannel interference (CCI) due to frequency reuse [1]. Space diversity techniques, which combine input signals from multiple receive antennas, are the well known techniques that can be used to alleviate the effects of these degradations [2]. The most popular diversity techniques are maximal-ratio combining (MRC), equal-gain combining (EGC) and selection combining (SC). The last one has the least implementation complexity since it processes only one of the diversity branches. Traditionally, SC receiver chooses the branch with the highest signal-to-noise ratio (SNR), or equivalently, with the strongest signal assuming equal noise power among the branches. In interference limited environment where the level of CCI is sufficiently high as compared with noise, SC receiver can employ one of decision power algorithms: the desired signal power algorithm, the total signal power algorithm and the signal-to-interference (SIR) power algorithm.

Several statistical models are used in communication systems to describe fading in wireless environment. For example, in microcellular environment, an undesired signal from distant cochannel cell may well be modelled by Rayleigh statistics, but Rayleigh fading may not be good assumption for desired signal since line-of-sight (LoS) path may exist within a microcell. Then, the Rician distribution is often used to model a propagation path consisting of one strong direct LoS signal and many randomly reflected and usually weaker signals. Therefore, different fading statistics are needed to characterize the desired signal and CCI in microcell systems [3].

There is set of performance measures that allow the system designer to compute the performance of different digital communication systems characterized by variety of modulation/detection types and fading channel models. Performance measures not only allow easy yet accurate performance evaluation but at the same time provide insight into the manner in which this performance depends on the key system parameter [2]. One of the most important performance measures is outage probability (OP). It is also essential to determine some other performance measures. OP for various diversity techniques with/without CCI and over independent and correlated channels can be found in [4]-[8]. In this paper, analytical expression for OP of dual SC system which applies desired signal power decision algorithm and operates over correlated Rician fading channels in the presence of Rayleigh distributed CCI is presented. Numerical results illustrate the proposed mathematical analysis and show the influence of system and channel parameters on the system performance. Moreover, presented numerical results are compared with results from [8] in order to compare performance of SC systems which apply different decision algorithms.

II. SYSTEM AND CHANNEL MODEL

Instead of simple case of the channels being independent, channels are correlated due to insufficient distance between diversity antennas. In that case, desired signal envelopes, r_1 and r_2, on two diversity branches follow correlated Rician distribution whose probability density function (PDF) is given in [9, Eq. (3.13)]. For identically distributed case, [9, Eq. (3.13)] simplifies to

[1]Aleksandra Panajotovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: aleksandra.panajotovic@elfak.ni.ac.rs.

[2]Nikola Sekulovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nikola.sekulovic@elfak.rs.

[3]Mihajlo Stefanovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: mihajlo.stefanovic@elfak.ni.ac.rs.

[4]Dragan Draca is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dragan.draca@elfak.ni.ac.rs.

[5]Dusan Stefanovic is with the High Technical School, Aleksandra Medvedeva 20, 18000 Nis, Serbia, E-mail: dusan.stefanovic@itcentar.rs.

$$p_{r_1r_2}(r_1,r_2)=\frac{r_1r_2}{\sigma^4(1-\rho^2)}\exp\left(-\frac{r_1^2+r_2^2+2b^2(1-\rho)}{2\sigma^2(1-\rho^2)}\right)$$
$$\times\sum_{k=0}^{\infty}\varepsilon_k I_k\left(\frac{r_1r_2\rho}{\sigma^2(1-\rho^2)}\right)I_k\left(\frac{br_1}{\sigma^2(1+\rho)}\right)I_k\left(\frac{br_2}{\sigma^2(1+\rho)}\right), \quad (1)$$
$$\varepsilon_k=\begin{cases}1, k=0\\ 2, k\neq 0\end{cases}$$

where ρ is branch correlation coefficient and $I_k(\cdot)$ is modified Bessel function of the first kind and *k*-th order. Rice factor, *K*, and average desired signal power, β, are defined as $K=b^2/(2\sigma^2)$, $\beta=\sigma^2(1+K)$.

In cellular mobile radio system with fading, an exact performance analysis is usually quiet complicated, and approximations are sometimes used to simplify the analysis. As in [6], [8] and [10], we consider the effect only of the strongest CCI and its envelope follows Rayleigh PDF expressed by

$$p_a(a)=\frac{a}{\sigma_a^{\,2}}\exp\left(-\frac{a^2}{2\sigma_a^{\,2}}\right), \quad (2)$$

where $\sigma_a^{\,2}$ is average CCI power.

The considered SC receiver uses the desired signal power decision algorithm. Actually, it selects the branch with the largest instantaneous desired signal power, i.e. $r^2=\max\left\{r_1^2,r_2^2\right\}$. The instantaneous SIR at the output of such interference-limited SC system is given by $\eta=\max\left\{r_1^2,r_2^2\right\}/a^2=r^2/a^2=R/A.$

III. Outage Probability

In cellular mobile communications systems subject to CCI, OP is defined as the probability that SIR being less than their respective predetermined threshold values. Mathematically speaking, the OP of dual SC system applying desired signal power algorithm, P_{out}, is given by [11]

$$P_{out}=1-\int_0^{\infty}\int_0^{R/\eta}p_A(A)\,p_R(R)\,dA\,dR, \quad (3)$$

where $p_R(R)$ and $p_A(A)$ are the PDFs of desired and interference signal power, respectively.

The PDF of desired signal envelope at considered dual SC receiver output can be obtained as [2]

$$p_r(r)=\int_0^{r}p_{r_1r_2}(r,r_2)\,dr_2+\int_0^{r}p_{r_1r_2}(r_1,r)\,dr_1. \quad (4)$$

Solving integrals in previous equation after substituting Eq. (1) into Eq. (4), the PDF of desired signal power at the SC receiver output is expressed in the form of following infinite-series [12]

$$p_R(R)=\sum_{k,p,n,l=0}^{\infty}\varepsilon_k\frac{\rho^{2p+k}K^{n+l+k}(1+K)^{p+k+1}}{(1-\rho)^p(1+\rho)^{2k+p+n+l}\beta^{p+k+1}}$$
$$\times\frac{R^{p+k}}{2^{p+k+1}n!p!l!\Gamma(n+k+1)\Gamma(p+k+1)\Gamma(l+k+1)}$$
$$\times\left\{(p+n+k)!(1-\rho)^n\left(\frac{(1+K)R}{2\beta(1+\rho)}\right)^l\right.$$
$$\times\left[\alpha_2-\alpha_1\sum_{i=0}^{p+n+k}\frac{1}{i!}\left(\frac{(1+K)R}{2\beta(1-\rho^2)}\right)^i\right]+(p+l+k)!(1-\rho)^l$$
$$\left.\times\left(\frac{(1+K)R}{2\beta(1+\rho)}\right)^n\left[\alpha_2-\alpha_1\sum_{j=0}^{p+l+k}\frac{1}{j!}\left(\frac{(1+K)R}{2\beta(1-\rho^2)}\right)^j\right]\right\}, \quad (5)$$

where $\Gamma(\cdot)$ is Gamma function and $\alpha_i=\exp\left(-\frac{R(1+K)}{i\beta(1-\rho^2)}-\frac{2K}{1+\rho}\right), i=1,2.$

Substituting PDF of interference signal power at the SC receiver output, derived applying [13, Eq. (4.10)] on Eq. (2), and Eq. (5) into Eq. (3), outage probability of SIR at the output of considered SC receiver is derived in the following form [12]

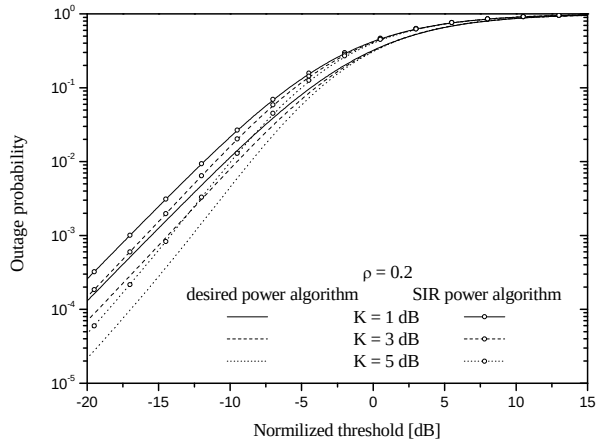
$$P_{out}=1-\exp\left(-\frac{2K}{1+\rho}\right)\sum_{k,p,n,l=0}^{\infty}\varepsilon_k\frac{\rho^{2p+k}(1-\rho)^{n+l+k+1}}{(1+\rho)^{n+l+k-1}}$$
$$\times\frac{K^{n+l+k}}{n!p!l!\Gamma(n+k+1)\Gamma(p+k+1)\Gamma(l+k+1)}$$
$$\times\left\{(p+n+k)!\left[(p+l+k)!\left(1-\frac{1}{\chi^{p+l+k+1}}\right)\right.\right.$$
$$\left.-\sum_{i=0}^{p+n+k}\frac{(p+l+k+i)!}{i!}\left(\frac{1}{2^{p+l+k+i+1}}-\frac{1}{(1+\chi)^{p+l+k+i+1}}\right)\right]$$
$$+(p+l+k)!\left[(p+n+k)!\left(1-\frac{1}{\chi^{p+n+k+1}}\right)\right.$$
$$\left.\left.-\sum_{j=0}^{p+l+k}\frac{(p+n+k+j)!}{j!}\left(\frac{1}{2^{p+n+k+j+1}}-\frac{1}{(1+\chi)^{p+n+k+j+1}}\right)\right]\right\}, \quad (6)$$

where average input SIR is defined as $S=\beta/\sigma_a^{\,2}$ and $\chi=1+\frac{(1-\rho^2)S}{(1+K)\eta}$.

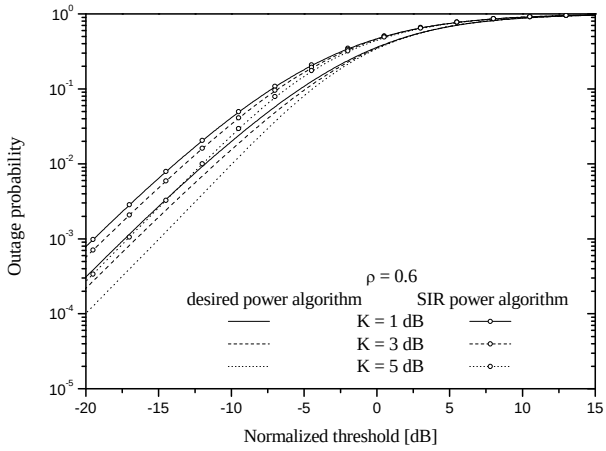
IV. Numerical Results

In this section, we present illustrative examples for the outage probability of spatially correlated dual SC system which applies different decision algorithms and which is exposed to the influence of Rayleigh CCI in Rician fading channels in order to both complement mathematical analysis provided in the previous section and compare outage performance of different SC systems.

All figures show outage probability in function of normalized threshold (η/S). The influence of Rice factor on outage system performance is presented in Fig.1, while



a)



b)

Fig. 1. Outage probability versus normalized threshold for dual SC receiver for several values of Rice factor: *a*) $\rho = 0.2$; *b*) $\rho = 0.6$.

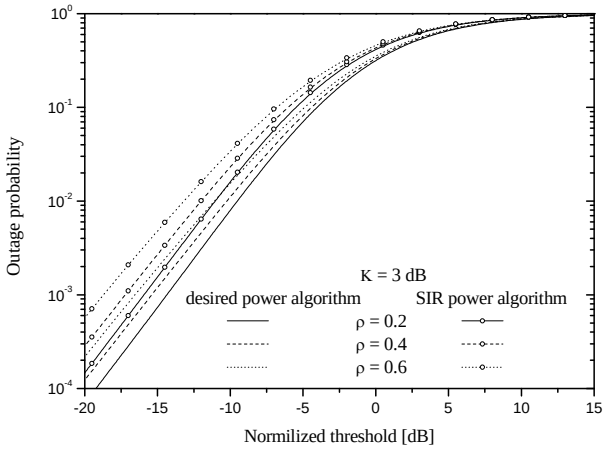


Fig. 2. Outage probability versus normalized threshold for dual SC receiver for several values of branch correlation coefficients.

influence of distance between diversity antennas is proposed in Fig. 2. The signal-to-interference power ratio decision algorithm requires more complex design of SC receiver and shows worst system performance than desired power algorithm. Presented results illustrate that. Regardless of applying decision algorithm system performance improves with increase of Rice factor. Decrease distance between diversity antennas produces increase of outage probability. Advantage of desired power decision algorithm is more evident for small value of Rice factor and great value of branch correlation coefficient.

V. CONCLUSION

In cellular land mobile radio, the received signal suffers from CCI, which also arises in mobile satellite communication channels. With the increasing demand for wireless systems and services, microcell and picocell structures have been proposed to increase system capacity. Propagation measurements in such environments have shown that the received signal envelope has a Rician distribution, but the interference envelope has Rayleigh distribution. In this paper, the performance of a dual SC system, operating over correlated Rician fading channels in the presence of Rayleigh distributed CCI, with desired signal power algorithm, i.e. SIR power algorithm has been studied. Numerical results for the outage probability are presented, describing its dependence on correlation coefficient and fading severity. They show that system's performance improves when the Rice factor increases (fading severity decreases) and/or correlation coefficient decreases. Moreover, presented results point out better outage system performance in case of applying desired power decision algorithm.

ACKNOWLEDGEMENT

This work has been funded by the Serbian Ministry for Education and Science under the projects TR-32052, III-44006 and TR-33035.

REFERENCES

[1] J. D. Parsons, *The Mobile Radio Propagation Channels*, 2nd ed., New York, Wiley, 2000.

