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

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

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

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

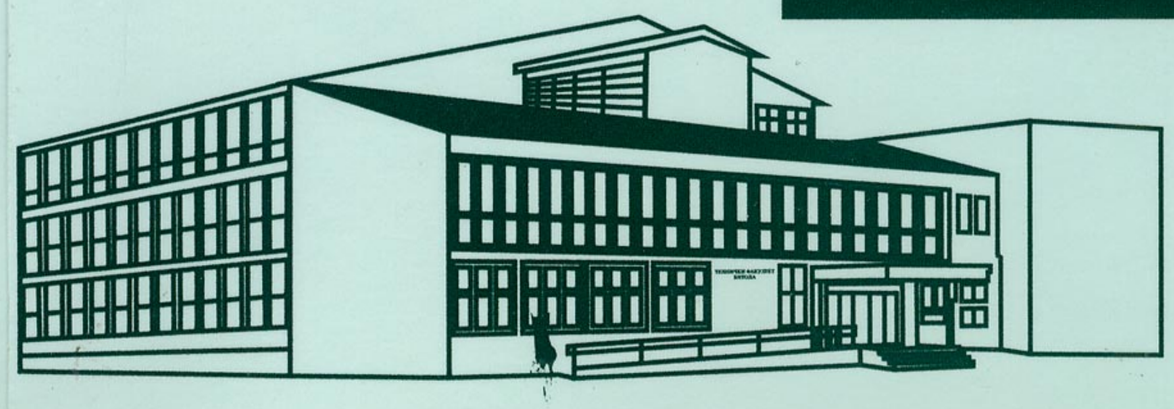
IEEE

ICEST

2007

ТЕХНИЧКИ-ФАКУЛТЕТ
БИТОЛА

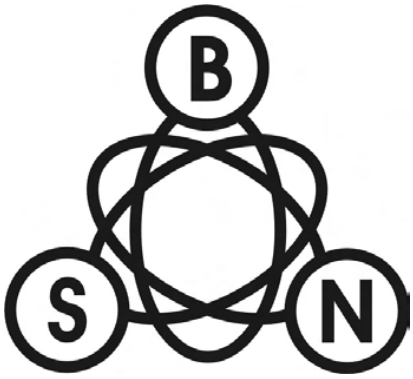
Proceedings of Papers
Volume 1



24 - 27 June, 2007

Hotel Granit, Ohrid, Macedonia

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



Proceedings of Papers

Volume 1

Bitola, 2007

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

-ICEST 2007-

Procedings of Papers – Volume 1 of 2 volumes

Editor: Prof. Dr. Cvetko Mitrovski

Technical Editor: Pargovski Jove

Published by: Faculty of Technical Sciences - Bitola

Printed by: Mikena - Bitola

Total print run: 200

ISBN: 9989 - 786 - 06 - 2

CIP - Каталогизација во публикација
Матична и универзитетска библиотека
"Св. Климент Охридски", Битола

621 . 391 : 004 (063) (082)
007 . 52 (063) (082)
004 . 738 . 52 : 004 . 89 (063) (082)
621 . 397 . 3 : 004 (063) (082)
621 . 396 . 67 (063) (082)

INTERNATIONAL Scientific Conference on Information, Communication and Energy Systems and Technologies (42 ; 2007 ; Bitola)
ICEST 2007: proceedings of papers / XLII International Scientific Conference on Information, Communication and Energy Systems and Technologies ; [Editor Cvetko Mitrovski]. - Bitola : Faculty of Technical Sciences, 2007. - 2 св. (XVI , 430 ; XVI, стр. 431 - 860) : илустр. ; 28 см

Фусноти кон текстот. - Библиографија кон трудовите. - Регистри

ISBN 9989 - 786 - 06 - 2
1. Gl. stv. nasl. 2. Mitrovski, Cvetko
а) Електроенергетски системи - Мониторинг - Анализи - Зборници б) Телекомуникациска техника - Сигнали - Анализи - Зборници в) Автоматски системи - Интелигентни системи - Зборници г) Компјутерски системи - Системи за менаџмент на бази на податоци - Интернет технологии - Зборници - д) Индустриска електроника - Контролни системи - Конвертери - Зборници
COBISS.MK - ID 18020161

WELCOME TO ICEST 2007

Dear Colleagues,

On behalf of the Technical Program Committee it is my pleasure to invite you to attend the XLII International Scientific Conference on Information, Communication and Energy Systems and Technology **ICEST-2007**, which will be held from 24 to 27 June at the Hotel "Granit" in Ohrid.

The ICEST Conference gained its acronym as a successor of EIST Conference in 2002, when the Faculty of Electronics from Niš, Serbia, joined the Faculty of Communication and Communication Technologies from Sofia, Bulgaria and the Faculty of Technical Sciences from Bitola, Macedonia, in organizing the series of Conferences.

As to the previous five ICEST Conferences, we have ever increasing number of papers and authors from institutions all over the world. This year we have 193 authors (and coauthors) from 15 countries with more than 250 scientific and application oriented papers out of which the Conference International Reviewer Board selected 216 papers for oral and poster presentations. Also, in the frame of the Conference there will be two invited papers from Prof. Gerard C.M. Meijer (TU-Delft, Netherlands) and from Prof. J. Gajda (AGH- Krakow, Poland).

In spite of the tight Conference schedule I hope that there will be enough time for social activities which will enable further strengthening of the professional and personal relations among the ICEST participants for the benefit of the institutions and the countries they are coming from.

At last, in the name of the Conference Technical and Program Committee I want to express my gratitude to the authors and reviewers who have contributed to maintain the high standards for ICEST, and to all others who were involved in organization of the Conference.

I am sure that the beautiful and picturesque Ohrid will give you all, an inspiration for fruitful work and pleasant time during the ICEST Conference.

I am looking forward for seeing you in Ohrid.

Prof. D-r Cvetko Mitrovski
Conference Chairman

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

ICEST 2007

Organized by

Faculty of Technical Sciences, Bitola, Macedonia

Faculty of Electronic Engineering, Niš,Serbia

Faculty of Communications and Communication Technologies, Sofia, Bulgaria

Under auspices of

Macedonian Ministry of Education and Science

In cooperation with:

Macedonian IEEE section

TECHNICAL PROGRAM COMMITTEE

Chairman:

Mitrovski C.	University of Bitola, Macedonia

Vice Chairmen:

Milovanović B.	University of Niš, Serbia
Dimitrov D.	Technical University of Sofia, Bulgaria

Members:

Acimović V.	University of Belgrade, Serbia
Altimirski E.	Technical University of Sofia, Bulgaria
Antić D.	University of Niš, Serbia
Arnaudov R.	Technical University of Sofia, Bulgaria
Arsov Lj.	University of Skopje, Macedonia
Bekjarski A.	Technical University of Sofia, Bulgaria
Bogdanov M.	University of Skopje, Macedonia
Deskovski S.	University of Bitola, Macedonia
Dimcev V.	University of Skopje, Macedonia
Dimitrijević B.	University of Niš, Serbia
Dobrev D.	Technical University of Sofia, Bulgaria
Gavrovski C.	University of Skopje, Macedonia
Gušev M.	University of Skopje, Macedonia
Kuncev R.	Technical University of Sofia, Bulgaria
Lutovac M.	University of Belgrade, Serbia
Manevska V.	University of Bitola, Macedonia
Markovski A.	University of Bitola, Macedonia
Milutinović V.	University of Belgrade, Serbia
Petrović Z.	University of Belgrade, Serbia
Popova A.	Technical University of Sofia, Bulgaria
Prancov R.	Technical University of Sofia, Bulgaria
Ranđić S.	University of Kragujevac, Serbia
Stojanov G.	Technical University of Sofia, Bulgaria
Stojcev M.	University of Niš, Serbia
Tasić D.	University of Niš, Serbia
Trpezanovski Lj.	University of Bitola, Macedonia

ORGANIZING COMMITTEE

Acevski N.	University of Bitola, Macedonia
Atanasovski M.	University of Bitola, Macedonia
Ceselkoska V.	University of Bitola, Macedonia
Docev I.	Technical University of Sofia, Bulgaria
Goleva R.	Technical University of Sofia, Bulgaria
Iliev G.	Technical University of Sofia, Bulgaria
Jolevski I.	University of Bitola, Macedonia
Kostov M.	University of Bitola, Macedonia
Lubih L.	Technical University of Sofia, Bulgaria
Nedelkovski I.	University of Bitola, Macedonia
Perić Z.	University of Niš, Serbia
Petkovski M.	University of Bitola, Macedonia
Radevska P. M.	University of Bitola, Macedonia
Sarevska M.	University of Niš, Serbia
Stanković Z.	University of Niš, Serbia
Stevanovski B.	University of Bitola, Macedonia

CORRESPONDENCE

Conference Chairman:
Prof. Dr. Mitrovski Cvetko
Faculty of Technical Sciences
Ivo Lola Ribar bb, 7000 Bitola, Macedonia
phone: +389 47 207 711
e-mail: cvetko.mitrovski@uklo.edu.mk

Organizing Committee Chairman:
Prof. Dr. Mitrevski Pece
Faculty of Technical Sciences
Ivo Lola Ribar bb, 7000 Bitola, Macedonia
phone: +389 47 207 149
fax: +389 47 203 370
e-mail: pece.mitrevski@uklo.edu.mk
e-mail: icest2007@uklo.edu.mk

Technical Secretary:
Assist. Pargovski Jove
Faculty of Technical Sciences
Ivo Lola Ribar bb, 7000 Bitola, Macedonia
phone: +389 47 207 711
fax: +389 47 203 370
e-mail: jove.pargovski@uklo.edu.mk
e-mail: icest2007@uklo.edu.mk

TECHNICAL SUPPORT

Pargovski J. University of Bitola, Macedonia
Kostov M. University of Bitola, Macedonia

CONFERENCE INTERNET SITE

http://icest2007@uklo.edu.mk

LIST OF ICEST 2007 REVIEWERS

Aleksandar Atanaskovic, Prof. Dr.
University of Niš, Serbia

Aleksandar Risteski, Prof. Dr.
University of Skopje, Macedonia

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

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

Boris Tsankov, Prof. Dr.
Technical University of Sofia, Bulgaria

Branislav Petrovic, Prof. Dr
University of Niš, Serbia

Cvetan Gavrovski, Prof. Dr.
University of Skopje, Macedonia

Danilo Stojanović, Prof. Dr.
Technical Faculty, Čačak, Serbia

Dejan Milić, Mr.
University of Niš, Serbia

Dejan Rančić, Prof. Dr.
University of Niš, Serbia

Dilyana Gospodinova, Prof. Dr.
Technical University of Sofia, Bulgaria

Dimitar Taskovski, Dr.
University of Skopje, Macedonia

Dimitar Todorov, Prof. Dr.
Technical University of Sofia, Bulgaria

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

Dragan Stojanović, Doc. Dr.
University of Niš, Serbia

Dragan Tasić, Prof. Dr.
University of Niš, Serbia

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

Emil Manolov, Prof. Dr.
Technical University of Sofia, Bulgaria

Emil Ratz, Prof. Dr.
Technical University of Sofia, Bulgaria

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

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

Goce Shutinoski, Prof. Dr.
University of Skopje, Macedonia

Goce L. Arsov, Prof. Dr.
University of Skopje, Macedonia

Goran Jovanović, Prof. Dr.
University of Niš, Serbia

Goran Lj. Djordjevic, Prof. Dr.
University of Niš, Serbia

Ilia G. Iliev, Prof. Dr.
Technical University of Sofia, Bulgaria

Ivailo Pandiev, Prof. Dr.
Technical University of Sofia, Bulgaria

Ivan Petković, Prof. Dr.
University of Niš, Serbia

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

Josif Kosev, Prof. Dr.
University of Skopje, Macedonia

Jovan Stefanovski, Dr.
JP Strezevo Bitola

Jugoslav Jokovic, Prof. Dr.
University of Niš, Serbia

Leonid Stoimenov, Doc. Dr.
University of Niš, Serbia

Liljana Gavrilovska, Prof. Dr.
University of Skopje, Macedonia

Liljana Zivković, Prof . Dr.
University of Niš, Serbia

Ljuben Janev, Prof. Dr.
University of Skopje, Macedonia

Ljupco Panovski, Prof. Dr.
University of Skopje, Macedonia

Marija Cundeva-Blajer, Doc. Dr.
University of Skopje, Macedonia

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

Mario Makraduli, Mr.
University of Skopje, Macedonia

Marko Porjazoski, Mr.
University of Skopje, Macedonia

Metodija Kamilovski, Prof. Dr.
University of Skopje, Macedonia

Mihail Antchev, Prof. Dr.
Technical University of Sofia, Bulgaria

Mile Stojcev, Prof. Dr.
University of Niš, Serbia

Miodrag Stojanović, Mr.
University of Niš, Serbia

Mirko Kosanovic, Prof. Dr.
High Technical School of Niš, Serbia

Momcilo Bogdanov, Prof. Dr.
University of Skopje, Macedonia

Natasa Males-Ilić, Dr.
University of Niš, Serbia

Nebojsa Doncov, Dr.
University of Niš, Serbia

Nebojsa Janković, Prof . Dr.
University of Niš, Serbia

Ognian Boumbarov, Prof. Dr.
Technical University of Sofia, Bulgaria

Oleg Panagiev, Dr.
Technical University of Sofia, Bulgaria

Olivera Pronic-Rancic, Dr.
University of Niš, Serbia

Pero Latkoski, Mr.
University of Skopje, Macedonia

Peter Goranov, Prof. Dr.
Technical University of Sofia, Bulgaria

Peter Petkov, Dr.
Technical University of Sofia, Bulgaria

Radojka Krneta, Prof. Dr.
Technical Faculty, Čačak, Serbia

Raina Tzeneva, Prof. Dr.
Technical University of Sofia, Bulgaria

Rosen Miletiev, Prof. Dr.
Technical University of Sofia, Bulgaria

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

Sasa Nikolić, Doc. Dr.
University of Niš, Serbia

Saša Stojković, Prof. Dr.
Technical Faculty, Čačak, Serbia

Siniša Ranđić, Prof. Dr.
Technical Faculty, Čačak, Serbia

Snejana Pleshkova – Bekiarska, Prof. Dr.
Technical University of Sofia, Bulgaria

Sofija Bogdanova, Prof. Dr.
University of Skopje, Macedonia

Svetlana Tzvetkova, Prof. Dr.
Technical University of Sofia, Bulgaria

Tania Vasileva, Prof. Dr.
Technical University of Sofia, Bulgaria

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

Tomislav Dzhekov, Prof. Dr.
University of Skopje, Macedonia

Valentin H. Videkov, Prof. Dr.
Technical University of Sofia, Bulgaria

Vera Markovic, Prof. Dr..
University of Niš, Serbia

Vladica Sark, Mr.
University of Skopje, Macedonia

Vladimir Atanasovski, Mr.
University of Skopje, Macedonia

Vladimir Dimcev, Doc. Dr.
University of Skopje, Macedonia

Zahari Zarkov, Prof. Dr.
Technical University of Sofia, Bulgaria

Zlatica Marinković, Mr.
University of Niš, Serbia

Zoran Hadzi-Velkov, Doc. Dr.
University of Skopje, Macedonia

Zoran Ivanovski, Prof. Dr.
University of Skopje, Macedonia

Zoran Stanković, Mr
University of Niš, Serbia

TABLE OF CONTENTS

VOLUME 1

PLENARY SESSION

PL.1 Smart Capacitive-Resistive Sensors 3
Gerard C.M. Meijer, Xiujun Li, Zu-yao Chang and Blagoy P. Iliev
Delft University of Technology, Netherlands

PL.2 Measurements of Road Traffic Parameters 13
Janusz Gajda1 and Ryszard Sroka1
University of Science and Technology, Krakow, Poland

TELECOMMUNICATIONS SYSTEMS AND TECHNOLOGY I

TST I.1 Performance of SOVA-based Turbo Decoders with Autostopping 23
N. T. Kostov, S. M. Yordanova, G. K. Marinova and N. A. Velikov*
Technical University of Varna, Bulgaria
*Naval Academy "N. Vaptzarov" of Varna, Bulgaria

TST I.2 The Blocking Probability Comparison for Some Routing and Wavelength Assignment Methods 27
G. Z. Marković and V. Aćimović-Raspopović
Faculty of Traffic and Transport Engineering, Belgrade

TST I.3 Turbo Codes in the CCSDS Standard for Wireless Data 31
T. B. Iliev and D. I. Radev
University of Rousse, Bulgaria

TST I.4 Construction and Simulation Analysis of Low-Density Parity-Check Codes 35
T. B. Iliev
University of Rousse, Bulgaria

TST I.5 New Intellectual Property Blocks for a Secure Modem on FPGA 39
G. I. Marinova, V. G. Guliashki* and M. Bellanger**
Technical University of Sofia
*Institute for Information Technologies – BAS, Sofia
**CNAM-Paris

TST I.6 Creation of Secrets Sharing Protocols Resistible Against Disloyal Participants 43
P. T. Antonov and V. R. Antonova
Technical University of Varna, Bulgaria

TST I.7 Comparison of Narrowband Interference Suppression Methods for OFDM Systems 47
M. Ovtcharov, V. Poulkov and G. Iliev
Technical University of Sofia, Bulgaria

TST I.8 The Linear Complexity of the LFSR Based Generalized Shrinking-Multiplexing Generator 51
T. D. Tashev, B. Y. Bedzhev and Z. N. Tasheva
University of Shumen, Bulgaria

TST I.9 Analysis of a Matching Network Effect on Noise 55
P. D. Petrova and D. P. Petrov
Technical University of Gabrovo, Bulgaria

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS I

RCMTA I.1 Design and Analysis of Wideband Antenna with An Application to Ground Penetrating System 61
G. E. Atteia, K. F. A. Hussein and A. A. Shaalan
University of Zagazig, Egypt
Electronics Research Inst. Cairo, Egypt

RCMTA I.2 Influence of the Feeding Position Over the Polarization Parameters of Patch Antennas with Different Geometric Shape 65
M. V. Stefanov
Technical University of Sofia, Bulgaria

RCMTA I.3 Novel Methods for Analysis of Smart Antennas for Wireless Communications 69
S. V. Savov and V. Y. Vassileva
Technical University of Varna, Bulgaria

RCMTA I.4 Investigation of Periodic EBG Structures with Increased Bandwidth 73
N. M. Stoyanov and Ch. P. Levchev
Ministry of Interior, Sofia, Bulgaria

RCMTA I.5 **Automated System for Antenna Parameters Measurements in Anechoic Chamber** **77**
Ch. P. Levchev and N. M. Stoyanov
Ministry of Interior, Sofia, Bulgaria
RCMTA I.6 **SDARS Antenna** **81**
P. Petkov, N. Dodov and N. Nikolova
Technical University of Sofia, Bulgaria
RCMTA I.7 **Investigation of a Wideband Foursquare Microstrip Antenna** **83**
S. R. Baev
Technical University of Sofia, Bulgaria
RCMTA I.8 **Investigation of Probe Correction Technique in Planar Near-Field Antenna Measurements** **87**
A. G. Toshev
Technical University of Sofia, Bulgaria
RCMTA I.9 **Coupled Linear Grounding Systems: Quasistationary Antenna Model** **91**
N. N. Cvetkovic and P. D. Rancic
University of Niš, Serbia

MEASUREMENT SYSTEMS

MS.1 **Metrological Support of the Distributed Power Quality Monitoring** **97**
B. R. Dimitrijević and M. M. Simić
University of Niš, Serbia
MS.2 **Power Resistors With Low Parasitic Inductance** **101**
P. Rahnev and S. Letskovska
Burgas Free University, Bulgaria
MS.3 **PC Based RLC Meter** **103**
M. Srbinovska, V. Dimcev, C. Gavrovski
and Z. Kokolanski
University of Skopje, Macedonia
MS.4 **Method for Calculating Residual Strains and Visual Diagnostics by a Specialized Software System** **107**
N. T. Kostov, S. M. Yordanova and G. K. Marinova
Technical University of Varna, Bulgaria
MS.5 **Prepar Measuring the Main Parameters of Optical Transfer Media – Optical Fibers and Cables** **111**
I. Balabanova
Technical University of Gabrovo, Bulgaria
MS.6 **Estimation of the Safety Vehicle Speed in Curve Using Inertial Sensors** **113**
E. Iv. Iontchev and R. G. Miletiev*
Higher Scool of Transport, Sofia, Bulgaria
*Technical University of Sofia, Bulgaria
MS.7 **Application of Ecological Monitoring in SMEs in Bulgaria** **117**
I. Nikolova-Jahn
Technical University of Sofia, Bulgaria
MS.8 **Uncertainties Estimation of Measurement Devices by Using Finite Element Method** **121**
M. Cundeva-Blajer and Lj. S. Arsov
University of Skopje, Macedonia
MS.9 **First Year of Wind Parameters Measurements in Macedonia** **125**
V. Dimcev, K. Najdenkoski and V. Stoilkov
University of Skopje, Macedonia

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES I

CSIT I.1 **Low-Power Multihop Routing for Wireless Sensor Networks** **131**
Z. Karakehayov
University of Southern Denmark
CSIT I.2 **Low Power Application Specific Processing Element** **135**
T. R. Nikolić, G. Lj. Djordjević and M. K. Stojčev
University of Niš, Serbia
CSIT I.3 **Throughput and Fairness in IEEE 802.16 Broadband Wireless Networks** **139**
R. A. Pachamanov and B. P. Tsankov
Technical University of Sofia, Bulgaria
CSIT I.4 **Implementation of TCP/IP Protocols in Wireless Sensor Networks** **143**
M. R. Kosanović and M. K. Stojčev*
High Technical School, Niš, Serbia
*University of Niš, Serbia
CSIT I.5 **Simulation Modelling of Overflow Probabilities in Tandem Queuing System** **147**
D. Radev, E. Rashkova and D. Stankovski
University of Rousse, Bulgaria

CSIT I.6 **The Effect of Load/Store Queue Size on System Performance** **151**
D. Curcievska, P. Mitrevski and M. Gusev*
University of Bitola, Macedonia
*University of Skopje, Macedonia
CSIT I.7 **Improving Mobile Complex Service Composition with Graph Technology** **155**
Lj. Antovski, M. Gusev and P. Mitrevski*
University of Skopje, Macedonia
* University of Bitola, Macedonia

TELECOMMUNICATIONS SYSTEMS AND TECHNOLOGY II

TST II.1 **Routing Simulation for Wireless Sensor Networks** **161**
Z. Monov and L. Karakehayova
Technical University of Sofia, Bulgaria
TST II.2 **Overloaded IP Node With Implemented Priority Queuing QoS Mechanism** **165**
I. Efnusev* and T. Janevski
*T-Mobile AD. Skopje, Macedonia
University of Skopje, Macedonia
TST II.3 **Planning and Pre-Operational Performance Optimization of WCDMA for Multimedia Wireless Networks** **169**
T. Janevski and V. Nikolic*
University of Skopje, Macedonia
*AD Makedonski Telekomunikacii, Skopje
TST II.4 **Signaling on the Most Critical Interfaces in GSM** **173**
Z. Bungurova-Kaziovska and T. Janevski
University of Skopje, Macedonia
TST II.5 **An Implementation of Adaptable "Bottom-up" Model for Calculating Interconnection Costs** **177**
A. Kostić-Ljubisavljević, V. Radojičić and
V. Aćimović-Raspopović
University of Belgrade, Serbia
TST II.6 **TCP Congestion Control and Asymmetric Networks** **181**
G. V. Hristov and M. P. Iliev
University of Rousse, Bulgaria
TST II.7 **Probability Density Function of M-ary FSK Signal in the Presence of Impulse Noise and Nakagami Fading** **185**
M. Stefanović, D. Krstić, N. Kapacinović, E. Mekić
and F. Dizdarević
University of Niš, Serbia
TST II.8 **Performances of Receiver with MRC, SC and SSC Combiner in the Presence of Fading and Gaussian Noise** **189**
M. Stefanović, J. Anastasov, B. Palović,
I. Temelkovski and V. Veličković
University of Niš, Serbia
TST II.9 **Evaluation of a Congestion Control Transport Protocol for Real Time Applications in NGN** **193**
S I. Statev and S. T. Mirtchev
Technical University of Sofia, Bulgaria

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS II

RCMTA II.1 **Resonance and Dispersion Characteristics of Microstrip Slow-Wave Open-Loop Resonator** **199**
M. V. Nedelchev and I. G. Iliev
Technical University of Sofia, Bulgaria
RCMTA II.2 **Relative Influence of Some Stochastic Factors on Bit-Error Rate of Ground-to-Ground Free Space Optics** **203**
B. G. Bonev
Technical University of Sofia, Bulgaria
RCMTA II.3 **Feed Type Influence Over the Bandwidth of Microstrip Patch Antennas** **207**
I. N. Stefanova
Technical University of Sofia, Bulgaria
RCMTA II.4 **Modeling of Multiconductor Shielded Microstrip Lines** **211**
S. M. Musa and M. N. O. Sadiku
Prairie View A&M University, Houston, USA
RCMTA II.5 **Substrate's Parameters Influence over the Directivity and Bandwidth at Microstrip Planar Arrays** **215**
N. I. Dodov and M. O. Krastev
Technical University of Sofia, Bulgaria

RCMTA II.6 **Modelling of Losses in a Microwave Applicator Using TLM Method** **219**
J. Joković, B. Milovanović and T. Dimitrijević
University of Niš, Serbia

RCMTA II.7 **Improved Algorithm for Time-Domain Electromagnetic Field Calculation at Points outside the TLM Workspace** **223**
N. Dončov and B. Milovanović
University of Niš, Serbia

RCMTA II.8 **Strip Lines on the Cylindrical Dielectric Substrate** **227**
Ž. J. Mančić, S. S. Ilić and S. R. Aleksić
University of Niš, Serbia

RCMTA II.9 **The Linearization of Multichannel Amplifiers Connected in Cascade with Enhanced Efficiency** **231**
N. Males-Ilić, A. Atanasković and B. Milovanović
University of Niš, Serbia

RCMTA II.10 **Software for Microwave Slotted Patch Antenna Design Based on Hybrid Empirical-Neural Model** **235**
B. Milovanović, Z. Stanković, M. Milijić and M. Sarevska
University of Niš, Serbia

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES II

CSIT II.1 **Simulation of Next Generation Network Applications** **241**
I. I. Atanasov
Technical University of Sofia, Bulgaria

CSIT II.2 **Pricing the Internet Services Using Extensive Game Solution** **245**
V. Radonjić and V. Aćimović-Raspopović
University of Belgrade, Serbia

CSIT II.3 **A Hybrid Population Based Method Solving Convex Integer Optimization Problems** **249**
V. G. Guliashki
Institute for Information Technologies – BAS, Sofia

CSIT II.4 **Technical Aspects of Interconnect Management of Multimedia IP-based Networks** **253**
A. Tsenov and T. Nikolov
Technical University of Sofia, Bulgaria

CSIT II.5 **Computer System for Deductive Issue from Natural Language Text through Interpretation in Semantic Resolution** **257**
I. Raeva, R. Raev and D. Stankovski
University of Rousse, Bulgaria

CSIT II.6 **Touch Interface Framework for GIS based C4I Systems** **261**
A. Milosavljević, D. Rančić, A. Dimitrijević and
V. Mihajlović
University of Niš, Serbia

CSIT II.7 **Full-text Search Over WebDAV Repository** **265**
V. M. Ognjanović and M. Lj. Gocić
Accordia Group, Niš, Serbia

CSIT II.8 **Models of Interconnect Management of Multimedia IPbased Networks** **269**
T. Ap. Nikolov
Technical University of Sofia, Bulgaria

TELECOMMUNICATIONS SYSTEMS AND TECHNOLOGY III

TST III.1 **Pulse Shape Influence on Optical System Performance in the Presence of Interference, Dispersive and Nonlinear** **275**
M. Č. Stefanović, D. Lj. Drača, A. S. Panajotović and D. M. Milović
University of Niš, Serbia

TST III.2 **Fiber Nonlinearity Limitations in WDM CATV Systems** **279**
L. T. Jordanova and D. M. Dobrev
Technical University of Sofia, Bulgaria

TST III.3 **Analysis of Blocking and Clipping for TDM in Satellite Communication Systems** **283**
S. M. Musa and N. F. Mir*
Prairie View A&M University, Texas, USA
*San Jose State University, CA, USA

TST III.4 **Influence of Optical Fiber Nonlinear Effects in HFC Television Networks with WDM Multiplexing** **287**
K. Angelov, K. Koitchev and S. Sadinov
Technical University of Gabrovo, Bulgaria

TST III.5 **Adaptive Packet Selection Algorithm for Bluetooth Data Packets** **291**
R. Sokullu and E. Karatepe
Ege University, Izmir, Turkey

TST III.6 **The Analysis of Rician PDF Integral Properties from the DSRC System Viewpoint** **295**
A. Mitić, D. Blagojević, D. Stefanović and V. Stanimir
University of Niš, Serbia

TST III.7 **Some Integral Properties of Nakagami-*m* distribution** **299**
H. Popović*, A. Mitić, I. Stefanović and D. Stefanović
*High School of Electrical Engineering, Beograd, Serbia
University of Niš, Serbia

TST III.8 **LAN Traffic Shaping Analyses in Metropolitan Area Networks** **303**
R. Iv. Goleva, D. K. Atamian, M. At. Goleva*,
and K. At. Golev**
Technical University of Sofia, Bulgaria
*University of Bremen, Germany
**Bianor Ltd., Bulgaria

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS III

RCMTA III.1 **Generating of Basic Wave Digital Elements for Modeling of Two-dimensional Planar Structures** **309**
B. P. Stošić and M. V. Gmitrović
University of Niš, Serbia

RCMTA III.2 **Analysis of the Class E Amplifier with Load Variation** **313**
I. V. Nedelchev and I. N. Nemigenchev
Technical University of Gabrovo, Bulgaria

RCMTA III.3 **Polar Diagram of Multiple Driver Loudspeaker Systems** **317**
E. S. Sirakov
Technical University of Varna, Bulgaria

RCMTA III.4 **Nonlinear Systems Modeling in Broadband Communications** **321**
O. B. Panagiev
Technical University of Sofia, Bulgaria

RCMTA III.5 **Intermodulation Composite Distortions Theoretically Research in HFC Networks** **325**
O. B. Panagiev
Technical University of Sofia, Bulgaria

RCMTA III.6 **Influence of the Refraction in Presence of Various Formed Obstacles** **329**
E. S. Altimirski and K. N. Angelov
Technical University of Sofia, Bulgaria

RCMTA III.7 **GPS/INS Automotive Tracking System** **333**
R. Miletiev and R. Arnaudov
Technical University of Sofia, Bulgaria

RCMTA III.8 **New Experiences in Simple Numerical Modelling of Sommerfeld's Integral Kernel** **337**
M. P. Rančić and P. D. Rančić
University of Niš, Serbia

RCMTA III.9 **Summing Signals of Push-Pull Broadband Amplifier Modules** **341**
B. D. Karapenev and I. N. Nemigenchev
Technical University of Gabrovo, Bulgaria

SIGNAL PROCESSING I

SP I.1 **Adaptive Notch Filters for Image Processing** **347**
G. L. Iliev
Technical University of Sofia, Bulgaria

SP I.2 **Efficient Compression for Contour Images** **351**
R. Kountchev*, V. Todorov, R. Kountcheva
*Technical University of Sofia, Bulgaria
T&K Engineering, Sofia, Bulgaria

SP I.3 **Using Associative Memories as a Mean for Image Accumulating** **355**
A. B. Bekiarski
Technical University of Sofia, Bulgaria

SP I.4 **Design of Huffman Decoder FPGA Core** **359**
I. Mezei and R. Struharik
Faculty of Technical Sciences, Novi Sad, Serbia

SP I.5 **An Algorithm for Accelerated Template Localization in an Image Using Correlation Coefficient** **361**
J. P. Petkova, V. J. Smarkov and M. N. Karova
Technical University of Varna, Bulgaria

SP I.6 **Processing of Sequences of Contour Images of Sign- Language Interpreter** **365**
R. Kountchev*, V. Todorov, R. Kountcheva
*Technical University of Sofia, Bulgaria
T&K Engineering, Sofia, Bulgaria

SP I.7 **Ordered Complex Hadamard Transform of Images** **369**
R. P. Mironov, R. K. Kountchev and V. S. Mirchev
Technical University of Sofia, Bulgaria
SP I.8 **Image Sculpture – From 2D Images to 3D Objects** **373**
M. Dontschewa and T. Sturn
UCT- Research, FHV, Dornbirn, Austria
SP I.9 **Analysis of the Realism of 3D Computer Animations in the Context of The Human Visual Perception** **377**
M. Dontschewa
UCT- Research, FHV, Dornbirn, Austria
SP I.10 **Analysis and Strategies for Fighting Against Attacks on Watermark Systems** **381**
J. Spasik, S. Bogdanova and D. Taskovski
University of Skopje, Macedonia

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES III

CSIT III.1 **Interpretational Approach to Service Logic Execution** **387**
I. I. Atanasov and E. N. Pencheva
Technical University of Sofia, Bulgaria
CSIT III.2 **Software for Estimation of xDSL Service Quality** **391**
M. Ranđelović, A. Atanasković, N. Dončov
and B. Milovanović
University of Niš, Serbia
CSIT III.3 **Efficiency of Parallel Graph Processing on Multicomputer Platform** **395**
P. I. Borovska and M. Lazarova
Technical University of Sofia, Bulgaria
CSIT III.4 **Parallel FFT Implementation on FPGA** **399**
P. I. Borovska and P. G. Manoilov
Technical University of Sofia, Bulgaria
CSIT III.5 **Architecture of HASIS-3D System Designed for Hail Suppression Purposes** **403**
V. Mihajlović, S. Đorđević-Kajan, D. Rančić,
B. Predić, I. Antolović, P. Eferica and Z. Babić*
University of Niš, Serbia
* Hydrometeorological Service of Serbia
CSIT III.6 **Low Cost-Interactivity for 3D-Computeranimation with Computer Game-Engines** **407**
A. Künz, M. Dontschewa and H. Weber
UCT- Research, FHV, Dornbirn, Austria
CSIT III.7 **B-tree Index Structures for Multimedia Data** **411**
V. T. Dimitrov
Technical University of Sofia, Bulgaria

TELECOMMUNICATIONS SYSTEMS AND TECHNOLOGY IV

TST IV.1 **Balancing the Reverse Path in CATV Systems** **417**
D. M. Dobrev and L. T. Jordanova
Technical University of Sofia, Bulgaria
TST IV.2 **Boundary Conditions when using Quadrature Modulations in Coaxial Cable Television Networks** **421**
K. Koitchev, K. Angelov and S. Sadinov
Technical University of Gabrovo, Bulgaria
TST IV.3 **Planning of VoIP System for Hybrid Fiber-Coaxial Television Networks** **425**
N. Varbanova, K. Koitchev and K. Angelov
Technical University of Gabrovo, Bulgaria
TST IV.4 **Estimation of the Effectiveness of Some of the Access Algorithms in Terms of Servicing Subscribers of the CATV Systems** **429**
L. T. Jordanova and J. I. Nenkov
Technical University of Sofia, Bulgaria
TST IV.5 **Defining the Channel Parameters and Their Influence on SER for OFDM Transmission** **433**
K. Koichev, S. Sadinov and G. Traikovski
Technical University of Gabrovo, Bulgaria
TST IV.6 **Cross-Phase Modulation Induced Nonlinear Phase Noise in Spectral Efficient WDM DPSK-OOK systems** **437**
G. K. Georgiev and N. M. Stoyanov
Ministry of the Interior of Bulgaria
TST IV.7 **Comparative Analysis of Cooperative Diversity and Classical Spatial Diversity** **441**
B. Dimitrijević, J. Spasić* and Z. Nikolić
University of Niš, Serbia
*Ministry of science and environmental protection, Serbia
TST IV.8 **Voice Over WLAN 802.11g in Presence of AWGN and Rayleigh Fading** **445**
I. Angelovski and L. Gavrilovska
University of Skopje, Macedonia

PLENARY SESSION

Smart Capacitive-Resistive Sensors

Gerard C.M. Meijer, Xiujun Li, Zu-yao Chang and Blagoy P. Iliev

***Abstract* – In this paper the issues around electrical signal processing of capacitive sensors and capacitive-resistive sensors are discussed. It is shown that in case of simple capacitive elements with simple electronic circuits a very high performance can be obtained. However, because of physical problems, the capacitive sensor elements make often part of a complex electrical network, which can complicate the measurement tasks considerably. As case studies, the signal processing circuitry for humidity sensors, blood-impedance measurement and water-content measurement are discussed. It is shown the design of such circuits should be based upon a good understanding, characterising and modelling of the physical phenomena.**

***Keywords* – Impedance sensors, capacitive sensors, smart sensors, universal sensor interfaces.**

I. Introduction

Capacitive sensors capacitive-resistive sensors are similar to thermal sensors in that the transduction of the physical input signal to the output signal is performed in two steps: firstly, by transducing a physical quantity into a change of electric capacitance; then, by measuring and converting the capacitive signal into an electric output signal. For this reason, many thermal sensors can be replaced by capacitive ones and vice versa. The majority of capacitive sensors can be found in applications for the detection of mechanical quantities of moving objects such as position, speed, and acceleration, as well as force and pressure [1]. Another important application area is the measurement of liquid content and level and dielectric properties of materials. For the latter group of sensors the measurement is still represented by capacitive elements. However, these elements make part of a complex network, which can make their measurement rather challenging. For instance, in the water-content measurements (Section IV.A), such a measurement needs a sinusoidal excitation with frequencies of 20 MHz. At these frequencies, physical effects such as the skin effect and the proximity effect make it difficult to perform a volumetric measurement. So another approach is required.

In a clean dry environment, the accuracy and resolution of capacitive sensors can be very high [2](Hicks and Atherton, 2000) and the capacitive signals can be processed with simple circuits, using square-wave excitation signals. These circuits have attractive features such as low energy consumption and simple structure. However, in case of capacitive sensors, such as humidity sensors (Section III), the physical phenomena of, for instance, unclean or wetted electrodes can serious problems for the sensor-system performance.

This chapter deals with a systematic approach to the design of reliable, high-performance, and low-cost capacitive and capacitive-resistive sensor systems. The relations between measurands and capacitances are presented together with a discussion on how to optimize the electrode structures. The physical and electrical effects of shielding and guard electrodes will be pointed out. The parasitic effects of contamination, condensation and shunting conductances will be discussed together with possible solutions on how to reduce their influence.

[1]The authors are with the Department of Micorelectronics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, E-mail: g.c.m.meijer@tudelft.nl

II. Basic Principles of Capacitive Sensors

For a simple flat-plate capacitor with two parallel-plate electrodes, the capacitance C_0 between the two electrodes with surface area S, which is separated by a distance d and a dielectric with dielectric permittivity ε, amounts to:

$$C_0 = \varepsilon \frac{S}{d}. \tag{1}$$

In this equation, the effects of field bending and non-homogeneity of the dielectric are neglected.

When such a capacitor is used as a sensor to convert a non-electrical quantity into a change of a capacitive quantity, we can distinguish three ways of doing this:

- The non-electrical quantity changes the dielectric properties (Fig. 1 (a)). This method is applied to sensors that can measure, for instance, humidity, liquid level and material properties.
- The non-electrical quantity changes the electrode distance (Fig. 1(b)). This method is applied to sensors that can measure mechanical quantities such as force, pressure, acceleration and distances.
- The non-electrical quantity changes the electrode area or shields a part of the electric field (Fig. 1(c) and Fig. 1(d)). This method is applied to sensors that can measure, for instance, speed, position, movement and liquid level.

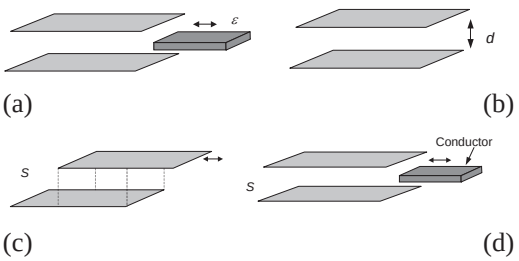


Fig. 1. Basic structure of capacitive sensors.

TABLE I. CAPACITIVE SENSORS, THEIR INPUT SIGNALS AND OPERATING PRINCIPLES

Name	Input signal	Change of
Angular encoder, Position sensor	Mechanical rotation and displacement	Effective electrode area
Force sensor Pressure sensor	Mechanical force or torque pressure	Electrode distance
Humidity sensor	Humidity	Dielectric constant
Accelerometer	Acceleration	Electrode distance
Liquid level gauge	Level of dielectric or conductive liquids	Dielectric constant or effective electrode area
Movement detection	Movement of objects or persons	Dielectric constant, effective electrode area or electrode distance
Property sensor	Material properties	Dielectric constant

Table I lists a number of examples in which the principles of capacitive sensors (as shown in Fig. 1) have been used. Some of these applications will be discussed below.

A. Humidity sensors

Humidity sensors are applied, for instance, in air conditioners, climate controllers, meteorological applications, food processing, and room-comfort control. A typical capacitive humidity sensor consists of a main (base) electrode (Fig. 2), a porous metal top electrode and an intermediate polymer dielectric layer which can absorb water molecules. Such absorption will result in an increase in the relative dielectric constant and thus also in the capacitance between the two connecting terminals.

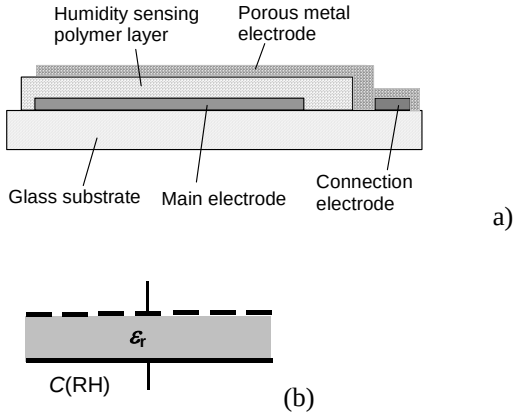


Fig .2. Capacitive humidity sensor

A capacitive humidity sensor can be made with, for instance, MEMS technology. Typical capacitance values of the sensor are in the range of 150 pF to 300 pF, with a 15% change of the capacitor values over the measurement range from about 5% RH to 95% RH.

One problem with capacitive humidity sensors is that their capacitance is shunted by a leakage resistance. When using (overly) simple electronics for the measurement of the capacitive impedance components, the leakage resistor can cause large measurement errors.

B. Liquid-level gauge

A capacitive liquid-level gauge can be used to measure the level of conductive and dielectric liquids. As an example, Fig. 3 shows the electrode structure of a liquid level gauge, which is composed of two vertical electrodes. One of them (the right-hand electrode) is segmented. After measuring all of the capacitances between the various segments with respect to the left-hand electrode, the level can be precisely measured in a two-step approach: with coarse signal processing the segment is found in the region in which the air-liquid interface exists; then, by using linear interpolation, the level position in this region is accurately calculated. In this way, the liquid level gauges can measure liquid levels in, for example, a range of 4 m with a resolution of 0.1 mm and an accuracy of 1.0 mm [3]. In such an application, to reduce the EMI effect, the guard and shielding electrodes (see the next section) should be carefully designed.

It is important to know whether or not the liquid is conductive or non-conductive; in the case of non-conductive liquids such as gasoline or oil, the presence of liquids with a

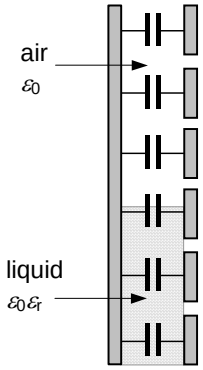


Fig. 3. Electrode structure of a liquid level gauge.

dielectric constant ε_r >1 will cause an increase in the capacitor values. In the case of conductive liquids such as water, the liquid will act as an electrode which is connected directly or indirectly to ground. In that case, the presence of (conductive) liquid will decrease the capacitance between the sensor electrodes. Moreover, the presence of conductive liquid will cause field-bending effects that can cause nonlinearity in the sensor characteristics. When all of these effects are carefully taken into account, capacitive level gauges can measure the levels of both conductive and non-conductive liquids.

C. The design of electrode configurations

A major drawback of capacitive sensors concerns their sensitivity to contamination and condensation, which can cause serious reliability problems. For instance, the measurement systems for capacitive sensors based on relaxation oscillators [4] are simple and offer a relatively high resolution. However, these capacitive measurement systems

cannot accurately measure the capacitance in the presence of shunting conductance.

Another drawback of capacitive sensors concerns the possible occurrence of electric-field-bending effects, which will cause inaccuracy in, for instance, capacitive displacement sensors. The use of guard electrodes is very important, as these reduce the influence of the electric-field-bending effect and also reduce the effect of external disturbing signals [1], [5] and [6]. However, in conventional capacitive sensors, the use of guard electrodes cannot eliminate the influence of electric-field-bending effects completely. Even when using well-designed guard electrodes, the electric-field-bending effect is still one of the major causes of residual nonlinearity of high-precision capacitive sensors.

In this section, the effects of EMI, electric-field bending, parasitic capacitors, and shunting conductance for the accuracy of capacitive sensors are discussed. Some solutions for reducing these effects are also presented.

C1. EMI effects

For easy understanding of the basic features of capacitive sensors, we use a simple structure with two parallel-plate electrodes (Fig(a)). The capacitance can be measured by applying an AC voltage V_m and measuring the resulting currents i_1 or i_2 (Fig. 4(a)). The signals to be measured are very small; capacitances are measured by calculating the charge displacement $\Delta Q = \Delta(CV)$. For instance, for a voltage swing equal to a supply voltage of 5 V and a desired resolution of 10 aF (= 10^{-17} F), the required resolution for charge displacement should be as high as 5×10^{-17} C. This small value equals the modulus of only 312 times the charge of a single electron. Therefore, reliable detection of such small capacitance changes requires integration over a number of these small charge displacements. Electro-Magnetic Interference (EMI) that is due to, for instance, an interfering voltage source V_{int}, can easily diminish the accuracy of the measured currents i_1 and i_2. Therefore, the sensitive structure requires electric shielding, as shown in Fig 4(b).

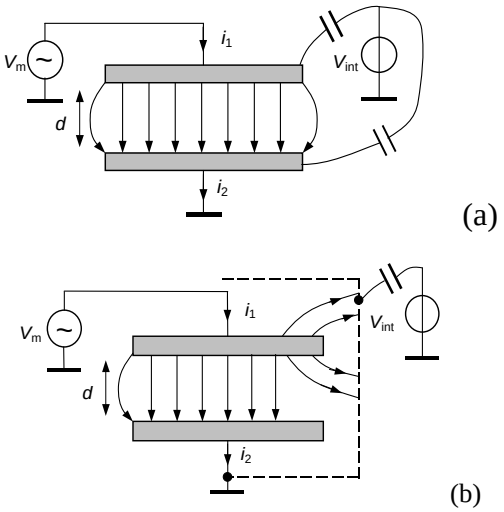


Fig. 4. To reduce the effect of the Interference source V_{int}, the capacitive structure is shielded.

C2. Electric-field-bending effect

In the electrode structure of Fig. 5(b), the electric field between the electrodes and the shielding directly affects the current i_1. Also, the current i_2 is slightly affected due to the electric-field-bending effect at the borders of the electrode. As a consequence, the measured position of the sensor electrodes is affected by that of the shield. Because of this problem, the current i_2 is a more accurate measurement for the capacitance between the electrodes than the current i_1. The effect of electric field-bending can be reduced by applying guard electrodes, as shown in Fig. 5. The bottom electrode is surrounded by grounded guard electrodes. Note that now the current i_2 is not influenced by the presence of the shielding, but the current i_1 still is. To enable the measurement of the current i_2 and to connect all of the electrodes properly, it is important to have access to all of the electrodes.

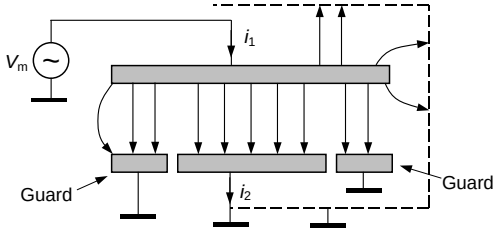


Fig. 5. Guard electrodes.

C3. Active-guard electrode

For practical reasons, sometimes only one of the two sensor electrodes is accessible while the other one is connected to ground. This is the case, for instance, when the bottom electrode is grounded in such a way that i_2 cannot be measured. In such a case, as an alternative, active guarding or shielding should be applied [7]. In this technique, a voltage follower (see Fig. 6) is applied to equalize the potential of the guard electrode with that of the active sensor electrode. By doing this, the voltage over the parasitic capacitance C_p equals zero so that no current will be drained through C_p. This technique can also be applied to eliminate the effects of parasitic resistors to the ground or to surrounding electrodes.

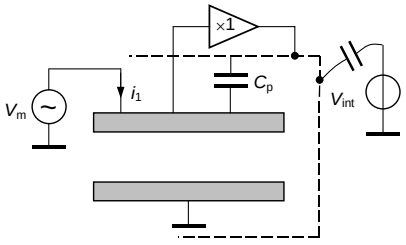


Fig. 6. Active-guard electrode.

C4. Floating electrodes

Floating electrodes are electrodes that are not connected to any well-known potential point in the system. The advantage of using a floating moving electrode (Fig. 1 (d)) is that, in a mechanical way, the measurement can be calculated without contact, which benefits the mechanical reliability. For

example, a capacitive encoder with a conductive rotor requires grounding of the sliding contact. This sliding contact can cause long-term mechanical reliability problems. To solve this problem, a floating-conductive or a dielectric rotor [8] can be used. With such a structure, the mechanical reliability is improved in exchange for a slight decrease in accuracy.

Another example of the use of floating electrodes is a capacitive personal detector [9], which is designed to detect the presence and movement of persons, animals, and objects. Although the idea is appealing and proof of this concept could be given for a laboratory environment, there is still a major reliability problem to be solved. This problem is due to the fact that electrodes that are not connected to a well-known potential, do have some parasitic, are not-well-controlled, and are coupled to the rest of the electrode structure. The influence of parasitics can heavily influence the measurement accuracy or even lead to a complete failure of the measurement.

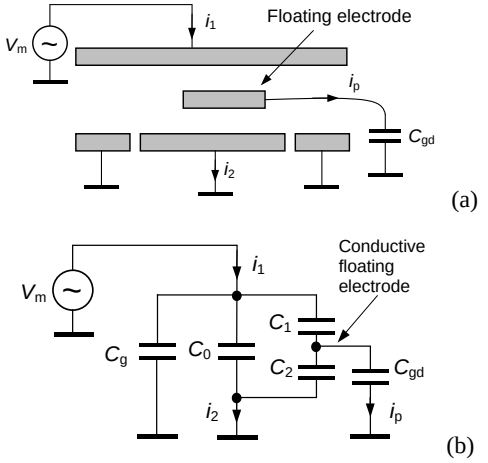


Fig. 7. (a) A floating electrode structure and (b) its electrical model.

In comparison to, for instance, magnetic sensors, a major drawback of capacitive sensors concerns their sensitivity to contamination and condensation. For a reliable measurement, the physical conditions should be well defined. However, in practical situations this is not always so easy to realize. Notorious examples of undefined conditions are those due to the occurrence of contamination and condensation. Contamination and condensation can form conductive layers that can be considered electrodes. Sometimes these undesired contamination electrodes are grounded and can attenuate the electric field within sensor capacitors. The opposite effect is also possible: when the undesired pollution electrodes enlarge the area of the transmitting electrodes, this will cause an increase in the sensor capacitance. It will be clear that the occurrence of undefined conductive layers is not acceptable. Therefore, the use of capacitive sensors is limited to those applications where a clean or dry environment can be guaranteed.

As an example, Fig. 7 shows a floating electrode structure and its electrical model in the case that the floating electrode is *conductive*. For the sensor structure of Fig. 7(a), the presence of the floating electrode can result in two opposite effects: depending on the value of the parasitic capacitance C_{gd}, the presence of the floating electrode can yield an *increase* or a *decrease* in the measured current i_2. Thus, it can be concluded that in order to obtain a reliable measurement, the capacitor C_{gd} from the floating electrode to its environment should be well defined.

In the case of floating dielectric electrodes, care should be taken to deal with problems caused by static charge. There are many examples of static charge at the dielectric surface of moving electrodes, which can cause an extreme amount of noise and related measurement inaccuracy and unreliability. Therefore, local static charge must be removed, for instance by applying a high-ohmic conductive layer on top of the dielectric electrode.

C5. Contamination and condensation

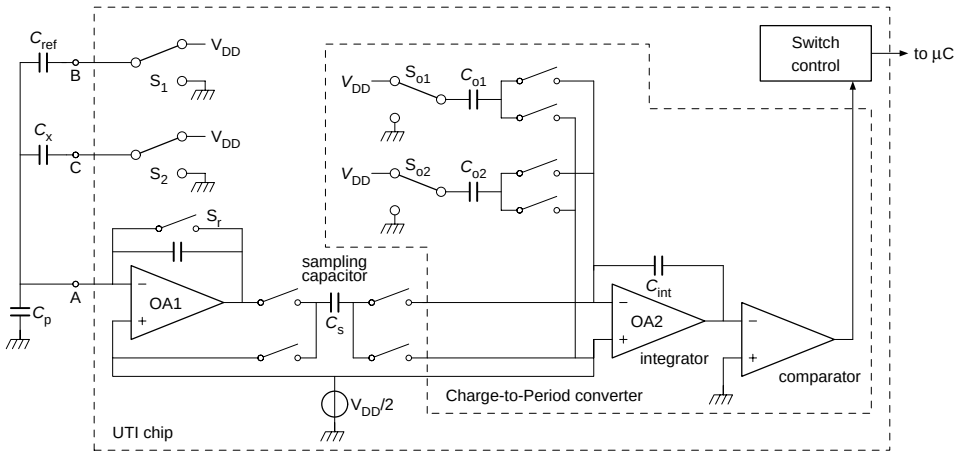


Fig. 8. Principle of the UTI system for capacitive-sensor

III. INTERFACE CIRCUITS

A main part of the interface is the electronic interface, which is called modifier. In [10] a very simple relaxation has been presented, which consist of a relaxation oscillator which generates a period-modulated output signal. The period of this signal can be easily be detected by a microcontroller. A similar circuit has been applied in a universal transducer interface (UTI) [11], [12], which has a number of capacitive modes to process the signals of capacitive sensors. Figure 8 shows the front-end circuit of the UTI that is used to determine the relative value of a sensor capacitor C_x with respect to a reference capacitor C_{ref}. These capacitors have one common electrode, thereby requiring three IC pins to connect the capacitors to the interface chip. In other modes, up to 4 capacitors and 1 reference capacitor can be connected, thereby requiring 6 pins. The modulator output controls the switches S_1 and S_2, so that a square-wave excitation voltage is generated controlled over the selected capacitor C_x or C_{ref}. This results in a square-wave output voltage for operational amplifier OA1, which is operated in its linear region. The amplitude of this output voltage is proportional to the value of the sensor capacitors C_x or C_{ref}. At the end of each half period, the capacitor C_s samples the magnitude of the square-wave output voltage of the front-end amplifier. After that, charge of C_s is dumped into the integrator capacitor C_{int}. This charge is removed by integrating the current I_{int}. As soon as the integrator output voltage exceeds the comparator reference voltage, the comparator switches into another state, starting the next step of the measurement. This relaxation process results in a periodic square-wave output signal of the comparator. The period length of the output signal is linearly related to the values of the sensor capacitors C_x or C_{ref}. A more detailed discussion on the operation of this circuit is presented in [11]

The sampling capacitor C_s and the charge-to-period converter (Fig. 8) belong to the UTI core, which is used in all modes. In this interface circuit a number of important measurement concepts have been implemented, such as:

- *The three-signal auto-calibration technique* [10], [11]. During the signal-measurement phase, the capacitor C_x is selected, during the reference phase, the capacitor C_{ref} is finally selected, and none of these two is selected during the offset phase. The three obtained signals are processed by the microcontroller in such a way that the scale and offset errors of the interface circuits are eliminated.
- *The two-port technique [10]*. Figure 9 (b) shows how this technique is used to measure a capacitor C_x, with a high immunity for the parasitic capacitances C_{p1} and C_{p2}.

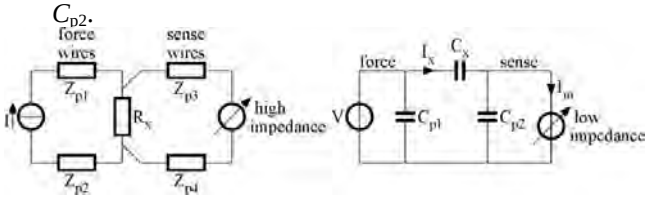


Fig. 9. The two-port technique.

- The transmitting electrode is driven from a low-ohmic voltage source and the receiving electrode is connected to virtual ground (terminal A of the interface chip). The application of this technique is important, because in sensor applications, the magnitude of the parasitic capacitances can exceed that of desired one with orders of magnitude.
- *An advanced chopper technique*, which makes the interface output signal rather insensitive to 1/*f* noise. This is achieved by modulating all relevant electrical signals at a frequency that is higher than the corner frequency of the 1/*f* noise. Because of this measure, it has been possible to implement the interface chip with low-cost CMOS technology, without problems caused by the strong 1/*f* noise of CMOS transistors.

The circuit of Fig. 8 has been designed for capacitive sensors that are free from leakage. In case of, for instance, humidity sensors a shunting leakage conductance 1 μS can cause a considerable measurement error. Therefore, Li and Meijer [13] designed a modified circuit (Fig. 10), in which the effect of a shunting conductance G_s is significantly reduced. This is achieved by minimizing the DC voltage over G_s and by discharging C_s as fast as possible.

In [13] it is reported that for a sensor capacitance of about 2 pF and shunting admittance of 1 μS causes an error of only 0.4%. Yet, in some application the shunting conductances can be much larger and can cause serious errors. To meet such a problem, another approach to measure capacitances will be required. Such an approach will be discussed in the next sections.

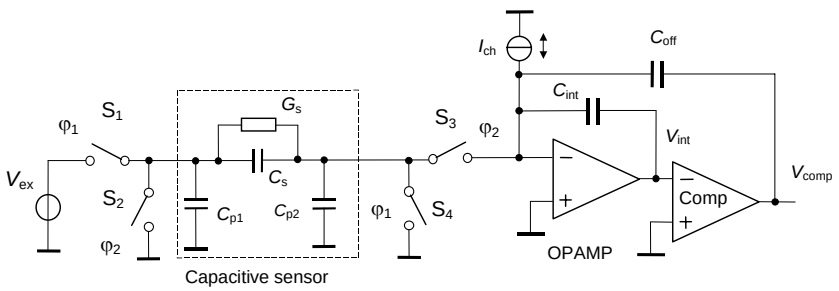


Fig. 10. The schematic diagram of the interface, with improved immunity for shunting conductance [13].

IV. IMPEDANCE SENSORS

When capacitive sensing elements are shunted by a relatively high parasitic conductance, other measurement techniques have to be applied. Firstly, the frequency of the excitation signal should be much higher than in the previous cases. Secondly, instead of square-wave signals, sinusoidal signals should be used. In that case, by measuring both the magnitude and the phase, it is possible to distinguish the capacitive impedance component from the resistive one. In this subsection, we will discuss this approach for two cases: a water-content sensor and a blood-impedance sensor.

A. Water content sensors

Nowadays, to grow crops in the agriculture and horticulture, instead of using natural soil, artificial soil is widely used. It prevents the spread of diseases and allows the natural soil to be able to recover. To optimize the crop-growing process the amount of water and nutrients have to be precise controlled.

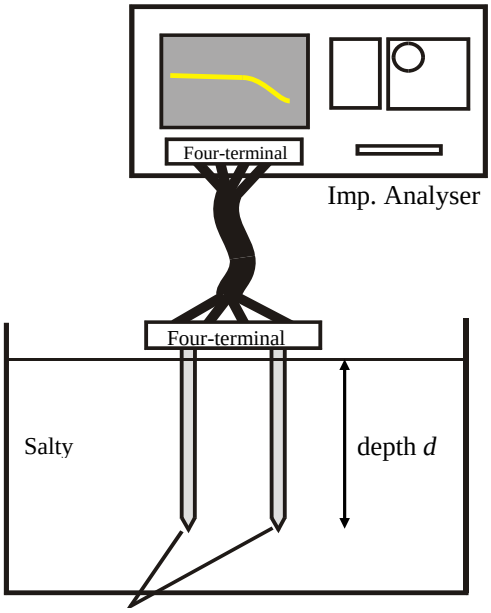


Fig. 11. Conventional-measurement setup. For easy testing of the set-up the artificial soil has been replaced by salty water/ethanol mixture [14]

Figure 11 shows a basic setup to measure the water content of artificial soil. Two electrodes are used to measure the impedance of soil [14]. For easy testing, in the setup of Fig. 11, the soil has been replaced by a mixture of salty water and ethanol. When the water content changes, both the capacitive and the resistive impedance components will change. However, the resistive component is also affected by many other physical effects, such as the salinity. Therefore, the capacitive component is used as a measure for water content. For a good measurement some important precautions have to be taken: Firstly, the electrodes have to be clean otherwise contaminations on the electrodes will obscure the measured results. In the experiments reported in [14], stainless-steel electrodes have been used, (being cheaper than gold or platinum). Secondly, to avoid the polarization effects the frequency of the excitation signals should be higher than at least 20 kHz. Thirdly, to measure the capacitive component in presence of a high shunting conductivity, the frequency of the excitation signals should be as high as is possible, but limited by the effects of a number of parasitic effects occurring at very high frequencies. In the experiments reported in [14], the excitation-signal frequency amounted to 20 MHz.

Figure 12 shows a simplified electrical model for the electrodes in a water/salt solution. It is a lumped model and the model for the polarization effects is left out, because at 20 MHz this effect can be neglected.

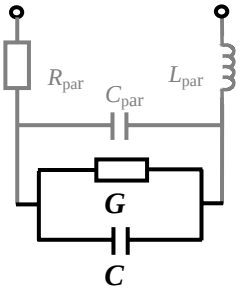


Fig. 12. Electrical model for a salt solution.

In this model the salty solution is modelled as a capacitance C in parallel with a conductance G. The element L_{par} represents the parasitic inductance of the metal rods and it varies with the length of the rods. The elements R_{par} and C_{par} represent respectively the parasitic resistance of the metal rods and the parasitic capacitance between the rods. Usually, the values of these elements are so small that they can be neglected. Unfortunately, the complexity of the model is increased by the occurrence of some physical effects, such as the skin effect and the proximity effect, which make that the values of the model elements are frequency dependent. Now, these effects will be considered in some more detail.

A. Skin and Proximity effects

When a current passes through a conductor, at high frequencies the main part of the current will flow along its surface. This well-known effect is called skin-effect. The depth δ of the skin where the amplitude has attenuated towards e^{-1} (37 %) equals:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}}\ , \qquad (2)$$

where $\omega = 2\pi f$, σ is the conductivity of the conductor and μ is the permeability of the conductor. The skin effect will also occur in the salty water. Table II shows the calculated skin depth δ for a signal frequency $f = 20$ MHz and for different conductivities of the water/salt solutions. From Table II it can be concluded that with rods of 6 cm in length, for $\sigma \geq 9.38$ mS/cm the attenuation at the end of the rods will be too large. Another effect that occurs during measurement with the rods at high frequency is the so-called proximity effect. This effect has the same origin as the skin effect, but occurs in the plane perpendicular to the electrode axis. With the skin as well as with the proximity effect, at higher frequencies the current loops tend to minimize their area. This will cause a frequency dependency of all of the model elements G and L_{par} in the model of Fig. 12.

TABLE II THE SKIN DEPTH OF SALTY WATER

σ [mS/cm]	δ [cm]
0.509	49.9
2.63	21.9
9.38	11.6
20.5	7.9

To reduce the influence of the skin effect, a special probe with segmented electrodes has been designed (Fig. 13). The reduction of the electrode sizes to much less than the skin depth δ, results in a much more uniform current distribution along the length of the electrodes. Moreover, a significant reduction in the value of the parasitic inductance has been obtained. With an experimental setup it has been found that with this probe, even at high conductivities of the water/salt solution, the dielectric capacitance of the salty water at different depths can still be measured. It is possible to design probes with a series of such electrodes over the desired measurement length, so that over the whole vertical region along the probe, the specific impedance can be measured. It has been found [14] that with such a probe in salty water with conductivity up to 2 S/m, the capacitive impedance component can be measured correctly.

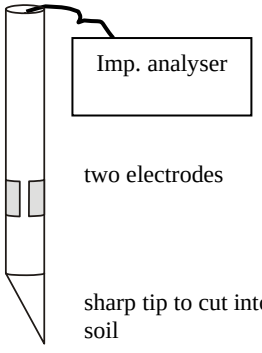


Fig. 13. Probe with segmented electrodes

B. A Blood-impedance characterization system

Blood impedance (Z_b) and particularly plasma resistance (R_p) and cell membrane capacitance (C_m) may be of a great importance for the medicine. They are related to hematocrit (H_t), which is the most valuable determinant of blood viscosity [15, 16]. By measuring blood viscosity various heart-related thrombotic events such as heart infarcts or strokes can be prevented. During animal experiments, a reduction of arterial thrombosis has been obtained by hemodilution therapy, which results in lower hematocrit. Fibrinogen, another determinant of blood viscosity, affects Z_b as well. In the view of the above-mentioned considerations our goal is to explore the diagnostic potential for measuring blood viscosity, hematocrit and fibrinogen levels and other blood parameters in-vitro and in-vivo by means of impedance measurement techniques. As compared to the previous case study, impedance measurement of blood is more complex, because of the more complex electrical behavior of the suspension, as will be discussed now.

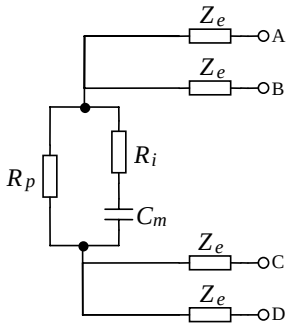


Fig. 14. Three-element model of blood including electrode polarization impedance Z_e.

Blood is a suspension of red cells, white cells, and platelets in plasma. A simplified three-element circuit model describes its properties (Fig. 14) [16, 17]. The concentration of red cells is called hematocrit and in this electrical structure is represented by R_p. R_i represents the cell interior resistance, as C_m is a measure for the cell membrane capacitance. The total complex impedance measured by a sensor includes the polarization effect of the electrodes (Z_e) too [17].

Though R_p is an accurate measure of H_t and therefore good indicator for the viscosity, it does not reflect fully viscosity because of the non-Newtonian characteristics of blood. The non-Newtonian behavior of blood means that it has higher viscosity at lower shear rate. Therefore, viscosity varies with the flow and shear rate eventually. To compensate for this, accurate measurement of C_m is needed. At a certain shear rate C_m is determined by the amount of cells (this is H_t), nevertheless approximately 10% of C_m is affected by the presence of plasma macromolecules between the cells. Consequently, R_p is considered to be measure for H_t and the combination of R_p and C_m is to be used to determine viscosity.

Fig. 15. HemoCard Vision measurement system.

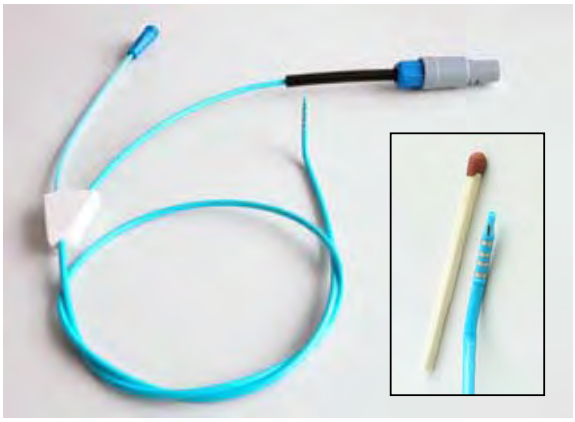


Fig. 16. Catheter for in-vivo characterization of blood in the right atrium.

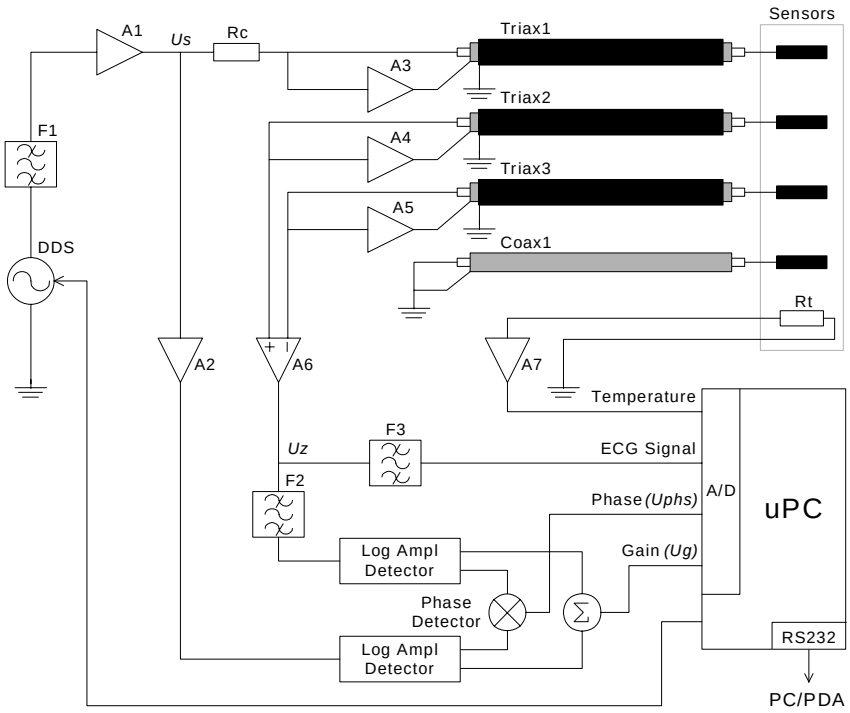


Fig. 17. Block diagram of the HCV interface electronics.

B1. Measurement system

A dedicated impedance-measurement system was designed to perform the characterization of blood. It works with sinusoidal signal in the frequency range from 20 kHz to 1.2 MHz and applies 10 uA current through the sensor. Figure 15 shows the measurement system named HemoCard Vision© (HCV). Its specially developed catheter can be seen in Fig. 16. It has four-electrode setup and a thermistor located at the distal end, near the tip. With the existing three lumens for administering of infusion solutions and/or measuring atrial pressure it enables all characteristic features of a central venous catheter (CVC) alongside with the impedance and temperature measurement capabilities. Four stainless steel electrodes, 0.8 mm wide, positioned equidistantly 2 mm center-to-center, sense the impedance. The outer pair is the excitation one and the inner is the sense pair. The thermistor is placed in a lumen, next to the electrodes. UV resin fills the cavity around the thermistor for better thermal contact. Three triaxial cables, 0.49 mm in diameter, and one coaxial cable, 0.33 mm in diameter, connect the electrodes to the interface electronics.

Regarding the equivalent electrical model of blood (Fig. 14), we set the following ranges for the HCV interface electronics: 20 Ω to 70 Ω for R_p and 0.2 nF to 2 nF for C_m. The applied current is limited by electromagnetic compatibility regulations to 10 uA. No DC current is allowed.