[2] M. K. Simon, M. -S. Alouini, *Digital Communications over Fading Channels*, 2nd ed., New York, Wiley, 2005.

[3] Y. -D. Yao, A. U. H. Sheikh, "Investigation into Cochannel Interference in Microcellular Mobile Radio Systems", IEEE Trans. Veh. Technol., vol. 41, no. 2, pp. 114-123, 1992.

[4] L. Yang, M. -S. Alouini, "Average Outage Duration of Wireless Communication Systems", ch. 8, *Wireless Communications Systems and Networks*, Springer, 2004.

[5] A. A. Abu-Dayya, N. C. Beaulieu, "Outage Probabilities of Diversity Cellular Systems with Cochannel Interference in Nakagami Fading", IEEE Trans. Veh. Technol., vol. 41, no. 4, pp. 343-355, 1992.

[6] S. Okui, "Effects of CIR Selection Diversity with Two Correlated Branches in the m-Fading Channels", IEEE Trans. Commun., vol. 48, no. 10, pp. 1631-1633, 2000.

[7] D. A. Zogas, G. K. Karagiannidis, "Infinite-Series Representations Associated with the Bivariate Rician Distribution and Their Applications", IEEE. Trans. Commun., vol. 53, no. 11, pp. 1790-1794, 2005.

[8] A. Panajotović, M. Stefanović and D. Drača, "Performance Analysis of System with Selection Combining over Correlated Rician Fading Channels in the Presence of Cochannel Interference", International Journal of Electronics and Communications-AEÜ, vol. 63, no. 12, pp. 1061-1066, 2009.

[9] M. K. Simon, *Probability Distributions Involving Gaussian random Variables - a Handbook for Engineers, Scientists and Mathematicians*, New York, Springer, 2002.

[10] G. K. Karagiannidis, "Performance Analysis of SIR-based Dual Selection Diversity over Correlated Nakagami-m Fading Channels", IEEE Trans. Veh. Tech., vol. 52, no. 5, pp. 1207-1216, 2003.

[11] H. Yang, M. -S. Alouini, "Outage Probability of Dual-Branch Diversity System in the Presence of Co-channel Interference", IEEE Trans. Wireless Commun., vol. 2, no. 2, pp. 310-319, 2003.

[12] A. Panajotović, M. Stefanović, D. Drača and N. Sekulović, "Average Fade Duration for Dual Selection Combining Diversity in Correlated Rician Fading with Rayleigh Cochannel Interference", under review.

[13] D. Zwillinger, S. Kokoska, *Standard Probability and Statistics Tables and Formulae*, Boca Raton, Chapman & Hall/CRC, 2000.

The Application of OSTBC with Alamouti Scheme in Spectrum-Sharing Cognitive Radio

Vesna Blagojević[1] and Predrag Ivaniš[1]

***Abstract* – In this paper the application of Alamouti scheme on the performance of secondary network in spectrum-sharing cognitive radio with limited average interference power at the primary user is analyzed. We considered the case when fading in all branches follows Nakagami-*m* distribution. Analytical results are confirmed using an independent simulation method.**

***Keywords* – Spectrum sharing, Cognitive radio, Orthogonal Space Time Block Codes, Alamouti scheme, Nakagami fading.**

I. INTRODUCTION

Radio spectrum is one of the most limited resources, so its efficient use is of highest importance for the further development of wireless communications. Regulatory bodies grant license for exclusive access to allocated spectrum bands with the strict regulations concerning interference protection. On the other hand, recent measurements have shown that during significant periods of time many allocated portions of spectrum are not occupied. As the majority of the spectrum has already been allocated, new approaches for its utilization have to be found [1].

The concept of cognitive radio (CR) was proposed by D. J. Mitola [2], as the further extension of software radio. These techniques are attracting more attention recently, as they can improve spectrum utilization through adaptive, dynamic and intelligent processes. One way of improving spectrum utilization is by permitting an unlicensed secondary user (SU) to access a spectrum hole unoccupied by the licensed primary user (PU). This function is often referred to as spectrum sensing. On the other hand, another important function of CR is spectrum sharing, where SUs can share spectrum with the PUs in a controlled fashion. SU is allowed to transmit under the condition that the level of interference that is caused to the primary user is not harmful [1], [3].

Furthermore, it is is well known that multiple input multiple output (MIMO) systems can achieve significant improvement in spectral efficiency and reliability of wireless communications over fading channels [4]. Orthogonal space-time block coding (OSTBC) represents a low-complexity technique, that although suboptimal, results in a maximum diversity gain. Within this class of codes, simple scheme proposed by Alamouti [5] has very important advantage of implementation simplicity.

[1]Vesna Blagojević is with the Faculty of Electrical Engineering, Bulevar kralja Aleksandra 73, 11020 Belgrade, Serbia, E-mail: vesna.golubovic@etf.rs.

[2]Predrag Ivaniš is with the Faculty of Electrical Engineering, Bulevar kralja Aleksandra 73, 11020 Belgrade, Serbia, E-mail: predrag.ivanis@etf.rs.

Channel capacity for spectrum-sharing communications is analyzed in [1], for various fading environments, under both scenarios of average and peak interference power constraints at the PU receiver. In [3] closed form expressions for ergodic, outage and minimum rate capacity are derived for the case of Rayleigh fading, under both average and peak interference power constraints caused at the PU. In [6] the secondary channel capacity is analyzed for the case of Rayleigh fading channel with maximal ratio combining (MRC) diversity on the secondary link and average interference power constraint.

In this paper we extend analysis to the case of both transmit and receive diversity. To the best of author's knowledge, the secondary link ergodic capacity of the spectrum-sharing system that employs OSTBC has not been analyzed.

We derive closed-form secondary link capacity expressions for spectrum-sharing communications, under the primary user interference power constraint. Fading on all links is assumed to be Nakagami distributed. The Alamouti scheme is employed in SU network. The analytical results are confirmed using an independent simulation method.

In Section II, system and channel model are presented. The analysis of secondary link ergodic capacity under the average interference constraint at the primary user side is given in Section III. Numerical results are presented in Section IV. Finally, conclusions are given.

II. CHANNEL AND SYSTEM MODELS

We consider radio network where PUs and SUs coexist in spectrum-sharing system. SUs are equipped with two antennas and PU receiver uses one antenna, as it is shown in Fig. 1. Further, it is assumed that channel fading is slow and perfect channel side information (CSI) of links to the both PU and SU receiver are available to the secondary user transmitter.

At the SUs, Alamouti code scheme is employed, presented in Fig. 2. At the given symbol period, two signals are simultaneously transmitted from the two antennas. The signal transmitted from first antenna (Tx_0) is denoted by x_0 and from the second one (Tx_1) by x_1. During the next symbol period signal $-x_1^*$ is transmitted from first antenna, and x_0^* from the second antenna. Complex conjugate operation is denoted by *. The complex channel fading envelope between transmit and receive antennas are denoted as h_i, i=0,1,2,3.

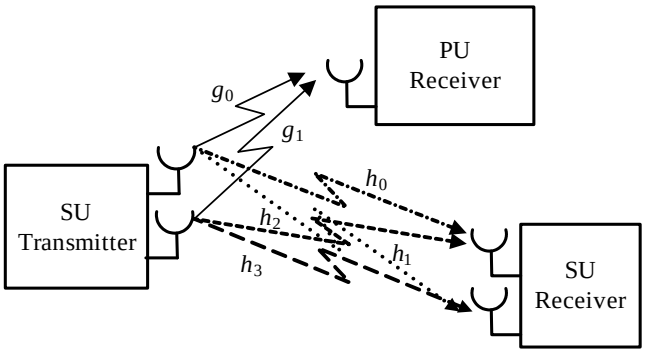


Fig. 1. Secondary user is sharing spectrum with the primary user.

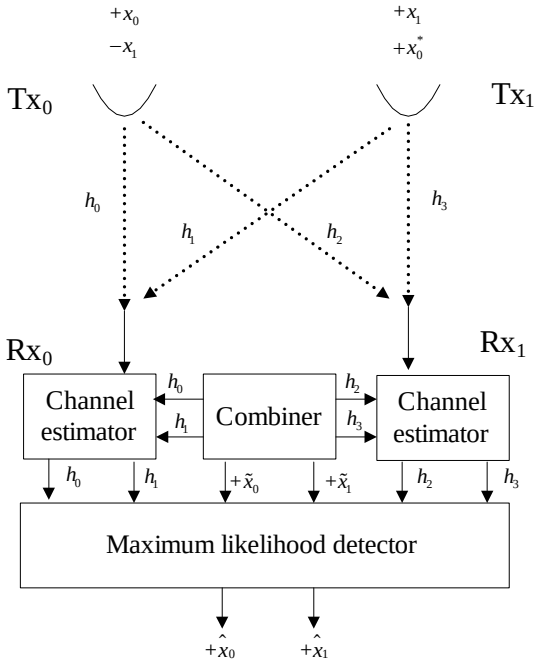


Fig. 2. The application of Alamouti scheme with two transmit and two receive antennas at SUs transmitter and receiver.

Then, the input signals to the maximum likelihood detector (MLLD) are equal [5]

$$\begin{aligned}\widetilde{x_0} &= \sum_{i=0}^{3}|h_i|^2 \cdot x_0 + h_0^* n_0 + h_1 n_1^* + h_2^* n_2 + h_3 n_3^* \\ \widetilde{x_1} &= \sum_{i=0}^{3}|h_i|^2 \cdot x_1 - h_0 n_1^* + h_1^* n_0 - h_2 n_3^* + h_3^* n_2,\end{aligned} \tag{1}$$

where n_i, i=0,1,2,3 are complex random variables (RVs) representing additive white Gaussian noise(AWGN).

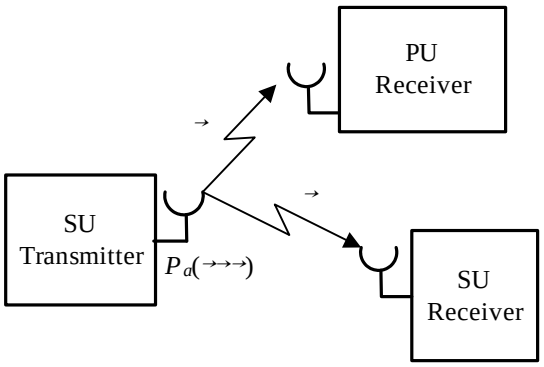


Fig. 3. Equivalent channel power gains for the spectrum-sharing scheme.

It is assumed that fading on links between SU transmitter and reciever antennas is Nakagami distributed with the same fading parameter m_s. The squared fading envelope α_i can be represented as the sum of squares of independent Gaussian zero-mean RVs, according to the expression [7, Eq. 7]

$$\alpha_i = |h_i|^2 = \begin{cases} \sum_{k=1}^{m_s}(c_{i,k}^2 + s_{i,k}^2), & 2m_s \text{ even}, \\ c_{i,0}^2 + \sum_{k=1}^{m_s-0.5}(c_{i,k}^2 + s_{i,k}^2), & 2m_s \text{ odd}. \end{cases} \tag{2}$$

Each pair $c_{i,k}$ and $s_{i,k}$ represents a scattered wave, consisting of an in-phase and quadrature component of a complex RV $c_{i,k} + js_{i,k}$ ($j = \sqrt{-1}$), with a Rayleigh distributed magnitude.

The equivalent channel power gain α between SU transmitter and the receiver is given with [8]

$$\alpha = \sum_{i=0}^{3}\alpha_i, \tag{3}$$

and its probability density function (PDF) is equal

$$f_\alpha(\alpha) = \frac{1}{(4m_s - 1)!}\alpha^{4m_s - 1}e^{-\alpha},\ \alpha > 0. \tag{4}$$

We further assume that fading on the links between SU transmitter antennas and PU receive antenna follows Nakagami distribution with the fading parameter m_p. As the signals coming from SUs antennas are orthogonal, we can consider incoming interference signals mutually independent. Therefore, the channel power gain β between SU transmitter and PU receiver is given with

$$\beta = |g_0|^2 + |g_1|^2, \tag{5}$$

and its PDF is given with

$$f_\beta(\beta) = \frac{1}{(2m_p - 1)!}\beta^{2m_p - 1}e^{-\beta},\ \beta > 0. \tag{6}$$

III. CAPACITY ANALYSIS

We evaluate the secondary link capacity, for a SU operating within licensed band of PU subject to the received interference power constraint. In this paper we suppose that the PU receiver's operation is limited by the average interference level.

Let us denote the transmit power of SUs antennas by P_a. In the fading environment, in order to maximize capacity, transmitted power is changing over time so the received interference constraint is fulfilled and $P_a = P_a(\alpha,\beta)$. Therefore, the following maximization problem should be solved

$$\frac{C}{B} = \max_{P(\alpha,\beta)} \int_{\beta}\int_{\alpha} \log_2\left(1+\frac{\alpha P_a(\alpha,\beta)}{N_0 B}\right) f_\alpha(\alpha)\, f_\beta(\beta)\, d\alpha d\beta, \quad (7)$$

where B is total available bandwidth and N_0 the power spectral density of the AWGN at the SU receiver.

As the links between SU transmit and PU receive antennas are independent, the received power at the PU equals

$$\left(|g_0|^2 + |g_1|^2\right)\cdot P_a = \beta P_a. \quad (8)$$

The average interference power constrained is given by

$$\int_{\beta}\int_{\alpha} \beta P_a(\alpha,\beta)\, f_\alpha(\alpha)\, f_\beta(\beta)\, d\alpha d\beta \le Q, \quad (9)$$

where Q is maximum average power, permitted at the primary user's receiver [1].

The optimal power allocation $P_a(\alpha,\beta)$ is given by [1]

$$P(\alpha,\beta) = \left(\frac{1}{\lambda_0\beta} - \frac{N_0 B}{\alpha}\right)^+, \quad (10)$$

where $(x)^+$ denotes $\max(0,x)$ and λ_0 is determined such that the average interference received power is equal to Q.

By substituting (4), (6) and (10) in (7), we obtain channel capacity

$$\frac{C}{B} = K \int_{\beta=0}^{+\infty} \int_{\alpha=\beta/\gamma_0}^{+\infty} \left[\log_2\left(\frac{\gamma_0\alpha}{\beta}\right)\alpha^{4m_s-1}e^{-\alpha}d\alpha\right]\beta^{2m_p-1}e^{-\beta}d\beta, \quad (11)$$

where $1/K = (4m_s-1)!(2m_p-1)!$ and $\gamma_0 = 1/(\lambda_0 N_0 B)$.

After substituting $t = \alpha\gamma_0/\beta$, we obtain

$$\frac{C}{B} = \frac{K}{\ln 2}\int_{\beta=0}^{+\infty} \beta^{2m_p-1}e^{-\beta}\left(\frac{\beta}{\gamma_0}\right)^{4m_s}\left[\int_{t=1}^{+\infty}\ln(t)\, t^{4m_s-1}e^{-\frac{\beta t}{\gamma_0}}dt\right]d\beta. \quad (12)$$

It is shown in [6] that the integral in square brackets is

$$\int_{t=1}^{+\infty}\ln(t)\, t^{4m_s-1}e^{-\frac{\beta t}{\gamma_0}}dt = \frac{(4m_s-1)!}{(\beta/\gamma_0)^{4m_s}}\sum_{k=0}^{4m_s-1}\frac{\Gamma(k,\beta/\gamma_0)}{k!}, \quad (13)$$

where $\Gamma(k,x)$ denotes complementary incomplete gamma function defined in [9, Eq. (8.350-2)].

$$\frac{C}{B} = \frac{K(4m_s-1)!}{\ln 2}\sum_{k=0}^{4m_s-1}\int_{\beta=0}^{+\infty}\beta^{2m_p-1}e^{-\beta}\frac{\Gamma(k,\beta/\gamma_0)}{k!}d\beta. \quad (14)$$

The integral with respect to β is solved using [9, Eq .(6.455)], and the final expression is given with

$$\frac{C}{B} = \frac{1}{\ln 2\cdot(2m_p)!}\sum_{k=0}^{4m_s-1}\frac{1}{k!}\frac{\Gamma(2m_p+k)\gamma_0^{2m_p}}{(\gamma_0+1)^{2m_p+k}}\times {}_2F_1\left(1;2m_p+k;2m_p+1;\frac{\gamma_0}{\gamma_0+1}\right), \quad (15)$$

where ${}_2F_1(a,b,c,z)$ denotes Gauss's hypergeometric function [10, 15.1].

The average interference power at the PU receiver is obtained by averaging over channel power gain realizations

$$\frac{Q}{N_0 B} = K\int_{\alpha=0}^{+\infty}\int_{\beta=0}^{\alpha\gamma_0}\left[\left(\gamma_0-\frac{\beta}{\alpha}\right)\beta^{2m_p-1}e^{-\beta}d\beta\right]\alpha^{4m_s-1}e^{-\alpha}d\alpha = K\int_{\alpha=0}^{+\infty}\alpha^{4m_s-1}e^{-\alpha}\left[\gamma_0\gamma(2m_p,\gamma_0\alpha)-\frac{\gamma(2m_p+1,\gamma_0\alpha)}{\alpha}\right]d\alpha. \quad (16)$$

and $\gamma(k,x)$ is incomplete gamma function [9, Eq. (8.350-1)]. Then, by using [9, Eq. (6.455)], we obtain the final expression

$$\frac{Q}{N_0 B} = \frac{\Gamma(4m_s+2m_p)}{(4m_s-1)!(2m_p-1)!}\frac{\gamma_0^{2m_p+1}}{(\gamma_0+1)^{4m_s+2m_p}}\times\left[{}_2F_1\left(1;4m_s+2m_p;2m_p+1;\frac{\gamma_0}{\gamma_0+1}\right)\cdot\frac{1}{2m_p} - {}_2F_1\left(1;4m_s+2m_p;2m_p+2;\frac{\gamma_0}{\gamma_0+1}\right)\frac{1}{2m_p+1}\right]. \quad (17)$$

IV. NUMERICAL RESULTS

A waveform sequence with $L=10^7$ samples is generated for every diversity branch by using an improved Jakes fading simulator [11]. For each realizations of the channel gain random variables α and β, the optimal transmit power is calculated according to Eq. (10). Channel capacity and average interference power are evaluated by averaging over random channel realizations. These results are compared with the analytical results obtained using (15) and (17). The obtained results are presented in Figs. 4-5.

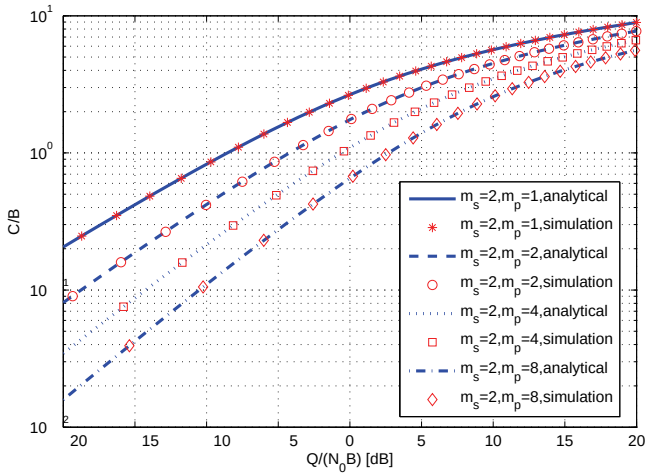


Fig. 4. Channel capacity vs. average interference power for fading parameters m_s=2 and various m_p values.

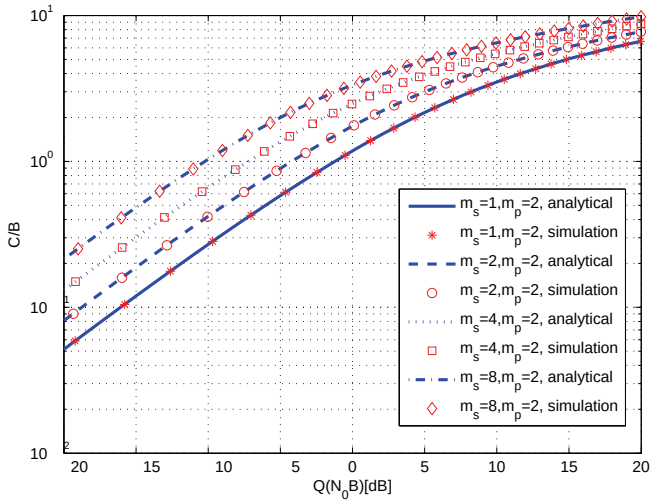


Fig. 5. Channel capacity vs. average interference power for fading parameter m_p=2 and various m_s values.

Channel capacity vs. average received interference power is presented in Fig. 4, for the case when Nakagami fading parameters in all branches of SU link are m_s=2, and values of fading parameter m_p=1, 2, 4, 8. Channel capacity is decreasing with the raise of parameter m_p, as the propagation characteristics on the links of the interference signals are improved.

In Fig. 5, channel capacity vs. average received interference power is presented, for the various secondary link fading parameter m_s=1, 2, 4, 8, and fading parameters of links between SU transmitter and PU receiver m_p=2. It can be noticed that for the fixed average interference power value, channel capacity is increasing with the improvement of propagation conditions in secondary link channel, that correspond to the raise of parameter m_s.

CONCLUSION

In this paper the closed-form expressions for the secondary link channel capacity is given, for the spectrum-sharing system with the limited average power at PU receiver.

We analyzed the case where OSTBC with Alamouti scheme is employed at the SUs, and fading on all links is Nakagami-*m* distributed. The analytical results are confirmed using an independent simulation method. This analysis can be further extended to the case of higher order diversity with OSTBC and correlation between branches [12].

ACKNOWLEDGEMENT

"This work was supported by the Serbian Ministry of Science under technology development project TR32028 - "Advanced Techniques for Efficient Use of Spectrum in Wireless Systems".

REFERENCES

[1] A. Ghasemi, E. S. Sousa, "Fundamental Limits of Spectrum-Sharing in Fading Environments," *IEEE Trans. on Wireless Communications*, Vol. 6, No.2, pp. 649-658, Feb. 2007.

[2] J. Mitola, "Cognitive Radio for Flexible Mobile Multimedia Communications," in Proc of *IEEE Int. Workshop on Mobile Multimedia Communications* (*MoMuC* '99), San Diego, 1999.

[3] L. Musavian, S. Aissa, "Capacity and Power Allocation for Spectrum Sharing Communications in Fading Channels," *IEEE Transactions on Wireless Communications*, Vol.8, No.1, 148-156, January 2009.

[4] B. Vucetic and J. Yuan, *Space-time coding*, 1st ed. John Wiley & Sons, 2003.

[5] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451-1458, Oct. 1998.

[6] R. Duan, M. Elmusrati, R. Jantti, R. Virrankoski, "Capacity for Spectrum Sharing Cognitive Radios with MRC Diversity at the Secondary Receiver under Asymmetric Fading", in Proc. of *Globecom* 2010, Miami, USA, 2010.

[7] M. D. Yacoub, J. E. V. Bautista, and L. G. de Rezende Guedes, "On higher order statistics of the Nakagami-m distribution," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 790 - 794, May 1999.

[8] V. Blagojevic, P. Ivanis, "Level Crossing Rate of MRC with Transmit Antenna Selection in Unequally Distributed Nakagami Fading Channels," in Proc. of *5th European Conference on Circuits and Systems for Communications (ECCSC'10)*, pp. 260-263, Belgrade, Nov. 2010.

[9] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed., Academic Press inc., 1994.

[10] M. Abramowitz , I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover, NewYork, 1972.

[11] Y. Zheng, C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, pp. 920-928, June 2003.

[12] P. Ivanis, V. Blagojevic, D. Drajic and B. Vucetic, "Closed-Form Level Crossing Rates Expressions of Orthogonalized Correlated MIMO Channels," *IEEE Transactions on Vehicular Technology*, paper accepted, DOI: 10.1109/TVT.2011.2129546, 2011.

Guiding Properties of the Polymer Optical Fibers

Valentina Markova[1], Boyka Ilieva[2] and Borislav Naydenov[3]

***Abstract* – In this paper the transmission possibilities of polymer optical fiber compared with its glass fiber counterpart are considered. It was confirmed that the PHFIP2 POF has the lowest chromatic dispersion compared to the PMMA POF and multimode glass optical fibers. It is also investigated the wavelength dependence of the refractive index of described fibers. As a result, the PHFIP2-FA is the best candidate for designing broadband local area networks.**

***Keywords* – Polymer optical fibers, material dispersion, refractive index, graded-index fiber.**

I. INTRODUCTION

The recent research interest has been focused on the development of polymer optical fibers that answer on growing requirements on up-to-date telecom applications.

Polymer optical fibers (POF) were first introduced several years ago for use in automobiles and sensors for on machine automation. Well known glass fibers are widely used in high-speed long-distance communication networks because of its high bandwidth and low attenuation. However, regarding the broadband local area networks, since the glass fibers require precise handling and connection, serious increase of the cost of the whole systems is expected. The POF is a promising candidate for providing high speed telecommunication services within the customer's premises. It is worth noting that termination and installation of POF are easier and faster compared to the single and multimode glass counterparts, especially in splicing. The typical large core of POF enables large tolerance on axial misalignment, which results in cheaper connectors.

POF are usually made of polymethylmethacrylate (PMMA) and perfluorinated (PF) polymer material. PMMA- POF with step index profile (SI) has been commercially available for many years. This fiber has high attenuation (approximately 100 dB/km) at the visible region and limited bandwidth [1], [2]. Slightly improvement in the bandwidth characteristics of POF has been obtained by grading the refractive index. A large core, high-bandwidth, and low-loss graded index polymer optical fiber (GI-POF) have been proposed in [3]. Additional reduction of transmission loss has been achieved by development of the PF GI-POF [4], [5]. The PF fibers are with lowest attenuation (40dB/km) due to the elimination of intrinsic absorption loss that exists in PMMA-POF.

It is well-known that transmission capacity of fiber data link is limited by it dispersion properties. Wavelength dependence of the refractive index of the polymer, the finite spectral width of the light source, material composition, index profile, strongly affects the dispersion properties of the fiber and should be taken into accounts.

Optimization of the attenuation and dispersion properties has opened the way for broadband transmission over POF based network. This article is comprehensive review on transmission possibilities of polymer fibers compared with its glass fiber counterpart.

The paper is organized as follows. Firstly, the Sellmeier coefficients determination for investigated POF will be described. Secondly, the transmission properties of designed POF such as dispersion and group delay will be discussed. Finally, optical field distribution in the polymer fibers will be investigated.

A. Dispersion relations for Refractive Index (RI)

The Refractive Index (RI) is a fundamental parameter for any kind of optical material.

If one treats the material as equivalent to a collection of *m* harmonic oscillators resonant to a radiation of various wavelengths A_j, one can derive the equation [6]

$$n^2(\lambda) = 1 + \sum_{i=1}^{m} \frac{A_i \lambda^2}{(\lambda^2 - \lambda_i^2)}, \qquad (1)$$

where λ is the wavelength of the incident radiation, and A_i is a constant that depends on the number of oscillators per unit volume and is called the *oscillator strength* of the oscillators resonant at wavelength λ_i. Eq. (1) is generally called the Sellmeier formula, since Sellmeier proposed it in 1871.