Figure 17 shows the block diagram of the interface electronics. It is battery-powered device using 3.7 V Li-Ion cell. Internal memory collects the measured data, which on later stage is transferred to a PDA or PC for further

processing. The RS232 connection is optically decoupled for safety reasons.

An excitation signal is generated by a direct digital synthesizer (DDS). Five discrete frequencies, 20 kHz, 200 kHz, 400 kHz, 600 kHz, and 1.2 MHz, are applied consecutively in the time. The excitation signal is filtered by F1, buffered by A1 and applied to the "high-potential" excitation electrode via clamp resistor R_c. The "low-potential" excitation electrode is connected to ground via decoupling capacitor (not shown in Fig. 17). Stray capacitance of 75 pF can be measured with each of the cables that connect the sensors to the electronics. Therefore, active guarding is applied in order to avoid phase and gain errors (A3, A4, A5). Additionally, third, grounded shield prevents emission or penetration of undesired signals.

High input-impedance differential amplifier (A6) senses the voltage, U_z, between the measurement electrodes. This signal is rather complex due to the presence of ECG component. The tip of the catheter is located inside the right atrium, between the sinoatrial node (SAN) and the atrioventricular node (AVN). Consequently, the measured signal is a mixture of weak impedance and strong ECG signals. The ECG component covers the frequency range from 0.5 Hz to 250 Hz, while the lowest impedance-signal frequency is 20 kHz. Two band-pass filters F2 and F3 ensure proper separation of the ECG and impedance signals.

Already filtered, the impedance signal is then compared with the source signal U_s by a phase gain analyzer and Z_b is calculated accordingly [20, 21].

With this setup a number of tests have been performed on swine and sheep [22]. These animals are known to have cardiovascular system similar to this of human, therefore these are used regularly during in-vivo experiments. During these tests R_p, C_m and ECG have been measured. These tests have proved that it is feasible to perform the described tests. It is shown, that prevention of growth of biolayers needs special care. Human trials will be performed in the near future.

V. Conclusion

This paper shows that capacitive sensors and capacitive-resistive sensors have to be designed as an overall system in its physical environment. A study of the physical environment should result in a characterization and modeling of the sensing elements and their parasitics. All of these elements and their mutual magnitudes should be taken into account when designing the interface system. For simple capacitive elements a very high accuracy (16 bits for a full range of 2 pF) can be obtained with simple interface circuits. To achieve such accuracy, powerful measurement techniques, such as autocalibration, chopping and two-port techniques should be applied. For easy use and rapid prototyping, it is convenient to use universal transducer interfaces, in which the subscribed measurement techniques have been implemented in the hardware of CMOS chips.

In case, that the sensor capacitances are shunted by a large parasitic conductance, the more complex impedance analyzers will be required. For this, two case studies have been presented. A first case study concerns a water-content sensor system, in which the capacitive impedance component represents the water content of artificial soil. For proper measurement a signal frequency of about 20 MHz is required. It has been shown that reduction of skin and proximity effects make it necessary to use special impedance probes. When using such probes water-content can be measured for water with conductivity up to 2 S/m.

As a second case-study a characterization system for blood impedance has been presented. It is shown that for electrical characterization of suspensions, such as blood, requires an even more complex measurement approach, using a range of signal frequencies. With the described impedance measurement technique it is possible to implement in-vivo monitoring of blood viscosity

References

[1] Baxter, L.K., "Capacitive Sensors: Design and Applications", IEEE press, NJ, 1997.

[2] Hicks, T.R. and Atherton P.D. "The nanopositioning book", Penton Press, UK, 2000.

[3] Toth, F.N., "A design methodology for low-cost, high-performance capacitive sensors", PhD. Thesis, Delft University of Technology, 1997.

[4] Toth, F.N., Meijer, G.C.M., "A low-cost smart capacitive position sensor", IEEE Transactions on Instrumentation and Measurement, vol. 41: 1041-1044, 1992.

[5] De Jong, G.W., "Smart capacitive sensor (physical, geometrical and electronic aspects)", Ph.D. Thesis, Electronics Research Lab., Delft Univ. of Technology, 1994.

[6] Heerens, W.Chr., "Multi-terminal capacitor sensors", Journal of Physics E: Sci. Instrum., vol. 15: 137-141, 1982.

[7] Huang, S.M., Beck, M.S., Green, R.G., Stott, A.L. "Electronic transducers for industrial measurement of low value capacitances", J. Phys. E: Sci. Instrum., vol. 21: 242-250, 1988.

[8] Gasulla, M., Li, X., Meijer, G.C.M., Van der Ham, L., Spronck, J.W., "A contactless capacitive angular-position sensor", IEEE Sensors Journal, vol. 3: 607-614, 2003.

[9] Toth, F.N., Meijer, G.C.M., Zapico, J.A., "A low-cost capacitive personnel detector with fuzzy-logic position estimation", Proceedings IMTC'96, Brussels: 1270-1274. 1996.

[10] Meijer, G.C.M., Van Drecht, J., De Jong, P.C., Neuteboom, H., "New concepts for smart signal processors and their application to PSD displacement transducers", Sensors and Actuators, vol. A35: 23-30, 1992.

[11] Van der Goes, F.M.L., Meijer, G.C.M., "A universal transducer interface for capacitive and resistive sensor elements", Analog Integrated Circuits and Signal Processing, vol.14, no.3: 249-260, 1997.

[12] Smartec, www.smartec.nl, May 2007.

[13] Li, X., Meijer, G.C.M., "An accurate interface for capacitive sensors", IEEE Transactions on Instrumentation and Measurement vol. 51, no. 5: 935-939, 2002.

[14] Chang, Z., Iliev, B.P., De Groot, F., Meijer G.C.M., "Extending the limits of a capacitive soil-water-content measurement", Proc. IMTC'05, Atlanta, 2005.

[15] G.A.M. Pop, W.J. Hop, et al., "Blood electrical impedance closely matches whole blood viscosity as parameter of hemorheology and inflammation", Applied Rheology, Vol 13, Issue 6.

[16] T. X. Zhao, B. Jacobson, Quantitative correlations among fibrinogen concentration, sedimentation rate and electrical impedance of blood, Medical & Biological Engineering & Computing, May 1997.

[17] L. A. Geddes, L. E. Baker, Principles in applied biomedical instrumentation, A Wiley-interscience publication 1989.

[18] H. P. Schwan, Electrode polarization impedance and measurements in biological materials. The more school of electrical engineering, Electromedical division, University of Pennsylvania, Philadelphia, 1988, Pennsylvania.

[19] Hugo Fricke and Sterne Morse, The electric resistance and capacity of blood for frequencies between 800 and 4½ million cycles, The Journal of General Physiology, Vol 9, 153-167, 1925.

[20] Blagoy P. Iliev and Gerard C. M. Meijer, An impedance-measurements system for electrical characterization of rockwool substrates**,** Sicon/04 - Sensors for Industry Conference, New Orleans, Louisiana, USA, 27-29 January, 2004.

[21] Tian-xian Zhao, Bertil Jacobson and Tommy Ribbe, Triple-frequency method for measuring blood impedance. Physiol. Meas., Vol. 14, 145-156, 1993.

[22] Iliev, B.P., Pop, G.A.M., and G.C.M., "In-vivo blood characterization system", Proc. IMTC'05, Atlanta, 2005.

Measurements of Road Traffic Parameters

Janusz Gajda[1] and Ryszard Sroka[1]

***Abstract* – The basic problems corresponding to the traffic measurements have been presented in the paper. Authors gave the general overview of the parameters and characteristics describing the vehicles and traffic flow. The current sensors' technologies used in measuring systems has been presented. Moreover the measuring systems used in traffic measurements have been described. Special attention has been paid to the more sophisticated problems like the vehicle's classification based on the analysis of its magnetic signature and to vehicles' weigh-in-motion technique.**

***Keywords* – Traffic parameters measurements, Single loop detectors, WIM systems, Multi sensor WIM systems, Vehicle classification, Vehicle's magnetic signature.**

I. INTRODUCTION

The story of the measurements of road traffic parameters is very long. The first system oriented towards road traffic control has been put into use in London, in the 1868 year. It was a simple traffic light system, equipped with gas lamps. Following attempts in this area have been undertaken in Ohio, USA, in the 1914 year.

Much more advanced control system, called „Green Wave", whose assignment was to increase traffic continuity, has been employed for the first time in Münich, Germany, in the 1953 year. Since then the number of vehicles participating in the traffic increases rapidly from year to year.

At present, the demand for the information concerning the basic traffic parameters and more sophisticated characteristics is still huge [27]. For example: traffic operators need these data for real-time traffic monitoring, transportation researches need these data to develop real-time incident detection algorithms, traffic engineers need these data as guidance of location of freight-oriented variable message signs, freight modellers need these data to provide calibration information for freight modelling, motor carrier dispatchers need these data for routing information, etc. The source of these data are measuring systems, which differ in their complexity and are equipped with different sensors kinds.

For optimal control the executive systems should be supplied with information - best of all in real time - about the situation taking place on the road. Therefore systems of traffic monitoring are coming into existence. The monitoring includes among others: presence and the number of vehicles in the detection area, speed of individual vehicle (or mean value of speed for specified number of vehicles), vehicle assortment, their load exerted upon road surface, occurrence of traffic jams, weather conditions (which have substantial influence upon traffic safety), etc.

The monitoring assignments are not merely aspects of direct traffic control, but its results may be of service for statistical studies, economic investigations, automatic toll collecting, highway infrastructure planning and reconditioning, as well as for protection against overloaded vehicles.

So monitoring prompts necessity to create measurement systems, capable of data acquisition and traffic parameters estimation. The basis of the latter are measurements of parameters of individual vehicle, participating in the traffic.

This paper addresses the basic problems corresponding to the measurements of the traffic parameters. The paper is organized as follows. The basic parameters characterising the individual vehicles and characteristics of the traffic stream are presented in chapter 2nd. The overview of sensors used in traffic measurement systems is given in chapter 3rd. In chapter 4th the systems designed for measurements of basic traffic parameters, characterising the individual vehicles are presented. Chapter 5th contains the discussion of the problem of automated vehicle classification. The very nowadays-important problem of weighing of the moving vehicles is discussed in chapter 6th. Chapter 7th presents the research problems corresponding to the control and maintains of the road traffic.

II. BASIC PARAMETERS AND CHARACTERISTICS OF THE ROAD TRAFFIC

The basic parameters, which are the subject of the direct measurement, describe the individual vehicle or the traffic at the specific point on the road. The characteristics describing the traffic system are estimated on the base of the directly measured parameters. These parameters are divided into two groups i.e. transient parameters and fixed parameters. The transient parameters group consists of the following: time of arrival, lane number, speed, time headway, space headway, time gap, space gap, travel direction, acceleration, total mass, axle load (static component), axle load (dynamic component), axle frequency, number of passengers, trailer presence. The group of fixed parameters contains: vehicle's class, number of axles, total wheel base, axle distances, length, width, area, height, front overhang, rear overhang, suspension height.

Considering individual vehicles, we can say that each vehicle i in a line of a traffic stream has the following

[1]Authors are with the Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, Department of Measurement and Instrumentation, AGH – University of Science and Technology, 30-059 Krakow, Poland, E-mail: jgajda@agh.edu.pl, rysieks@agh.edu.pl

informational variables: a length, denoted by l_i, a longitudinal position, denoted by x_i, a speed, denoted by $v_i = \frac{dx_i}{dt}$ and acceleration, denoted by $a_i = \frac{dv_i}{dt} = \frac{d^2 x_i}{dt^2}$. In this first approach, the other spatial characteristics of the vehicle such as its width, height and line number are neglected [21].

Describing the traffic flow characteristics, we can consider two consecutive vehicles in the same lane in a traffic stream: a follower i and its leader $i+1$. The vehicle i has certain space headway h_{si}, to its predecessor, the space gap g_{si} to its leader and its own length l_i:

$$h_{si} = g_{si} + l_i \tag{1}$$

Taking the rear bumper as a vehicle's position, the space headway $h_{si} = x_{i+1} - x_i$. The space gap is thus measured from vehicle's front bumper to its leader's rear bumper.

Analogously to equation (1) each vehicle also has a time headway h_{ti} (expressed in seconds), consisting of a time gap g_{ti} and an occupancy time ρ_i:

$$h_{ti} = g_{ti} + \rho_i \tag{2}$$

When the vehicle's speed is constant, the time gap is the amount of time necessary to reach the current position of the leader when traveling at the current speed. Similarly, the occupancy time can be interpreted as the time needed to traverse a distance equal to the vehicle's own length at the current speed, i.e., $\rho_i = l_i / v_i$. Both equations (1) and (2) are furthermore linked to the vehicle's speed as follows:

$$\frac{h_{si}}{h_{ti}} = \frac{g_{si}}{g_{ti}} = \frac{l_i}{\rho_i} = v_i \tag{3}$$

As the above definitions deal with what is called single-line traffic, we can easily extend them to multi-lane traffic. In some traffic flow literature, other nomenclature is used: space for the space headway, distance or clearance for the space gap, and headway for the time headway.

When considering many vehicles simultaneously, the macroscopic flow characteristics are used to represent all traffic faithfully.

The macroscopic characteristic, called density, allows us to get an idea of how crowded a certain section of a road is. It is typically expressed as the number of vehicles per kilometer. When density can not be exactly measured or computed, it has to be estimated. Using the spatial region R_s (corresponding to measurements at a single instant in time), the density k for single-line traffic is defined as:

$$k = \frac{N}{K} \tag{4}$$

with N the number of vehicles presents on the road segment of length K. In general, density can be defined as the total time spent by all the vehicles in the measurement region, divided by the area of this region. This generalization allows us computing the density at a point using the measurement temporal region R_t (corresponding to measurements at single fixed location in space, during a certain time period T_{mp}):

$$k = \frac{\sum_{i=1}^{N} T_i}{T_{mp}\, dx} = \frac{1}{T_{mp}\, dx} \sum_{i=1}^{N} \frac{dx}{v_i} = \frac{1}{T_{mp}} \sum_{i=1}^{N} \frac{1}{v_i} \tag{5}$$

with T_i the travel time and v_i the speed of the i^{th} vehicle.

Flow can be considered as a temporal measurement and it is typically expressed as a number of vehicles per hour. Sometimes other synonyms such as intensity, flux, throughput, current or volume are used. Measuring the flow q in region R_t for single-line traffic is done using the following equation:

$$q = \frac{N}{T_{mp}} \tag{6}$$

with N the number of vehicles that has passed the detector's site.

When using single inductive loop detectors, the vehicle speeds are not available. The detector's logic therefore resorts to a temporal measurement called the occupancy ρ, which corresponds to the fraction of time the measurement location was occupied by a vehicle:

$$\rho = \frac{1}{T_{mp}} \sum_{i=1}^{N} o_{ti} \tag{7}$$

with o_{ti} the on-time of the i^{th} vehicle, i.e., the time period during which it presents above the detector.

The final macroscopic characteristic to be considered is the mean speed of a traffic stream. It is expressed in kilometers per hour. If we base our approach on direct measurements of the individual vehicles' speed, we can generally obtain the mean speed as the total distance traveled by all the vehicles in the measurement region, divided by the total time spent in this region. This gives the following formulas for the spatial and temporal regions, R_s and R_t respectively:

$$\bar{v}_s = \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} T_i} = \begin{cases} \dfrac{\sum_{i=1}^{N} v_i\, dt}{N\, dt} = \dfrac{1}{N} \sum_{i=1}^{N} v_i & (region\ R_s) \\ \dfrac{N\ dx}{\sum_{i=1}^{N} \frac{dx}{v_i}} = \dfrac{1}{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{v_i}} & (region\ R_t) \end{cases} \tag{8}$$

with X_i and T_i the distance, respectively time, traveled by the i^{th} vehicle and N the number of vehicles presents during the measurement. The mean speed computed by the previous

equations, is called the average travel speed, which is more commonly known as the space-mean speed.

It is interesting to see that the spatial measurement is based on an arithmetic average of the vehicles' instantaneous speeds, whereas the temporal measurement is based on the harmonic average of the vehicles' spot speed. If we take the arithmetic average of the vehicles' spot speeds in the temporal measurement region R_t instead, this would lead to what is called the time-mean speed:

$$\bar{v}_t = \frac{1}{N}\sum_{i=1}^{N} v_i \qquad (region\ R_t) \tag{9}$$

The traffic parameters and characteristics defined above are measured by using specialized measurement systems based on different sensor technologies.

III. Traffic Sensor Technologies

There are many technologies available for collecting traffic data. Although inductive loop detectors (ILD) have been used more than any other method up to now, the other technologies are beginning to replace or complement loop detectors in many applications last years. Many of these new technologies are roadside (or vertical sensors), which do not require pavement cuts or the disruption of traffic for installation. Many of the new sensors are cost competitive with ILD systems.

This section provides descriptions of currently available detectors. Road traffic detectors can be divided into two groups [8]:

- Embedded Detectors,
- Non-intrusive Detectors.

An embedded detector system consists of sensors in or below the surface of the roadway. These detectors are currently the most widely used form of vehicle detection. The main detectors being used include the following: rubber tubes, inductive loops, magnetometers, piezoelectric sensors, capacitive sensors, fiber optic sensors, strain gauge sensors, resistive sensors.

Non-intrusive detectors are also known as the above ground detectors. They are mounted on a structure above the surface of the pavement.

Advantages of that non-intrusive detectors over embedded detectors include the following: modifications to pavement are not required for installation, detectors can be moved or replaced more easily, lane closure may not be required during installation and maintenance, detectors can be used during and after any reconstruction or maintenance activities.

A disadvantage of non-intrusive detectors is that they may produce unreliable results during adverse weather conditions.

Non-intrusive detectors include the following [16]: microwave radar, laser detector, passive infrared detector, ultrasonic detector, passive acoustic detector, light curtain, video image processing.

The detailed discussion concerning the properties of each kind of traffic detectors exceeds the frames of this paper. The summarized information on chosen sensors is collect in Table I.

TABLE I
ADVANTAGES AND DISADVANTAGES OF SOME TRAFFIC DETECTORS

Detector Type	Advantages	Disadvantages
Inductive Loop	Flexible design. Wide range of applications. Provides basic traffic parameters (e. g. volume, speed, presence, occupancy).	Installation requires pavement cuts. Installation and maintenance requires lane closure. Detectors subject to stresses of traffic.
Magnetometer	Can be used in situations where loops are not feasible (e. g., bridge decks). Less sensitive than loops to stresses of traffic.	Installation requires pavement cuts. Installation and maintenance require lane closure. Small detection zone. Typically used only to provide count and occupancy.
Microwave Radar	Generally insensitive to weather conditions. Provides day and night operation.	Requires license for operation and maintenance. May lock on to the strongest signal (e. g., large truck).
Infrared	Active detector emits narrow beam allowing for accurate determination of vehicle position. Provides day and night operation. Provides most basic traffic parameters. Passive detectors can be used for strategic loop replacement.	Operation affected by precipitation (e. g., rain, fog, etc.). Difficulty in maintaining alignment on vibrating structures.
Ultrasonic	Provides most basic traffic parameters.	Environmental conditions (e. g., temperature, humidity, air turbulence, etc.) can affect performance. Snow covered vehicles are difficult to detect. High level of special maintenance capability is required.
Acoustic	Completely passive. Generally insensitive to weather conditions. Provides day and night operation.	Relatively new technology for traffic surveillance.
Video Image Processing	Location or addition of detector zones can easily be done. Provides basic traffic parameters. Provides wide-area detection.	Inclement weather, shadows, and poor lighting can affect performance. May require significant processing power and a wide communication bandwidth.

IV. Measurements of Basic Traffic Parameters

Figure 1 presents the measuring systems designed for measurement of basic traffic parameters. The systems are equipped with different sensor kinds and are able to measure different sets of parameters.

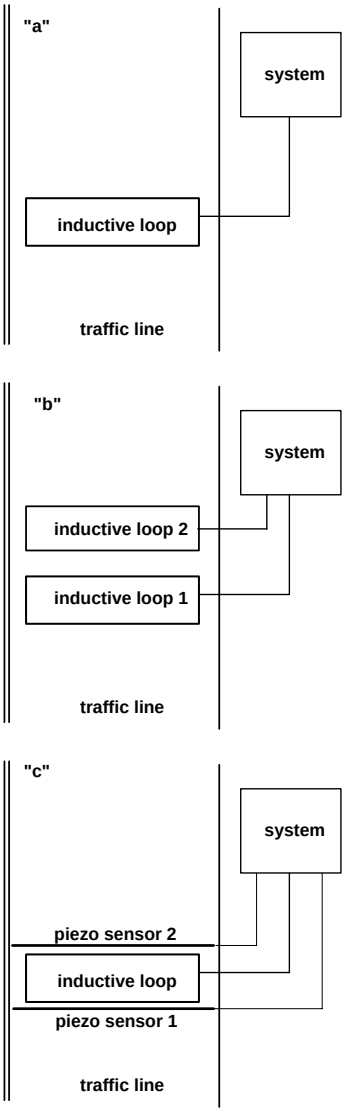


Fig. 1. Diagrams of the measuring systems

The system equipped with single inductive loop is the most simple one. In general, the measured parameters are as follows: volume, magnetic signature of the vehicle, that allows the vehicles classification, occupancy, time intervals between successive vehicles. The more sophisticated analysis of the magnetic signature allows estimation of individual vehicle speed, trailer detection, as well as counting the number of axles, their reciprocal distances and vehicle length [5, 6, 9, 10, 24, 27]. Uncertainty of the speed estimation depends on the loop width and vehicle class. In the class of personal cars this uncertainty, measured by standard deviation of the measurement results, is equal to 6% in the velocity range from 20km/h to 90km/h. The problem of vehicle classification**,** based on the analysis of its magnetic signature, is discussed in chapter 5th.

The second system is classical. It is equipped with two inductive loop sensors, what allows measurement of the following parameters: volume, occupancy, speed, individual vehicle length, trailer presence, time intervals between successive vehicles and vehicles class according to the vehicle length and presence of trailer.

The length measurement uncertainty in this system was tested using the reference vehicle 1250cm long, which traveled 4000 times through the tested measurement site. The standard deviation of the error of length measurement results was estimated as equal to 20cm. The maximum error did not exceed 60cm.

The error analysis of speed measurements in system with two inductive loop sensors was conducted by comparison its results with the reference results obtained in the system with two piezoelectric axle sensors. Than the speed was measured at the same moment by two systems i.e. by tested and reference. The speed of 800 vehicles was measured by compared systems. The mean value of the error was equal to zero and its standard deviation was 1.3 km/h.

The most advanced system is equipped with single inductive loop sensor and two piezoelectric axle sensors. The traffic parameters estimated in this system are as follows: presence detection, volume, occupancy, speed, vehicle length, number of axles, distances between successive axles, magnetic signature, trailer presence, time intervals between successive vehicles. The system is also able to classify vehicles on the base of the number of axles and their distances. Typical measurement signals of the system are presented in Fig. 2. Such a system allows differentiating between 13–14 vehicle classes [25].

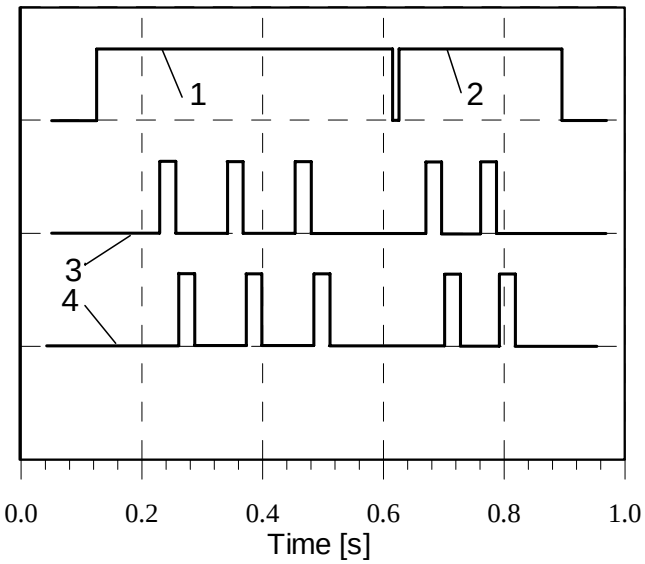


Fig. 2. Measurement signals from presented in Fig. 1c system, taken from a three-axle vehicle with two-axle trailer. 1 – vehicle presence signal; 2 – trailer presence signal; 3,4 – signals from axle load sensors

The uncertainty of axle counting in this system is very difficult to estimate. In the set of 400 vehicles passed the tested site under different metrological and traffic conditions none case of erroneous axle counting was detected.

The error and uncertainty of the axle distance measurement of the moving vehicles was tested by using the multi-axle reference vehicle, which distance between first and second axles is equal to 360cm. The mean value of the measured distance was equal to 364 cm with the standard deviation 1.1cm. The mean value of the measured distance between

axles number 4 and 5 was equal to 129.1 cm (true value 132cm) with standard deviation 0.83cm.

V. VEHICLE CLASSIFICATION

Classifying an automotive vehicle means determining to which of the selected classes the vehicle belongs. Classification methods are depend on the vehicle parameters that can be determined in a given measuring system and on the classification purpose.

The most simple classification method, often used in practice, is based on measuring the vehicle length. Not more than three classes are defined then. The method can be applied in a very simple measuring system, e.g., in a single-sensor system with inductive loop utilizing only the signal of vehicle occurrence above the sensor (Fig.1a).

When a necessity of defining more (four or five) classes arises, it is possible to use a system with inductive sensor and process the obtained magnetic profile of the vehicle [10, 23]. The shape of a time signal generated by the interaction of a vehicle passing over an induction loop sensor with the magnetic field produced by the sensor is the basis for the recognition of vehicle class. Signals generated by different vehicles differ in their parameters such as e.g. amplitude, frequency spectrum, statistical parameters. These signals are called “vehicle magnetic signature” (or magnetic profile). The exemplary magnetic signatures of basic vehicle classes are presented in Figure 3.

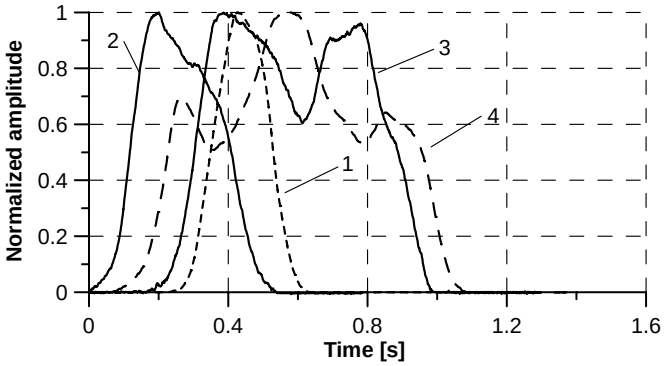


Fig. 3. Magnetic signatures of the basic vehicle classes.
1 – passenger car, 2 – van, 3 – lorry, 4 – bus

The essence of the problem of the vehicle class recognition lies in the analysis of the recorded measurement signals concerning the above mentioned parameters and the comparison of the results of this analysis to the results of the analysis of the reference signals representing selected vehicle classes. This vehicle class is adopted as a classification result, which is represented by a reference signal the closest one in the sense of the analyzed properties, to the recorded, measured signal generated by the vehicle being identified. Reference signals are generated as a result of averaging many measurement signals corresponding to the selected vehicle.

An essential problem here is to make the measurement signal parameters (and, therefore, the results of classification) independent of the vehicle velocity, vehicle suspension height and other disturbing factors. This aspect of the problem was discussed e.g. in [11]. There is also a tendency to ensure the satisfactory selectivity of the classification process in order to differentiate among the required number of vehicle classes.

One method out of the magnetic profile preprocessing methods consists in transforming the profile into the vehicle length domain [11]. This operation results in that, the profile contains the combined information on the shape and length of the primary profile, what enables more selective classification to be made (Fig. 4). To carry out such transformation the information about the vehicle speed is also necessary. In specific cases, amplitude standardization can be abandoned, gaining this way additional information about the vehicle suspension height.

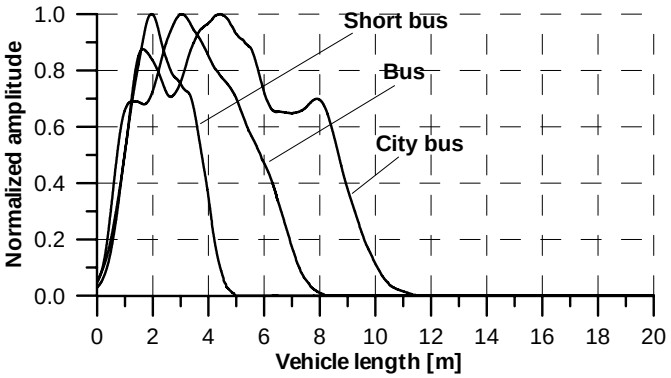


Fig. 4. Bus magnetic profiles in the vehicle length domain

Nonparametric classification methods consist in comparing directly the profile generated by the vehicle being classified (after transformation) with the reference profiles representing each of the defined classes. Depending on vehicle class, the effectiveness of such classification ranges from 67% to 100%.

Parametric methods consist in comparing the profile parameters of the vehicle being classified and reference vehicle. The effectiveness of the classification based on individual profile parameters is unsatisfactory [11, 23] (depending on the selected parameter, the effectiveness gained is 60%–70% for one of the classes and considerable worse for the others).

Combined utilization of various parameters is much more effective: the classification effectiveness in all classes that were under consideration is then increased and equalized [11], [12]. It can be implemented basing on voting or weighted voting methods, or hierarchical methods. Depending on the class, the classification effectiveness of the voting methods is in the interval of 50% to 97%. The classification effectiveness of the hierarchical methods ranges from 77% to 96%.

However, such classification may not be enough selective. It is then necessary to measure the inter-axle distances. Taking into account this parameter improves considerably the classification selectiveness although requires the vehicle speed to be measured. In such a system also vehicle length is measured and a trailer is detected. Such classification is performed in the measuring system equipped with single inductive loop sensor and two piezoelectric axle detectors. For such classification systems used in Europe *European Classification of Road Vehicle* is recommended, which

distinguishes 13 classes. The axle distance measured using the piezoelectric axle sensors is strongly affected by the vehicle mass, because the length of the tire footprint depends on the axle load.

The erroneous classification of the vehicle on the base of its axle distances occurs first of all in the population of two-axle vehicles. The error in the distance measurement may cause the classification of the vehicle to one of the neighbor class, for instance the car is classified as the van and so. The estimated probability of erroneous classification of two-axle vehicles is presented in Table II.

TABLE II
ESTIMATED PROBABILITY OF THE ERRONEOUS CLASSIFICATION OF TWO-AXLE VEHICLES, BASING ON THE AXLE DISTANCE MEASUREMENT

Vehicle class	Probability of the erroneous classification (%)
Car classified as van	0.24
Van classified as car	4.8
Van classified as lorry	0.5
Lorry classified as van	1.2
Lorry classified as bus	5.2
Bus classified as lorry	46.0

VI. WEIGH IN MOTION SYSTEMS (WIM)

The term *weigh-in-motion* (WIM) means a process of measuring the dynamic wheel forces of a moving vehicle onto the road and estimating the corresponding static loads or/and total vehicle weight. The lack of significant limitations posed on the vehicle speed is a characteristic feature of such weighing systems. At present, the WIM systems complement the static vehicle weighing stations, playing the role of preselection systems. Limited accuracy of these systems is the reason of such situation.

Classic WIM preselection systems are based on an inductive sensor and two load sensors (Fig. 1c). Such a system configuration allows estimation of the static loads of individual axles and total weight. Moreover it provides classification of the vehicle, based on the number of its axles. The pavement temperature measurement is necessary for correction of weighing results, which depend on the thermal and mechanical properties of the pavement and sensors.

The basic reasons limiting the accuracy of the WIM systems are as follows:

- vertical balancing of the suspended mass (car body) and hopping of the unsuspended mass (wheels),
- influence of the temperature on the sensitivity of the piezoelectric load sensors,
- changes in sensitivity along the sensors.

The load exerted by the wheels of weighed vehicle on the road surface contains both the constant, as well as dynamic components. The amplitude and spectral properties of the load dynamic component depend on the specific vehicle mechanical parameters, its speed and roughness of the road surface. The load signal is sampled at the moments when the weighed axle is passed over the load sensor (Fig. 5). It means that in preselective WIM systems equipped with two load sensors the axle static load is estimated on the base of two samples only. The High Speed WIM systems equipped with two load sensors allow estimation of the total weight of a moving vehicle with error not less than 10–15%.

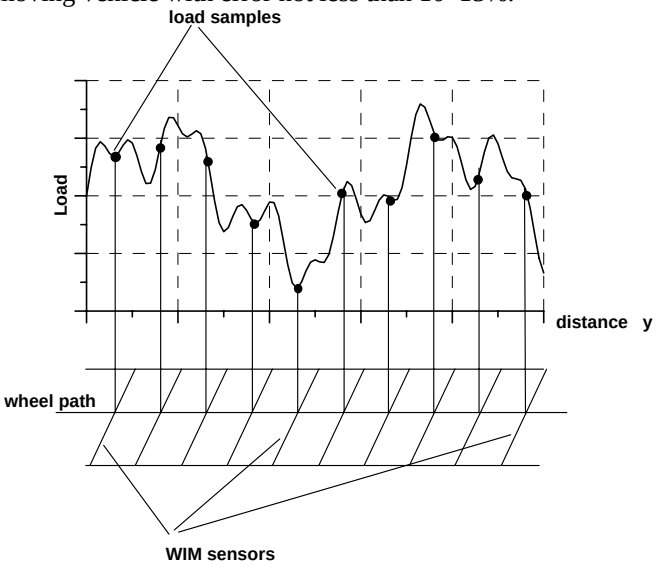


Fig. 5. Sampling of the load force on the MS-WIM site

The piezoelectric, strip sensors are one of the load sensors frequently used in WIM systems. They are mounted under the asphalt surface. Than asphalt transfers the load from the wheel to the sensors. Changes in the asphalt mechanical parameters vs. temperature cause correlated changes in the sensor sensitivity. From this point of view the quartz load sensors have better metrological properties but are far more expensive. Fig. 6. presents the temperature characteristic of the WIM system equipped with piezoelectric load sensors.

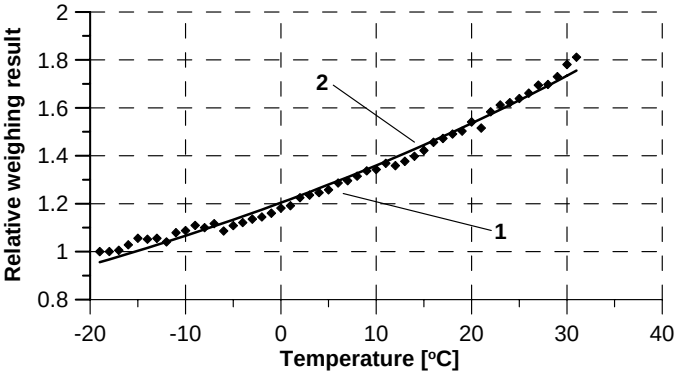


Fig. 6. Weighing result vs. asphalt temperature

This characteristic is very well described by model (10).

$$C(Ta) = k_1 \cdot 10^{w1(To-Ta)} \qquad (10)$$

where: k_1 = 3.8702, $w1$ = -0.0053 [1/°C]–constant coefficients, To=45 [°C]–reference temperature, Ta- current asphalt temperature.

Change in the sensitivity along the load sensor is the third reason limiting the accuracy of the WIM systems. The sensitivity of the exemplary piezoelectric load sensor is presented in Fig. 7. Its sensitivity changes in the range ±7.5%. It means that the weighing results of the same vehicle passing through the weighing site many times may differ in these limits.