The usual form of the equation for glasses is

$$n^2(\lambda) - 1 = \frac{A_1 \lambda^2}{\lambda^2 - \lambda_1^2} + \frac{A_2 \lambda^2}{\lambda^2 - \lambda_2^2} + \frac{A_3 \lambda^2}{\lambda^2 - \lambda_3^2}, \qquad (2)$$

where *n* is the RI, λ is the wavelength, and $A_{1,2,3}$ and $\lambda_{1,2,3}$ are experimentally determined Sellmeier coefficients. For common optical glasses, the refractive index calculated with the three-term Sellmeier equation deviates from the actual refractive index by less than 5×10^{-6} over the wavelengths range of 365 nm to 2.3 μm.

These coefficients are determined experimentally for common optical materials.

Actually there exists a great diversity of polymeric materials. The most frequently used material for polymer fibers is the thermoplastics PMMA (Polymethylmethacrylate), better know as Plexiglas and perfluorinated polymers

To change the refractive index of optical fiber, pure polymer is often doped with dopants. For example, adding

[1]Valentina Markova , [2]Boyka Ilieva and [3]Borislav Naydenov are with the Faculty of Electronics, Telecommunications Department, Technical University of Varna, 1 Studentska str., 9010 Varna, Bulgaria.

[1]E-mail: valliq@abv.bg

[2]E-mail: boykailieva@abv.bg

[3] E-mail: borna@abv.bg

germanium can result in an increase in the refractive index, while adding fluorine reduces it. The refractive index of doped material can be determined by Sellmeier equations. Corresponding coefficients for fibers proposed in [7] are given in Table 1. PHFIP 2-FA is one of the fluorinated polymers whose monomer unit contains only three carbon-hydrogen bonds. As dopants to control the refractive-index profile, benzyl benzoate BEN for PMMA and dibutyl phthalate DBP forPHFIP 2-FA were used.

For comparison, the Sellmeier's coefficients of silica-based fiber are also considered.

TABLE 1
SELLMEIER'S COEFFICIENTS

material	A_1	A_2	A_3	λ_1^*	λ_2^*	λ_3^*
PMMA	0.4963	0.6965	0.3223	0.0718	0.1174	9.237
PMMA-BEN	0.4855	0.7555	0.4245	0.1043	0.1147	4.934
PHFIP2-FA	0.4200	0.0461	0.3484	0.0587	0.0878	0.0927
PHFIP2-FA-DBP	0.2680	0.3513	0.2498	0.0791	0.0838	0.1062
Silica	0.6968	0.4082	0.8994	0.0685	0.1161	9.9140

II. DESIGN AND INVESTIGATIONS OF POF

A. Design of Polymer Optical Fibers

There are two important limitations for applications of polymer optical fiber systems – high attenuation and limited bandwidth. Choosing appropriate fiber parameters is an important issue for a given optical system. Reduction of transmission loss depends on material used for the fiber core. Cross-sectional dimensions, material composition, and refractive index profile all influence the losses, dispersion and the nonlinearities of the fiber and must be chosen carefully to achieve a satisfactory tradeoff for a given application. To fit user requirements all these parameters of existing polymer fiber samples are experimentally measured.

Design of new optical fiber usually starts with defining refractive index profile. This step includes fixing the geometry and the material composition of the fiber.

The material dispersion and dispersion of the profile are calculated from the known Sellmeier coefficients of the used material, which are taken from table1.

The typical characteristics of investigated POF consistent with IEC standard are given in table 2.

TABLE 2
THE CHARACTERISTICS OF INVESTIGATED POF CONSISTENT WITH IEC STANDARD

	IEC standard	Material used	Core/ Cladding diameter [µm]	RI Core/ Cladding	NA	Operating wavelength [µm]
GOF-SI	607932-20 Cat.A2	Silica	62.5/125	1.48/1.45	0.18	0.85
GOF-GI	607932-10 Cat.A1	Silica	62.5/125	1.48/1.46	0.27	0.85
PMMA-SI	60793240 Cat.A4a	PMMA	975/1000	1.49/1.42	0.5	0.65
PMMA-GI	607932-40 Cat.A4e	PMMA	500/750	1.49/1.45	0.25	0.65
PMMA-BEN-SI	607932-40 Cat.A4a	PMMA-BEN	975/1000	1.51/1.34	0.5	0.65
PMMA-BEN-GI	607932-40 Cat.A4e	PMMA-BEN	500/750	1.51/1.47	0.25	0.65
PHFIP2-FA-GI	607932-40 Cat.A4g	PHFIP2-FA	120/490	1.35/1.32	0.19	0.65
PHFIP2-FA-DBP-GI	607932-40 Cat.A4g	PHFIP2-FA-DBP	120/490	1.37/1.34	0.19	0.65

*GOF - Glass Optical Fiber

B. Transmission Properties of Polymer Optical Fibers

It is well known that dispersion in POF can be divided into two main types – chromatic (material and waveguide) dispersion and modal dispersion. The modal dispersion has dominant influence on bandwidth of SI POF. Grading the index profile significantly enhanced transmission capacity of polymer fibers.

The material dispersion is defined as the wavelength dependence of refractive index of the polymer. Fig.1 displays the refractive index dependence of PMMA, PMMA-BEN-doped, PHFIP 2-FA, PHFIP-2-DBP-doped and glass optical fiber (GOF) in respect to wavelength. It is evident from results that the PMMA-BEN polymer has the highest deviation of refractive index versus wavelength, while PHFIP 2- FA – lowest.

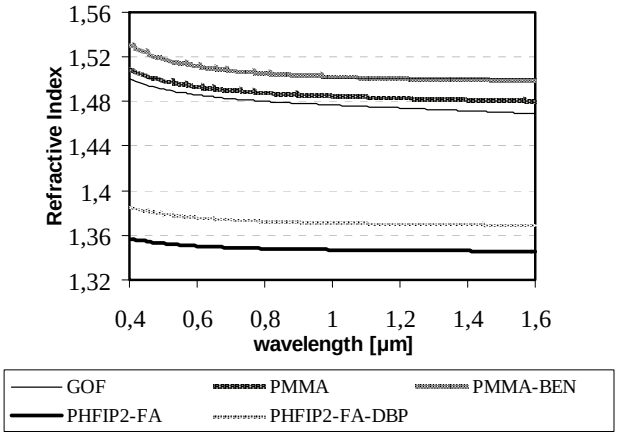


Fig.1. Wavelength dependence of Refractive index of PMMA, PHFIP2_FA polymer and silica

The described fibers are optimized to operate at 0,65μm and 0,85μm wavelength because the low attenuation in these windows. Low attenuation at visible and near infrared region is advantageous even with the dispersion limitation because the material dispersion decreases with increasing wavelength.

In a fiber, the materials of core and cladding are different. If there are L layers in the fiber cross-section, each layer has different refractive index.

The total material dispersion of a fiber is calculated by:

$$D(\lambda) = -\frac{\lambda z}{c}\sum_{i=1}^{L}\Gamma_i \frac{d^2 n_i}{d\lambda^2} \qquad (3)$$

where the confinement factor of each layer is Γ_i. The confinement factor is the portion of total power guided in the i-th layer.

The typical characteristics of PMMA and PHFIP2-FA fibers with SI and GI are summarized in table 3. For comparison, the transmission parameters of glass fibers at 0,65μm and 0,85μm wavelengths are also considered.

Chromatic dispersion, considering material and waveguide dispersions, as described in detail in this section. The waveguide dispersion of investigated fibers is almost zero at 0,65μm and 0,85μm wavelength, so the corresponding material dispersion strongly affects the bit rate performance of POF link.

Fig. 2 shows the experimental results of the chromatic dispersion of the PMMA GI POF, PHFIP-2-FA GI POF and graded index GOF. The PHFIP-2-FA GI polymer fiber has lowest material dispersion in wide wavelength range from visible to near infrared region.

The material dispersion of Perfluorinated (PF) polymer-based GI POF at 0,85μm wavelength is 58.995 ps/km.nm which is much lower than 133.532 ps/km.nm of the PMMA at the same wavelength. Therefore, it is expected that PF polymer-based GI POF will have the highest bandwidth from the visible to the near infrared region. Moreover in case of glass fiber, the material dispersion at 0,85μm wavelength is almost the same as that of PMMA POF and two times higher than that of PHFIP-2-FA GI POF.

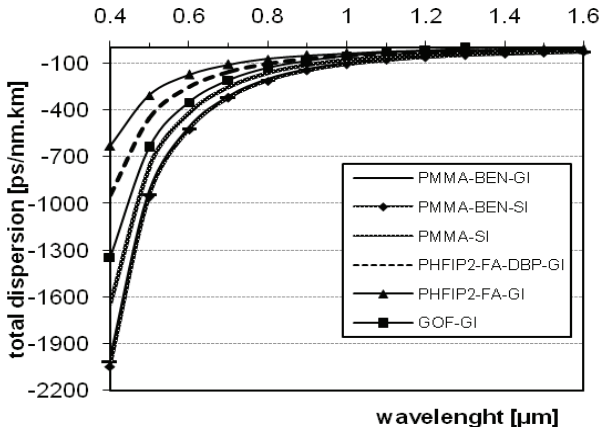


Fig.2. Chromatic dispersion of PMMA-based, PHFIP2_FA-based POF and glass optical fiber.

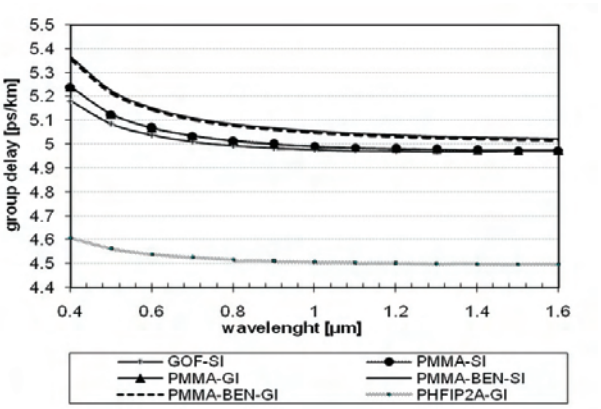


Fig.3. Group delay of PMMA-based, PHFIP2_FA-based POF and glass optical fiber.

It is observed in all optical fibers that the speed of propagation of light differs for different wavelength. Group delay gives the propagation group delay for the selected fiber mode versus the wavelength.

The group delay T_g is defined as:

$$T_g = z\frac{d\beta}{d\omega} = z\frac{d\lambda}{d\omega}\frac{d(nk_0)}{d\lambda} \qquad (4)$$

where n is the refractive index, $\lambda = 2\pi c/\omega$, $k_0 = 2\pi/\lambda$

The group delay of PMMA, PMMA-BEN with SI and GI, PHFIP2A – GI POF and glass optical fiber are indicated in fig.3. As expected, with increasing wavelength the delays become smaller, which means greater speed. Since, fluorinated materials have the lowest values of dispersion loss, therefore PHFIP2A-FA had the lowest values of group delay ($4{,}513\text{x}10^{-6}$,ps/km at 0,850μm wavelength).

It was confirmed from fig.2 and fig. 3 that PHFIP2A – GI POF shows better performance compared to PMMA and GOF with respect to dispersion at discussed wavelengths.

It is well-known that GOF connection operations require specific setups and skilled technicians. POF's have larger core diameters (0,125-2mm) than GOF's (9μm and 62,5 μm). POFs are easier to connect, because of the existing technology allowing customers to do the connection themselves.

The values of the mode field diameters of investigated OF are summarized in table 3. PMMA-SI and PMMA-BEN-SI have the largest diameters of the intersection of OF, within which the main power of optical radiation (mode field) is concentrated. The effective mode field for PMMA-SI is ~670μm and PMMA-BEN-SI is ~680μm, which is result of the larger diameters of the core and cladding of the fibers. On the other hand, increasing the core diameter causes a decrease in the speed of data transmission.

POFs have large numerical aperture related with large effective mode field, which eliminate the necessity to use precise light sources.

TABLE 3
CHARACTERISTICS OF PMMA POF, PHFIP2-FA POF AND GOF

	wavelength [μm]	RI	Dispersion [ps/km.nm]			Near field MFD [μm]	Far field MFD [μm]	Effective MFD [μm]	Group delay x10^{-6} [ps/km]
			Waveguide	Material	Chromatic				
GOF-SI	0.652	1.458	0.237	–269.289	-268.837	41.763	37.409	43.657	5.021
	0.854	1.479	0.209	–104.574	-104.148	41.290	37.619	43.821	4.986
GOF-GI	0.652	1.484	–0.019	–268.798	–268.982	10.225	10.211	10.202	5.021
	0.854	1.479	0.026	–104.280	–104.384	11.796	11.789	11.786	4.986
PMMA-SI	0.652	1.492	0.166	–322.026	–321.948	647.307	573.833	677.055	5.049
	0.854	1.487	0.142	–133.698	–133.844	643.069	573.985	672.064	5.006
PMMA-GI	0.652	1.4892	0.028	–321.707	–321.610	35.896	33.268	37.497	5.048
	0.854	1.487	0.303	–133.532	–133.329	36.880	34.559	38.490	5.006
PMMA-EN-SI	0.652	1.512	0.255	–408.630	–408.392	651.129	573.503	681.640	5.129
	0.854	1.501	–5.312x10^{-6}	–176.152	–176.121	648.606	573.738	678.589	5.074
PMMA-BEN-GI	0.652	1.510	0.190	–401.172	–401.047	34.278	31.101	35.829	5.120
	0.854	1.505	0.238	–172.419	–172.123	34.684	31.647	36.251	5.067
PHFIP2-FA-GI	0.652	1.350	–0.149	–133.972	–134.265	13.335	13.253	13.346	4.531
	0.854	1.348	–0.132	–58.639	–58.995	15.481	15.425	15.405	4.513
PHFIP2-FA-DBP-GI	0.652	1.375	–0.145	–195.858	–196.209	13.310	13.228	13.323	4.628
	0.854	1.372	–0.071	–84.846	–85.183	15.437	15.380	15.362	4.602

III. CONCLUSION

In this paper the guiding properties of polymer optical fibers compared with its glass fibers counterpart are considered. It was investigated the wavelength dependence of the refractive index of PMMA POF, PMMA-BEN POF, PHFIP2-FA POF, PHFIP2-FA DBP POF and GOF. It was confirmed that the PHFIP2 POF has the lowest chromatic dispersion compared to the PMMA POF and multimode glass optical fibers. As a result, the PHFIP2-FA is the best candidate for providing high speed telecommunication services within the customer's premises.