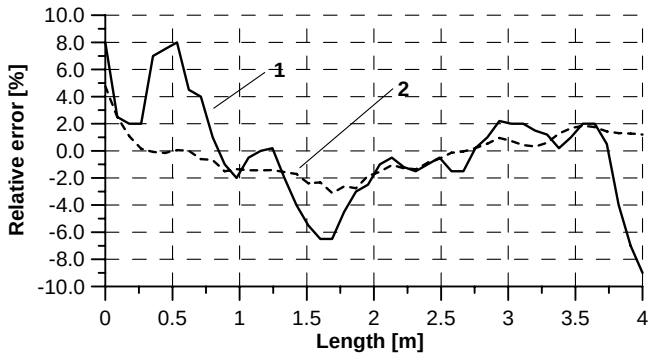


Fig. 7. Internal errors of the exemplary piezoelectric load sensor (1) and mean error of the set of 16 sensors (2)

The improvement of measuring accuracy of the total weight and static axle loads of vehicles moving with road speed up to the acceptable range of a few percent is possible when building multi-sensor weigh-in-motion systems (MS-WIM), developing static load estimation algorithms, and applying suitable methods for calibrating these systems. The greater number of load signal samples (number of samples corresponds to sensor number) allows limiting the influence of the dynamic component, as well as sensitivity change along the sensors.

Equation (11) describes the measurement result W_y obtained from single load sensor placed at the specific point y.

$$W_y = P(t) + e_y = P_0 + F(t) + e_y \tag{11}$$

where: $P(t) = P_y$ - load force at the time, in which axle crosses point y, P_0 - static load of the axle, $F(t)$ - load dynamic component, e_y - internal error of the sensor.

In WIM system equipped with p load sensors, the static load is frequently estimated as a mean value:

$$W_{ms} = \frac{1}{p}\sum_{i=1}^{p} W_i = P_0 + \frac{1}{p}\sum_{i=1}^{p} F(t_i) + \frac{1}{p}\sum_{i=1}^{p} e_i \tag{12}$$

where: W_i - load measured at *i*-th sensor, e_i - internal error of the i^{th} sensor, t_i - time moment at which weighed axle passes i^{th} sensor.

It is fulfilled under additional condition, that mean value of the dynamic component on sufficiently long distance is equal to zero.

The mean characteristic of the set of 16 sensors is presented in Fig. 7. The mean error was reduced few times and is in the range ±3%.

In the same way proper sampling of the load signal also reduces the dynamic component $F(t)$. Designing the MS-WIM system depends on the choice of p coordinates y_i.

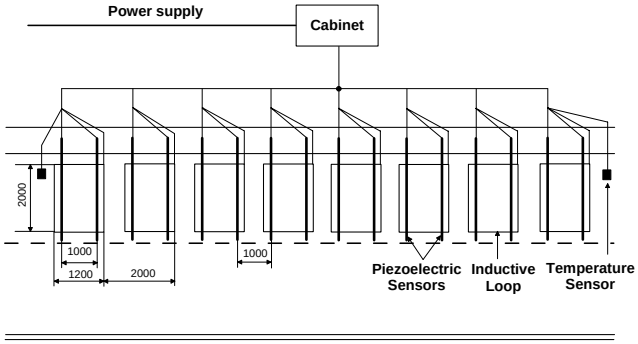


Fig. 8. Scheme of the MS-WIM system

The MS-WIM system was designed and constructed in Department of Measurement and Instrumentation, AGH-UST in Krakow (Poland). Its scheme is presented in Fig. 8. The system is equipped with 8 inductive loop sensors, 16 piezoelectric load sensors and 2 temperature sensors. The following algorithms were adopted for estimation of the static load of successive axles and total mass: mean value (**Mean**), usually calculated from the measurement results of instantaneous load on successive sensors, maximum likelihood estimator (**ML**) [3], modified nonlinear least-squares estimate (**ML/NLS).** The modification depends on connection between the ML estimator, generating an initial estimate of the parameters to be determined and the nonlinear least-squares estimator.

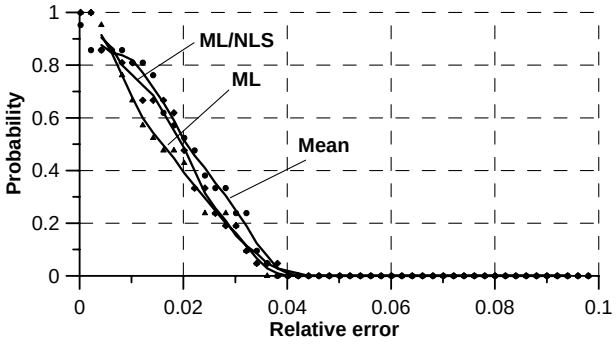


Fig. 9. Reliability characteristic of the WIM system. These characteristics correspond to different algorithms of static load estimation: mean value of the measurement results of separate sensors (Mean), maximum likelihood estimator (ML), joined maximum likelihood and nonlinear least square estimators (ML/NLS)

The metrological properties of the system were experimentally examined and evaluated using characteristic (13).

$$\Pr(\delta) = 1 - P(\delta) \tag{13}$$

where: $\delta = \left|\left(\hat{P}_0 - P_0\right)/P_0\right|$ is the absolute value of the relative estimation error of the static component P_0, $\hat{P}_0$ is an estimate of the static component, $P(\delta)$ is the cumulative probability distribution function of error δ.

This characteristic specifies the probability of occurrence of error greater than δ and it is called reliability characteristic. The characteristic (13) determined for described MS-WIM system is presented in Fig. 9. The conclusion, which could be draw out from the obtained results says, that none vehicle was weighed with the error greater than 4%.

VII. CONCLUSIONS

The basic problems corresponding to the road traffic measurements have been addressed in the paper. In correlation with the growth of vehicle transport the research projects are still conducted and new technical tools are offered for practical applications in this area. A lot of effort has been put into the solution of the following problems corresponding to the traffic measurements and monitoring during last years:

- new technologies of the traffic detectors and sensors (wireless sensors),
- design of the new data processing algorithms, using the classical sensors for obtaining the new traffic parameters, which were not available in classical solutions (estimation of the individual vehicle speed in the system with single loop detector, counting the axles on the base of vehicle magnetic signature),
- design of the MS-WIM systems allowing to estimate the static axle load and vehicle total mass with uncertainty limited to several percentage (multi-sensor site with uniformly distributed WIM sensors, more sophisticated algorithms of static load estimation),
- automatic weighing and charging of the overloaded vehicles,
- design of the traffic monitoring and information system that visually conveys the traffic flow information to general public through the client/server computer architecture.

REFERENCES

[1] Athol P. „Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream", Highway Research Record, Vol. 72, 1965.

[2] Cebon D., *Handbook of Vehicle-Road Interaction*, Swets&Zeitlinger B.V., Lisse, the Netherlands 1999.

[3] Cebon D.: "Design of multiple-sensor weigh-in-motion systems", Journal of Automobile Engineering, Proc. I. Mech. E., 204, 1990, pp. 133 – 144.

[4] Cebon D., Winkler CB.; „Multiple-Sensor WIM: Theory and experiments", Transportation Research Record, TRB, 1311, 1991, pp. 70 -78.

[5] Dailey D.J.; "Travel Time Estimates Using a Series of Single Loop Volume and Occupancy Measurements", Transport Research Board 76th Annual Meeting, Washington, D.C., 1997.

[6] Dailey D.J., Harn P., Lin P.; *ITS Data Fusion*, Final research report, Research project T9903, Department of Transportation, Washington State Transportation Commission, April 1996.

[7] Dolcemascolo V., Jacob B.: "Multiple sensor Weigh-In-Motion: Optimal Design and Experimental Study", Pre-proceedings of 2nd European Conference of Weigh in Motion of Road Vehicles, Lisbon, 14/16 September 1998, pp.129-138.

[8] Dukkipati R.V.: *Vehicle Dynamics*, Alpha Science International Ltd. 2000.

[9] Gajda J., Sroka R., Stencel M., Żegleń T.; "Measurement of Road Traffic Parameters Using an Inductive Single-Loop Detector", 9th International Symposium on Electrical Instruments in Industry, Glasgow 1997.

[10] Gajda J., Stencel M.; "Determination of Road Vehicle Types Using an Inductive Loop Detector", XIV IMEKO World Congress, Tampere, Finland, 1997.

[11] Gajda J., Sroka R.: "Vehicle classification by parametric identification of the measured signals", Proceedings of XVI IMEKO World Congress, Vienna 2000, vol. IX, pp. 199-204.

[12] Gajda J., Sroka R., Stencel M., Wajda A., Zeglen T.: "A vehicle classification based on inductive loop detectors", Proceedings of the 18th IEEE, IMTC, Budapest 2001, pp. 460-464.

[13] Gajda J.; "Statistical calibration of WIM systems", Scientific Series of Rzeszów Politechnic, Electrotechnic, nr 27, Rzeszów, 2004r, (in Polish).

[14] Gajda J., Burnos P.: "Self-calibration of the weigh-in-motion systems". Proceedings of XV Symposium Modelling and Simulation of Measurement Systems, Krynica 2005, (in Polish).

[15] Gonzales A., Papagiannakis A.T., O'Brien E.: *Evaluation of an Artificial Neural Network Technique Applied to Multiple Sensor Weigh-n-Motion Systems*, University College Dublin, Ireland.

[16] Huhtala M.: *Factors Affecting Calibration Effectiveness*, Proceedings of the Final Symposium of the Project WAVE, Paris 1999.

[17] Klein L.A.: *Sensor Technologies and Data Requirements for ITS*, Artech House, London 2001.

[18] Karlsson B.: "Fuzzy Measures for Sensor Data Fusion in Industrial Recycling", Measurement Science and Technology, Vol. 9, 1998, pp. 907-912.

[19] Ki Y.K., Baik D.K.: "Vehicle Classification Algorithm for Loop Detectors using Neural Network", IEEE Transactions on Vehicular Technology, vol. 55, Nr 6, 2006, pp. 1704-1711.

[20] Mangeas M., Glaser S., Dolcemascolo V.: "Neural networks estimation of truck static weights by fusing weight-in-motion data", Proceedings of Eurofusion 2000.

[21] Maerivoet S., De Moor B.; "Traffic Flow Theory", internet paper, 2006.

[22] Stanczyk D.: "New Calibration Procedure by Axle Rank", Proceedings of the Final Symposium of the Project WAVE, Paris 1999, pp. 307-316.

[23] Sun C., Ritchie G., Tsai K., Jayakrishnan R.,; "Use of vehicle signature analysis and lexicographic optimisation for vehicle reidentification on freeways". Transportation Research Part C 7 (1999), pp. 167-185.

[24] Wang Y., Nihan N.; „Can Single-Loop Detectors Do the Work of Dual-Loop Detectors?", internet paper.

[25] *Weigh-in Motion of Road Vehicle*, Final Report of COST 323 action, Ver. 3.0, 1999.

[26] *Weigh-In-Motion of Axles and Vehicles for Europe (WAVE)*, General Report of 4th Programme Transport, Laboratoire Central des Pontes et Chaussees, 2001.

[27] Zhang X., Wang Y., Nihan N.; „Monitoring a Freeway Network in Real Time Using Single - Loop . Detectors", Int. J. Vehicle Information and Communication Systems, Vol. 1, Nos. 1/2 2005.

SESSION TST I

Telecommunications Systems and Technology I

Performance of SOVA-based Turbo Decoders with Autostopping

Nikolay T. Kostov[1], Slava M. Yordanova[2] , Ginka K. Marinova[3] and Nikolay A. Velikov[4]

***Abstract* - A class of convolutional turbo codes with soft output Viterbi algorithm (SOVA) decoding is considered in this paper. Short-framed SOVA-based turbo decoders with autostopping on a Gaussian environment are studied. Bit error rate (BER), frame error rate (FER) and average number of decoding iterations versus the signal-to-noise ratio (SNR) are presented for the considered turbo decoders. The obtained results could be of interest regarding applications of these decoders in low-complexity, real-time digital communications services.**

***Keyword* – Turbo codes, soft output Viterbi algorithm (SOVA) decoding, simulations.**

I. INTRODUCTION

Turbo codes (TC) are a relatively new class of error control technique that was proposed by Berrou et. al. [1], [2]. They have been acknowledged as an extremely powerful coding scheme that achieves near-Shannon limit performance. The original rate $R_c = 0.5$ convolutional turbo code [1], [2] require a signal-to-noise ratio (SNR) of $E_b / N_0 = 0.7$ dB for bit error rate (BER) of 10^{-5} which is only 0.7 dB above the Shannon limit. However, the excellent performance of the original turbo code was achieved for a data frame size of 65536 bits and 18 decoding iterations per data frame. Because of prohibitively long latency involved with such large frame sizes, the original turbo code is not well suited for real-time multimedia communications.

There are many factors that affect the performance of TC. The most important parameters on TC's performance are the interleaver size and the decoding algorithm [1]-[3]. As the interleaver size increases, performance improves. For example, the original TC [1], [2] gains more than 1 dB at a bit error rate (BER) of 10^{-5} as the interleaver size increases from 1024 bits to 4096 bits. However, as the interleaver size increases so does the overall encoding/decoding latency. Soft output Viterbi algorithm (SOVA)-based decoders are attractive for practical implementations due to their low complexity and relatively high speed.

[1]Nikolay T. Kostov is with the Technical University of Varna, Department of Radio engineering, Studentska Street 1, Varna 9010, Bulgaria, E-mail: n_kostov@mail.bg

[2]Slava M. Yordanova is with the Technical University of Varna, Department of Computer Science and Technologies, Studentska Street 1, Varna 9010, Bulgaria, E-mail:slava_yordanova@mail.bg

[3]Ginka K. Marinova is with the Technical University of Varna, Department of Computer Science and Technologies, Studentska Street 1, Varna 9010, Bulgaria, E-mail: gin_kaleva@abv.bg

[4]Nikolay A. Velikov is with the Naval Academy "N. Vaptzarov" of Varna, Department of Electronics, Studentska Street 1, Varna 9010, Bulgaria, E-mail:na_velikov@mail.bg

However, there is a gap of 0.5 dB or even higher between their performance and that of maximum *a posteriori* (MAP)-based turbo decoders. Thus, TC's possess an inherent tradeoff between performance and latency as well as between performance and decoding complexity.

In the present paper short-framed SOVA-based turbo decoders with autostopping on a Gaussian environment are studied. The considered system model along with the obtained results is presented in the next sections.

II. SYSTEM MODEL

Let us now consider the encoding/decoding processes associated with the considered SOVA-based turbo decoders. The data frame to be transmitted is first encoded by a punctured TC using two identical recursive systematic convolutional codes. A uniform pseudorandom interleaver between convolutional codes is used. For an additive white Gaussian noise (AWGN) channel model the input/output relationship can be expressed as

$$r_i = \sqrt{E_s}\,(2b_i - 1) + n_i\,, \tag{1}$$

where r_i is the received symbol, E_s is the energy per code symbol, $(2b_i - 1)$ is the binary phase shift keying modulated code symbol and n_i is a zero-mean Gaussian variable with variance $\sigma^2 = N_0 / 2$.

The log-likelihood ratio (LLR) L_i at the output of a SOVA-based decoder can be represented as [3]

$$L_i = y_i + g_i + l_i\,, \tag{2}$$

where $y_i = \dfrac{4E_s}{N_0} r_i$ is the weighted channel observation, g_i is the *a priori* information and l_i is the so-called extrinsic information gained by the current stage of decoding. Both of the two SOVA-based elementary decoders use SOVA to form an estimate of the LLR of each bit encoded by the TC. The essence of SOVA is finding the most likely transmitted sequence of bits along with reliability values for the bits [4], [5]. Let us define the likelihood ratio or "soft" value of the binary path decision at time *i* as

$$\Delta_i^0 = \frac{1}{2}\left(M_i^{m_0} - M_i^{\tilde{m}_0}\right), \tag{3}$$

where $M_i^{m_0}$ and $M_i^{\tilde{m}_0}$ are the path metrics of the survivor and competitor path, respectively. Now, the SOVA output LLR of the δ-delayed decision $\hat{b}_{i-\delta}$ can be expressed as

$$L(\hat{b}_{i-\delta}) \approx \hat{b}_{i-\delta} \cdot \min_{l=0,\ldots,\delta} \Delta_i^l . \quad (4)$$

For detailed explanation of SOVA see [4], [5]. The extrinsic information from the first elementary decoder is used as *a priori* information by the second elementary decoder during the next halfiteration and so on. After a predetermined maximum number of iterations, the final estimate of the message bits $\hat{a}_i, i = 1,\ldots,M$, is found by hard-limiting the output of one of the elementary decoders:

$$\hat{a}_i = \begin{cases} 1 \text{ if } L_i^2 \geq 0 \\ 0 \text{ if } L_i^2 < 0 \end{cases} . \quad (5)$$

It is possible to improve both the average decoding speed and power consumption of the turbo decoder if an "early stopping" rule is applied. Various "early stopping" rules for TCs was considered in [6] – [8]. Most of the presented results are for optimum MAP-based turbo decoders with large interleavers. In the present work we further study the performance of SOVA-based turbo decoders with small interleavers and simple "early stopping" rules.

III. The Simulation Code

Performance of two SOVA-based TCs with data frame sizes of N = 128 bits and N = 256 bits was studied through simulations in MATLAB. The first TC, denoted as TC1, uses convolutional codes with generators $g_0 = (07)_8$ and $g_1 = (05)_8$ (in octal notation) and the second TC, denoted as TC2, uses convolutional codes with generators $g_0 = (13)_8$ and $g_1 = (15)_8$. In fact, the considered TC2 is a punctured version of the Third Generation Partnership Project (3GPP) turbo code with overall rate $R_c = 0.5$. The information bits are obtained using uniformly distributed pseudorandom data. A uniform pseudorandom interleaver is used in both TC1 and TC2. The channel output is unquantized throughout and obtained according to (1). The turbo decoding is based on SOVA as described above. The SOVA decoding window was set to 24 bits. Soft iterative decoding with up to 12 iterations and a hard-decision or a soft-decision autostopping rule (e.g. "early stopping") is used to decode the TC schemes. The hard-decision "early stopping" rule is as follows [9]: stop iterations if both elementary decoders output identical sets of hard-limited extrinsic values at a given full iteration, e.g. stop iterations if

$$\text{sign}(Le_{i,1}) = \text{sign}(Le_{i,2}), \quad \forall i, 1 \leq i \leq N \ , \quad (6)$$

where $\text{sign}(Le_{i,1})$ and $\text{sign}(Le_{i,2})$ are the *i*-th hard-limited extrinsic values at a given iteration of the first and second SOVA-based decoder, respectively. The soft-decision rule is based on comparing a metric on extrinsic values reliability (the minimum absolute value) with a threshold value θ, e.g. stop iterations if

$$\min_{1 \leq i \leq N} [|Le_{i,1}|, |Le_{i,2}|] > \theta . \quad (7)$$

The set of these threshold values as a function of the channel signal-to-noise ratio (SNR) is obtained through simulations and shown in Table I.

TABLE I
The minimum absolute values of extrinsic information for early stopping of TC1 and TC2

	N = 128 bits	N = 256 bits
SNR = 1 dB	min(|Le|) = 0.2	min(|Le|) = 0.25
SNR = 1.5 dB	min(|Le|) = 0.33	min(|Le|) = 0.5
SNR = 2 dB	min(|Le|) = 0.5	min(|Le|) = 1
SNR = 2.5 dB	min(|Le|) = 1	min(|Le|) = 1.5
SNR = 3 dB	min(|Le|) = 1.5	min(|Le|) = 2.5

The MATLAB simulation code of the considered turbo codes is given below.

```
clear; close all;
% initialization
rand('state', 123789);
randn('state', 432654);
EbN0db = 2.0;
Ntot = 256;
g = [1 1 1; 1 0 1]; [n,K] = size(g); m = K - 1; nstates = 2^m;
rate = 1/2;
nframe = 0;  frerr_max=50; niter_max = 12;
nferr = 0; final_err = 0; sum_iter = 0;
BER = 0; FER = 0; avr_iter = 0;
EbN0lin = 10^(EbN0db/10);   % convert Eb/N0 from db
Lc = 4* EbN0lin *rate;     % reliability value of the channel
sigma = 1/sqrt(2*rate* EbN0lin);
% simulation core
  while (nframe < frerr_max)
     nframe = nframe + 1;
     x = round(rand(1, Ntot-m));  % info bits
     [temp, gamma] = sort(rand(1,Ntot)); % interleaver
     enc_output = turbo_code(x, g, gamma); % encoder
     r = enc_output+sigma*randn(1,Ntot*2); % received bits
     y = demultiplex(r,gamma);
     rec = 0.5*Lc*y; % scaling the received bits
     Le = zeros(1,Ntot); % initialize the extrinsic information

     for iter = 1:niter_max
        % Decoder one
        La(gamma) = Le;  % a priori info
        Lall = sova_dec(rec(1,:), g, La, 1);
        Le = Lall - 2*rec(1,1:2:2*Ntot) - La;  % extrinsic info
        Le = Le*0.7; % scaling
        Le1 = Le(gamma);
        % Decoder two
        La = Le(gamma);  % a priori info
        Lall = sova_dec(rec(2,:), g, La, 2);
        Le = Lall - 2*rec(2,1:2:2*Ntot) - La;  % extrinsic info
        Le = Le*0.7; %scaling
        Le2 = (Le);
        % Estimate the info bits and errors
```

```
        a(gamma) = (sign(Lall)+1)/2;
        err = length(find(a(1:Ntot-m)~=x));
       % stop rule 1 – a "hard decision" rile
       % stop = nnz(sign(Le1)~=sign(Le2));
       % stop rule 2 – a "soft decision" rule
         stop1 = min(abs(Le1));  stop2 = min(abs(Le2));
         stop = min([stop1 stop2]);
         threshold = 1;

       if stop > threshold
            break
         end

    end  % iter

   if err > 0
      nferr = nferr + 1;
    end

    % BER
    final_err =  final_err + err;
    BER =   final_err/(nframe*(Ntot-m));

    % FER
    FER = nferr/nframe;

   % Average number of iterations
    sum_iter =  sum_iter + iter;
    avr_iter =  sum_iter/nframe;

  end   % while

% Save results
   save turbo_sim EbN0db Ntot g BER FER avr_iter
```

In the MATLAB code above the initialization stage includes definitions of the SNR in decibels (denoted as **EbN0db**), the interleaver size (denoted as **Ntot**), the convolutional code generators (denotes as **g**) etc. For a given SNR the simulation is executed by means of a **while-end** construction. The simulation runs until a predetermined number of frame errors (denoted as **frerr_max**) is encountered. The inner loop (e.g. the **for-end** construction) is used for iterative SOVA-based turbo decoding and runs until the "early stopping" rule is satisfied (thanks to the **if-break-end** construction) or the maximum number of decoding iterations (denoted as **niter_max**) is executed. Finally the bit error rate (BER), frame error rate (FER) and average number of decoding iterations are obtained.

IV. Performance Results

Performance results of the considered TC schemes are shown in Table II, Fig.1 end Fig.2. In Table II the required SNR for a BER of 10^{-4} is given. The BER and FER of the considered TC1 scheme are shown in Fig. 1 and the corresponding performance of the TC2 scheme is shown in Fig.2. A typical performance of the average number of decoding iterations versus the signal-to-noise ratio (SNR) for both the hard-decision and soft-decision autostopping rules applied in TC2 is shown in Fig.3 (for the case of 256 bits frame size). In Fig.3 the "soft decision+" denotes performance of the SOVA-based turbo decoding with a stopping threshold values 50% above the nominal values listed in Table I. There was no noticeable performance difference (in terms of BER and FER) between the considered hard-decision and soft-decision autostopping rules, so in Fig.1 and Fig.2 results only for the hard decision rule are presented.

TABLE II

The required SNR for a $BER \approx 10^{-4}$ of TC1 and TC2

	TC 1	TC 2
N = 128 bits	3.7 dB	3.4 dB
N = 256 bits	3.25 dB	3 dB

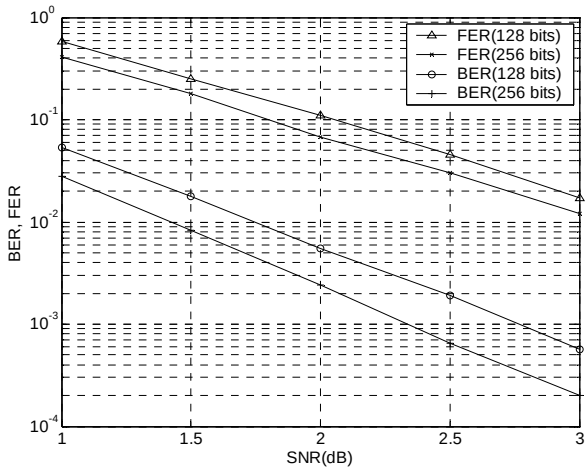


Fig.1. BER and FER of SOVA-based TC1 with 128 and 256 bits frame sizes.

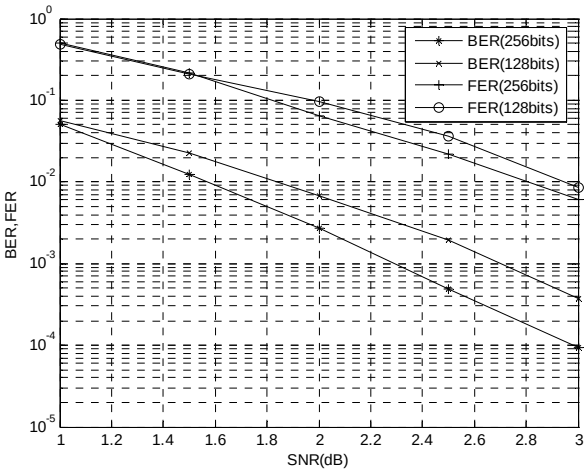


Fig.2. BER and FER of SOVA-based TC2 with 128 and 256 bits frame sizes.

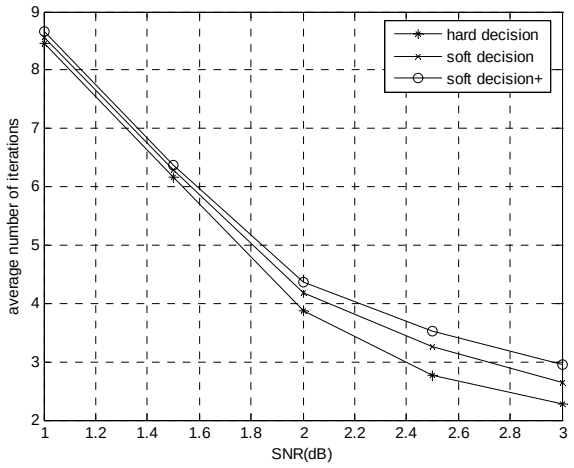


Fig.3. Average number of decoding iterations versus the SNR for TC2 with 256 bits frame size.

V. CONCLUSION

Short-framed SOVA-based turbo decoders with autostopping on a Gaussian environment are studied in this paper. Performance results in terms of BER, FER and average number of iterations versus the SNR are presented. As could be expected the performance of both TC1 and TC2 is dominated by the data frame size (e.g. by the interleaver size). The obtained results show no noticeable performance degradation of the considered "early stopped" SOVA-based decoders compared to the turbo decoding with a fixed number of iterations (at least for $BER's < 10^{-5}$). According to the Fig.3 both the hard-decision and soft-decision autostopping rules results in a significant reduction of average number of iterations performed of the turbo decoder. Thus, it is possible to improve both the average decoding speed and power consumption of the considered SOVA-based turbo decoders. The considered hard-decision autostopping rule should be preferred to soft-decision autostopping rule on non-stationary channels (like fading channels) because it is not SNR-dependent.

REFERENCES

[1] C. Berrou, A. Glavieux and P. Thitimasjshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes", In Proceedings of IEEE International Conference on Communications, Geneva (Switzerland), pp. 1064-1070, 1993.

[2] C. Berrou, and A. Glavieux, "Near optimum error-correcting coding and decoding: Turbo-codes", IEEE Transactions on Communications, vol. 44, no. 10, pp. 1261–1272, 1996.

[3] S. Benedetto, and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes", IEEE Transactions on Information Theory, vol. 42, no. 2, pp. 409–428, 1996.

[4] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes", IEEE Transactions on Information Theory, vol. 42, no. 2, pp. 429–445, 1996.

[5] J. Hagenauer and P.Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications", In Proceedings of IEEE Global Communications Conference. Dallas (USA), pp. 47.1.1-47.1.7., 1989.

[6] A. Matache, S. Dolinar and and F. Pollara, "Stopping rules for turbo decoders", NASA TMO Progress report 42-142, pp. 1-22, August 15, 2000.

[7] R. Shao, S. Lin and M. Fossorier, " Two simple stopping criteria for turbo decoding", IEEE Transactions on Communications, vol. 47, pp. 1117–1120, 1999.

[8] A. Shibutani, H. Suda and F. Adachi, " Reducing average number of turbo decoding iterations", Electronics letters, vol. 35, pp. 701-702, 1999.

[9] N. Kostov, " Convolutional/single parity check turbo codes for multimedia communications" Journal of Radioengineering, vol.13, No 4, pp. 18-21, 2004.

The Blocking Probability Comparison for Some Routing and Wavelength Assignment Methods

Goran Z. Marković[1] and Vladanka Aćimović-Raspopović[2]

Abstract **– Optical networks employing wavelength division multiplexing (WDM) technique are becoming increasingly popular for high bandwidth connections. With the advent of the cross-connect technology, optical connections or lightpaths can be established fully in the optical domain without optical to electronic conversion at intermediate network nodes. Routing and wavelength assignment (RWA) are important issues in such networks. This work focuses on the RWA problem in optical networks with dynamic traffic demands. One of the primary design objectives of wavelength-routed all-optical networks in such situation is to minimize the blocking probability. Two routing algorithms together with four wavelength assignment methods are tested and compared according to the blocking probability performance metric. Simulations are performed on a given optical network topology with 8 network nodes in a case of Poisson traffic arrivals and exponential call holding times.**

Keywords **– Routing and wavelength assignment (RWA), optical network, lightpath, blocking probability**

I. INTRODUCTION

Recent advances in wavelength division multiplexing (WDM) and optical switching technique make it possible to consider the deployment of wavelength routing optical networks (WRON) that will provide backbone connectivity over a wide-area distances at very high data rates. A WRON consists of wavelength routers and the fiber links that interconnect them. Wavelength routers are optical switches capable of routing optical signal at a given wavelength from any input port to any output port, making it possible to establish end-to-end lightpaths. A lightpath is direct all-optical connection between two end nodes in physical network topology that is established without any intermediate electro-optic conversion. Such networks are also referred to as all-optical networks *(AON).*

To establish a lightpath in a WDM network, it is necessary to determine the route over which the lightpath should be established and the wavelength to be used on all the links along the route. This problem is known as the *routing and wavelength assignment* (RWA) problem. Routing and wavelength assignment requires that two lightpaths on a given link do not share the same wavelength. In addition, in networks without the wavelength converter in network nodes, lightpaths must satisfy the wavelength continuity constraint, that is, the same wavelength must be used on all the links along the selected route. The RWA algorithm is critically important to increase the efficiency of wavelength routed networks. It is responsible for selecting a suitable route and wavelength among the many possible choices for establishing the lightpath.

Acording to the traffic assumptions, the RWA problem can be classified into two types: the *static* RWA problem and the *dynamic* RWA problem. In the case of static RWA problem, the set of connections is known in advance and the problem is to set up lightpaths for the connections while minimizing network resources such as the number of wavelengths or the number of fibers. Alternatively, one may attempt to set up as many lightpaths as possible for a given number of wavelengths. Dynamic RWA tries to perform routing and wavelength assignment for connections that dynamically arrive to and depart from the network one by one in a random manner. The objective of dynamic RWA problem is typically to minimize the connection blocking probability, that is, the probability that a lightpath cannot be established in the network due to lack of resources (e.g., link capacity or free wavelengths). Blocking is highly undesirable in WRON networks due to the very high data rates carried on a single wavelength channel.

We consider dynamic RWA problem in this paper. Because of the real time nature of the problem, RWA algorithm in a dynamic traffic environment must be very simple. Since combined routing and wavelength assignment is a hard problem, a typical approach to designing efficient algorithms is to decouple the problem into two separate sub-problems: the routing problem and the wavelength assignment problem.We simulated and compared two routing algorithms: fixed routing and fixed-alternate routing, jointly with the four wavelength assignment methods, which will be described in more detail in the next section.

The rest of the paper is organized as follows. In Section II we describe the problem we study. In Section III, the simulation results related to the connection blocking probability for different routing and wavelength assignment methods and traffic scenarious are given. Finally, the section IV concludes the paper.

II. PROBLEM STATEMENT

In this paper, we consider a wavelength-routing network without the wavelength conversion operating in circuit-switched mode under dynamic traffic scenarious. When

[1]Goran Marković is with the Faculty of Traffic and Transport Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: g.markovic@sf.bg.ac.yu

[2]Vladanka Aćimović-Raspopović is with the Faculty of Traffic and Transport Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: v.acimovic@sf.bg.ac.yu

connections are established and released dynamically, RWA algorithms must be used to assign resources dynamically to the connection requests. In case when the network does not have sufficient resources to support a connection, the connection is *blocked*. In networks without the wavelength convertors, a connection may also be blocked if there does not exist any common wavelength on all the links along a path. The routing and wavelength assignment problem for dynamic traffic is called Dynamic Lightpath Establishment (DLE) problem. Given a physical network topology, the DLE problem can be defined as the RWA problem with maximization of accepted connection requests or equivalently with minimization of blocked requests in a given time period.

Our objective is to study the performances for some dynamic RWA algorithms, in terms of their call blocking probabilities. We considered two routing methods: *fixed* routing (FR) and *fixed-alternate* routing (FAR).

In *fixed* routing, for every source-destination *(s,d)* node pair a fixed route (typically the shortest path) is computed offline. When a connection request arrives, a free wavelength is searched along the corresponding route to establish the connection. The request is blocked if there does not exist any free wavelength along this route. This technique addresses the problem by minimizing the amount of resources used to establish a connection.

In *fixed-alternate* routing, for every *(s,d)* node pair, a set of *k* alternate paths ($k>1$) is computed off-line. When the connection request arrives, the candidate routes are searched in the sequential order and the first available (if it exists) is choosen for the establishing the lightpath between coresponding node pair. The candidate routes in our simulations are ordered according to the total number of their physical hops (links).

Let us discuss the considered wavelength assignment methods, according to the manner in which the wavelength list is ordered. For a given candidate path, the wavelengths are used in the order in which they appear in the list to find a free wavelength for the connection request. To evaluate the blocking probability of the considered routing algorithms, we used next four wavelength assignment techniques:

- *Random* (RN) - using random strategy, a set of available wavelengths that can be used to establish the connection is determined. After that, one wavelength is randomly selected with uniform probability distribution from this set.
- *First-fit* (FF) - in the first-fit scheme, all the wavelengths are numbered. The lowest numbered wavelength that can be used to establish a connection is used for the connection.
- *Most-used* (MU) - in this scheme, all the available wavelengths that can be used to establish a connection are considered. The wavelength that has been used the most in a network is selected for a lightpath. Usage can be defined either as the number of links in the network in which a wavelength is currently used or as the number of active connections using a wavelength.
- *Least used* (LU) – in this method wavelengths are tried in the order of increasing usage in a network.

III. Simulation results

The routing and wavelength assignment methods described in the previous section were applied to the 8 node optical WDM network topology, given in Fig. 1. We assumed that the network is single fiber, which means that each physical link has only one optical fiber for each direction. Also, we assumed that each fiber has the same available number of wavelengths W.

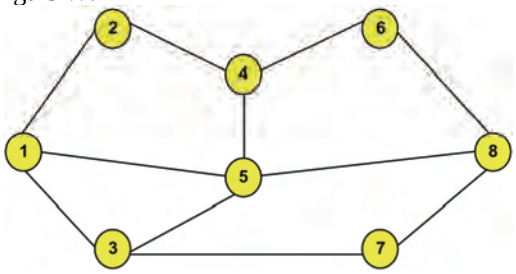


Fig. 1. The considered optical WDM network topology.