REFERENCES

[1] T. Kaino, M. Fujiki, and K. Jinguji, "Preparation of plastic optical fibers," Rev. Electron. Commun. Lab., vol. 32, pp. 478–488, 1984.

[2] T. Monroy, H.P.A. vd Boom, A. Koonen, G. Khoe, Y. Watanabe, Y. Koike, T. Ishigure, "Data transmission over polymer optical fibers", Optical Fiber Technology 9, Academic press, pp. 159-171, 2003, available at www.sciencedirect.com .

[3] Y. Koike, T. Ishigure, and E. Nihei, "High-bandwidth graded-index polymer optical fiber," J. Lightwave Technol., vol. 13, pp. 1475–1489, July 1995.

[4] N. Yoshihara,"Low-loss, high-bandwidth fluorinated POF for visible to 1.3-mm wavelength," in Proc. Optic. Fiber Conf. (OFC'98), San Jose,

[5] T. Ishigure, Y. Koike, and J. W. Fleming, "Optimum index profile of the perfluorineted polymer-based GI polymer optical fiber and its dispersion properties", J. Lightwave Technol., vol. 18, pp. 178–184, Feb. 2000.

[6] E. Palik, "Handbook of thermo-optic coefficients of opticl materials with applications", Japan, Academic Press, 1998.

[7] T. Ishigure, E. Nihei and Y. Koike, "Optimum refractive-index profile of the graded-index polymer optical fiber, toward gigabit data links", Applied Optics, vol.35, No 12, 20 April 1996.

Design of a TDMA-based Multi-Channel MAC Protocol for Wireless Sensor Networks

Milica D. Jovanovic[1] and Goran Lj. Djordjevic[2]

***Abstract* – In this paper, we propose a TDMA-based multi-channel Multiple Access Control (MAC) protocol for Wireless Sensor Networks (WSNs) and study its performance. The proposed protocol builds on TFMAC, a MAC protocol proposed in our previous work, but extends it in several ways to improve self-configuration capabilities and autonomy of operation, as follows: (a) protocol operation is divided into two main phases: initialization phase, where sensor nodes first discover their neighbours, and then cooperatively establish TFMAC-like collision-free multichannel TDMA schedule without the need for any local or global master nodes, and active phase, when regular data communication is taking place, and (b) a new hierarchical TDMA scheme is introduced which eases both the node synchronization and the adaptation to the incremental topology changes during the active period. We demonstrate by simulations that the proposed multi-channel protocol achieves reasonable performances during self-configuration process, in terms of the duration of the initialization phase.**

***Keywords* – wireless sensor networks, medium access control, energy efficiency, TDMA-based protocol**

I. INTRODUCTION

A wireless sensor network (WSN) is made of a number of autonomous and inexpensive sensor nodes each of which composed of sensors, a low-power radio transceiver, small amount of memory and processing capability as well as limited battery power supply. The common vision is to create a large WSN through *ad-hoc* deployment of hundreds or thousands of such tiny devices able to sense the environment, compute simple task and communicate with each other in order to achieve common objective [1].

Media Access Control (MAC) is a key component to ensure the successful operation of WSNs and it has obtained intensive research attention [2]. A MAC protocol decides when competing nodes could access the shared medium to transmit their data and tries to ensure that no collisions occur. MAC protocol controls the activity of nodes' radio transceiver directly, and therefore makes a strong impact to the overall network performance and energy efficiency.

Most of the MAC protocols proposed for WSNs assume the use of simple, low-cost transceivers that can operate on a single channel (frequency), only. On the other hand, the current commercial, low-power transceivers, such as CC1100 [3], already provide the basic functions required to support multiple channels. Such transceiver cannot transmit and receive at the same time, but it can switch the operating frequency dynamically. Availability of multiple channels adds one more degree of freedom to wireless communications that can be exploited to increase the spatial reuse by providing more simultaneous transmissions than is possible in single-channel WSNs. Thus, network throughput can potentially be increased.

There exist several proposals for multi-channel usage in WSN [4]. MMSN [5] is a slotted CSMA protocol that assigns channels to the receivers. At the beginning of each beacon interval, potential senders for the same receiver contend for the medium on receiver's channel. Y-MAC [6] is based on TDMA access where timeslots are assigned to the receivers. In this protocol, potential senders contend for the medium on default channel at the beginning of the timeslot owned by the common intended receiver. If multiple packets need to be transmitted, then the senders and the receiver hop to a new channel according to a predetermined sequence. TFMAC is a TDMA-based multi-channel MAC protocol, proposed in our previous work [7], which divides each channel into time slots and assigns each transmitter one time slot on each channel. A distributed time slot/channel allocation scheme employed in TFMAC guaranties collision-free data communication among neighboring nodes.

In this paper, we propose a TDMA-based multi-channel MAC protocol for WSNs which builds on TFMAC but extends it in several ways to improve self-configuration capabilities and autonomy of operation. The protocol operation is divided into two main phases: *initialization phase*, where sensor nodes first discover their neighbors, and then cooperatively establish TFMAC-like collision-free multi-channel TDMA schedule, and *active phase*, when regular data communication is taking place. Also, we introduce a new hierarchical TDMA scheme which eases both the node synchronization and the adaptation to the incremental topology changes during the active period. The rest of the paper is organized as follows. In Section II, we describe the design of the enhanced TFMAC protocol. In Section III, we provide performance evaluations and analysis. Finally in Section III, we conclude the paper.

II. THE PROTOCOL DESCRIPTION

The protocol assumes a randomly deployed WSN with a single base station. The protocol operation comprises three main phases: (a) inactive phase, (b) initialization phase, and (c) active phase (Fig. 1)). After deployment, all nodes are in the inactive phase. During this period, nodes regularly sample the medium to check for activity on default channel. In this context, sampling means periodically measuring the received signal strength. The transition from inactive to initialization phase is initiated by the base station which transmits a wake-

[1,2] Milica D. Jovanovic and Goran Lj. Djordjevic are with UNIVERSITY OF NIŠ, FACULTY OF ELECTRONIC ENGINEERING, Aleksandra Medvedeva 14, P.O. Box 73 18000 Niš, Serbia, E-mails: {milica.jovanovic, gdjordj}@elfak.ni.ac.rs

up tone - an unmodulated continuous signal at frequency of the default channel of duration equal to the sampling period. Nodes that detect activity on the channel, first retransmit the wake-up tone, and then switch to the initialization phase. During the initialization phase, the network prepares for its working life by going through a self-organization process which aim is to synchronize the nodes, set up a multi-channel TDMA schedule, and configure the nodes with the correct control parameters. The working life i.e. the active phase of the protocol starts as soon as the organized structure is established. The network toplogy may change over the active phase: existing nodes can fail, and new nodes may be introduced. Once the network is too altered to function correctly, return to inactive phase may be necessary in order to re-initialize the network and rebuild the self-organization. The phase transition from initialization to active phase, as well as from active to inactive phase is initiated by the activation and deactivation command messages, respectively. These messages are broadcasted by the base station and then flooded through the network.

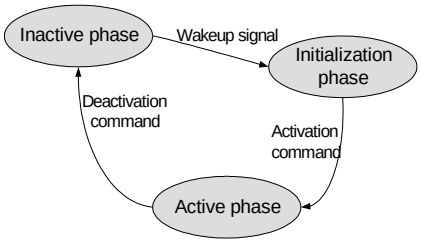


Fig. 1. The protocol phases

A. Active phase

We assume that all nodes are synchronized during the active period, and time is divided into cyclic intervals of fixed length, so called *epochs*, each having N super-frames. Each super-frame begins with a control slot (CS), followed by a sequence of N_F frames. CSs are used for conflict-free transmission of control packets which include information needed for the nodes to get synchronized and join the network. Each frame is divided into the N time slots, during which the nodes send data messages without contention and sleep when they do not have a message to send or receive. Time slot duration is suitably chosen to accommodate the transmission of one fixed-size data message. The hierarchical chart which describes the organization of time in the protocol is shown in Fig. 2. Symbols on edges in this chart indicate the number of sub-intervals contained within the parent interval. Symbols which denote the duration of the various intervals are written in brackets.

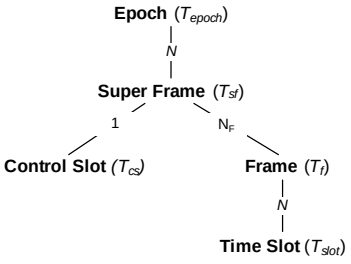


Fig. 2. The protocol time organization

The protocol requires that each node is assigned two identifiers: the pre-assigned permanent identifier (ID), which is used as a network-wide unique node's MAC address, and the local identifier (LID), which is dynamically assigned to the node during network initialization phase. LID is an integer from interval [0, N-1] which should be unique in node's 2-hop neighborhood. LID indicates CS in the epoch which is owned by the node. The protocol also requires each node to maintain a simple local lookup table, called *LID table*, which will associate 1-hop neighbor nodes IDs with their assigned LIDs.

The protocol provides conflict-free transmission during a particular CS slot by scheduling access among 2-hop neighboring nodes according to their LIDs. A node transmits control packet in its own CS; listens to the CSs that are owned by its neighbors, and eavesdrops the unused ones, to hear the newly joined nodes. A control packet includes the following information about the sending node: (a) both identifiers (ID and LID) (b) a bit-vector detailing which LIDs are occupied by the 1-hop neighbors, (c) status of each slot in the frame and (d) timestamp with local time.

In a network with N_F available frequencies, each node is assigned N_F transmission slots, and each transmission slot is assigned a different channel for data transmissions. On the other hand, each node is assigned a single receiving channel that it uses to receive data packets during its reception slots. Every node has one reception slot for each neighbor.

A TFMAC slot schedule is correct if for any given node n_i the following two conditions are satisfied:

- The transmission slot when n_i transmits on channel frequency f_j overlaps with reception slots of all its neighbors with assigned receiving frequency f_j.
- During the transmission slot when n_i transmits on channel frequency f_j, n_i is the only transmitter on frequency f_j in its two-hop neighborhood.

The first condition is necessary to provide a bidirectional data link between any two neighboring nodes, while the second one ensures collision-free communication.

B. Initialization Phase

Network initialization phase is divided into three time-separated stages: (a) neighborhood discovery, when nodes collect IDs of all nodes within their 2-hop neighborhoods, (b) LID and channel allocation stage, when each node is allocated a receiving channel and unique LID within the scope of its 2-hop neighborhood, and (c) time slot allocation stage, when each node is allocated one time slot for data transmission in the frame for every channel.

Essential requirement for all three stages is the ability to efficiently disseminate information, in a form of so called *initialization messages*, among nodes within 2-hop neighborhoods. During the first two stages, the protocol uses a simple, randomized (asynchronous) information dissemination scheme, and then, for the third phase, switches to deterministic (synchronous) TDMA-based transmission schedule. In randomized information dissemination scheme, each node transmits its initialization messages at randomly chosen times. When its transmission finishes at time t_i, the node schedules new transmission at time $t_{i+1} = t_i + T_d$, where T_d

is picked uniformly at random within the interval [0, T_{dmax}]. At time t_{i+1} the node sense medium and skip this transmission trail if medium is busy. In the deterministic information dissemination scheme, nodes are synchronized and transmit their initialization messages according to a collision-free transmission schedule. TDMA frame consists of N slots, and each node is allocated one transmission slot in the frame according to its LID. Two-hop uniqueness of assigned LIDs ensures conflict-free transmission.

During the node discovery stage, the initialization message contains ID of the transmitting node and IDs of all its currently discovered neighbors. Using the content of the received initialization message, the node updates its neighborhood list. In the second initialization stage, the initialization message contains not only IDs, but also LIDs and receiving channels of the transmitting node and all its neighbors with known LID and channel. Node chooses LID (so it's unique in its 2-hop neighborhood), when all its 2-hop neighbors with smaller ID have already chosen its LID. During the third stage, the squentiality is provided in the same manner as in the previous stage: the node chooses transmitting slots in the frame only if all of its 2-hop neighbors with smaller ID have already chosen their transmitting slots.

III. Performance Evaluation

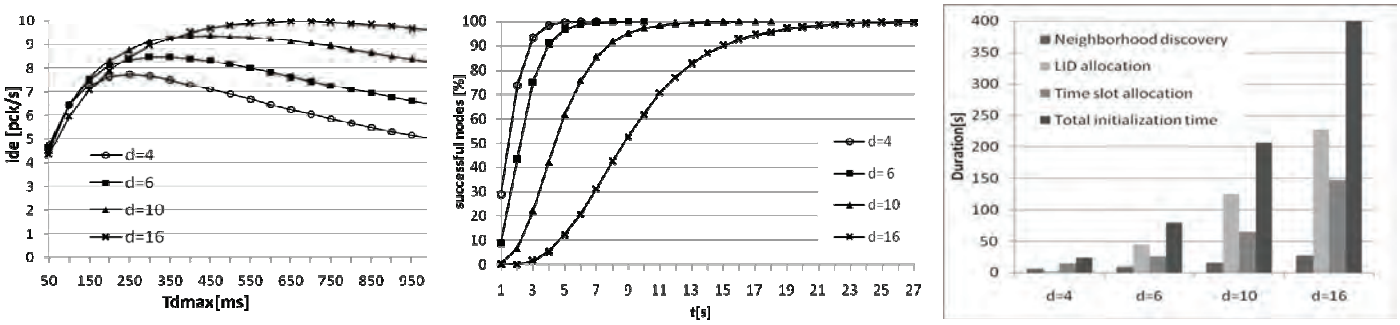
We implement enhanced TFMAC protocol in a custom WSN simulator build in C++, and conduct several experiments to evaluate its performances. All our evaluations are based on the simulation of the same WSN composed of 200 nodes randomly placed at fixed positions within an area of 100x100 m^2. The node density, which is defined as the average size of 1-hop neighborhood in the WSN, was varied indirectly, by varying radio transmission range. All nodes are equipped with single half-duplex transceivers with multi-channel capability. We used transceiver CC1100 as the hardware reference. Transfer rate of 20 Kbps is assumed, and packet length is fixed to the value of 64 bytes.

In the first set of experiments, we investigate the performances of the proposed three-stage self-organization mechanism in terms of time efficiency. The time needed to complete each individual stage of the initialization phase is primarily affected by the efficiency of the information dissemination scheme employed in the stage. This time is deterministic for the third stage, where TDMA-based information dissemination is used, and depends on the number of time slots per frame, only. Contrary, the duration of both neighborhood discovery and LID allocation stage is nondeterministic due to random nature of the asynchronous dissemination scheme used in these stages. This means that time periods reserved for these two stages, T_{ND} and T_{LID}, should be large enough to allow each node in the network to successfully finish its task with high probability. In order to estimate appropriate values of parameters T_{ND} and T_{LID} we rely on simulations. First we define the information dissemination efficiency (*ide*) as the average number of packets received per second at each node under asynchronous dissemination scenario. Intuitively, as *ide* is higher, the initialization process will be completed in less time and with less consumed power. The *ide* is influenced by the value of T_{dmax} which defines the rate at which nodes transmit their initialization messages. Choosing a small value of T_{dmax}, which means higher rate of initialization message transmission, would lead to a smaller *ide* due to frequent message collisions. With a large value of T_{dmax}, the probability of message collision will be small, but *ide* will be reduced again due to low rate of initialization message transmissions. The goal then is to choose T_{dmax} so as to maximize the *ide*. Fig 2(b) shows the *ide* as a function of T_{dmax} for four different node density values. As can be seen from Fig. 3(a), the optimal value of T_{dmax} depends on the node density (d), and ranges from 300 ms for d = 4 up to 750 ms for d = 16.

Fig. 3(b) provides insight on how the percentage of nodes that successfully finish neighborhood discovery stage changes as time progresses. Note that these simulations are carried out assuming the optimal value of T_{dmax} for each of four different node density values. According to Fig. 3(b), the adequate value of T_{ND} is highly dependent on node density. For example, when d=4, T_{ND} should be set to at list 5s to offer each node in the network to fully discover its neighborhood with high confidence. As node density increases, the time that should be reserved for neighborhood discovery stage increases also, reaching T_{ND}=25s for d=16.

Fig. 3(c) shows the estimated duration of each initialization stage as well as the total duration of the initialization phase for different node density values. As can be seen in Fig. 3(c), the LID allocation is the most time consuming initialization operation which takes more than 50% of total initialization time. This can be contributed to the relatively high complexity of the distributed LID allocation algorithm which requires several iterations of information exchange among 2-hop neighboring nodes. As a result of deterministic information dissemination, the time slot allocation stage lasts for about 30% less then LID allocation stage although basically the same algorithm is used in both stages. The neighborhood discovery is the least time consuming stage since it require one complete information exchange within each 2-hop neighborhood, only. As expected, the total initialization time is greatly dependent on node density and ranges from 25s, for low density network (d=4), up to 400s, for network with high node density (d=16).

The second set of experiments relates to the active phase of enhanced TFMAC protocol. Our focus in this study is on the benefits of multiple channels in terms of throughput and latency. As a performance metrics we adopt: the packet delay, and throughput. The packet delay is the average time, in seconds, from the arrival of a packet at the buffer of the source node to the arrival of the packet to the destination neighboring node. The throughput is defined as the ratio of the average number of packets that are successfully received at each node to the total simulation time. We run simulations by varying the following three traffic/network parameters: traffic load, number of available channels, and node density. Node traffic is statistically generated with packet inter-arrival time chosen from an exponential distribution with rate λ, i.e., with the average inter-arrival time 1/λ. The traffic load was varied by changing λ. Also, we assume the gossip traffic pattern, in which each node only communicates with its 1-hop neighbors.



(a) Information dissemination efficiency (b) Successfulness of the neighborhood discovery (c) Duration initialization stage

Fig. 3. Protocol performance in initializaton phase.

Fig. 4 and 5 show throughput and delay characteristics in network with node density d=6, when enhanced TFMAC protocol is configured with different number of channels. These simulation results confirm TFMAC's scalability in term of number of channels. When the number of channels increases from 1 to 8, the maximum throughput (i.e. the largest admissible traffic load yielding a finite average packet delay) increases from 1.68 pck/s to 5.46 pck/s. Also, with larger number of channels the protocol transfers packets between nodes with smaller delay. For example, when λ=0.8 pck/s, the average packet delay decreases from 0.53s to 0.32s when the number of channels increases from 1 to 8.

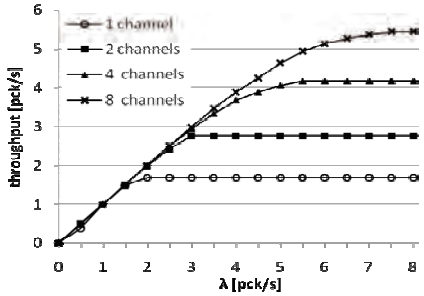


Fig. 4. Throughput over varying traffic load.

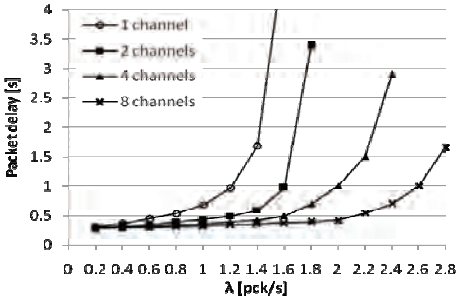


Fig. 5. Packet delay over varying traffic load.

IV. Conclusion

In this paper we have presented a schedule-based multi-channel MAC protocol designed for WSNs with high throughput requirements. The protocol adopts a channel/time slot allocation method of TFMAC, a multi-channel MAC protocol introduced in our previous work. In contrast to TFMAC, the protocol described in this paper is completely self-organizing in the sense that nodes are able to autonomously establish collision-free multi-channel schedules in a distributed manner. The self-organization is implemented as a three-state process that is carried out during initialization phase of the protocol. Simulation results show that (a) the initialization process completes in reasonable time even in high density networks, and (b) the multi-channel operation of the protocol greatly improves the throughput compared to a single-channel TDMA.

Acknowledgement

This work was supported by the Serbian Ministry of Science and Technological Development, project No. TR – 32009 - "Low-Power Reconfigurable Fault-Tolerant Platforms".

References

[1] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey", Computer Networks, vol. 38, no. 3, pp. 393–422, 2002.

[2] I. Demirkol, C. Ersoy, F. Alagoz, "Mac protocols for wireless sensor networks: a survey", IEEE Communications Magazine 44 (4) pp.115–121, 2006.

[3] "CC1100 Low-Power Sub- 1 GHz RF Transceiver", http://www.ti.com,

[4] M. Wang, L. Ci, P. Zhan, Y. Xu, "Multi-channel MAC Protocols in Wireless Ad Hoc and Sensor Networks," 2008 ISECS International Colloquium on Computing, Communication, Control, and Management, cccm, vol. 2, pp.562-566, 2008.

[5] G. Zhou, C. Huang, T. Yan, T. He, J. Stankovic and T.Abdelzaher, "MMSN: Multi-Frequency Media Access Control for Wireless Sensor Networks", In IEEE Infocom, April 2006.

[6] Y. Kim, H. Shin, and H. Cha, "Y-mac: An energy-efficient multi-channel mac protocol for dense wireless sensor networks," in Proceedings of IPSN '08, pp. 53–63, April 2008.

[7] M. Jovanovic and G. Djordjevic, "TFMAC: Multi-channel MAC Protocol for Wireless Sensor Networks", in Proc. of 8-th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS) , pp. 23-26, Nis, Serbia, September 2007.

Composite Third Order Intermodulation Products in HFC/CATV Systems

Oleg Panagiev[1] and Valentin Hristov[2]

***Abstract* – This paper presents research on the influence of composite third order nonlinear products on the quality of the transmitted signals in HFC/CATV systems. The study was made of the impact between the transmitted signals. Created are algorithm and block diagram for determining the number of nonlinear composite products. The results are presented in graphical and table form.**

***Keywords* – third order composite nonlinear products, three-component beat, two-component beat, HFC/CATV, algorithm.**

I. INTRODUCTION

Third order intermodulation is the beating of one signal carrier with the second harmonic of another signal carrier ($2f_i \pm f_j$) or the beating of three signal carriers together ($f_i \pm f_j \pm f_k$) with or without modulation in a broadband multichannel system (HFC/CATV). A brief mathematical analysis of third order components will help to establish the relationship between the fundamental and spurious signals. When two, three or more sinusoidal voltages of different frequencies are applied to an amplifier or/and laser diode with distributed feedback (DFB), Mach-Zehnder modulator (MZM), etc. - several third order components are generated [1], [2].

II. MATHEMATICAL ANALYSIS

In the general case the combination frequencies are determined by the formula:

$$f_{NP} = r_1 f_1 + r_2 f_2 + r_3 f_3 + ... = \sum_{i=1}^{N} r_i f_i \text{ , where} \quad (1)$$

r_i are arbitrary integers, possibly equal to zero. If the transmission characteristic of the system is of n-th order, then the coefficients $r_1, r_2, r_3...$ need to satisfy the inequality $|r_1|+|r_2|+|r_3|+... \le n$. Since subject of this research are composite nonlinear products (NP) from third order, it is possible for Eq. (1) to be given as follows

$$f_{NP} = r_1 f_i + r_2 f_j + r_3 f_k . \quad (2)$$

Here $|r_1|+|r_2|+|r_3| = 3$ and f_i, f_j, f_k are the input signals' frequencies for the respective device. The number of the transmitted in the system signals is N, where $i=1 \div N$, $j=1 \div N$, $k=1 \div N$. If $i \neq j \neq k$ and $r_1=r_2=r_3=\pm 1$ is obtained a three-component beat ($f_i \pm f_j \pm f_k$). If $i \neq j$, $k=0$ or $i \neq k$, $j=0$ or $j \neq k$, $i=0$ and respectively $r_1=\pm 1/\pm 2$, $r_2=\pm 2/\pm 1$ or $r_1=\pm 1/\pm 2$, $r_3=\pm 2/\pm 1$ or $r_2=\pm 1/\pm 2$, $r_3=\pm 2/\pm 1$ a two-component beat ($2f_i \pm f_j$, $2f_i \pm f_k$, $2f_j \pm f_k$, etc.) is obtained.

In the following mathematical analysis is taken that $f_i < f_j < f_k$ and also a unmodulated signal is applied on the system input:

$$x(t) = \sum_{i=1}^{N} A_i \cos(2\pi f_i t + \theta_i) . \quad (3)$$

as for the D/K standard in range 111MHz÷862MHz Eq. (3) can be written in the following way:

$$x(t) = \sum_{i=1}^{N} A_i \cos[2\pi.((i-1).8 + f_1)t + \theta_i] . \quad (4)$$

Let the nonlinearity is described by the polynomial

$$y(t) = a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) , \quad (5)$$

$$y(t) = y_1(t) + y_2(t) + y_3(t) . \quad (5a)$$

The nonlinear products from third order in Eqs. (5) and (5a) are generated from the third term, and the nonlinear products of second order are not a subject of this paper, the output signal can be shortened to:

$$y(t) = y_1(t) + y_3(t) = a_1 x(t) + a_3 x^3(t) . \quad (6)$$

After a substitution with Eq. (3) in Eq. (6) and getting done the respectively mathematical operations, it comes to the following expression:

a) In case of three-component beat

$$y_3^{(3)}(t) = A_{NP}^{(3)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N} \cos[2\pi(f_i \pm f_j \pm f_k).t + \theta_i \pm \theta_j \pm \theta_k] , \quad (7)$$

$$y_3^{(3)}(t) = A_{NP}^{(3)} . \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N} \cos[2\pi f_{NP}^{(3)}).t + \theta_{NP}^{(3)}] , \text{ where} \quad (7a)$$

$A_{NP}^{(3)} = G_{NP}^{(3)} . A_i . A_j . A_k = 3/2 a_3 . A_i . A_j . A_k$ is amplitude, $G_{NP}^{(3)}$ is gain, $f_{NP}^{(3)}$ - frequency and $\theta_{NP}^{(3)}$ - phase of the nonlinear product at three-component beat.