A sequence of lightpath requests arrives over time and each lightpath has a random holding time. These lightpaths need to be set up dynamically by choosing a route across the network connecting the source and destination node and assigning a free wavelength along the path. At a given time, only one lightpath can use a specific wavelength in a fiber. The existing lightpaths cannot be re-routed to accommodate the new lightpath requests until they are released. We consider a dynamic traffic model, in which the connection requests arrive according to Poisson distribution with mean arrival rate of λ requests per unit time. The connections have exponential holding time with mean $\bar{t} = 1/\mu$ time units. Thus, the total network load is equal $\rho = \lambda/\mu$ and is measured in Erlangs. The total number of connections simulated in the network during each simulation run is assumed to be 10000. The source-destination *(s,d)* nodes for each connection are chosen (with equal probability) from uniform distribution. In all performed simulations, we assumed that the mean connection holding time is $\bar{t} = 1$ time unit and we only varied the mean arrival rate λ as the number of connection requests per time unit.

Firstly, we will analyze the simulation results for the blocking probabilities in a case of fixed routing (FR) algorithm with different considered wavelength assignment methods applied. Fig. 2 illustrates the relative number of blocked requests per 10000 requests totally generated during the simulations in given network, in a case when network load is $\rho = 60\text{Erl}$ for different number of available wavelengths per fiber links. It can be seen that when the number of wavelengths is small ($W<8$), the performance of the four wavelength assignment schemes are nearly identical, indicating that the blocking probability is determined mainly by the limited resources and not by the wavelength assignment scheme implemented. However, as the number of

wavelengths increase (W>8), the blocking probability values decrease significantly. In such situation, it can also be seen that MU method outperforms the other considered schemes, but also that the FF method gives the similar results as the MU method. The LU method gives the worst performances in terms of call blocking probability.

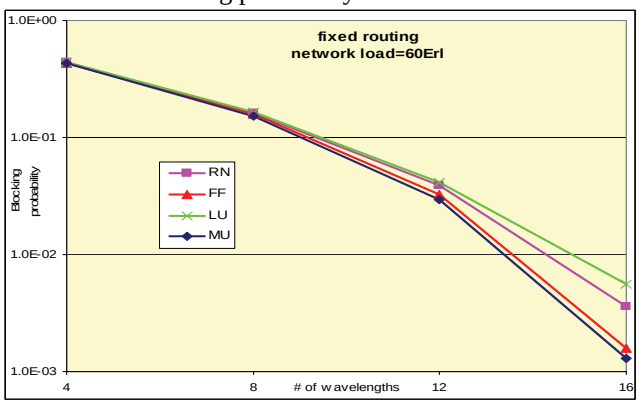


Fig. 2 The blocking probability for different wavelength assignment methods as a function of wavelength number -FR

Let us now consider the performances of fixed-alternate routing (FAR) algorithm. In Fig. 3 the simulation results for blocking probabilities are presented in optical network (Fig.1) for the case of network load value of $\rho = 60\text{Erl}$ and W=12 wavelengths, obtained by applying different wavelength assignment methods and considered routing algorithms with various number of alternate routes, beginning from *k*=1 (case of FR) to maximally *k*=5 (in a case of FAR).

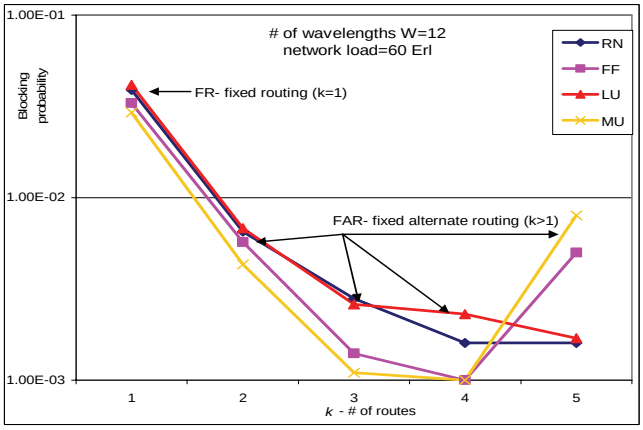


Fig. 3 The blocking probability as a function of the number of routes for various wavelength assignment methods – FR/FAR

The results given by Fig. 3 show that fixed routing (*k*=1) is significantly poorer compared to the fixed alternate routing (FAR) algorithm when the network performance is measured in terms of call blocking probability. It can also be seen, that the greatest improvements in blocking probabilities compared to fixed routing (*k*=1) could be achieved in the cases when the number of alternative routes is *k*=2 or at most *k*=3, but when the number of routes is larger than *k*=3, the improvements are not so significant. Moreover, if a number of alternate routes *k* is more than 4, the results for blocking probabilities can be worse, because in this case the longer candidate routes can be chosen and so more resources in a network could be consumed. Therefore, for each node pair *(s,d)* in considered network topology, it is adequate to predefine only few alternate routes in order to obtain much better performances compared to fixed routing algorithm.

When FR or FAR is used, with increasing the traffic load in network, the blocking probability of the network increases significantly. This can be seen from the results given in Fig. for different wavelength assignment methods in a case of FR and FAR algorithms.

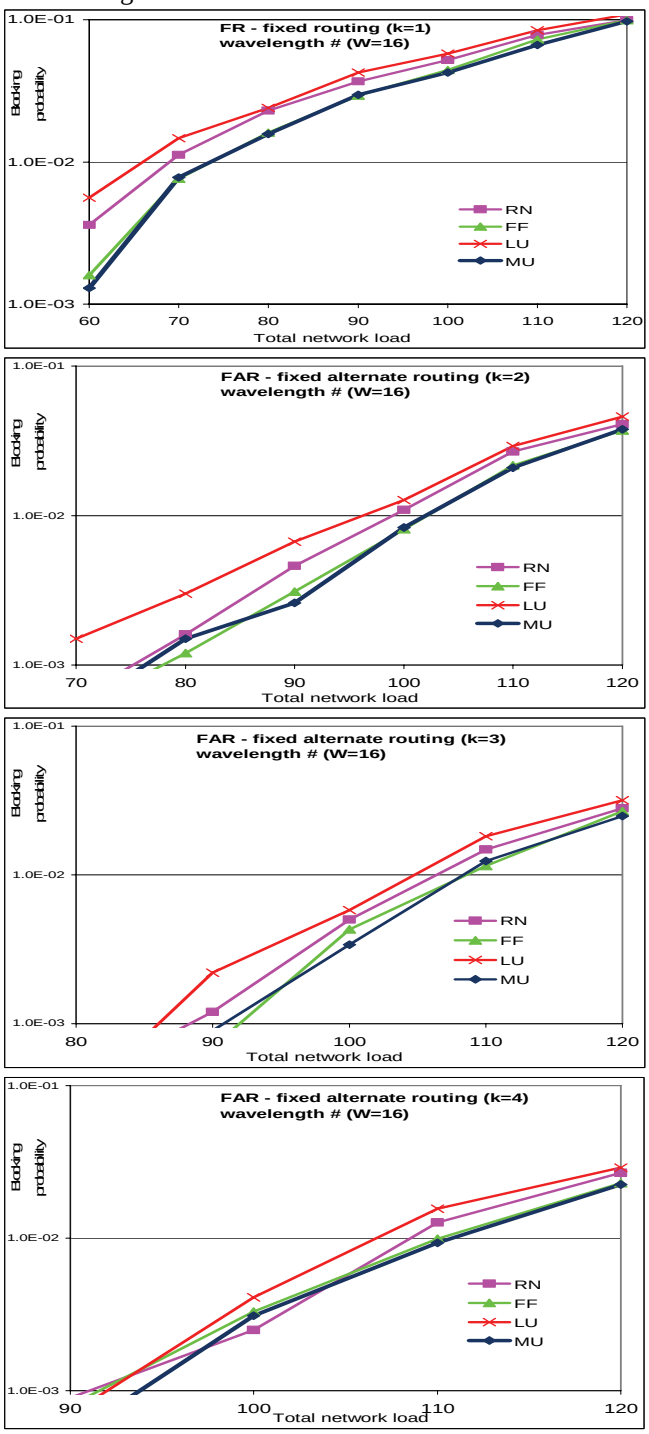


Fig. 4 The blocking probability for different wavelength assignment methods as a function of network load – FR/FAR

From Fig. 4 it can be also seen that the blocking probabilities (for the same network loads) in a case of FAR algorithm are much lower than those for FR algorithm.

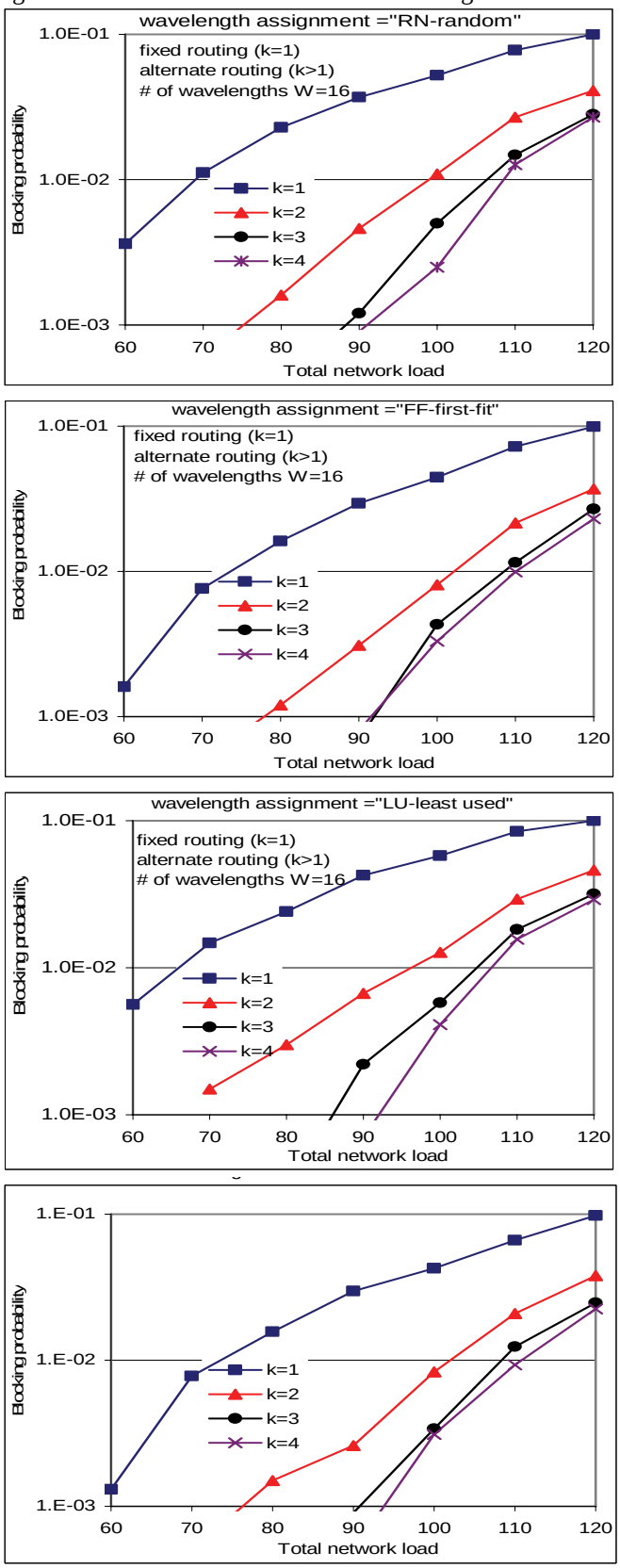


Fig. 5 Blocking probability vs. network load for different wavelength assignment techniques – FR/FAR

Fig. 5 depict the blocking probability versus network load for the considered optical WDM mesh network topology employing *random* (RN), *first-fit* (FF), *least-used* (LU) and *most-used* (MU) wavelength assignment techniques, when fixed and fixed-alternate routing algorithms are performed. The results for most used wavelength assignment are similar to the first fit wavelength assignment technique The random the least used techniques gives poorer performances than the most used and first fit wavelength assignment for different network loads.

IV. Conclusion

The routing and wavelength assignment procedure in a case of dynamic traffic demands has the goal to minimize the number of blocked connection requests in a given network. In this paper, we developed a software module to simulate two routing and four wavelength assignment methods in order to obtain the call blocking performance in a concrete optical WDM network topology. We performed a lot of simulations in order to compare the efficiencies of the considered RWA algorithms.

Results show that the fixed alternate routing with no more than three alternate routes performs much better than the fixed routing algorithm for various network loads. Also, the obtained results show that the most-used wavelength assignment technique outperforms the other three considered methods, but also the first-fit method is comparable to the most-used in terms of call blocking probability.

Acknowledgments

This paper resulted from the researching project *Planning, Forecasting and Pricing in Telecommunication Networks* that is supported by Serbian Ministry of Science and Environmental Protection (project designation TP-6106A).

References

[1] C. S. Ram Murthy, M. Gurusamy, *WDM Optical Networks - Concepts, Design and Algorithms*, Prentice Hall, 2002.

[2] V.Acimović-Raspopović, G.Marković »Routing in Optical WDM Networks«, Invited paper, The XI TELFOR Conference, Belgrade 2004, available at www.telfor.org

[3] G. Marković, V. Aćimović-Raspopović, "Optimal Solution for Routing and Wavelength Assignment Problem in Optical WDM Networks," *Proc. of ICEST Conference*, pp.289-292, June 2005.

[4] G.Marković, »A Contribution to Optimization of Telecommunication Network Resources Usage by Applying Dynamic Traffic Routing«, M. Sc Thesis, Belgrade 2002.

[5] M. Pioro, D. Medhi, *Routing, Flow, and Capacity Design in Ccommunication and Computer Networks*, Elsevier, 2004.

Turbo Codes in the CCSDS Standard for Wireless Data

Teodor B. Iliev[1] and Dimitar I. Radev[2]

***Abstract* – A turbo code has been included in the CCSDS channel coding standard for space telemetry. Many future missions with critical link budgets will benefit of its large coding gain. This paper will review the performance of the proposed Turbo Code configurations of the CCSDS telemetry channel coding standard, and we will discuss how this may be implemented in reconfigurable hardware.**

***Keywords* – Turbo code, Interleaver, Code rate, Convolutional code.**

I. INTRODUCTION

Turbo codes, introduced in 1993 [1], represent a great advancement in coding theory. Their excellent performances, especially at low and medium signal-to-noise ratios, have raised up an enormous interest for applications. Currently, even if many research issues are still open, the success of turbo codes is growing, and their introduction in many international standards is in progress (among them, the UMTS standard for third generation personal communications, and the DVB-T ETSI standard for Terrestrial Digital Video Broadcasting). This is also the case of the CCSDS (Consultative Committee for Space Data Systems) standard for space telemetry, which has often represented a benchmark for new technologies. Recently, the CCSDS channel coding standard has been updated to include turbo codes [2]. Turbo codes can now be employed as an alternative to the old standard codes: a (255,223) Reed-Solomon code, a 64-state rate-1/2 convolutional code, and their serial concatenation through an interleaver. A number of missions have already manifested their intend to use turbo codes, such as Rosetta, Lunarsat, and Mars Express. The BER/FER performances of the CCSDS turbo code have been largely studied: it has been pointed out that an additional coding gain of 2.5 dB can be achieved with respect to the serial concatenated scheme of the standard.

II. THE CCSDS TURBO CODE

The CCSDS Telemetry Channel Coding standard [3] uses a turbo code with two component codes with selectable rates and block lengths. The encoder for this code is shown in Figure 1. Two rate R=1/4 recursive convolutional encoders are used to generate R=1/2, 1/3, 1/4, and R=1/6 turbo codes. The two switching matrices combine these rates, where a solid circle means every symbol is taken, and an open circle means every other symbol is taken. The feedback polynomial $h_0 = D^4 + D + 1 = 23$ of both encoders is a primitive polynomial of degree 4, and the feedforward polynomials are $h_1 = 33$, $h_2 = 25$, and $h_3 = 37$. The various rates are achieved by using the connections shown on the right-hand side of Figure 1. For example, to achieve R=1/4 the four outputs are the systematic bit, the second and third parity bits from encoder 1, and the first parity bit from encoder 2.

Trellis termination is accomplished by first terminating code number 1, then code number 2, by equating the input bit to the feedback bit for four symbol clock ticks. This causes the encoders to be loaded with the all-zero state.

The interleaver is a block permutation interleaver that has good spreading factors and, unlike the S-random interleavers, has an algorithmic implementation and is scalable. The interleavers are fashioned according to Berrou's analytical algorithm [1]. The CCSDS standard allows five interleaver lengths $K_1 = 1784$, $K_2 = 3568$, $K_3 = 7136$, $K_4 = 8920$, and $K_5 = 16384$. The first four block lengths are chosen to be compatible with an outer Reed–Solomon code with interleaving depths of 1, 2, 4, and 5, respectively. The largest block length is allowed for those users requiring the most power efficiency and able to tolerate a larger decoder latency.

The corresponding codeblock lengths in bits, n=(k+4)/R, for the specified code rates are shown in Table 1.

TABLE I
CODEBLOCK LENGTHS FOR SUPPORTED CODE RATES

Information block	Codeblock length n			
length k	R=1/2	R=1/3	R=1/4	R=1/6
1784	3576	5364	7152	10728
3568	7144	10716	14288	21432
7136	14280	21420	28560	42840
8920	17848	26772	35696	53544
16384	32776	49164	65552	98328

[1]Teodor B. Iliev is with the Department of Communication Technique and Technologies, 8 Studentska Str., 7017 Rousse, Bulgaria, E-mail: tiliev@ecs.ru.acad.bg

[2] Dimitar I. Radev is with the Department of Communication Technique and Technologies, 8 Studentska Str., 7017 Rousse, Bulgaria, E-mail: dradev@abv.bg

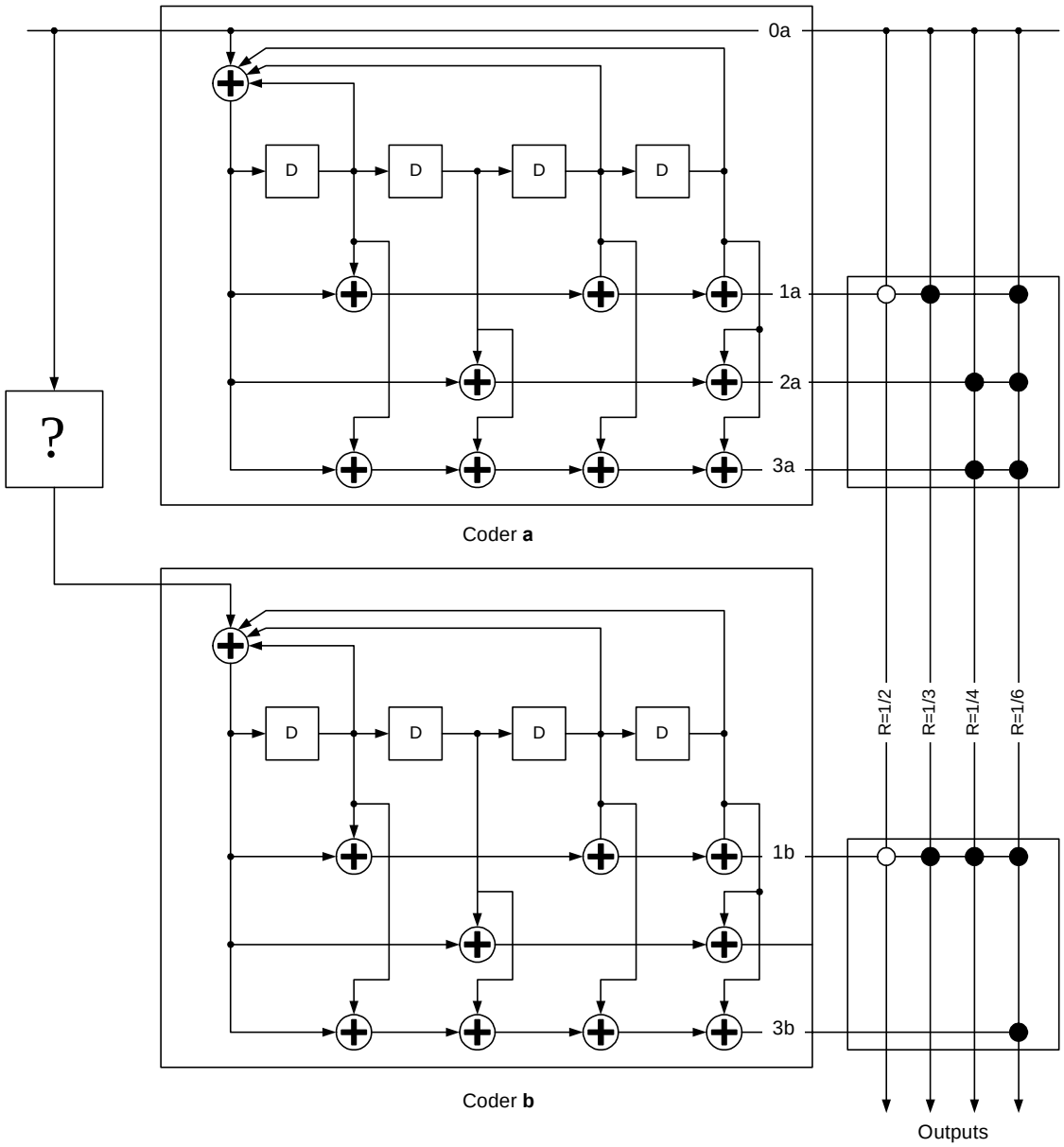


Fig. 1. Turbo Encoder Block Diagram

III. TURBO CODEBLOCK SPECIFICATION

Both component encoders in figure 1 are initialized with 0s in all registers, and both are run for a total of k+4 bit times, producing an output codeblock of (k+4)/r encoded symbols, where r is the nominal code rate. For the first k bit times, the input switches are in the lower position (as indicated in the figure) to receive input data. For the final 4 bit times, these switches move to the upper position to receive feedback from the shift registers. This feedback cancels the same feedback sent (unswitched) to the leftmost adder and causes all four registers to become filled with zeros after the final 4 bit times. Filling the registers with zeros is called terminating the trellis. During trellis termination the encoder continues to output nonzero encoded symbols. In particular, the 'systematic uncoded' output (line 'out 0a' in the figure) includes an extra 4 bits from the feedback line in addition to the k information bits.

In figure 1, the encoded symbols are multiplexed from top-to-bottom along the output line for the selected code rate to form the Turbo Codeblock. For the rate 1/3 code, the output sequence is (out 0a, out 1a, out 1b); for rate 1/4, the sequence is (out 0a, out 2a, out 3a, out 1b); for rate 1/6, the sequence is (out 0a, out 1a, out 2a, out 3a, out 1b, out 3b). These sequences are repeated for (k+4) bit times. For the rate 1/2 code, the output sequence is (out 0a, out 1a, out 0a, out 1b), repeated (k+4)/2 times. Note that this pattern implies that out 1b is the first to be punctured, out 1a is the second, and so forth. The turbo codeblocks constructed from these output

sequences are depicted in figure 2 for the four nominal code rates.[4]

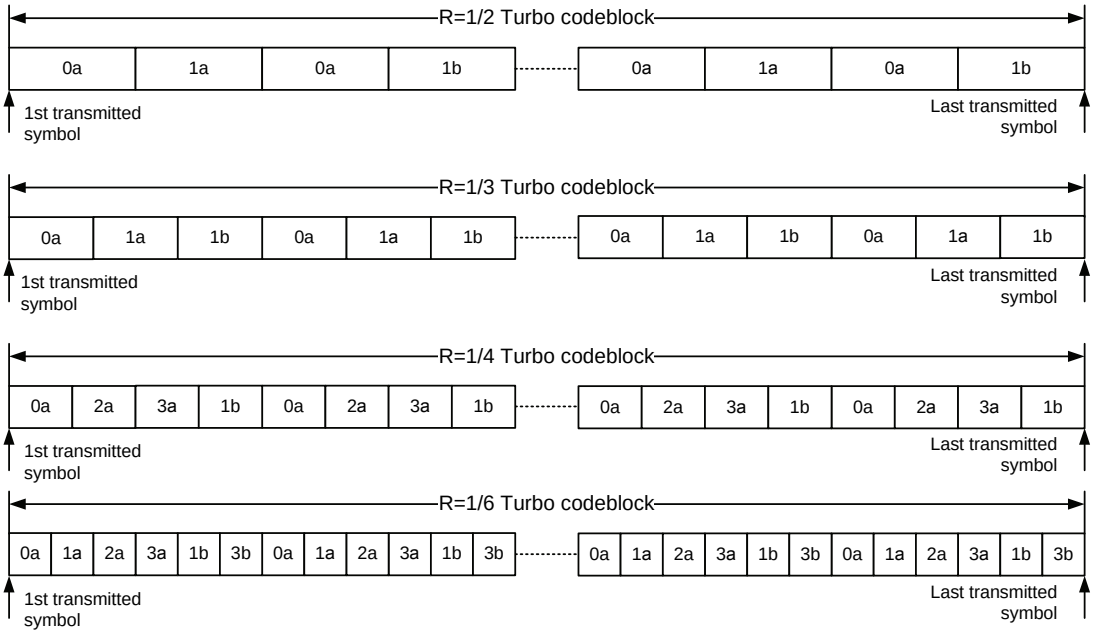


Fig. 2. Turbo Codeblocks for Different Code Rates

IV. Code performance

This section will review the performance of the CCSDS coding standard. Figures 3, and 4 show the performance of the codes in a different light. Here, the block lengths are kept at 1784 bits and 7136 bits respectively, but the rates have been allowed to vary. Common for all these graphs is that the R=1/2 code is substantially worse than the others, and that, as expected, the R=1/6 code is superior. More than 1dB asymptotic coding gain is obtained with the R=1/6 code compared with the R=1/2 code for a block length of 7136 bits, which is also consistent with the other block lengths.

Figure 5 shows the performance of the R=1/2 code of the CCSDS standard, where the block length is 16384 bits long. A varying number of iterations were used to obtain the curves. It is noticeable from this graph that a substantial coding gain is obtained when increasing the number of iterations from one to four. Furthermore, coding gains are consistently obtained when increasing the number of iterations beyond 4, albeit only minor ones after about 8 iterations. It will depend heavily on the application whether or not increasing the number of iterations beyond 4-8 would be of interest.

Figures 6, 7 and 8 show the performance of the rate R=1/2, R=1/3, and R=1/6 codes respectively, for a varying number of block lengths. The number of iterations was kept constant at 10 iterations for all these codes. As is seen from all these curves, the higher the block length, the better and the performance. A considerable coding gain is obtained by going from a block length of 1784 to 16384, particularly at low BER values. As an example, a coding gain of about 0.5dB is obtained when using a block length of 16384 instead of 1784 for the R=1/2 code, when seen at a BER of 10^{-5}. This coding gain reduces for the lower rate codes, albeit it is still around 0.35dB for the 1/6 rate code.

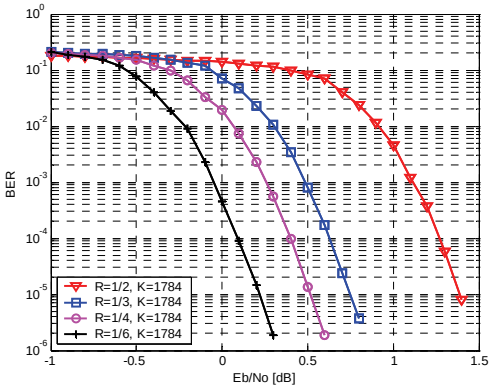


Fig. 3. Performance of different rate CCSDS codes of information block lengths 1784 bits

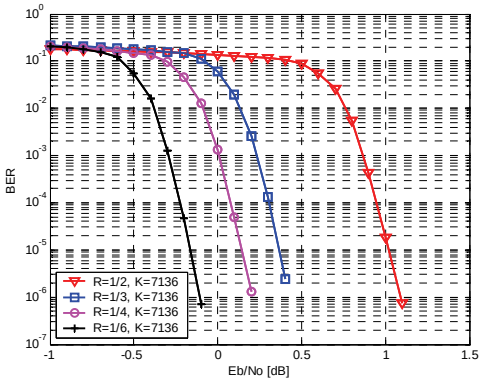


Fig. 4. Performance of different rate CCSDS codes of information block lengths 7136 bits

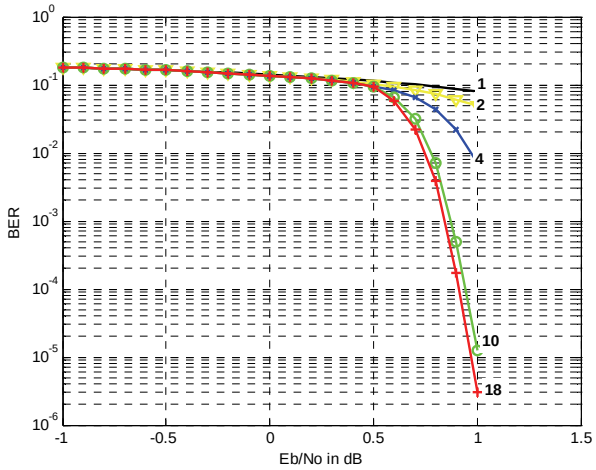


Fig. 5. Performance of CCSDS R=1/2 Turbo code of block length 16284 bits, with varying number of iterations in the decoder

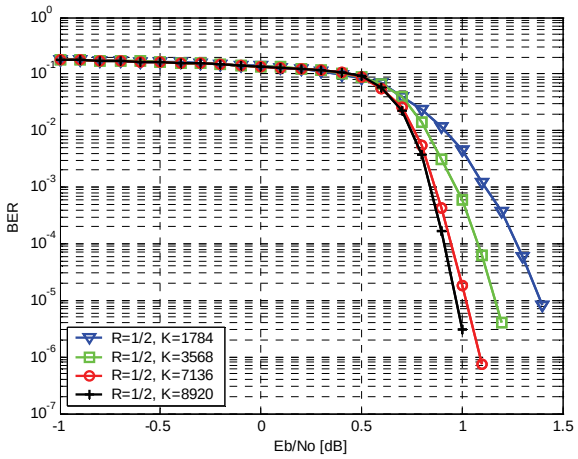


Fig. 6. Performance of CCSDS rate R=1/2 for different block lengths

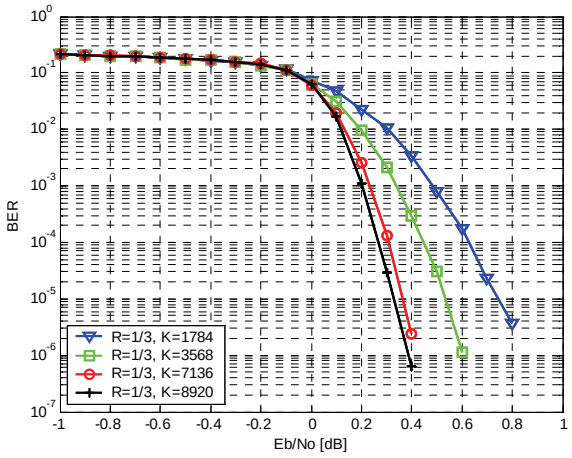


Fig. 7. Performance of CCSDS rate R=1/3 for different block lengths

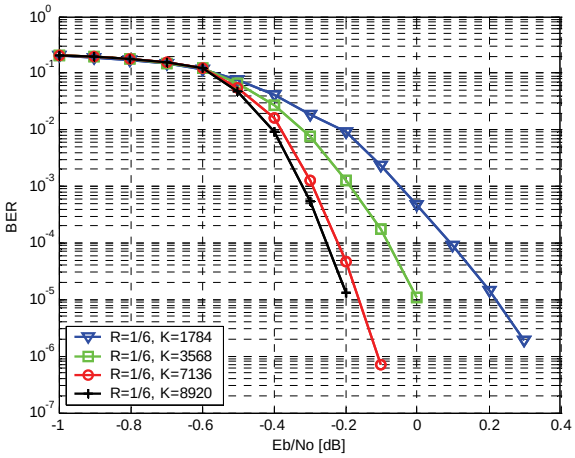


Fig. 8. Performance of CCSDS rate R=1/6 for different block lengths

V. Conclusion

In this paper we have reviewed the performance of the CCSDS turbo coding standard. The codes have been presented in general terms, and it has been argued that this standard may easily be used and adapted by other applications, where UMTS and DVB were explicitly mentioned. Some notes about implementation aspects were given.

A remaining challenge with turbo codes, is to find a suitable form of implementation that will allow for these various standards to the one turbo codec and to be of the sufficiently high speed which is required by modern data communications. Big steps have been made already, e.g. [5], and more will arrive as more applications are found for these codes.

Acknowledgement

The authors are grateful to Georgi Petkov for his helpful comments.

References

[1] C. Berrou, A. Glavieux, P. Thitmajshima, “Near Shannon limit error – correction coding and decoding: turbo–codes”, in Proc. IEEE International Conf. on Commun. (ICC), Geneva, Switzerland, pp.1064–1071, 1993.

[2] C. Schlegel, L. Perez, Trellis and turbo coding, IEEE Press, 2004.

[3] Consultative Committee for Space Data Systems, Recommendation for space data system standards, TM synchronization and channel coding, CCSDS 131.0-B-1,Blue Book, 2003

[4] M. R. Soleymani, Y. Gao, and U. Uilaipornsawai, Turbo Coding for Satellite and Wireless Communications, Kluwer Academic Publishers, Boston, 2002.

[5] F. Chiaraluce and R. Garello, “Highly efficient channel codes for high data rate missions,” presented at CCSDS Sub-Panel 1B Spring 2001 Meeting, Pasedena, CA, May 2001.

Construction and Simulation Analysis of Low-Density Parity-Check Codes

Teodor B. Iliev[1]

***Abstract* – Low – Density Parity Check codes were invented in 1960 by R. Gallager. They were largely ignored until the discovery of turbo codes in 1993. Since then, LDPC codes have experienced a renaissance and are now one of the most intensely studied areas in coding. In this article we review the basic structure of LDPC codes and the iterative algorithms that are used to decode them.**

***Keywords* – Parity-check matrix, low–density parity check code**

I. INTRODUCTION

Low-Density parity-check (LDPC) codes, discovered by Gallager in 1962 [1], were rediscovered and shown to form a class of Shannon-limit-approaching codes in the late 1990 [2, 3]. These codes, decoded with iterative decoding based on belief propagation, such as the sum-product algorithm (SPA) [3, 4, 7], achieve amazingly good error performance. Ever since their rediscovery, design, construction, decoding, efficient encoding, performance analysis, and applications of these codes in digital communication and storage systems have become the focal points of research. Various methods for constructing LDPC codes have been proposed. Based on the methods of construction, LDPC codes can be classified into two general categories: 1) random (or random-like) codes generated by computer search based on certain design guidelines and required structural properties of their Tanner graphs 6], such as the girth and degree distributions [3, 5]; and 2) structured codes constructed based on algebraic and combinatorial methods, such as those given in [8, 9], and many others.

II. GRAPHICAL REPRESENTATION OF LDPC CODES

A linear block code C of rate $R = k/n$ can be defined in terms of a $(n-k)\times n$ parity-check matrix $H = [h_1, h_2,\ldots,h_n]$. Each entry h_{ij} of H is an element of a finite field GF(p). We will only consider binary codes in our work, so each entry is either a '0' or a '1' and all operations are modulo 2. The code C is the set of all vectors x that lie in the (right) nullspace of H, that is, $H_x=0$. Given a parity-check matrix H, we can find a corresponding $k\times n$ generator matrix G such that $GH^T=0$. The generator matrix can be used as an encoder according to $x^T = u^TG$.

In its simplest guise, an LDPC code is a linear block code with a parity-check matrix that is "sparse"; that is, it has a small number of nonzero entries. Gallager proposed constructing LDPC codes by randomly placing 1's and 0's in a $m\times n$ parity-check matrix H subject to the constraint that each row of H had the same number d_c of 1's and each column of H had the same number d_v of 1's.[9] On Fig. 1 is shown the m=15×n=20 parity-check matrix with d_v=3 and d_c=4 and defines an LDPC code with length n=20. Codes of this form are referred to as regular (d_v, d_c) LDPC codes of length n. In a (d_v, d_c) LDPC code each information bit is involved in d_v parity checks and each parity-check bit involves d_c information bits. The fraction of 1's in the parity-check matrix of a regular LDPC code is $\frac{md_c}{mn}=\frac{d_c}{n}$, which approaches zero as the block length gets large and leads to the name low-density parity-check codes.

H_1=

$$\left[\begin{array}{cccccccccccccccccccc}
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
\hline
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&0&0&0&1&0&0&1\\
0&0&1&0&0&0&1&0&0&0&0&0&0&1&0&0&0&1&0&0\\
0&0&0&1&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
\hline
1&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&1&0&0\\
0&1&0&0&0&0&1&0&0&0&1&0&0&0&0&1&0&0&0&0\\
0&0&1&0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&1&0\\
0&0&0&1&0&0&0&0&1&0&0&0&0&1&0&0&1&0&0&0\\
0&0&0&0&1&0&0&0&0&1&0&0&0&0&1&0&0&0&0&1
\end{array}\right]$$

Fig.1 Gallager – type low density parity check H_1

We must determine the rate of the code defined by H_1. Since the parity check matrix is randomly constructed, there is no guarantee that the rows are linearly independent and the matrix is full rank. Indeed, the rank of H_1 is 13<m=15, and this parity-check matrix actually defines a code with rate R=7/20. In general, such randomly constructed parity-check matrices will not be full rank and m≠n−k. We could eliminate linearly dependent rows to find a $(n-k)\times n$ parity-check matrix, but the new matrix would no longer be regular. For LDPC codes with large n, it is convenient to retain the original parity-check matrix even if it is not full rank, and it is also convenient to refer to $1-\frac{m}{n}=1-\frac{d_v}{d_c}$ as the designed rate of the code.

[1]Teodor B. Iliev is with the Department of Communication Technique and Technologies, 8 Studentska Str., 7017 Rousse, Bulgaria, E-mail: tiliev@ecs.ru.acad.bg

Having defined a (d_v, d_c) regular LDPC code, we are now left to construct a particular instance of a code. To do this requires the choice of d_v, d_c, n, and k, which are constrained by the relationship that, for a regular code, $md_c = nd_v$ and that d_v must be less than d_c in order to have a rate less than R=1. Assuming that the block length n and the code rate are determined by the application, it remains to determine appropriate values for d_v and d_c. In [1] Gallager showed that the minimum distance of a typical regular LDPC code increased linearly with n, provided that $d_v \geq 3$. Therefore, most regular LDPC codes are constructed with d_v and d_c on the order of 3 or 4, subject to the above constraints. For large block lengths, the random placement of 1's in H such that each row has exactly d_c 1's and each column has exactly d_v 1's requires some effort, and systematic methods for doing this have been developed [3].

Tanner graph

An important advance in the theory of LDPC codes occurred when Tanner [6] used bipartite graphs to provide a graphical representation of the parity-check matrix. The bipartite graph is a graph in which the nodes may be partitioned into two subsets such that there are no edges connecting nodes within a subset. In the context of LDPC codes, the two subsets of nodes are referred to as variable nodes and check nodes. There is one variable node for each of the n bits in the code, and there is one check node for each of the m rows of H. An edge exists between the i^{th} variable node and the j^{th} check node if and only if h_{ij} =1. The bipartite graph corresponding to the parity-check matrix H_1 is shown in Fig. 9. In a graph, the number of edges incident upon a node is called the degree of the node. Thus, the bipartite graph of a (d_v, d_c) LDPC code contains n variable nodes of degree d_v and m check nodes of degree d_c.

It is clear that the parity-check matrix can be deduced from the bipartite graph, and thus the bipartite graph can be used to define the code C. We can therefore start talking about codes as defined by a set of variable nodes, a set of check nodes, and set of edges. This is the current approach to addressing LDPC codes. Note that the pair (d_v, d_c), together with the code length n, specifies an ensemble of codes, rather than any particular code. This ensemble is denoted by $C^n(d_v, d_c)$. Once the degrees of the nodes are chosen, we are still free to choose which particular connections are made in the graph.

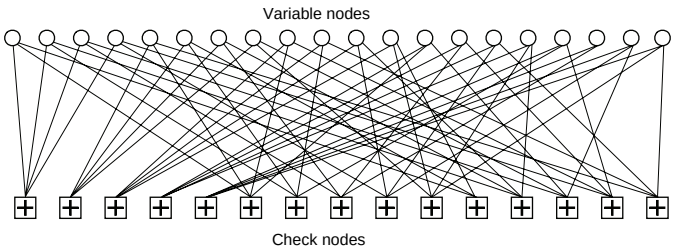


Fig.2 Bipartite graph for a (3,4) regular LDPC code with parity check matrix H_1

A socket refers to a point on a node to which an edge may be attached. A variable node has d_v sockets, meaning d_v edges may be attached to that node. There will be a total of nd_v sockets on variable nodes and md_c sockets on parity-check nodes. Clearly the number of variable-node sockets must be equal to the number of check-node sockets and a particular pattern of edge connections can be described as a permutation π from variable-node sockets to check-node sockets. An individual edge is specified by the pair (i, $\pi(i)$), which indicates that the i^{th} variable node socket is connected to the $\pi(i)^{\text{th}}$ check-node socket.

Selecting a random code from the ensemble $C^n(d_v, d_c)$ therefore amounts to randomly selecting a permutation on nd_v elements. Many permutations will result in a graph that contains parallel edges—that is, in which more than one edge join the same variable and parity-check nodes. Note that in the parity-check matrix, an even number of parallel edges will cancel. If they are deleted from the graph, then the degrees of some nodes will be changed and the code will cease to be a regular LDPC code. If they are not deleted, their presence renders the iterative decoding algorithms ineffective. We must therefore make the restriction that permutations leading to parallel edges are disallowed.

An irregular LDPC code cannot be defined in terms of the degree parameters d_v and d_c. We must instead use degree distributions to describe the variety of node degrees in the graph. A degree distribution $\gamma(x)$ is a polynomial:

$$\gamma(x) = \sum_i \gamma_i x^{i-1} \tag{1}$$

such that $\gamma(1)=1$. The coefficients γ_i denote the fraction of edges in the graph which are connected to a node of degree i. We will also use the notation $\int\gamma$ to denote the inverse average node degree, given by

$$\int\gamma = \int_0^1 \gamma(x)dx = \sum_i \frac{\gamma_i}{i} \tag{2}$$

Let d be the average node degree corresponding to the degree distribution γ. To show that $\int\gamma = \frac{1}{d}$, let n be the total number of nodes, and let n_i be the number of nodes of degree i. The total number of edges in the graph is $d \cdot n$. Because γ_i is the fraction of edges connected to a node of degree i, we conclude that $\frac{\gamma_i}{i} = \frac{n_i}{d \cdot n}$. Completing the sum of Eq. (2), we arrive at $\int\gamma = \frac{1}{d}$.

The code length n and two degree distributions - $\lambda(x)$ and $\rho(x)$ for the variable and check nodes, respectively – are sufficient to define an ensemble $C^n(\lambda, \rho)$ of irregular LDPC codes. A graph G from this ensemble will have n variable nodes. The number of check nodes m, is given by

$$m = n\frac{\int\rho}{\int\lambda} \tag{3}$$

The number of degree-*i* variable nodes in *G* is

$$n\tilde{\lambda}_i = n\frac{\lambda_i}{i\int\lambda} \tag{4}$$

where $\tilde{\lambda}_i$ denotes the fraction of variable nodes of degree *i*. Similarly, the number of degree-*i* check nodes in *G* is

$$m\tilde{\rho}_i = m\frac{\rho_i}{i\int\rho} \tag{5}$$

where $\tilde{\rho}_i$ denotes the fraction of check nodes of degree *i*. And the design rate of the code represented by *G* is

$$r = \frac{n-m}{n} \tag{6}$$

We can enumerate the variable-node and check-node sockets of the irregular code graph. Selection of a code from the ensemble is a selection of a permutation on N_E elements, where N_E is the number of edges in the graph, given by

$$N_E = \frac{n}{\int\lambda} \tag{7}$$

III. MESSAGE PASSING DECODING ALGORITHMS

It was stated in the introduction that a principal advantage of low-density parity check codes is that they can be decoded using an iterative decoding algorithm whose complexity grows linearly with the block length of the code. Belief propagation is one instance of a broad class of message passing algorithms on graphs as discussed in [7, 10]. Similar decoding algorithms for the binary erasure channel and the binary symmetric channel are discussed in [11]. All message passing algorithms must, however, respect the following rule, which was introduced with turbo codes.

Rule (Extrinsic Information Principle) *A message sent from a node n along an edge e cannot depend on any message previously received on edge e.*

Before stating the algorithm, it is necessary to formulate the decoding problem. An LDPC code is constructed in terms of its parity-check matrix *H* which is used for decoding. To encode the information sequence *u*, it is necessary to derive a generator matrix *G* such that $GH^T=0$. Finding a suitable *G* is greatly simplified if we first convert *H* into the equivalent systematic parity-check matrix $H_S = [A|I_{n-k}]$. The systematic generator matrix is now given by

$$G_s = \left[I_k \middle| A^T\right]$$

and the information sequence is encoded as $x^T=u^TG_s$. It is worth noting that encoding by matrix multiplication has complexity $O(n^2)$ and that, in general, LDPC codes have linear decoding complexity, but quadratic encoding complexity.

The codeword *x* is transmitted across the additive white Gaussian noise (AWGN) channel, using BPSK modulation resulting in the received sequence $r=(2x-1)+n$. The optimal decoder for the AWGN channel is the maximum a posteriori (MAP) decoder that computes the log-likelihood ratio (LLR):

$$\Lambda(x_r) = \log\left(\frac{P(x_r=1|y)}{P(x_r=0|y)}\right)$$

and makes a decision by comparing this LLR to the threshold zero. The belief propagation on a graph with cycles can closely approximate the MAP algorithm, and we can state the decoding algorithm for LDPC codes on the AWGN channel using these results.

Let A={−1, +1} denote the message alphabet, let $r_i \in R$ denote the received symbol at variable node *i*, and let $\Lambda_i \in R$ denote the decision at variable node *i*. A message from variable node *i* to check node *j* is represented by $\mu_{i\to j} \in R$, and a message from check node *j* to variable node *i* is $\beta_{j\to i} \in R$. Let $C_{j\backslash i}$ be the set of variable nodes which connect to check node *j*, excluding variable node *i*. Similarly, let $V_{i\backslash j}$ be the set of check nodes which connect to variable node *i*, excluding check node *j*. The decoding algorithm is then as follows:

Step 1: Initialize $\Lambda_i = \frac{2}{\sigma^2}r_i$ for each variable node. ($\sigma^2=N_0/2$).

Step 2: Variable nodes send $\mu_{i\to j}=\lambda_i$ to each check node $j \in V_i$.

Step 3: Check nodes connected to variable node *i* send

$$\beta_{j\to i} = 2\tanh^{-1}\left(\prod_{l\in C_{j/i}}\tanh\left(\frac{\Lambda_l}{2}\right)\right) \tag{8}$$

Step 4: Variable nodes connected to check nodes *j* send

$$\mu_{i\to j} = \sum_{l\in V_{i/j}}\beta_{l\to i} \tag{9}$$

Step 5: When a fixed number of iterations have been completed or the estimated code word $\hat{x}$ satisfies the syndrome constraint $H\hat{x}=0$, stop. Otherwise return to Step 3.

The check node's rule (8) is fairly complex. But for quantized messages it is possible to map LLRs to messages in such a way that the check-node rule can be implemented with some extra combinational logic and an adder. Such decoders provide very good performance with only a few bits of precision. If we want to keep the complexity even simpler, we can use the max-Log approximation, in which we can replace the above rule with:

$$\beta_{j\to i} \approx \min_{l\in C_{j/i}}\left(\left\{\left|\lambda_l\right|\right\}\right)\prod_{l\in C_{j/i}} sign(\Lambda_l) \tag{10}$$

Some performance loss will result from this approximation, but it may be justified by the overall savings in decoder complexity.

IV. SIMULATION RESULT

We perform simulation on an optimized regular (3,6) LDPC code with 128x256 parity check matrix over an additive white Gaussian noise channel (AWGN) with binary phase – shift keying (BPSK) modulation. In our simulation, we stop iteration as soon as a codeword is detected or when a

maximum number of iteration is reached. The maximum number of iteration is set to 10.

Fig. 3 shows the frame error rate as a function of the energy per bit (E_b) to the spectral noise density (N_0) – E_b/N_0 and 30 codewords errors. Fig. 4 shows the bit error rate as a function of E_b/N_0.

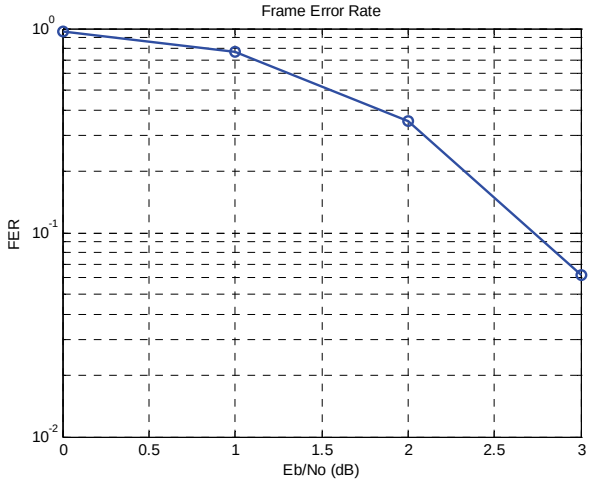


Fig.3 Frame error rate (FER) for LDPC codes over the additive white Gaussian noise channel

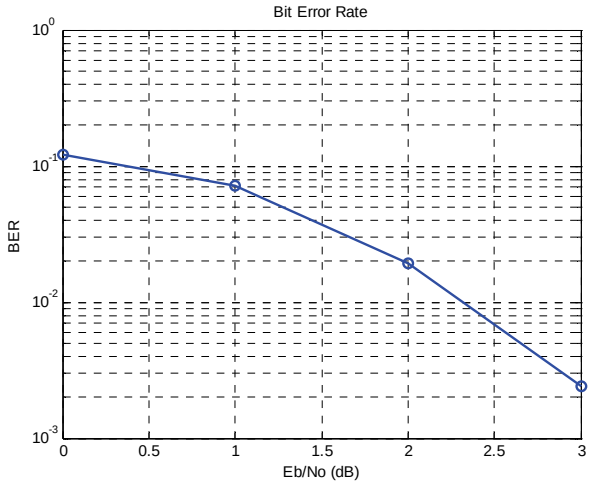


Fig.4 Bit error rate (BER) for LDPC codes over the additive white Gaussian noise channel

V. Conclusion

Analysis techniques formulated for LDPC codes are generic and can be used for other coding schemes based on iterative decoding methods. Each LDPC code is characterized by a number of fixed parameters that include the following: check node degree, variable node degree and blocksize *N*. These parameters are used to determine the nodes in the LDPC factor graph and a collection of permissible edges between the nodes. Given these parameters, there exists a finite number of possible ways in which edges can be connected between nodes in a factor graph. By varying the parameters of the graphs to increase their girth one can improve the distance properties of the code without increasing the number of parity checks in which each bit is involved. However the size of the graph and hence size of the code, required grows rapidly. Some modifications to standard decoding algorithms for LDPC codes have also been presented.

Finally, the goal of this paper was to present the main constructive principle of low–density parity check codes, as well as a design methodology, that gives good performance for many communication systems like latest DVB satellite communications standard.

Acknowledgement

The author is grateful to Georgi Petkov for his helpful comments.

References

[1] R. G. Gallager, “Low density parity check codes,” IRE Trans. Inf. Theory, vol. IT-8, no. 1, pp. 21–28, Jan. 1962.

[2] D. J. C. MacKay and R. M. Neal, “Near-Shannon-limit performance of low density parity check codes,” Electron. Lett., vol. 32, pp. 1645–1646, Aug. 1996.

[3] D. J. C. MacKay, “Good error-correcting codes based on very sparse matrices,” IEEE Trans. Inf. Theory, vol. 45, no. 3, pp. 399–432, Mar. 1999.

[4] T. Richardson and R. Urbanke, “The capacity of low-density parity check codes under message-passing decoding,” IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[5] T. Richardson, A. Shokrollahi, and R. Urbanke, “Design of capacity approaching low density parity check codes,” IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[6] R. M. Tanner, “A recursive approach to low complexity codes,” IEEE Trans. Inf. Theory, vol. IT-27, no. 9, pp. 533–547, Sep. 1981.

[7] F. R. Kschischang, B. J. Frey, and H. -A. Loeliger, “Factor graphs and the sum-product algorithm,” IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[8] R. M. Tanner, “Spectral graphs for quasi-cyclic LDPC codes,” in Proc. IEEE Int. Symp. Inf. Theory, Washington, DC, Jun. 2001, p. 226.

[9] S. Lin and D. J. Costello, Jr., Error Control Coding: Fundamentals and Applications, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2004.

[10] G. D. Forney, “Codes on graphs: Normal realizations,” IEEE Trans. Inform. Theory, pp. 520–548, Feb. 2001.

[11] C. Schlegel, L. Perez, Trellis and turbo coding, IEEE Press, 2004

New Intellectual Property Blocks for a Secure Modem on FPGA

Galia I. Marinova[1] , Vassil G. Guliashki[2] and Maurice Bellanger[3]

***Abstract* – The paper presents new Intellectual Property (IP) blocks realized for completing a Data Base with IPs for secure multicarrier modem design on FPGA. The modems considered are designed on the Filter Banc principle. The new IP blocks described are – equalizer for ADSL and WiFi applications with channel coefficient estimation and crypto-cores based on DES (Data Encryption Standard), 3DES (Triple DES) and AES (Advanced Encryption Standard) algorithms. IPs are developed in VHDL and their realizations are estimated for time and surface area parameters on FPGA from the Virtex-4 Xilinx circuit family.**

***Keywords* – multicarrier modem; crypto-processors; DES, AES, 3DES; FPGA; equalizer;**

I. INTRODUCTION

The paper presents results from our recent work on creating an environment for design of Filter Banc based multicarrier modems with some emphasis on security functions. We present here new IP blocks developed in VHDL for some DSP functions – equalizer for applications following ADSL and WiFi standards and for security functions – crypto-cores based on DES (Data Encryption Standard), 3DES (Triple DES) and AES (Advanced Encryption Standard) algorithms. We present the structure of the Data Base with IP blocks for Secure multicarrier modem design, completed with the new IP blocks. Then we consider more in details the equalizers and the crypto-core IP blocks specifications, the FPGA based realizations of those blocks and finally we give results for time and surface area estimations of the IPs realized on a concrete FPGA from Xilinx circuit family.

II. DATA BASE WITH IP BLOCKS FOR SECURE MULTICARRIER MODEM DESIGN

The specification of a multicarrier Filter Banc based modem is presented in [5,8]. Our research during the last three years aimed the development of a Data Base with Intellectual Property (IP) block for the main functions in multicarrier modem specification which permit a flexible approach to design different modem applications following the design methodology described in [7]. The basic multicarrier modem core functions were presented in [6].

In this paper we present our work on completing the Data Base with IPs for equalization and a group of crypto-cores for secure modem designs. Fig. 1 presents the structure of the Data Base with IP blocks for Filter Banc based multicarrier modem design with emphasis on new DSP functions and secure modem crypto-cores.

III. NEW IP BLOCKS IN THE DATA BASE

We describe here the new IPs that we developed, adapted and added in the Data Base for Secure multicarrier modem design on FPGA.

A. New IPs for DSP blocks

The principle of equalization in the multicarrier modem is described in [2]. The equalization is realized in two stages – learning or initialization stage and equalization stage. Two cases are studied – equalization of a multicarrier modem core for ASDL standard where 230 over 256 sub-channels are used for data transmission and equalization for wireless LAN modem application where 115 over 128 sub-channels are used for data transmission.

The transfer function of the transmission multipath channel in both cases is considered as:

$$C(z)=1-0.5z^{-1}+0.3z^{-2}$$

The initial values of the coarse equalizer coefficients are obtained through interpolation with the help of 4 pilot tones uniformly distributed in the transmission bandwidth, which is efficient in keeping synchronization. They are used to track the residual carrier frequency offset that remains after the frequency correction during the training phase.

The simplest form of the fine equalizer with just one coefficient h_i, i=1,115/i=1,230 for each sub-channel is realized. The value of each h_i is updated in a decision directed mode, through least square type of algorithm. The structure is able to track the slow evolutions of the complex sub-channel gain over the duration of the packet.

The input data at the modem transmitter is a known sequence of random data d(i), $i=1,10^5$; $d(i)\in\{+1,-1\}$, the data coming out from the analysis Filter Banc in the modem receiver are x(i), $i=1,10^5$ and the output error for sub-channel i, with sub-channel equalizer, at time n is:

$$e(n)=d_i(n)-(c_i+h_i)x_i(n)$$

[1] Galia I. Marinova is with the Faculty of Telecommunications, Technical University – Sofia, 8, bul. "Kliment Ohridski", Sofia 1000, Bulgaria, e-mail: gim@tu-sofia.bg

[2] Vassil G. Guliashki is with the Institute for Information Technologies – BAS, "Acad. G. Bonchev" Str., bl. 29A, Sofia 1113, Bulgaria, e-mail: vggul@yahoo.com

[3] Maurice Bellanger is with CNAM-Paris, 2, Rue Conte, Paris 75003, France, e-mail: bellang@cnam.fr

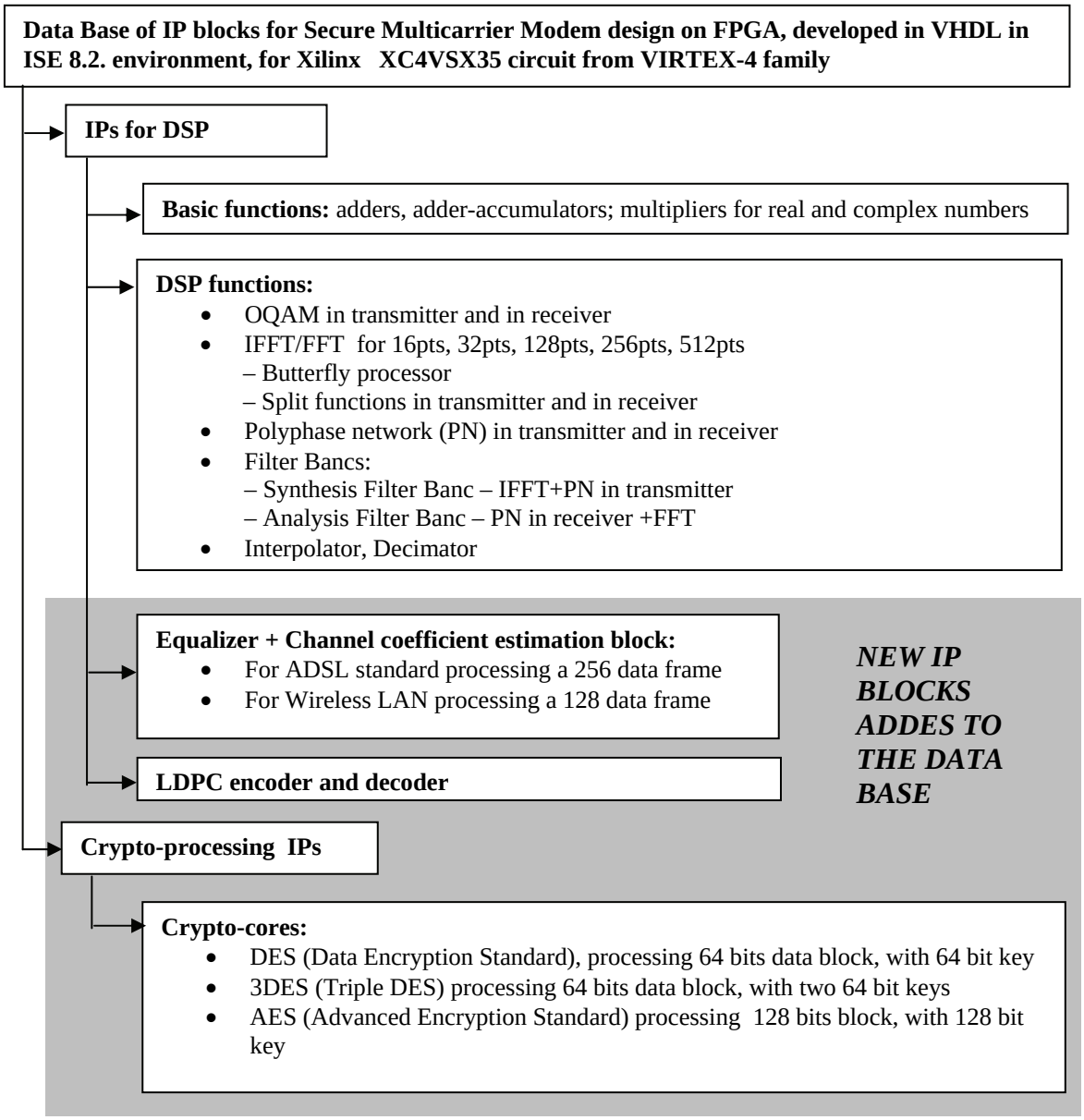


Fig. 1. Data Base with IP blocks for Secure modem design

These output errors serve to update the equalizer coefficients - h_i, i=1,115/230 and to compute in real time the SNR in the corresponding sub-channel. The benefit of the fine spectral analysis of the radio-channel performed by the AFB is that a single complex coefficient can be sufficient for each equalizer, since the distortion in the sub-channels can be approximated reasonably close by a flat gain. If the radio channels are severely distorted, more coefficients might be needed to cope with the residual timing offset and to reduce the interference level. Four-bit symbols are transmitted by sub-channel.

The IP-oriented specification of the equalizer in the multicarrier modem is presented on Fig. 2.

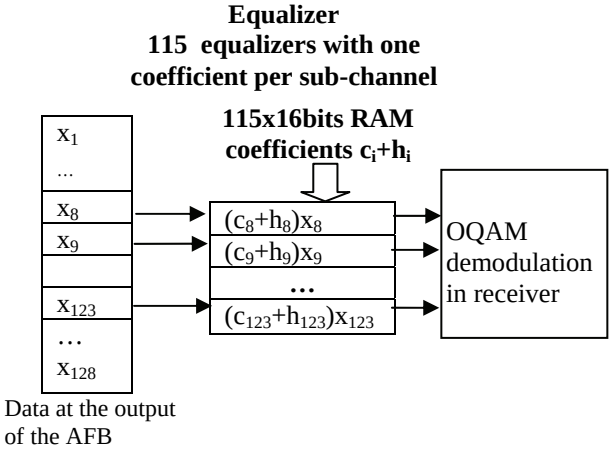


Fig. 2. IP-oriented specification of the equalizer block

The equalizer IPs are integrated in the multicarrier modem cores and they are realized on FPGA from XILINX family - VIRTEX-4 , the circuit XC4VSX35 with frequency f=400MHz and surface area of 10^7gates.The ISE development system from [10] is used for the synthesis and the programming of the FPGA.

The equalization IP block for ADSL application is realized with fully serial architecture using one real multiplier. It treats a 256 data frame. The equalizer IP realizes 115 serial complex multiplications.

The architecture of the modem for wireless LAN uses 4 parallel multipliers and it treats a 128 data frame. The equalizer IP realizes 230 complex multiplications parallelized on 4 real number multipliers.

Table I presents time estimation of the modem core including equalizer IP in two applications and Table II presents the surface area taken by the equalizer IP blocks on the FPGA.

Table I. TIME ESTIMATION OF THE EQUALIZER IP BLOCKS

IP block	Time per frame	
	Modem for ADSL Fully serial architecture with one multiplier Frame of 256 data	Modem for wireless LAN Architecture with 4 parallel multipliers Frame of 128 data
Equalizer	3.6μs	0.6μs

Table II. SURFACE AREA ESTIMATION OF THE EQUALIZER IP BLOCKS

Equalizer IP block	BRAM	TBUF	MULT	Number of Slices
Modem for ADSL	6%	1%	1%	4%
Modem for wireless LAN	10%	1%	4%	6%

The LDPC encoder and decoders we used are described in [11].

B. New crypto-processing IP cores

In previous work [9] we gave results from the integration of commercial DES and 3DES crypto-cores in a secure multicarrier modem design. But in order to improve the flexibility of our environment we developed proprietary crypto-cores and/or adapted existing open cores of the encryption/decryption blocs. We developed three types of crypto-cores: DES, 3DES and AES crypto-cores.

- DES crypto-core – The principle of DES algorithm consists in an initial permutation, followed by 16 rounds (iterations) and a final permutation at the end. The DES crypto-core we adapted is from [4]. It uses a 64-bit key and it treats a 64-bit data block. The encryption and the decryption follow the same algorithm, only the key processing steps are inverted. The choice of encryption or decryption mode is done through the signal E_D which is "1" for encryption and "0" for decryption. The DES crypto-core IP treats a 64-bit data block in 16 clock cycles.

- 3DES crypto-core – The 3DES crypto-core is developed on the base of the DES crypto-core. In our case, it supports two independent 64-bit keys. A triple DES encryption operation with 2 independent keys consists of the transformation of a 64-bit data block I into a 64-bit data block O, defined as follows:

$$O = E_{K1}(D_{K2}(E_{K1}(I))),$$

where $E_K(I)$ and $D_K(I)$ represent the DES encryption and decryption of I, using DES key Kn (where n=1,2).
A triple DES decryption operation with 2 independent keys consists in the transformation of a 64-bit data block I into a 64-bit data block O, defined as follows:

$$O = D_{K1}(E_{K2}(D_{K1}(I)))$$

Compared to the DES algorithm, the triple DES algorithm provides a much higher level of security. The 3DES crypto-core IP treats a 64-bit data block into 48 clock cycles.

- AES crypto-core – It implements the Advanced Encrypting Standard, based on the cryptographic algorithm, created by Rijndael [1, 3]. In the presented secure modem application the plain text data are encrypted/decrypted in blocks of 128 bits, using 128-bit key size. The AES algorithm consists of a complex non-linear function, which is iterated multiple times (rounds) starting from the incoming plain text data block. There is an initial pre-processing round at the start of every encryption. The number of rounds required depends on the selected key size – in our case with 128-bit key size 10 rounds are necessary, or together with the initial pre-processing round 11 rounds in total. Each round requires an unique 128-bit round key schedule. The necessary schedules are generated by means of a key expansion algorithm using the supplied initial 128-bit key. Eleven key schedules are necessary for this key size. They can be generated in real time, when they are required by the encryption algorithm. They can also be generated off-line and can be stored in an internal RAM. We realized the last possibility in this application by means of AES cores, which cover both encryption/decryption functions and key expansion for 128-bit key size. The cores implement all the building blocks of AES algorithm individually and are easily integrated in the created VHDL code. The AES crypto-core IP treats a 128-bit data block into 11 clock cycles. In AES decryption algorithm the basic transformations used in AES encryption algorithm are inverted. The sequence of these transformations differs in the straightforward AES decryption algorithm from that one of the AES encryption algorithm. However, by means of a change in the key schedule an equivalent AES decryption algorithm, having the same order of transformations as the encryption algorithm, is obtained. This decryption algorithm has a more efficient structure than that one of the straightforward AES decryption algorithm. In our application we implemented the equivalent AES decryption algorithm. The selection of encryption or decryption mode is done through the signal E_D, which is "1" for encryption and "0" for decryption.

Table III gives time parameters for the crypto-cores realized and Table IV presents data for the surface area taken

by the three different types of crypto-cores – DES, 3DES and AES. The estimation is made for a XC4VSX35 circuit from the Xilinx VIRTEX-4 family.

Table III. TIME PARAMETERS OF THE CRYPTO-CORES

IP block	Time per frame		
Crypto-processing core	DES	3DES	AES
	480ns	1.44μs	178ns

Table IV. SURFACE AREA OF THE MODEM CRYPTO-CORES ON XC4VSX35 CIRCUIT

Crypto-processing IP	GCLK	LUT	Number of Slices Flip-Flops	Number of Slices
DES	1%	4%	1%	4%
3DES	1%	5%	1%	5%
AES	1%	4%	1%	10%

IV. CONCLUSION

The Data Base with Intellectual Property Blocks that we presented here is a part of an environment which permits to design Filter Banc based multicarrier modems for different applications: for example applications following ADSL or WiFi standards. The emphasis of the new blocks we developed is on crypto-cores DES, 3DES and AES that allow the design of different secure modem couples performing encryption in transmitters and decryption in receivers with previous key exchange. We also added equalizer blocks for two applications and LDPC encoding/decoding IPs blocks. A future work is foreseen on more precise modem synchronization algorithms and IP blocks, as well as further optimization of the Filter Banc based modem designs and realizations. The new IPs complete the Data Base and create an useful environment for teaching and self-education in the area of Filter Banc – based mulcarrier modem design and Secure modem design.

ACKNOWLEDGEMENT

The research work described in this paper is partly supported by NATO grant: CLG 981551/2005.

The authors thank Milen Vassilev Simeonov for his work on DES core design and testing during his final project work.

REFERENCES

[1] AES (Rijndael) IP-Cores for Encryption/Decryption and Key Expansion, ErSt Electronic GmbH, Switzerland, April 2006, http://www.opencores.org

[2] Bellanger M., Increasing the data throughoutput of wireless LAN with OFDM/OQAM, March, CNAM-Paris, 2006

[3] Daemen J. and Rijmen V., AES Proposal: Rijndael, "AES Algorithm Submission", September 3, 1999, available at: http://www.nist.gov/CryptoToolkit

[4] Lagger A., "Implementation of DES Algorithm Using FPGA Technology", 2003, available at: http://lsmwww.epfl.ch/Education/reports/lagger_report_2003.pdf

[5] Marinova G., C. Fernandes, M. Bellanger, "Specification of multicarrier modem aimed to intellectual property and FPGA implementation", Proceedings of the *International Conference on Basic Technologies for E-business'2002*, 15-18 09.2002, Albena, Bulgaria, pp. 203-210.

[6] Marinova G. and C. Fernandes, "Data base of IP blocks developed in VHDL for multicarrier modem realization on FPGA", *MELECON 2004* Proceedings, Volume I, *The 12th IEEE Mediterranean Electrotechnical Conference*, May 12-15, 2004, IEEE Catalog Number 04CH37521, Dubrovnik, Croatia, pp. 217-220.

[7] Marinova G. and C. Fernandes, "Study on the realization with FPGA of a multicarrier modem", Proceedings of the *2004 International TICSP Workshop on Spectral Methods and Multirate Signal Processing, SMMSP2004*, Vienna, Austria, September 11-12,2004, Edited by Jaako Astola, Karen Egiazarian and Tapio Saramaki, TICSP Series 25, pp. 115-122.

[8] Marinova G., V. Guliashki, D. Le Ruyet, M. Bellanger, Multicarrier modem core on FPGA, *MELECON'2006*, Proceedings of *IEEE Mediterranean Electrotechnical Conference*, May 16-19, Benalmadena (Malaga), Spain, Volume 1. pp. 66-69.

[9] Marinova G., Guliashki V., "Security Solutions for Modem Communications", Proceedings of *National Conference with international participation ELECTRONIKA'2006*, Sofia, Bulgaria, June 1-2. 2006, pp. 287-292.