In this case there are 7 different products generated. The strongest and most important third order composite products are the result of three frequencies. These can be expressed as:

- $f_{NP}^{(3)} = f_i - f_j + f_k$. $f_{NP}^{(3)}$ values are between f_i and f_k;
- $f_{NP}^{(3)} = f_j + f_k - f_i$. $f_{NP}^{(3)} > f_k$;
- $f_{NP}^{(3)} = f_i + f_j - f_k$. For close values of f_i, f_j and f_k the values of NP are under f_i. When $f_k >> f_j > f_i$ they are negative;

[1]Oleg Panagiev is with the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com.

[2]Valentin Hristov is with the SWU"N. Rilski"-Blagoevgrad, Bulgaria, E-mail: v_hristov@aix.swu.bg .

- $f_{NP}^{(3)} = f_i + f_j + f_k$. $f_{NP}^{(3)} >> f_k$. If f_i, f_j and f_k are in the UHF band, the values of $f_{NP}^{(3)}$ go outside it.

The nonlinear products of $f_i+f_j+f_k$ type, resulted out of the beating between analogue signals, are fallen onto the sound's carrier frequencies and on 3,75MHz over the digital signal's carrier frequencies (Fig.1). Only for frequencies from the upper part of the 110MHz÷470MHz range, NP fall out of the UHF band.

The nonlinear products of $f_i+f_j-f_k$ type resulted out of the beating between digital signals, fall on 2,75MHz over the image's carrier frequencies.

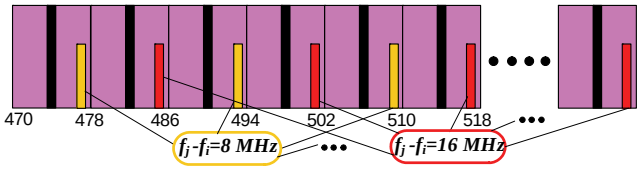


Fig.1. Influence of $f_i+f_j+f_k$

Nonlinear products from $f_i-f_j+f_k$ and $f_j+f_k-f_i$ type, influence substantially the signals of the whole working frequency range 110MHz÷862MHz (analog over analog, digital over digital, analog over digital, digital over analog: AM-VSB ↔ AM-VSB; M-QAM ↔ M-QAM; AM-VSB ↔ M-QAM), Table 1 and Fig.2.

b) In case of two-component beat

$$y_3^{(2)}(t) = A_{NP}^{(2)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \cos[2\pi(2f_i \pm f_j).t + 2\theta_i \pm \theta_j], \quad (8)$$

$$y_3^{(2)+}(t) = A_{NP}^{(2)} . \sum_{i=1}^{N} \sum_{j=i+1}^{N} \cos[2\pi f_{NP}^{(2)}).t + \theta_{NP}^{(2)}], \text{ where} \quad (8a)$$

$A_{NP}^{(2)} = G_{NP}^{(2)}.A_i.A_j = 3/4a_3.A_i.A_j$ is amplitude, $G_{NP}^{(2)}$ is gain, $f_{NP}^{(2)}$ - frequency and $\theta_{NP}^{(2)}$ - phase of the nonlinear product at two-component beat. $f_{NP}^{(2)}$ adopts and the following values (depending on the type of two-component nonlinear product): $2f_i \pm f_k$; $2f_j \pm f_k$; $f_i \pm 2f_j$; $f_i \pm 2f_k$; $f_j \pm 2f_k$. $\theta_{NP}^{(2)}$ adopts and the following values (depending on the type of two-component nonlinear product): $2\theta_i \pm \theta_k$; $2\theta_j \pm \theta_k$; $\theta_i \pm 2\theta_j$; $\theta_i \pm 2\theta_k$; $\theta_j \pm 2\theta_k$.

In this case there are 12 different products generated. The strongest and most important third order composite products are the result of one signal carrier with the second harmonic of another signal carrier. These can be expressed as:

- $f_{NP}^{(2)} = 2f_i - f_j$. $0 > f_{NP}^{(2)} < f_i < f_j$ and influences on channels from both (VHF and UHF) bands. This depends from values of f_i and f_j. Where $f_i << f_j$, the values of NP is negative;
- $f_{NP}^{(2)} = 2f_j - f_i$. $f_i < f_j < f_{NP}^{(2)}$ and influences on channels from both (VHF and UHF) bands. This depends from values of f_i and f_j. Where $f_i << f_j$, the values of $f_{NP}^{(2)}$ go outside UHF band.
- $f_{NP}^{(2)} = 2f_i + f_j$. $f_i < f_j << f_{NP}^{(2)}$.
- $f_{NP}^{(2)} = 2f_j + f_i$. $f_i < f_j << f_{NP}^{(2)}$.

Note: For last two kind nonlinear products:

- If f_i and f_j are in the VHF band, the values of $f_{NP}^{(2)}$ influence to M-QAM channels, distributed in UHF band.
- If $f_j > f_i$ >287,25MHz, the values of $f_{NP}^{(2)}$ go outside UHF band.

Nonlinear products from $2f_i-f_j$ and $2f_j-f_i$ type, influence mostly the frequency ranges, in which are f_i and f_j located.

Nonlinear products of $2f_i+f_j$ and $2f_j+f_i$ type, derived from the beating between the analog signals, influence into the working frequency range 340MHz÷862MHz and those of the beating between digital signals, are going out of the UHF range (Table 1 and Fig.3).

III. Determination of Number of the Nonlinear Products from Composite Triple Beat

According to the complicity of the problem with studying of the influence of nonlinear products from a composite third beat, is necessary to be defined the number of nonlinear products, going onto or around the channel's carrying frequency. The methods presented in [2], [3] and [4] have some restrictions which do not allow a fully and comprehensive definition of the nonlinear products' number. We suggested here an algorithm (Fig.4) define the exact number of nonlinear products, and the results are to be given in a table or/and in a graphic type. Here f_r, f_b and f_h are respectively the carrying frequency, the lowest and the highest frequency of the studied signal. For example their values for a RVII channel are: f_r=183,25MHz, f_b=182MHz and f_h=190MHz. For the defining the number of nonlinear products, included in the studied channel ($f_{NP} \neq f_r$), is used a step by step changing of f_r to the left or to the right of her. The step is $\pm k.f_o$, where f_o=0,25MHz and k=1÷16. In Table 2 are given the results for the number of nonlinear products for a transmitting of between 3 and 30 channels in a CATV system. The channels are spread according to the D/K standard. In Fig.5 are graphically presented the results for the same system, but now transmitting 35 channels.

TABLE 1

Input frequencies MHz			Nonlinear products MHz							
f_i	f_j	f_k	$f_i - f_j + f_k$	$f_j + f_k - f_i$	$f_i + f_j - f_k$	$f_i + f_j + f_k$	$2f_i - f_j$	$2f_j - f_i$	$2f_i + f_j$	$2f_j + f_i$
111,25	127,25	215,25	199,25	231,25	23,25	453,75	95,25	143,25	349,75	365,75
119,25	135,25	215,25	199,25	231,25	39,25	469,75	103,25	151,25	373,75	389,75
127,25	143,25	215,25	199,25	231,25	55,25	485,75	111,25	159,25	397,75	413,75
159,25	175,25	215,25	199,25	231,25	119,25	549,75	143,25	191,25	493,75	509,75
167,25	183,25	215,25	199,25	231,25	135,25	565,75	151,25	199,25	517,75	533,75
175,25	191,25	215,25	199,25	231,25	151,25	581,75	159,25	207,25	541,75	557,75
183,25	**199,25**	215,25	199,25	231,25	167,25	597,75	167,25	215,25	565,75	581,75
191,25	207,25	215,25	199,25	231,25	183,25	613,75	175,25	223,25	589,75	605,75
199,25	**215,25**	215,25	199,25	231,25	**199,25**	629,75	183,25	231,25	613,75	629,75
474	490	674	658	690	290	1638	458	506	1438	1454
482	498	674	658	690	306	1654	466	514	1462	1478
490	506	674	658	690	322	1670	474	522	1486	1502
530	546	674	658	690	402	1750	514	562	1606	1622
538	554	674	658	690	418	1766	522	570	1630	1646
546	562	674	658	690	434	1782	530	578	1654	1670
626	642	674	658	690	594	1942	610	658	1894	1910
642	**658**	674	658	690	626	1974	626	674	1942	1958
658	**674**	674	658	690	**658**	2006	642	690	1990	2006