[10] www.xilinx.com

[11] Yang Sun, M. Karkooti and J. R. Cavallaro, "High Throughput, Parallel, Scalable LDPC Encoder/Decoder Architecture for OFDM Systems". Fifth IEEE Dallas Circuits and Systems Workshop (DCAS-06) , Dallas, Oct 2006, pp 39-42.

Creation of Secrets Sharing Protocols Resistible Against Disloyal Participants

Petar T. Antonov[1] and Valentina R. Antonova[2]

***Abstract* – This paper reviews a perfected variant of the secrets sharing protocol, bearing in mind the loyalty level of the separate participants. A classification of the possibilities for the grouping of the participants in the protocol is offered, and recommendations are brought out about creation of resistible against disloyal participants secrets sharing protocol.**

***Keywords* – Secret, security, secrets sharing protocol, grouping of the participants, loyalty.**

I. INTRODUCTION

It is well known that the approach to dividing a secret to parts and giving these parts to chosen separate participants is used for a long time in order to increase the security during preserving and subsequent usage of this secret [1,2 etc.]. The security in this case is expressed by the level of prevention against eventual abuse of secrets, for example, independent access and subjective disposition of a common bank account, independent access and illegal consumption of common resources, personal activation of military destructive capacities, etc.

For the practical realization of the approach, mentioned above, it is necessary to have a corresponding *secrets sharing protocol* (SSP) among the participants, in which two interconnected components are differentiated: *secrets sharing scheme* (SSS) and *secrets restoring scheme* (SRS). In SSP it is supposed that all or at least a certain minimum number of the participants of the protocol will be if loyal behaviour and will be able to restore successfully the secret when that occurs to be necessary.

To designate such a scheme of realization the term *threshold structure of restoration* (n, k) is proposed in [1]. In [1] is also offered a perfection of the considered protocol SSP, taking account of the introduced for that purpose *loyalty level of the participants* and using on that basis a *secrets restoring probability scheme* (SRS).

Furthermore, in [1] an analysis is carried out concerning the *probability for successful restoration of secrets* at a threshold structure (n, k), which supposes loyal behaviour of a minimum of *k* out of the included in the protocol *n* participants.

It should be mentioned, however, that the *threshold structure of restoration* introduced above might be considered also as a *structure for grouping of the participants* in SSP by a *parameter of grouping (threshold of restoration) k.* This means that for the completeness of study upon SSP, it is necessary to determine and consider all possibilities of such a grouping. In this connection, in the present report, which is an extension and a development of [1], a *classification of the possibilities for grouping of the participants* in SSP is offered and recommendations are brought out about the choice of parameters of grouping that would guarantee some preliminary given resistibility against disloyal behaviour of the participants in the protocol.

For the purposes of the further consideration we shall introduce and determine the following designations:

- n - number of participants in SSP;
- A_i - i-th participant in SSP;
- k - threshold of restoration;
- S - secret;
- s_i - one i-th part of the secret, that is given to participant A_i;
- R - dealer (distributor);
- p_i - probability for loyal behaviour of the participant A_i;
- q_i - probability for disloyal behaviour of the participant Ai (q_i=1- p_i);
- P - probability for successful restoration of the secret.

II. CLASSIFICATION OF THE POSSIBILITIES FOR GROUPING OF THE PARTICIPANTS

The possibilities for grouping of the participants in SSP or, in other words, the varieties of the threshold structure of restoration (*n, k*), can be classified as follows:

- **arbitrary grouping** (*threshold structure with arbitrary grouping*);
- **orderly grouping** (*threshold structure with arranged grouping*);
- **consecutive – parallel grouping** (*threshold structure with consecutive – parallel grouping*);
- **parallel – consecutive grouping** (*threshold structure with parallel – consecutive grouping*);
- **combined grouping** *(threshold structure with combined grouping*).

In the case of *arbitrary grouping* the separate participants Ai are not interconnected among themselves and an arbitrary

[1]Petar T. Antonov is with the Department of Computer Science and Engineering, Technical University of Varna, 1 Studentska str., 9010 Varna, Bulgaria, E-mail: peter.antonov@ieee.org

[2]Valentina R. Antonova is with the Department of Computer Science and Engineering, Technical University of Varna, 1 Studentska str.,9010 Varna, Bulgaria, E-mail:valyvarna@yahoo.com

subset of them with dimension $l \geq k$ is in position to restore successfully the secretes.

The analysis carried out in [1] concerns this variant of the threshold structure, where ($k \leq l \leq n$) and ($1 \leq k \leq n$). In this case the location of the chosen subset of loyal participants in SSP, respectively, of the parts s_i of the secret granted by these participants, in the *chain of restoration of the secret* is accidental and does not reflect in any way upon the result of the restoration.

Under *orderly grouping* not only the threshold *k* should be taken into account, but also the successive order and the exact location of the separate participants in the *chain of restoration of the secret*. For the case in consideration this chain might be schematically represented as in Fig.1, where the dark squares depict the participants with loyal behaviour, whose number l must not be less than k, and the contents of these squares corresponds to the granted by these participants parts of the secret (s_1, s_3 , … , s_l). It is possible for the chain restoration to be with dimension *n*, as shown in Fig.1, but it might be with reduced dimension *l* as well, if the empty squares, corresponding to the disloyal participants, drop off and the location of the loyal participants, respectively of their parts of the secret, moves forward, keeping their succession.

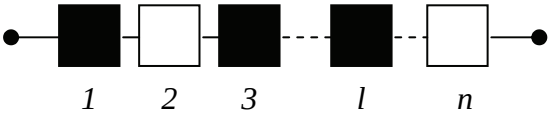


Fig. 1

The scheme for *consecutive – parallel grouping* of the participants in SSP is given in Fig.2, where *m* consecutive groups are shown, connected in parallel, and in each of the consecutive groups there are included *n* participants. At that, the first part s_1 of the secret *S* is given to all the first participants in the consecutive groups, the second part s_2 – to all the second participants and so on, up to the last part s_n, which is given to the last of the participants in each group. The participants in the separate consecutive groups are not interconnected among themselves and can assist in the restoration of the secret only within their own group.

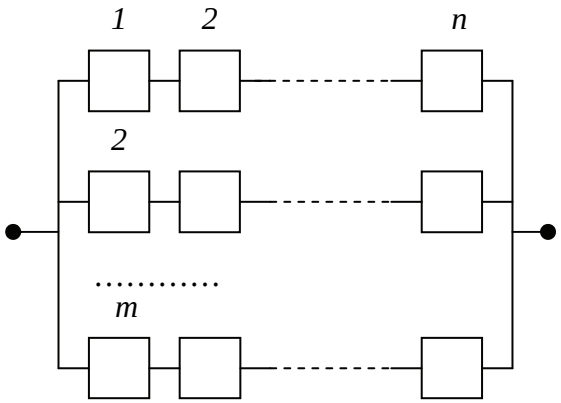


Fig. 2

In this case the total number of the participants is $N = n.m$, and for the successful restoration of the secret it is necessary to have loyal behaviour of not less than *k* participants in at least one of the consecutive groups, connected in parallel. As in each of the consecutive groups there might be included either arbitrary or orderly grouping, we should distinguish *consecutive – parallel arbitrary*, and respectively *consecutive – parallel orderly grouping*.

Parallel – consecutive grouping might be represented as in Fig.3, where *n* parallel groups are consecutively connected, and in each of the parallel groups there are included *m* participants. All the participants in the i-th parallel group receive from the dealer the corresponding part of the secret s_i, so for the successful restoring back the secret are necessary the parts of not less than *k* parallel groups.

It is evident that if the arrangement of the parallel groups in the process of restoration of the secret can be accidental, then the parallel – consecutive grouping will be *arbitrary*, otherwise – *orderly*.

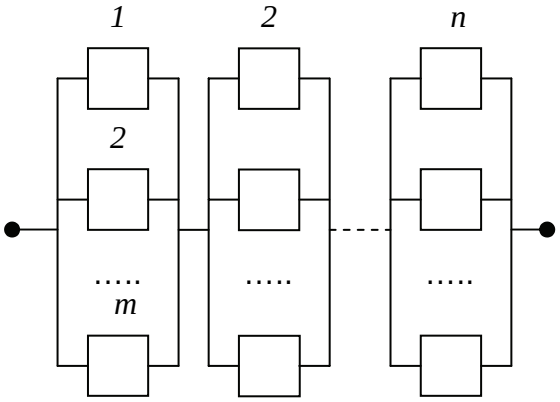


Fig. 3

Combined grouping represents combined usage of the mentioned above possibilities of basic grouping of the participants in SSP. It is evident that the possible variants for such a grouping are many, which gives the opportunity of finding an optimum solution for every particular case in practice.

III. Analysis of the Probability for Successful Restoration of the Secret

In the case of *arbitrary* and *orderly* grouping and provided the separate participants A_i can be considered equally loyal, i.e.

$$\forall i,\ p_i = p = const = 1 - q, \tag{1}$$

then the probability P can be determined using one of the following two equivalent correlations:

$$P = \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i} = \sum_{i=0}^{n-k} C_n^i q^i (1-q)^{n-i}. \tag{2}$$

In Fig.4 there are given exemplary curves of the changes of *P*, depending on *p*, for different values of *n* and *k*.

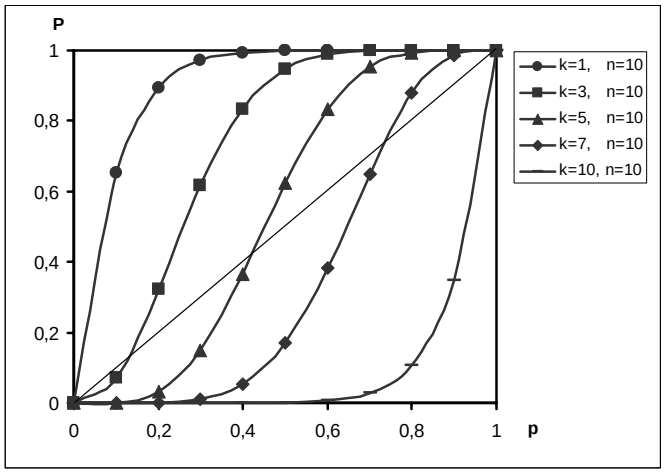


Fig. 4

On the basis of the introduced definitions, in the cases of *consecutive – parallel* and *parallel – consecutive grouping*, the following correlations can be deduced for the probability *P*:

$$P = 1 - [1 - \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i}]^m \quad \text{and} \tag{3}$$

$$P = \sum_{i=k}^{n} C_n^i [1-(1-p)^m]^i [(1-p)^m]^{n-i} \tag{4}$$

The correlations (2), (3) and (4) given above allow a comparative analysis to be made for the varieties of threshold structure for restoration and recommendations to be given about optimum choice of the *scheme of grouping*, depending on the specific conditions and requirements for resistibility against disloyal behaviour of the participants.

In case a minimum admissible value of *P* is assigned as P_{min} , and using the correlations (2), (3) and (4), analysis can be carried out and can be chosen optimum for the case (with respect to the expenses) scheme of grouping with determination of the parameters *n*, *k* and *m*. Here we ought to mention that in some cases the threshold k and/or the number of the participants *n* will be predetermined by the specific conditions and will not be possible to be decreased.

For example, if we suppose that the threshold $k = n$, then the correlations (2), (3) and (4) will be written respectively in the following form:

$$P = p^n = P_* \tag{5}$$

$$P = 1 - (1 - p^n)^m = P_{**} \tag{6}$$

$$P = [1 - (1-p)^m]^n = P_{***} \tag{7}$$

Let $n = 3$, $p = 0.8$ and $P_{min} = 0.92$. It can be seen that the *arbitrary* and *orderly* grouping in this case are misplaced, as

$$P_* = p^n = 0.512 < P_{\min} \tag{8}$$

For *consecutive – parallel* and respectively, *parallel – consecutive grouping*, we get:

$$P_{**} = 1 - (1 - p^3)^m \geq P_{\min} = 0.92 \tag{9}$$

$$P_{***} = [1 - (1-p)^m]^3 \geq P_{\min} = 0.92 \tag{10}$$

The solutions of the inequalities (9) and (10) are respectively $m = 4$ and $m = 3$, so $P_{**} = 0.9433$ and $P_{***} = 0.9762$.

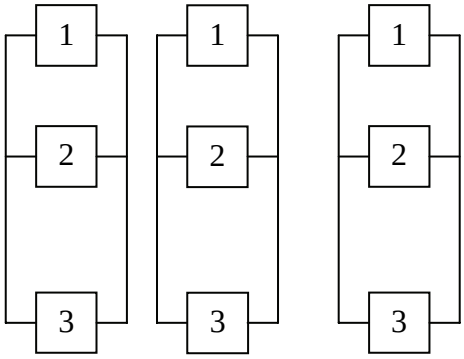


Fig. 5

Therefore in this case it is more expedient to use *parallel – consecutive grouping* with parameter $m = 3$, as at that the expenses on realization will be smaller and the probability $P = 0.9762$ for successful restoration of the secret – considerably bigger. The scheme of this case is given in Fig. 5.

In conclusion, the presented approach can be used to creation a secrets sharing protocols with different resistible level against disloyal participants.

REFERENCES

[1] П. Антонов, В. Антонова, "Подход за усъвършенстване на протокола за разпределяне на секрети", Компютърни науки и технологии, бр. 1, с. 4 – 8, изд. ТУ-Варна, 2005.

[2] В. В. Ященко, Введение в криптографию, СПб.: Питер, 2001. – 288 с.

This page intentionally left blank.

Comparison of Narrowband Interference Suppression Methods for OFDM Systems

Miglen Ovtcharov[1], Vladimir Poulkov[1] and Georgi Iliev[1]

***Abstract* – In this paper we compare two schemes for suppression of narrowband interference (NBI) in OFDM wideband systems. The first one is based on a frequency excision method and the second one is based on adaptive narrowband filtering for the detection and suppression of the interfering signal. The study shows that the two schemes give slightly different performance depending on the type of the channel. It could be seen that it is recommended both methods to be applied together in order to combat effectively the NBI and to obtain better performance for the wideband OFDM system.**

***Keywords* – Narrowband interference suppression, OFDM systems, Adaptive filtering.**

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is becoming leading technology for many applications, such as various kinds of broadband communication systems and recently for Ultra Wide-Band (UWB) systems [1]. Due to their relative low transmission power such systems are very sensitive to Narrowband Interferences (NBI). Because of the spectral leakage effect caused by DFT demodulation at the OFDM receiver, many subcarriers near the interference frequency will suffer serious SINR degradation, which could deteriorate and even block communications [2]. Similar is the case in cable communications systems over unshielded pairs with radio frequency interference, which usually could be treated as NBI for broadband OFDM system. Therefore, NBI suppression is of primary importance for such systems.

The issue of NBI suppression for wideband OFDM systems has been studied extensively in the last years and mainly two types of general approaches are proposed. The first one concerns various frequency excision methods, where the affected frequency bins of the OFDM symbol are excised or their usage avoided. The second approach is related to the so called cancellation techniques, aimed at elimination or mitigation of the effect of the NBI on the received OFDM signal. In most cases the degradation in a wideband OFDM based receiver is beyond the reach of the frequency excision method when the SIR is less than 0dB. Thus, mitigation techniques employing cancellation methods, one of which is based on adaptive notch filtering, are recommended as an alternative [3], [4],[5].

In this paper two schemes for suppression of NBI are compared. The first is a frequency excision method and the second is a cancellation method based on adaptive digital filtering using the LMS algorithm to adapt to the central frequency of the NBI [6]. The study shows that the two schemes give slightly different performance depending on the type of the channel. It could be seen that it is recommended both methods to be applied simultaneously in order to combat effectively the NBI, and to obtain better performance for the wideband OFDM system.

The paper is organized as follows. In section II and III the frequency excision method and the adaptive filtering scheme are considered. In section IV the simulation model used for the comparison is described. Section V presents the computer simulation results. Finally some conclusions are made.

II. FREQUENCY EXCISION METHOD

In the current scheme, a FFT based frequency-domain excision method is used to remove narrowband interferences [7]. The discrete Fourier transform output of each block of N_{FFT} samples, $r_{m,n}$ is given by:

$$r_{m,n} = \sum_{k=1}^{N_{FFT}} r_{m,k} e^{-j2\pi kn/N_{FFT}}, \quad k = 1,....N_{FFT} \qquad (1)$$

In the frequency domain, the narrowband interference manifests itself as a peak in the spectra. By comparing the magnitude of each frequency bin to a threshold and limiting those bins within the threshold, interferences can be excised. The effectiveness of the FFT based method depends on the selection of the threshold. The following method is used for the determination of the threshold. The mean value of the logarithm amplitude of the frequency bins and its variance are computed:

$$T_{mean} = \sum_{n=1}^{N_{FFT}} \frac{10 \log_{10} |r_{m,n}|}{N_{FFT}} \qquad (2)$$

$$T_{\text{var}} = \frac{1}{N_{FFT}} \left[\sum_{n=1}^{N_{FFT}} \left(10\log_{10}|r_{m,n}|\right)^2 - \frac{1}{N_{FFT}} \left(\sum_{n=1}^{N_{FFT}} \left(10\log_{10}|r_{m,n}|\right) \right)^2 \right] \qquad (3)$$

The threshold is determined according to the mean value and variance and is given by:

[1]Miglen Ovtcharov is with the Department of Telecommunications, Technical University of Sofia, Sofia 1000, 8 Kl. Ohridski St., Bulgaria, E-mail: mov@sparnex-coc.com

[1]Vladimir Poulkov is with the Department of Telecommunications, Technical University of Sofia, Sofia 1000, 8 Kl. Ohridski St., Bulgaria, E-mail: vkp@tu-sofia.bg

[1]Georgi Iliev is with the Department of Telecommunications, Technical University of Sofia, Sofia 1000, 8 Kl. Ohridski St., Bulgaria, E-mail: gli@tu-sofia.bg

$$T_{excision} = Tmean - \alpha T_{\mathrm{var}}^{1/2} \quad (4)$$

The scale factor α in the above equation is adjusted to maintain the threshold at some value of the noise floor. Each frequency bin is compared to the threshold and if it exceeds the threshold, its value is held at the threshold. After IFFT, the signal is much less contaminated with narrow band interferences.

III. Adaptive Narrowband Filtering

Compared with the desired wideband signal the interference occupies a much narrower frequency band, but with a higher power spectral density. On the other hand the wideband signal usually has autocorrelation properties quite similar to that of AWGN, so filtering in the frequency domain could be realized. The filtering is performed at the input of the OFDM demodulator. To do this, a simple variable filter section with independent tuning of the central frequency and the bandwidth is used which is then turned into adaptive to implement it in an OFDM receiver. A realization based on second-order Gray-Markel lattice circuit is used [8] – Fig. 1. Using that circuit it becomes possible to implement a second-order notch/bandpass filter illustrated in Fig. 2.

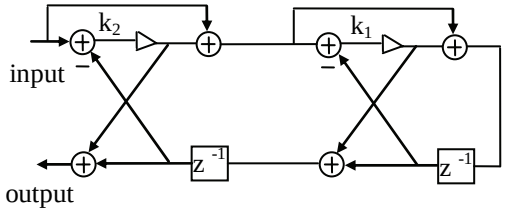


Fig. 1. Second-order lattice Gray-Markel circuit realizing all-pass function A(z)

This implementation has two very important advantages: first extremely low passband sensitivity that means resistance to quantization effects, second independent control of central frequency and filter bandwidth.

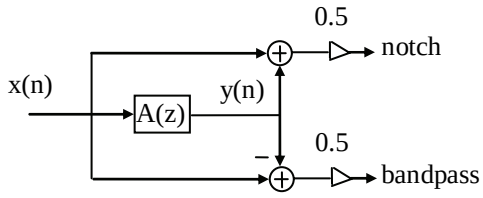


Fig. 2. Second-order notch/bandpass filter

Thus if the allpass function A(z) is

$$A(z) = \frac{k_2 + k_1(1+k_2)z^{-1} + z^{-2}}{1 + k_1(1+k_2)z^{-1} + k_2 z^{-2}} \quad (5)$$

then k_1 controls the central frequency ω_0 while k_2 is related to the bandwidth BW via

$$k_1 = -\cos\omega_0 \quad (6)$$

$$k_2 = \frac{1 - \tan(BW/2)}{1 + \tan(BW/2)}. \quad (7)$$

BW is directly connected to the distance from the pole to the unity-circle and transforming the structure in Fig.4 into an adaptive filter, it is possible to fix the bandwidth and implement an adaptive IIR filter free of stability problems. Adapting k_1 the central frequency can be shifted around the unity-circle.

For the adjustment of filter coefficient a Least Mean Squares (LMS) algorithm is applied as follows

$$k_1(n+1) = k_1(n) - \mu e(n)[x(n-1) - y(n-1)]; \quad (8)$$

where e(n) is the error signal and μ is the step size controlling the convergence speed.

In order to ensure the stability of the adaptive algorithm the range of the step size μ should be set according to [6]

$$0 < \mu < \frac{K}{L\sigma^2}. \quad (9)$$

In this case L is the filter order, σ^2 is the power of the signal [x(n-1) – y(n-1)] and K is a constant depending on the statistical characteristics of the input signal. In most of the practical situations K is approximately equal to 0.1.

IV. Simulation Model

To evaluate the performance of the NBI suppression methods simulations relative to baseband are conducted assuming standard OFDM receiver.

The information source is modeled by a generator of uniformly distributed random integers based on the modified version of Marsaglia's "Subtract with borrow algorithm" [9]. This classes of non-linear random number generators, called add-with-carry (AWC) and subtract-with-borrow (SWB), are capable of quickly generating very long-period pseudo-random number sequences using very little memory. These sequences are essentially equivalent to linear congruential sequences with very large prime moduli. So, the AWC/SWB generators can be viewed as efficient ways of implementing such large linear congruential generators. This method can generate all the double-precision values in the closed interval $[2^{-53}, 1\text{-}2^{-53}]$. Theoretically, it can generate over 2^{1492} values before repeating itself.

The channel encoder is implemented as a convolutional encoder. In the simulation, the code rate: Rc = 1/2 is chosen. In the receiver, a Viterbi hard threshold convolutional decoder is implemented.

A block interleaver - deinterleaver is implemented in the simulation which chooses a permutation table randomly using the initial state input that is provided.

The digital modulator is implemented as 256-point IFFT. The OFDM symbol consists of 128 data bins and 2 pilot tones. Each OFDM data can use different modulation formats. In the experiments Grey encoded 64-QAM modulation format

is used. After the IFFT process, the prefix and suffix guard intervals are added.

The output signal s(t) in the transmitter is a complex OFDM symbol starting at time $t = t_s = kT_s$ [10]:

$$s_k(t) = \begin{cases} w(t-t_s)\sum_{i=Ns/2}^{Ns/2-1} d_{i+Ns(k+1/2)} e^{j2\pi\left(fc-\frac{i+0.5}{Ts}\right)(t-ts-Tprefix)} \\ \qquad\qquad ,t_s \le t < t_s + T_s(1+\beta) \\ 0, \quad t < t_s \wedge t > t_s + T_s(1+\beta) \end{cases} \quad (10)$$

To minimize the spectrum leakage and limit the frequency bandwidth, windowing w(t) is applied to the individual OFDM symbols. A commonly used widow type is Raised Cosine Window, defined as:

$$w(t) = \begin{cases} 0.5 + 0.5\cos\left(\pi + \frac{t\pi}{\beta Ts}\right), & 0 \le t \le \beta Ts \\ 1, \quad \beta Ts \le t \le Ts \\ 0.5 + 0.5\cos\left((t-Ts)\pi + \frac{t\pi}{\beta Ts}\right), & Ts \le t \le (1+\beta)Ts \end{cases} \quad (11)$$

In the simulations a roll-off factor of 0.025, is used.

For the wireless channel a multi-ray model with direct and delayed (reflected) components is used. The delayed components are subject to fading, while the direct one is not. To preserve total signal energy, the direct and delayed signal components are scaled by the square roots of $K/(K + 1)$ and $1/(K + 1)$, respectively. The delay τ is the difference between the propagation time of the delayed component and that of the direct one. To simplify simulations, a complex baseband representation of the system is used [10], [11]. Moreover, to keep simulation memory and computational loads to a minimum, it is desirable to sample at two times modulation symbol rate. This requires delay τ to be a multiple $\boldsymbol{\kappa}$ of the symbol period T_s. With $\tau = kT_s$, the discrete equivalence of the wireless channel simulation model, can be written:

$$r_i = y_i + n_i \quad (12)$$

Where r_i is the complex baseband signal at the receiver side, s_i is the transmitted symbol, y_i is the fading sample and n_i is the complex noise sample.

The fading channel is represented by a FIR filter, where the subscript i indicates that the sample is taken at time $t = iT_s$ and with tap weights given by h_k

$$y(i) = \sum_{j=0}^{N-1} s(i-j)h(j) \quad (13)$$

and

$$h_n = \sum_k \operatorname{sinc}\left(\frac{\tau_k}{T} - n\right) g_k . \quad (14)$$

Here, N is the number of major paths, $\{\tau k\}$ is the set of path delays, T is the input sample period, {gk} is the set of complex path gains, which are not correlated with each other. To generate a particular path gain gk, the model performs the following steps. First, white Gaussian noise is generated. Then the noise is passed through a filter whose transfer function corresponds to the Jakes Doppler spectrum and the output values are interpolated so that the sample period is consistent with that of the signal. The filter is adjusted accordingly to obtain the correct average path gain.

The excision method is applied to the OFDM signal with a NBI at the input of the demodulator. As mentioned in section II the signal is converted into the frequency domain by FFT and the noise peaks in the spectra of the signal are limited to the determined threshold. After this the signal is converted back in the time domain and applied to the input of the demodulator. It should be noted that for more precise frequency excision, FFT of higher order than the one in the demodulator is applied.

For the realization of the suppression method the adaptive notch filter is connected at the receiver's input. An adaptation algorithm tunes the filter in such a way that its central frequency and bandwidth match to the NBI signal spectrum. In the simulations, the central frequency of the notch filter is chosen in such a way that it is equal to the NBI central frequency, while its bandwidth is equal to 20% of the bandwidth between two adjacent OFDM sub-carriers.

In the OFDM demodulator the guard prefix and suffix intervals are removed and 256-point FFT is applied. The pilot tones are removed and respective channel equalization of the OFDM symbol is performed. Finally, corresponding 64-QAM demodulation is done.

V. Experimental Results

Using this general simulation model different experiments are performed, estimating the bit error ratio (BER) as a function of the Signal to Interference Ratio (SIR). The NBI is modeled as a single tone the frequency of which is located in the middle between two adjacent OFDM sub-carriers. Three types of channels are considered: AWGN, Rayleigh and Rician. The Rayleigh and Rician channels are subject to strong fading and additionally background AWGN is applied, so that the signal to AWGN ratio at the input of the OFDM receiver is 20dB. In Fig. 3, a standard Gaussian channel is considered. The SIR is varied from -20 dB to 10 dB. It could be seen that for high NBI, where the SIR is less than 0 dB both suppression methods lead to a significant improvement in performance. The filtering scheme gives better performance than the frequency excision method. This could be explained with the NBI spectral leakage effect caused by DFT demodulation at the OFDM receiver, when many sub-carriers near the interference frequency suffer degradation. Thus filtering out the NBI before demodulation is better than frequency excision. Additional improvement is achieved by employing both methods together. In the case of Rayleigh and Rician types of channels Fig.4 and Fig.5 it could be seen that similar results are obtained. It should be noted that the filtering scheme leads to a degradation of the overall performance when SIR >0, which is due to the amplitude and phase distortion of the filter. The degradation could be reduced by the implementation of a higher-order notch filter or avoided by simply by switching off the filter when SIR>0.

Such a scheme is easily realizable as the amplitude of the NBI can be monitored at the bandpass output of the code.

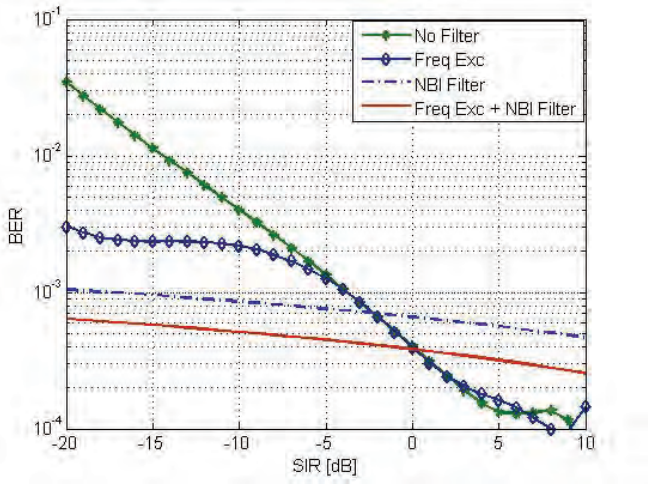


Fig. 3. BER as a function of SIR for AWGN channel

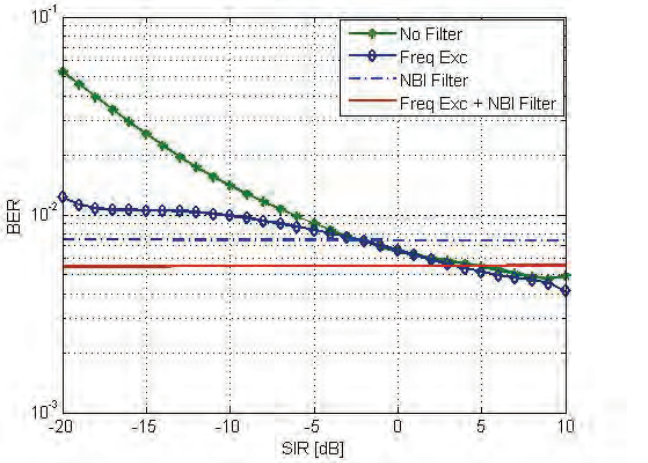


Fig. 4. BER as a function of SIR for Rayleigh channel

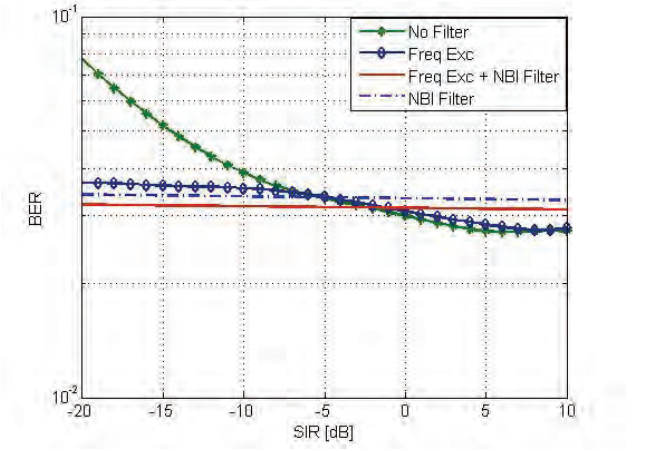


Fig. 5. BER as a function of SIR for Rician channel

VI. Conclusions

In this paper, a comparison of two schemes for suppression of NBI is performed. The first is a frequency excision method and the second is a cancellation method based on adaptive digital filtering using the LMS algorithm to adapt to the central frequency of the NBI. The experiments show that for high NBI, where the SIR is less than 0 dB both suppression methods lead to a significant improvement in performance. With SIR<0 the filtering scheme gives better performance than the frequency excision method. The performance of the two schemes is similar when using different types of channels. It could be seen that it is recommended both methods to be applied simultaneously in order to combat effectively the NBI, and to obtain better performance for the wideband OFDM system.

References

[1] A. Bara et al., “Multi-band OFDM Physical Layer Proposal” IEEE P802.15-04/0493r1-TG3a. Sept. 2004.

[2] A. Giorgetti, M. Chiani, M.Z.Win, “The Effect of Narrowband Interference on Wideband Wireless Communication Systems”, IEEE Trans. Communications, vol. 53, no.12, pp.2139-2149, Dec. 2005.

[3] C. Carlemalm, H.V. Poor, A. Logothetis, “Suppression of multiple narrowband interferers in a spread-spectrum communication system”, IEEE J. Select. Areas Communications, vol.3, no. 5, pp.1431-1436, Sept 2004.

[4] L. B. Milstein, “Interference rejection techniques in spread spectrum communications”, Proc. IEEE, vol. 76, no. 6, pp. 657-671, June 1988.

[5] J. D. Laster and J. H. Reed, “Interference rejection in digital wireless communications”, IEEE Signal Proc. Mag., vol. 14, no. 3, pp. 37 - 62, May 1997.

[6] S. Douglas, Adaptive filtering, in Digital Signal Processing Handbook, D. Williams and V. Madisetti, Eds., Boca Raton: CRC Press LLC, 1999, pp. 451-619.

[7] Jyh-Ching Juang, Chung-Liang Chang, Yu-Lung Tsai. An Interference Mitigation Approach Against Pseudolites, The 2004 International Symposium on GNSS/GPS, 6–8 December 2004, Sydney, Australia

[8] P. Regalia, S. Mitra, P. Vaidyanathan, “The all-pass filter: a versatile signal processing building block”, Proc. IEEE, pp. 19-37, Jan. 1988.

[9] Shu Tezuka , Pierre L'Ecuyer , Raymond Couture, On the lattice structure of the add-with-carry and subtract-with-borrow random number generators., ACM Transactions on Modeling and Computer Simulation (TOMACS, Volume 3 , Issue 4 (October 1993), Pages: 315 – 331.

[10] Richard van Nee, Ramjee Prasad, “OFDM for Wireless Multimedia Communications”, Artech House, 2000.

[11] Henrik Shulze, Christian Luders, “Theory and Applications of OFDM and CDMA”, John Wiley, 2005.

The Linear Complexity of the LFSR Based Generalized Shrinking-Multiplexing Generator

Todor D. Tashev [1], Borislav Y. Bedzhev[2] and Zhaneta N. Tasheva[3]

***Abstract* – An architecture of Generalized Shrinking-Multiplexing Generator (*GSMG*), based on Linear Shift Feedback Registers (*LFSRs*), is investigated in the paper. The linear complexity of its output binary pseudo random sequences is established. Some linear complexity analysis is given. The established *GSMG* properties show that the proposed architecture allows producing binary pseudo random sequences with good properties like uniform distributions of 1s and 0s, unpredictable nonlinearity, enormous period and large linear complexity.**

***Keywords* – Cryptography, Pseudo Random Number Generator, Stream Cipher, Clock Controlled Generators, *LFSR*, Linear Complevity.**

I. INTRODUCTION

Nowadays, the clock controlled Pseudo Random Number Generators (*PRNGs*) are an important tool for development of stream ciphers, applied in the communication information systems. On the one hand, their high performance velocity and cost-effective implementation is based on their simple architecture which combines fast and cheap elements like Linear Feedback Shift Registers (*LFSRs*) and Feedback with Carry Shift Registers (*FCSRs*) with some nonlinear functions [2, 3, 5]. On the other hand, the performance quality of the clock controlled *PRNGs* [6, 7, 8] depends on their crypto resistance, which is connected with its ability to generate nonlinear Pseudo Random Sequence (*PRS*) with enormous period, uniform distribution and large linear complexity.

Due to this reason the aim of this paper is to investigate the linear complexity of a *LFSR* based *Generalized Shrinking-Multiplexing Generator* (*GSMG*). The paper is organised as follows. First, the *LFSR* based *GSMG* architecture is described. Second, the linear complexity of the *LFSR* based *GSMG* is established. After that some linear complexity analysis and a comparison with the Shrinking Generator are given. Finally, the advantages and possible application areas of the *LFSR* based *GSMG* are discussed.

[1] Eng. Todor D. Tashev, PhD. Student in Communication Network and Systems, Shumen University, Shumen, Bulgaria, Phone: +359 54 980 422, e-mail: todor_tashev@yahoo.com.