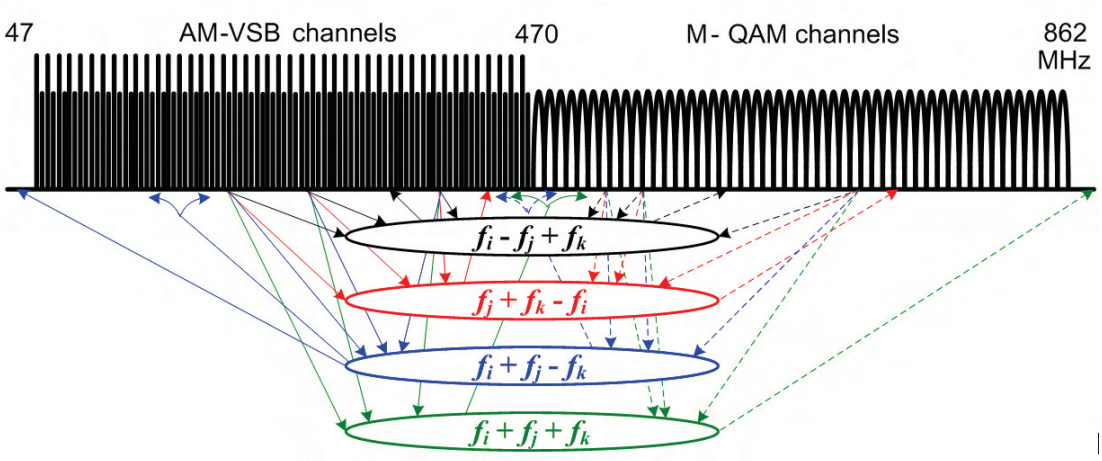


Fig.2. Influence of three-component beat

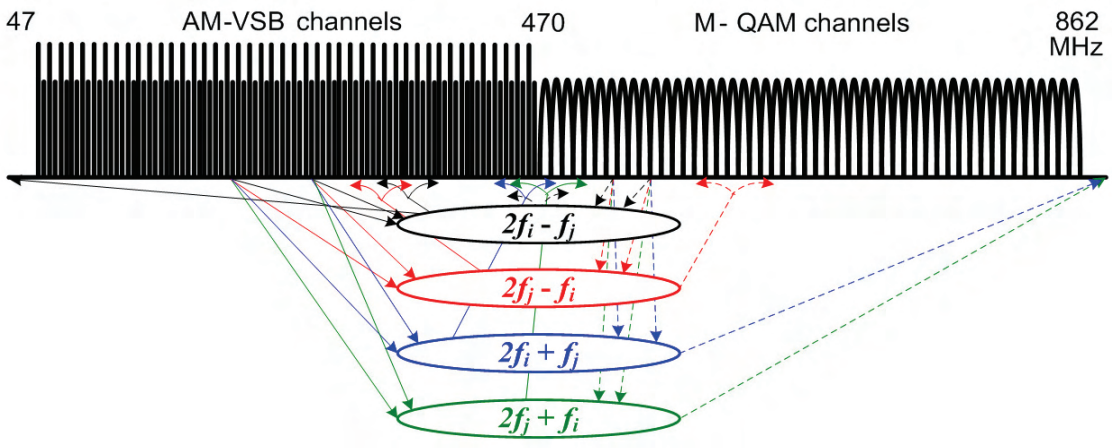


Fig.3. Influence of two-component beat

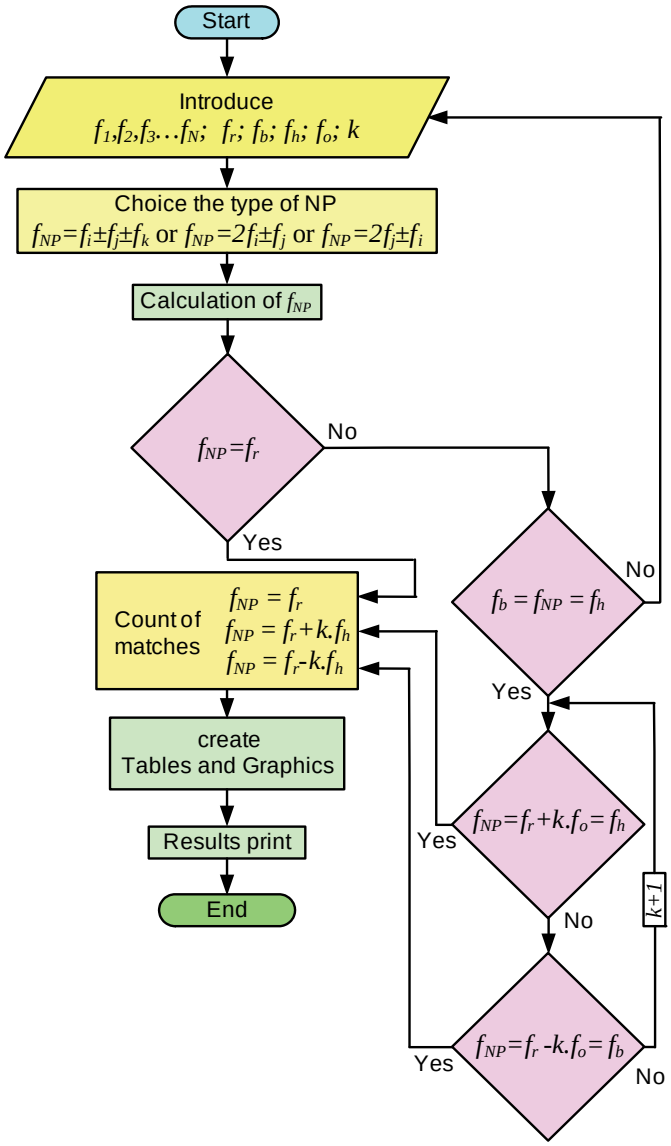


Fig.4. Block diagram of the algorithm

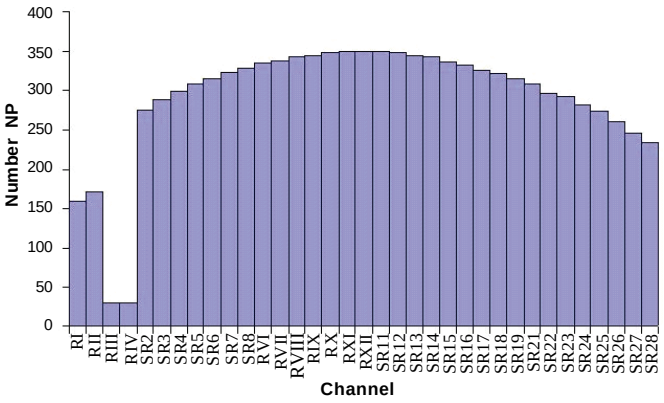


Fig.5. Distribution of third order beats for a 35 channel system

TABLE 2

No. of Channels	Channels	Central Channel	Max. No. on central channel	
			$2f_i\pm f_j$	$f_i\pm f_j\pm f_k$
3	RX-RXII	RXI	0	1
4	RIX-RXII	RXI	1	2
5	RVIII- RXII	RX	2	4
6	RVII- RXII	RX	2	7
7	RVI- RXII	RIX	2	11
8	SR8- RXII	RIX	3	15
9	SR7- RXII	RVIII	4	20
10	SR6- RXII	RVIII	4	26
11	SR5- RXII	RVII	5	33
12	SR4- RXII	RVII	5	40
12 STD	RI- RXII	RIX	2	19
13	SR3- RXII	RVI	6	47
14	SR2- RXII	RVI	6	56
15	SR1- RXII	RVI	7	65
16	RV- RXII	SR8	7	77
17	RV- RXI	SR8	8	88
18	RV-SR12	SR8	8	100
19	RV-SR13	SR8	9	112
20	RV-SR14	RVI	9	125
21	RV-SR15	RVI	10	139
22	RV-SR16	RVII	10	157
23	RV-SR17	RVII	11	170
24	RV-SR18	RVIII	11	187
25	RV-SR19	RVIII	11	204
26	RV-SR21	RIX	12	204
27	RIV-SR21	RIX	12	206
28	RIII-SR21	RVIII	12	212
29	RII-SR21	RVIII	12	219
30	RI-SR21	RVII	13	226

IV. CONCLUSION

The presented mathematical analysis and algorithm for determining the number of nonlinear products of composite triple beat, make it possible to explore these nonlinear products not only by the D/K standard. Furthermore, the distribution of channels might be different than presented above. Analog and digital channels can be carried through the whole range of 47MHz to 862MHz in the desired order and number. This flexibility allows a frequency planning of HFC/CATV system with a minimum number of intermodulation products and provides a quality and reliable transmitting of the signals.

REFERENCES

[1] Y.-I. Kim, J. H. Kim, S. Lee, D. H. Woo, S. H. Kim, and T.-H. Yoon, "Broad-band all-optical flip-flop based on optical bistability in an integrated SOA/DFB-SOA", IEEE Photon. Technol. Lett., vol. 16, no. 2, pp. 398–400, Feb. 2004.

[2] L.T. Jordanova, V. I. Topchiev, "Reduction of the Signal Nonlinear Distortion in CATV Systems applying Dual Mach-Zehnder Modulators', J. Opt. Commun. 30, pp.74-79, 2009.

[3] "Some Notes on Composite Second and Third Order Intermodulation Distortions" (2005) http://www.matrixtest.com/literat/MTN108.pdf

[4] O. B. Panagiev, "Determinating the amplitudes of intermodulation products of higher order by means Volterra kernels", ICEST, Proc. of Papers, vol.1, Bitola, pp. 215-216, 16-19 June 2004.

Radio Coverage Planning with Small-Scale Fading

Dimitar G. Valchev

***Abstract* – This paper describes an aspect not traditionally accounted for in the radio coverage planning methodology. It is shown that when the radio coverage is planned to ensure enough power, there are points in vicinity of the user's receiver that have essentially no power due to the multipath effects. Thus, the multipath channel modelling should play a significant role in the radio coverage planning process, where not only the landscape properties are taken into account, but also the spatial properties of the indoor propagation environment surrounding the antenna of the user's receiver.**

***Keywords* – Fading channels, Propagation, Scattering.**

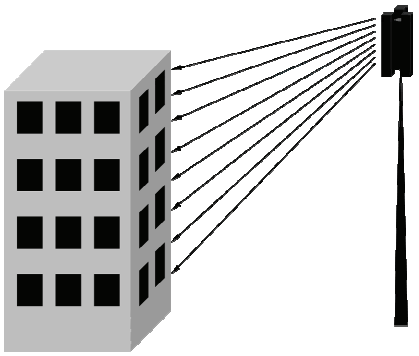
I. INTRODUCTION

Planning of the radio coverage is an important part of the wireless network design in wireless local area networks (WLANs) as well as in cellular networks. It ensures availability of radio signals to the terminal equipment within the area or volume covered by the corresponding access point or base station. Without loss of generality, the base station and the access point will be united in this paper in the common concept of base station (BS). The planning process ensures enough radio power in any point within the covered by the BS volume.

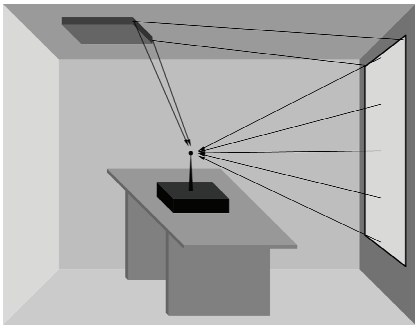
After the radio signal leaves the BS, power attenuation is in effect. It is caused mainly by three factors: path loss, shadowing and multipath fading [1]. The path loss determines the natural attenuation of the radio signal with distance from the transmitter. The dependence of the radio power on the distance is inversely proportional with some power factor depending on the properties of the propagation environment [1, 2]. The shadowing effect is caused by the presence of randomly placed obstacles within the volume covered by the transmitter of the BS. The shadowing effect causes power variation over distances comparable to the size of the surrounding blocking and reflecting objects.

The combined effect of path loss and shadowing forms the basis of the ray tracing models. The ray tracing models are particularly suited for computer aided radio coverage planning in various sites, by using geographical information systems (GIS) showing the corresponding landscape and buildings [2].

What is not accounted for in the ray tracing models forming the basis of the traditional radio coverage planning process, is the effect of multipath or small-scale fading. This effect is caused by multiple reflections and scattering of the radio signal in the vicinity of the receiving antenna, causing a standing wave interference pattern within the local volume of the user's receiver. Fig. 1 shows how different propagating radio waves can superimpose on the receiving antenna to create constructive and destructive interference



(a) Outdoor propagation



(b) Indoor propagation

Fig. 1. Radio signal propagation from the BS to the user's receiver

Even when the radio coverage is planned to deliver enough radio power to the receiver, the multipath fading itself might cause total absence of radio signal in particular points within the local volume! While the outdoor propagation is planned for a particular power level for the indoor users, the indoor propagation causes multipath by reflections of the radio signal. This paper stresses the importance of considering the multipath fading in the entire process of coverage planning.

The rest of the paper is organized as follows. Section II describes the multipath mitigation measures. Section III defines multipath shape factors, characterizing the local volume around the receiver. Upon them the fading signal coherence distance is defined. Section IV concludes the paper.

II. MULTIPATH FADING MITIGATION

The effects of multipath fading can be mitigated generally using a spatial diversity scheme in which multiple receiving antennas are combined. Two receiving antennas should be separated at least by the coherence distance of the fading signal for achieving maximum diversity gain [2]. If the multipath is concentrated around a relatively narrow beam, the coherence distance can be too large. Depending on the

Dimitar G. Valchev is with the Department of Radioengineering, Technical University of Varna, Varna 9010, Bulgaria, E-mail: D.Valchev@tu-varna.bg

This work is supported by the state budget of Technical University of Varna, research project # NP4/2011

multipath angular power distribution (APD) around the user's receiver $p(\theta,\phi)$, with θ being the azimuth and ϕ being the elevation, the coherence distance has angular dependence [3]. In the spatial diversity system the multiple antennas should be placed such that the total occupied space is minimized and/or the diversity gain is maximized. Therefore, it is crucial for the system designer to get the APD around the receiver and based on it, to derive expressions that determine the angular dependence of the fading signal coherence distance within the particular indoor environment. The indoor environment itself determines the APD through the presence of scattering and reflecting objects. The coherence distance in different directions within the local volume is derived in terms of the so called multipath shape factors [3] that relate the APD around the receiver to the second-order fading statistics of the received radio signal. The next section will give approximate expression for the coherence distance that can be used in system design for the purposes of multipath fading mitigation.

III. Multipath Shape Factors

In [3] the concept of three-dimensional multipath shape factors is developed to derive the important fading statistics in non-omnidirectional wireless channels. The definitions of the shape factors are based on the *l*-th degree, *m*-th order spherical harmonic coefficients S_l^m of the APD $p(\theta,\phi)$ as follows:

Angular Spread, ranging from 0 to 1:

$$\mathrm{Y} = \sqrt{1 - \frac{S_1^{0^2} + \left|S_1^1\right|^2}{S_0^{0^2}}}\,, \tag{1}$$

Elevational Constriction, ranging from –0.5 to 1:

$$\xi = \frac{1.5 S_2^0 S_0^0 - \left(S_1^{0^2} - 0.5\left|S_1^1\right|^2\right)}{S_0^{0^2} - \left(S_1^{0^2} + \left|S_1^1\right|^2\right)}, \tag{2}$$

45°-*Inclined Constriction*, ranging from 0 to 1:

$$\chi = \frac{2\left|S_2^1 S_0^0 - S_1^0 S_1^1\right|}{S_0^{0^2} - \left(S_1^{0^2} + \left|S_1^1\right|^2\right)}, \tag{3}$$

Azimuthal Constriction, ranging from 0 to 1:

$$\chi = \frac{\left|S_2^2 S_0^0 - S_1^{1^2}\right|}{S_0^{0^2} - \left(S_1^{0^2} + \left|S_1^1\right|^2\right)}, \tag{4}$$

Azimuthal Direction of Maximum Fading at 45° *Elevation*:

$$\theta_{\phi 45}^{\max} = \arg\left\{S_2^1 S_0^0 - S_1^0 S_1^1\right\}, \tag{5}$$

Azimuthal Direction of Maximum Fading at Zero Elevation:

$$\theta_{\phi 0}^{\max} = \arg\left\{S_2^2 S_0^0 - S_1^{1^2}\right\}. \tag{6}$$

The spherical harmonic coefficients are calculated by the following integrations [4]:

$$S_0^0 = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} p(\theta,\phi)\cos\phi\, d\phi\, d\theta\,, \tag{7}$$

$$S_1^0 = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} p(\theta,\phi)\sin\phi\cos\phi\, d\phi\, d\theta\,, \tag{8}$$

$$S_1^1 = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} p(\theta,\phi)\cos\phi e^{j\theta}\cos\phi\, d\phi\, d\theta\,, \tag{9}$$

$$S_2^0 = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} p(\theta,\phi)\left(\sin^2\phi - 1/3\right)\cos\phi\, d\phi\, d\theta\,, \tag{10}$$

$$S_2^1 = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} p(\theta,\phi)\cos\phi\sin\phi e^{j\theta}\cos\phi\, d\phi\, d\theta\,, \tag{11}$$

$$S_2^2 = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} p(\theta,\phi)\cos^2\phi e^{j2\phi}\cos\phi\, d\phi\, d\theta\,, \tag{12}$$

The coherence distance by which the multiple antennas should be separated in the spatial diversity scheme is given in terms of the multipath shape factors as [3]:

$$d_c = \frac{\lambda}{\mathrm{Y}}\left\{15.33\left[1 + 1.5\left(\xi\left(2\sin^2\phi - 0.67\right) + \chi\sin 2\phi\cos\left(\theta - \theta_{\phi 45}^{\max}\right) + \zeta\cos^2\phi\cos 2\left(\theta - \theta_{\phi 0}^{\max}\right)\right)\right]\right\}^{-\frac{1}{2}}. \tag{13}$$

It is seen that the coherence distance is a function of the azimuth and elevation and therefore has a minimum at a particular direction. This is the direction at which the multiple elements of the receiver diversity scheme should be aligned in order to achieve minimum occupied space and/or maximum diversity gain for the received signal.

IV. Conclusion

This paper augments the traditional radio coverage planning methodology by introducing the small-scale fading into the designing process. It is shown that even when the radio coverage is planned for a sufficient power level, there are points within the local volume in vicinity of receiver antenna where the signal is absent. Therefore, a spatial diversity scheme at the receiver is needed in order to ensure a reliable communication link. The paper shows the importance of analyzing the environment around the user's receiver in order to optimally plan the radio coverage.

References

[1] A. J. Goldsmith, *Wireless Communications*, Cambridge, 2005.

[2] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 2 ed, 2001.

[3] D. G. Valchev and D. Brady, "Three-dimensional multipath shape factors for spatial modeling of wireless channels", *IEEE Trans. Wireless Commun.*, vol.8, no.11, November 2009, pp. 5542–5551.

[4] W. W. Bell, *Special Functions for Scientists and Engineers*, Dover, 2004.

AUTHOR INDEX