[2] Assoc. Prof. DSc. Eng. Borislav Y. Bedzhev, NMU "V. Levski", Faculty of Artillery, Air Defence and Communication-Information Systems, Shumen, Bulgaria, Phone: +359 54 88 64 38, e-mail: bedzhev@mail.pv-ma.bg.

[3] Assoc. Prof. PhD. Eng. Zhaneta N. Savova-Tasheva, NMU "V. Levski", Faculty of Artillery, Air Defence and Communication-Information Systems, Shumen, Bulgaria, Phone: +359 54 980 422, e-mail: tashevi86@yahoo.com.

II. THE LFSR BASED GSMG ARCHITECTURE

The proposed general architecture of the *GSMG* [9] can be realized by means of linear or nonlinear pseudo random sequences. There are eight possible variants of the *GSMG* architecture depending on the linear constructive elements, which most often are fast and cheap *LFSRs* and *FCSRs*. Most of these *GSMG* architectures are statistically analyzed by the authors of [11, 12, 13] but strong mathematical analysis have not been made yet. Here the fifth architecture, proposed in [9], is analyzed.

The *LFSR* based *GSMG* architecture (Fig. 1) uses as building modules *LFSRs*.

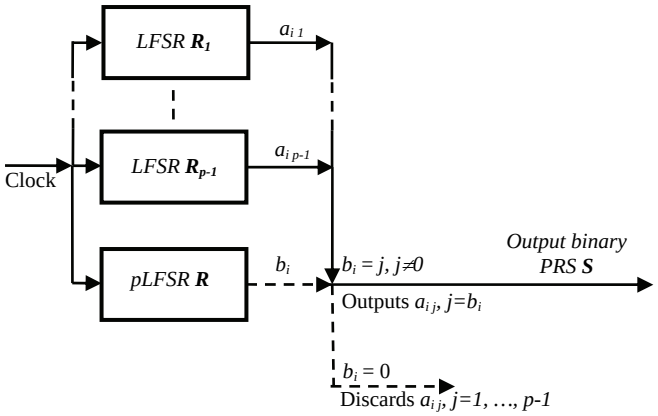


Fig. 1. The LFSR Based Generalized Shrinking – Multiplexing Generator

Definition 1: A *LFSR* based *GSMG* comprises a *pLFSR* $\boldsymbol{R}$ of length L which produces one p-ary number in a time and $p-1$ slaved *LFSRs* of length $L_1, L_2, \ldots, L_{p-1}$. The clock controls the movement of a data in all used *LFSRs*.

The algorithm of *LFSR* based *GSMG* consists of the following steps:

1. All slaved *LFSRs* and control *pLFSR* are clocked.

2. If the p-ary output of the control *pLFSR* $\boldsymbol{R}$ at moment i is non-zero ($b_i = j, j \neq 0$), the binary output of the slaved *LFSRs* $\boldsymbol{R_j}$ forms a part of the *LFSR* based *GSMG* output *PRS* $\boldsymbol{S}$.

3. Otherwise, if the output of the control *pLFSR* $\boldsymbol{R}$ is equal to 0 ($b_i = 0$), the outputs of all slaved *LFSRs* $\boldsymbol{R_1}, \boldsymbol{R_2}, \ldots, \boldsymbol{R_{p-1}}$ are discarded.

Therefore, the produced binary *PRS* is a *shrunken* version of the slaved binary *PRSs*, generated by the *LFSRs* $\boldsymbol{R_1} \div \boldsymbol{R_{p-1}}$, when the output of the control *pPRS* $\boldsymbol{B}$ is zero, and a *mixed*

version of the slaved *PRSs*, when the output is nonzero. Due to this reason the output *LFSR* based *GSMG* sequence $\boldsymbol{S}$ is nonlinear and unpredictable with more complexity than the sequences, produced by the slaved *LFSRs*. The nonlinearity in the *LFSR* based *GSMG* architecture is a result of the fact that the linear algebraic structure of the slaved *LFSR* sequences is destroyed by means of the shrinking and multiplexing.

III. The Linear Complexity of the Sequences Generated by the LFSR Based GSMG

In this section it is proved the exponential bounds of the linear complexity of sequences, generated by the *LFSR* based *GSMG*. The importance of the exponentially large *PRS* linear complexity follows from the strong necessity of avoiding some popular attacks on *PRSs* or stream chippers. There is no need to know the way a *PRS* is generated in order to break it through its linear complexity. In fact, any *PRS* with linear complexity λ can be easily reconstructed if 2λ bits are known by the Berlekamp-Massey algorithm [1, 4], which in time $O(\lambda^2)$ finds the shortest *LFSR* generating this *PRS*.

Here it should be mentioned that the high linear complexity is only necessary but not sufficient condition *PRNG* to have good cryptographic properties. There are many others conditions like period; uniform distribution of *d*-tuples for a large range of *d*; good, usually lattice-liked, structure in high dimensions; good statistical properties; resistance to known attacks and so on.

A. Theoretical analysis

The following symbols are used when the linear complexity of the *LFSR* based *GSMG* sequence is established:

- L_i, $i = 1, 2, \ldots, p-1$ – length of the slaved *LFSR* $\boldsymbol{R_i}$;
- L - length of the control *pLFSR* $\boldsymbol{R}$;
- T_i, $i = 1, 2, \ldots, p-1$ – period of the slaved *LFSR* $\boldsymbol{R_i}$;
- T – period of the control *pLFSR* $\boldsymbol{R}$;
- T_s – period of the generated by the *LFSR* based *GSMG* sequence $\boldsymbol{S}$;
- $N_{(\neq 0)}$ – quantity of the nonzero elements in a period of the control *p*PRS sequence;
- $a_j(i)$ – the i-th element of the sequence $\boldsymbol{A_j}$, $j = 1, \ldots, p-1$, generated by the slaved *LFSR* $\boldsymbol{R_j}$;
- $b(i)$ – the i-th element of the sequence $\boldsymbol{B}$, generated by the control *pLFSR* $\boldsymbol{R}$;
- $s(i)$ – the i-th element of the sequence $\boldsymbol{S}$, generated by the *LFSR* based *GSMG*;
- k_{ij} – i-th position with value j in the sequence $\boldsymbol{B}$, generated by the control *pLFSR* $\boldsymbol{R}$.

Also the following Theorem 1 determining the period T_s of the generated by the *LFSR* based *GSMG* sequence $\boldsymbol{S}$ is used. Refer to [10] to see the proof of the Theorem 1.

Theorem 1: If the generated by the slaved *LFSR* $\boldsymbol{R_j}$ sequence $\boldsymbol{A_j}$, $j = 1, \ldots, p-1$ and the generated by the control *pLFSR* $\boldsymbol{R}$ sequence $\boldsymbol{B}$ have maximal length (i.e. have primitive connections) and all periods T_i are co-prime with the period T, i.e. the greatest common devisor is $(T_i, T) = 1$ for $i = 1, 2, \ldots, p-1$, then the output shrinking and multiplexing sequence, generated by the *LFSR* based *GSMG*, has a maximal period defined by the equation:

$$T_S = N_{(\neq 0)} \cdot \prod_{i=1}^{p-1} T_i = N_{(\neq 0)} \cdot \prod_{i=1}^{p-1} 2^{L_j} - 1 \qquad (1)$$

The linear complexity λ_s of the sequence $\boldsymbol{S}$ generated by the *LFSR* based *GSMG* satisfies the following Theorem 2.

Theorem 2: If the generated by the slaved *LFSR* $\boldsymbol{R_j}$ sequence $\boldsymbol{A_j}$, $j = 1, \ldots, p-1$ and the generated by the control *pLFSR* $\boldsymbol{R}$ sequence $\boldsymbol{B}$ have maximal length (i.e. have primitive connections) and all periods T_i are co-prime with the period T, i.e. the greatest common devisor is $(T_i, T) = 1$ for $i = 1, 2, \ldots, p-1$, then the output shrinking and multiplexing sequence $\boldsymbol{S}$, generated by the *LFSR* based *GSMG*, has a linear complexity λ_s satisfied the inequality

$$\frac{(p-1)p^{L-1}}{2} \prod_{i=1}^{p-1} L_i < \lambda_S \le (p-1)p^{L-1} \prod_{i=1}^{p-1} L_i \,. \qquad (2)$$

The next proposition, which follow from the definition of the *LFSR* based *GSMG*, is used in the proof.

Proposition 1: The integers $N_{(\neq 0)}$ and T are connected by equation

$$s(i + n.N_{(\neq 0)}) = a_j(k_i + nT), \quad \text{for } n = 0, 1, \ldots \,. \qquad (3)$$

Proof: To determine an *upper bound* on the linear complexity λ_s of the sequence $\boldsymbol{S}$, it is sufficed to find a polynomial $P(.)$, for which $P(s) = 0$, i.e. the coefficients of $P(.)$ represents a linear dependency satisfied by the elements of a sequence $\boldsymbol{S}$. Let $s^{N(\neq 0)}$ be the sequence $s(nN_{(\neq 0)}), n = 0, 1, \ldots$, i.e. the sequence $\boldsymbol{S}$ is decimated by $N_{(\neq 0)}$.

Proposition 1 states that this decimation results in transformations of every slaved sequence of the form $a_j(i + nT)$, $j = 1, 2, \ldots, p-1$. Since $(T_j, T) = 1$, $j = 1, 2, \ldots, p-1$, the above sequences $a_j(i + nT)$, $j = 1, 2, \ldots, p-1$ have maximal length and have the same linear complexity as the original sequences $a_j(i)$, $j = 1, 2, \ldots, p-1$. Therefore, the polynomials $Q_j(.)$ of degree L_i exist for which $Q_j(a_j) = 0$. But then the decimated sequence $s^{N(\neq 0)}$ satisfies polynomials $Q_j(.)$, i.e.

$$Q_j(s^{N(\neq 0)}) = 0, \; j = 1, 2, \ldots, p-1. \qquad (4)$$

Hence, a polynomial

$$P(s) = \prod_{j=1}^{p-1} Q_j(s^{N(\neq 0)}) \qquad (5)$$

of degree $N_{(\neq 0)}\prod_{i=1}^{p-1} L_i$, such that $P(s) = 0$, is found.

Consequently, the linear complexity λ_s of the sequence $\boldsymbol{S}$ generated by the *LFSR* based *GSMG* is at most

$$\lambda_S \le N_{(\neq 0)}\prod_{i=1}^{p-1} L_i = (p-1)p^{L-1}\prod_{i=1}^{p-1} L_i \ . \qquad (6)$$

To determine an *lower bound* on the linear complexity λ_s of the sequence $\boldsymbol{S}$, it is necessary to find the minimal polynomial $M(s)$ for which $M(s) = 0$. Since the sequence $\boldsymbol{S}$ satisfied the equation (4), then the polynomial $M(s)$ divides each polynomial $Q_j(s^{N_{(\neq 0)}})$, $j = 1, 2, \ldots, p-1$. After putting the equation

$$N_{(\neq 0)} = (p-1)p^{L-1} \qquad (7)$$

in (4), the following equations are obtained

$$Q_j(s^{N_{(\neq 0)}}) = Q_j(s^{(p-1)p^{p-1}}) =$$

$$= (Q_j(s))^{(p-1)p^{p-1}},\ j = 1, 2, \ldots, p-1. \qquad (8)$$

Therefore, the minimal polynomial $M(s)$ must be in the form $(Q_j(s))^r$ for $r \le (p-1)p^{p-1}$. The p is a prime number and hence, $p - 1$ is even.

The following assumptions will be made

$$r \le \frac{(p-1)p^{p-1}}{2}. \qquad (9)$$

Then the minimal polynomial $M(s)$ divides each polynomial $(Q_j(s))^{\frac{(p-1)p^{p-1}}{2}}$, $j = 1, 2, \ldots, p-1$. Since $Q_j(s)$, $j = 1, 2, \ldots, p-1$ are irreducible polynomials of degree ca L_j, they divide the polynomials $1 + x^{T_j}$ respectively.

TABLE I
LINEAR COMPLEXITY OF THE PRSS, GENERATED BY THE LFSR BASED GSMG, WITH P = 3

№	Used *PRSs*	Primitive Polynomials	Length	Period T_S	Linear Complexity of the *PRSs*, generated by the *LFSR* based *GSMG*		
					Lower Bound	Upper Bound	Real
1	*PRS* $\boldsymbol{A_1}$	$1 + x + x^3$	$L_1 = 2$	7.15.18 = 1890	$3^2.3.4 = 108$	$108.2 = 216$	126
	PRS $\boldsymbol{A_2}$	$1 + x + x^4$	$L_2 = 3$				
	3PRS $\boldsymbol{B}$	$1 + 2x^2 + x^3$	$L = 2$				
2	*PRS* $\boldsymbol{A_1}$	$1 + x + x^2$	$L_1 = 2$	3.15.18 / 3 = 270	$3^2.2.4 = 72$	$72.2 = 144$	108
	PRS $\boldsymbol{A_2}$	$1 + x + x^4$	$L_2 = 4$				
	3PRS $\boldsymbol{B}$	$1 + 2x^2 + x^3$	$L = 3$				
3	*PRS* $\boldsymbol{A_1}$	$1 + x + x^3$	$L_1 = 3$	7.3.18 = 378	$3^2.2.3 = 54$	$54.2 = 108$	90
	PRS $\boldsymbol{A_2}$	$1 + x + x^2$	$L_2 = 2$				
	3PRS $\boldsymbol{B}$	$1 + 2x^2 + x^3$	$L = 3$				
4	*PRS* $\boldsymbol{A_1}$	$1 + x + x^3$	$L_1 = 3$	7.15.6 = 630	$3^1.3.4 = 36$	$36.2 = 72$	42
	PRS $\boldsymbol{A_2}$	$1 + x + x^4$	$L_2 = 4$				
	3PRS $\boldsymbol{B}$	$2 + x + x^2$	$L = 2$				
5	*PRS* $\boldsymbol{A_1}$	$1 + x + x^2$	$L_1 = 2$	3.15.6 / 3 = 90	$3^1.2.4 = 24$	$24.2 = 48$	36
	PRS $\boldsymbol{A_2}$	$1 + x + x^4$	$L_2 = 4$				
	3PRS $\boldsymbol{B}$	$2 + x + x^2$	$L = 2$				
6	*PRS* $\boldsymbol{A_1}$	$1 + x + x^3$	$L_1 = 3$	7.3.6 = 126	$3^1.2.3 = 18$	$18.2 = 36$	30
	PRS $\boldsymbol{A_2}$	$1 + x + x^2$	$L_2 = 2$				
	3PRS $\boldsymbol{B}$	$2 + x + x^2$	$L = 2$				
7	*PRS* $\boldsymbol{A_1}$	$1 + x + x^3$	$L_1 = 3$	7.3.54 = 1134	$3^3.2.3 = 162$	$162.2 = 324$	270
	PRS $\boldsymbol{A_2}$	$1 + x + x^2$	$L_2 = 2$				
	3PRS $\boldsymbol{B}$	$2 + x + x^4$	$L = 4$				
8	*PRS* $\boldsymbol{A_1}$	$1 + x + x^3$	$L_1 = 3$	7.31.6 = 1302	$3^1.3.5 = 45$	$45.2 = 90$	48
	PRS $\boldsymbol{A_2}$	$1 + x^2 + x^5$	$L_2 = 5$				
	3PRS $\boldsymbol{B}$	$2 + x + x^2$	$L = 2$				
9	*PRS* $\boldsymbol{A_1}$	$1 + x + x^2$	$L_1 = 2$	3.31.6 = 558	$3^1.2.5 = 30$	$30.2 = 60$	42
	PRS $\boldsymbol{A_2}$	$1 + x^2 + x^5$	$L_2 = 5$				
	3PRS $\boldsymbol{B}$	$2 + x + x^2$	$L = 2$				

Consequently, polynomial $M(s)$ divides

$$(1+x^{T_j})^{\frac{(p-1)p^{p-1}}{2}} = 1+x^{T_j \cdot \frac{(p-1)p^{p-1}}{2}}. \qquad (10)$$

But then the period of the sequence $\boldsymbol{S}$, generated by the *LFSR* based *GSMG*, is at most

$$T_S = \frac{(p-1)p^{L-1}}{2} \cdot \prod_{i=1}^{p-1}(2^{L_i}-1). \qquad (11)$$

This contradicts to the Theorem 1, because

$$\frac{(p-1)p^{L-1}}{2} < N_{(\neq 0)} = \frac{(p^L-1)(p-1)}{p}. \qquad (12)$$

Therefore, the assumption isn't true and $r > \frac{(p-1)p^{p-1}}{2}$, i.e. the lower bound on the linear complexity λ_s of the sequence $\boldsymbol{S}$ is

$$\lambda_S > \frac{(p-1)p^{L-1}}{2} \prod_{i=1}^{p-1} L_i. \qquad (13)$$

This conclusion ends the proof of the Theorem 2.

B. Practical analysis

The *LFSR* based *GSMG* architecture is modelled in Visual C++ environment. The linear complexity of the generated by the *LFSR* based *GSMG* sequences is practically analyzed by means of the Berlekamp-Massey algorithm. The theoretical lower and upper bounds of the linear complexity λ_s, given by the Theorem 2 and the found by the Berlekamp-Massey algorithm real λ_s are given in Table 1. The period of the output shrinking and multiplexing sequence $\boldsymbol{S}$ also is shown in the Table 1.

The practical analysis of the linear complexity and period of the sequences $\boldsymbol{S}$, generated by the *LFSR* based *GSMG*, confirm the theoretical results given by Theorem 1 and Theorem 2, i.e. the exponential period and exponential bounds of the linear complexity.

IV. CONCLUSION

In this paper the linear complexity of the *LFSR* based Generalized Shrinking-Multiplexing Generator is investigated mainly through algebraic techniques. It is proved the exponential lower and upper bounds of the linear complexity λ_s. Thus, the proposed *LFSR* based *GSMG* architecture allows to produce binary pseudo random sequences with good properties like uniform distributions of 1s and 0s, unpredictable nonlinearity, enormous period and large linear complexity. This shows that the elemental goals of the pseudo random number generators are achieved by *LFSR* based *GSMGs*. Consequently, they can be used as a part of a stream ciphers in the height-speed communication applications.

ACKNOWLEDGEMENT

We will be glad to thanks everyone who helps us to make some strong cryptanalysis of the *LFSR* based *GSMG*.

REFERENCES

[1] N. B. Atti, G. M. Diaz–Toca, H. Lombardi, The Berlekamp-Massey Algorithm revisited, Springer Berlin, Heidelberg, Volume 17, Number 1, April, 2006, http://www.springerlink.com/content/hk46h32538075152/.

[2] D. Coppersmith, H. Karwczayk, Y.Mansour, The Shrinking Generator, *Crypto'93*, http://imailab-www.iis.u-tocio.ac.jp/limit/Papers/Crypto_Eurocrypt/ HTML/PDF/C93/22.pdf

[3] M. Goresky, A. Klapper, Fibonacci and Galois Representations of Feedback-With-Carry Shift Registers, *IEEE Trans. on Inform. Theory*, vol. 48, pp. 2826−2836, November 2002.

[4] S.Greenberg, N. Feldblum, G. Melamed, Implementation of the Berlekamp-Massey algorithm using a DSP, Proceedings of the 2004 11th IEEE International Conference Electronics, Circuits and Systems, ICECS 2004, ISBN: 0-7803-8715-5, pp. 358- 361, 2004.

[5] W. Meier, O. Staffelbach, The Self-Shrinking Generator, *Advances in Cryptology*, Eurocrypt 1994 (LNCS 950), 205-214, 1995, http://homes.esat.kuleuven.be/~jlano/stream/papers/SSGms.pdf.

[6] A. Menezes, P. Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, p. 780, 1997, www.cacr.math.uwaterloo.ca/hac.

[7] R. Ruepel, *Analysis and Design of Stream Ciphers*, Springer Verlag, N. Y., 1986.

[8] B. Schneier, *Applied Cryptography*, John Wiley & Sons, New York, 1996.

[9] T. Tashev, B. Bedzhev, Zh. Tasheva, The Generalized Shrinking-Multiplexing Generator, International Conference on Computer Systems and Technologies - *CompSysTech*'07, June 22-26, 2007, Bulgaria, under printing.

[10] T. Tashev, The Period of the LFSR Based Generalized Shrinking-Multiplexing Generator, International Conference on Computer Systems and Technologies - *CompSysTech*'07, June 22-26, 2007, Bulgaria, under printing.

[11] Zh. Tasheva, B. Bedzhev, V. Mutkov, A Shrinking Data Encryption Algorithm with p-adic Feedback with Carry Shift Register, *Conference Proc. of XII Int. Symp. on Theoretical Electrical Engineering*,ISTET'03,July 6-9, 2003, Warsaw, Poland, Volume II, pp.397–400, 2003.

[12] Zh. Tasheva, B. Bedzhev, B. Stoyanov, N-adic Summation-Shrinking Generator. Basic properties and empirical evidences, *Cryptology ePrint Archive*, Co-Editors: Mihir Bellare, UCSD Christian Cachin, IBM Zurich, Accepted and posted with Number 2005/068, http://eprint.iacr.org/2005/068.pdf.

[13] Zh. Tasheva, B. Bedzhev, B. Stoyanov, P-adic Shrinking–Multiplexing Generator, IEEE Third International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications IDAACS'2005, September 5-7, 2005 Sofia, Bulgaria, 2005.

Analysis of a Matching Network Effect on Noise

Pesha D. Petrova[1] and Dimitar P. Petrov[2]

***Abstract* – Effect of a matching network on noise is analyzed. Analysis is applied to an amplifier model with a Thévenin source. Noise characteristics for amplifier model without and with two classes of lossless matching networks between the source and the amplifier input are studied. Some numerical simulations for a matching network effect on noise characteristics demonstrating the effectiveness of the proposed analysis are presented.**

***Keywords* – Amplifier, matching network, noise, Thévenin source, simulation.**

I. INTRODUCTION

Noise behavior is very important characteristic of electronic circuits, such as amplifiers, filters etc, as it usually determines the fundamental limit of the performance of circuits [1]. When designing amplifiers for specific applications, there are many conflicting requirements, to be met and decisions to be made [2] - [4].

As we know, the noise factor for the amplifier is defined as

$$F = 1 + \frac{P_{na}}{GP_{ni}} > 1 \qquad (1)$$

where P_{na} is the noise power added by the amplifier, P_{ni} is the input signal power, and G is the gain of the amplifier.

A common way of presenting the noise factor is the noise figure, i.e.

$$NF = 10 \log F \qquad (2)$$

These equations give a hint of why a low noise figure is considered important, and why a high gain is often associated with a low noise figure. Noise figure is, however, often overstated in its importance in a low noise design system. It is the correct balance between noise figure and gain that is important rather than a low noise figure itself. Because of the conflicting requirements, some compromises are often made in radio frequency amplifier design between lowest noise and largest gain. Indeed, it can be proven [4], that there exists a combination of components which could match the signal source impedance to the amplifier input. To study how the addition of a matching network between the source and the amplifier effects on noise characteristics, an amplifier without and with a matching network is analyzed.

II. MATCHING NETWORK EFFECT ON NOISE

The amplifier noise model with a Thévenin source and a lossless input matching network is shown in Fig. 1.

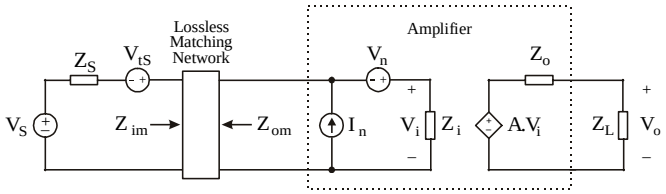


Fig. 1. Amplifier noise model with an input matching network.

The noise source V_{ts} models the thermal noise generated by the signal source V_s having output impedance $Z_s = R_s + jX_s$. The amplifier input impedance is $Z_i = R_i + jX_i$ and its output impedance $Z_o = R_o + jX_o$, respectively. The matching network has an input impedance $Z_{im} = R_{im} + jX_{im}$ and output impedance $Z_{om} = R_{om} + jX_{om}$. The noise generated by the amplifier is modeled by the noise sources V_n and I_n.

The effect of a matching network between the source and the amplifier on noise can be studied analyzing the model in Fig. 1.

A. Equivalent amplifier noise input voltage

The total power delivered to the matching network by the source is

$$P_{im} = \left| \frac{V_s + V_{ts}}{Z_s + Z_{im}} \right|^2 Re(Z_{im}) = \frac{v_s^2 + 4kTR_s\Delta f}{|Z_s + Z_{im}|^2} R_{im} \qquad (3)$$

where T is the absolute temperature in K, k is Boltzmann's constant, and Δf is the bandwidth over which the noise is measured.

Let V_{is} and V_{its}, respectively, be the voltages at the amplifier input due to V_s and V_{ts}. The output power from the matching network is

[1]Pesha D. Petrova is with the Department of Communication Equipment and Technologies, Technical University of Gabrovo, 5300 Gabrovo, 4, Hadji Dimitar St., Bulgaria, E-mail: daneva@tugab.bg

[2]Dimitar D. Petrov is with the Department of Electrical Engineering, Technical University of Gabrovo, 5300 Gabrovo, 4 Hadji Dimitar St., Bulgaria, E-mail: dpetrov@tugab.bg

$$P_{om} = \left|\frac{V_{is}+V_{its}}{Z_i}\right|^2 Re(Z_i) = \frac{v_{is}^2 + v_{its}^2}{|Z_i|^2} R_i \ . \quad (4)$$

Because the matching network is lossless, it follows that $P_{im} = P_{om}$. This leads to the relation

$$v_{is}^2 + v_{its}^2 = \frac{R_{im}}{R_i}\left|\frac{Z_i}{Z_s + Z_{im}}\right|^2 \left(v_s^2 + 4kTR_s\Delta f\right) . \quad (5)$$

Equation (5) can be rewritten as

$$v_{is}^2 + v_{its}^2 = \left|\frac{Z_i}{Z_i + Z_{om}}\right|^2 \left[\frac{R_{im}}{R_i}\left|\frac{Z_i + Z_{om}}{Z_i + Z_{im}}\right|^2 \left(v_s^2 + 4kTR_s\Delta f\right)\right]$$

$$= \left|\frac{Z_i}{Z_i + Z_{om}}\right|^2 \left[v_{is(oc)}^2 + v_{its(oc)}^2\right] \quad (6)$$

where $v_{is(oc)}^2$ and $v_{its(oc)}^2$ are the open-circuit values v_{is}^2 and v_{its}^2. The latter is given by

$$v_{its(oc)}^2 = \frac{R_{im}}{R_i}\left|\frac{Z_i + Z_{om}}{Z_i + Z_{im}}\right|^2 4kTR_s\Delta f \ . \quad (7)$$

This equation must be of the form $v_{its(oc)}^2 = 4kTR_{om}\Delta f$, where $R_{om} = Re(Z_{om})$. In this case, the output resistance of the matching network can be expressed as

$$R_{om} = \frac{v_{its(oc)}^2}{4kT\Delta f} = \frac{R_{im}}{R_i}\left|\frac{Z_i + Z_{om}}{Z_s + Z_{im}}\right|^2 R_s \ . \quad (8)$$

When this equation is solved for $\frac{R_{im}}{R_i}$ and the result used in Eq. (5), it follows that

$$v_{is}^2 + v_{its}^2 = \left|\frac{Z_i}{Z_i + Z_{om}}\right|^2 \frac{R_{om}}{R_s}\left(v_s^2 + 4kTR_s\Delta f\right) . \quad (9)$$

It might seem a contradiction that R_{om} in Eq. (8) is a function of Z_i. However, the dependence cancels because Z_{im} is also a function of Z_i. It can be concluded from Eq. (9) that the mean-square open-circuit output voltage from the matching network is given by

$$v_{is(oc)}^2 + v_{its(oc)}^2 = \frac{R_{om}}{R_s}\left(v_s^2 + 4kTR_s\Delta f\right) . \quad (10)$$

To obtain the total mean-square open-circuit voltage at the input to the amplifier, the contributions of the noise sources V_n and I_n must be added to Eq. (10). The addition of noise signals is accomplished mathematically by simply adding the functions, which describe these signals. Taking the effect of the noise sources into consideration [5], the following result is obtained

$$v_{i(oc)}^2 = v_{is(oc)}^2 + v_{its(oc)}^2 + v_n^2 + 2v_n i_n Re\left(cZ_{om}^*\right) + i_n^2|Z_{om}|^2$$

$$= \frac{R_{om}}{R_s}\left(v_s^2 + 4kTR_s\Delta f\right) + v_n^2 + 2v_n i_n Re\left(cZ_{om}^*\right) + i_n^2|Z_{om}|^2 \quad (11)$$

where $c = c_r + jc_i$ is the correlation coefficient between the noise sources V_n and I_n.

This result shows that the mean-square equivalent amplifier noise input voltage in series with the signal source V_s can be expressed as

$$v_{ni}^2 = 4kTR_s\Delta f + \frac{R_s}{R_{om}}\left[v_n^2 + 2v_n i_n Re\left(cZ_{om}^*\right) + i_n^2|Z_{om}|^2\right]. \quad (12)$$

B. Signal - to - Noise Ratio

When the source is modeled by a Thévenin equivalent circuit as in Fig.1, the signal – to – noise ratio (SNR) is $SNR = v_s^2 / v_{ni}^2$. Substituting Eq. (12) for v_{ni}^2 leads to the equation

$$SNR = \frac{v_s^2}{4kTR_s\Delta f + \frac{R_s}{R_{om}}\left[v_n^2 + 2v_n i_n Re\left(cZ_{om}^*\right) + i_n^2|Z_{om}|^2\right]}. \quad (13)$$

C. Noise Factor

The noise factor for the amplifier model is

$$F = \frac{v_{ni}^2}{v_{ts}^2} = \frac{v_{ni}^2}{4kTR_s\Delta f} \ . \quad (14)$$

Substituting Eq. (12) into Eq. (14) gives

$$F = 1 + \frac{v_n^2 + 2v_n i_n Re\left(cZ_{om}^*\right) + i_n^2|Z_{om}|^2}{4kTR_{om}\Delta f}. \quad (15)$$

This is the same as the noise factor calculated at the output of the matching network. The basic reason that the noise factors at the source and at the output of the matching network are equal is because a lossless matching network cannot add noise. Thus it follows that the signal-to-noise ratio is also the same at the input to the matching network as it is at the input to the amplifier. However, these conclusions do not hold for a lossy matching network. Equation (15) can be used to predict the noise factor for any matching network.

D. Signal Power Delivered to Z_i

It follows from Eq. (10), that the open-circuit value v_{is}^2 can be expressed by

$$v_{is(oc)}^2 = \frac{R_{om}}{R_s} v_s^2 . \quad (16)$$

The signal power delivered to the amplifier input is given by

$$P_i = i_i^2 \, Re(Z_i) = \frac{v_{is(oc)}^2}{|Z_i + Z_{om}|^2} R_i = \frac{v_s^2}{|Z_i + Z_{om}|^2} \frac{R_i R_{om}}{R_s} . \quad (17)$$

E. Noise characteristics for amplifier with different matching networks

Based on the above expressions and on the model in Fig. 1, a noise analysis for arbitrary matching network between the signal source and the amplifier input can be performed. In order to estimate the effect of a matching network on noise, the amplifier without and with a match is analyzed bellow.

Case 1 *No matching network is used at the amplifier input*

If a matching network is not used, $Z_{om} = Z_s$, and [5]

$$v_{ni}^2 = 4kTR_s \Delta f + v_n^2 + 2v_n i_n \, Re(cZ_s^*) + i_n^2 |Z_s|^2 . \quad (18)$$

Under this condition, the *SNR*, and the noise factor can be expressed as follows

$$SNR = \frac{v_s^2}{4kTR_s \Delta f + v_n^2 + 2v_n i_n \, Re(cZ_s^*) + i_n^2 |Z_s|^2} , \quad (19)$$

$$F = 1 + \frac{v_n^2 + 2v_n i_n \, Re(cZ_s^*) + i_n^2 |Z_s|^2}{4kTR_s \Delta f} . \quad (20)$$

The power delivered to Z_i can be determined by

$$P_i = \frac{v_s^2}{|Z_i + Z_s|^2} R_i . \quad (21)$$

Case 2 *Matching network with output impedance chosen to be optimum source impedance*

When $Z_{om} = Z_{opt}$, where Z_{opt} is given by [5]

$$Z_{opt} = \left[\sqrt{1 - c_i^2} - jc_i \right] \frac{v_n}{i_n} , \quad (22)$$

the voltage v_{ni}^2 is

$$v_{ni}^2 = 4kTR_s \Delta f + \frac{R_s}{R_{opt}} \left[v_n^2 + 2v_n i_n \, Re(cZ_{opt}^*) + i_n^2 |Z_{opt}|^2 \right] , \quad (23)$$

In this case, the *SNR* is defined by

$$SNR = \frac{v_s^2}{4kTR_s \Delta f + \frac{R_s}{R_{opt}} \left[v_n^2 + 2v_n i_n \, Re(cZ_{opt}^*) + i_n^2 |Z_{opt}|^2 \right]} . \quad (24)$$

This matching network minimizes the noise factor so that

$$F_{min} = 1 + \frac{v_n i_n \left(1 + c_r \sqrt{1 - c_i^2} - c_i^2 \right)}{2kT \sqrt{1 - c_i^2} \Delta f} . \quad (25)$$

If the matching network is designed so that the amplifier sees its optimum source impedance, Eq. (17) reduces to

$$P_i = \frac{v_s^2}{|Z_i + Z_{opt}|^2} \frac{R_i R_{opt}}{R_s} . \quad (26)$$

Case 3 *Conjugate impedance matching network*

For a conjugate match, the condition $Z_{om} = Z_i^*$ must hold [4], and that's way $R_{om} = R_i$. Under this condition, the expression for v_{ni}^2 in Eq. (12) reduces to

$$v_{ni}^2 = 4kTR_s \Delta f + \frac{R_s}{R_i} \left[v_n^2 + 2v_n i_n \, Re(cZ_i) + i_n^2 |Z_i^*|^2 \right] . \quad (27)$$

The corresponding *SNR* and noise factor are, respectively

$$SNR = \frac{v_s^2}{4kTR_s \Delta f + \frac{R_s}{R_i} \left[v_n^2 + 2v_n i_n \, Re(cZ_i) + i_n^2 |Z_i^*|^2 \right]} , \quad (28)$$

$$F = 1 + \frac{v_n^2 + 2v_n i_n \, Re(cZ_i) + i_n^2 |Z_i^*|^2}{4kTR_i \Delta f} . \quad (29)$$

For a conjugate match, the maximum value of P_i occurs and it is given by

$$P_{imax} = \frac{v_s^2}{4R_s} . \quad (30)$$

III. Numerical Simulations

In order to illustrate a matching network effect on noise, the noise characteristics for the amplifier having a resistive input impedance, driven from a Thévenin source with a resistive output impedance, have been simulated. The simulation for the above analyzed cases has been implemented. The bellow presented characteristics illustrate the matching network effect on noise with $\Delta f = 1Hz$, $R_i = 25\Omega$, $i_n / \sqrt{\Delta f} = 31pA / \sqrt{Hz}$, $v_n / \sqrt{\Delta f} = 0{,}447nV / \sqrt{Hz}$, and $c = 0{,}12 - j0{,}44$.

Fig.2 shows the equivalent noise voltage change versus the source resistance. It is clear that if a matching network is used, the voltage is lower than in case1. For example, if $R_S = 50\Omega$, then the noise voltage in case1 is 1,271 times grater than that in case2, and 1,178 times grater than the voltage in case3. The results also show that the matching network, whose output impedance is equal to the optimum impedance of the source, provides the lowest noise voltage.

The effect of a match between the source and the amplifier on the noise figure is illustrated in Fig.3.

It follows from the simulated waveforms that as the source resistance increases, the noise figure increases in case1, only. In the other two cases the noise figures do not change. The comparison between the noise figures with $R_S = 50\,\Omega$, shows

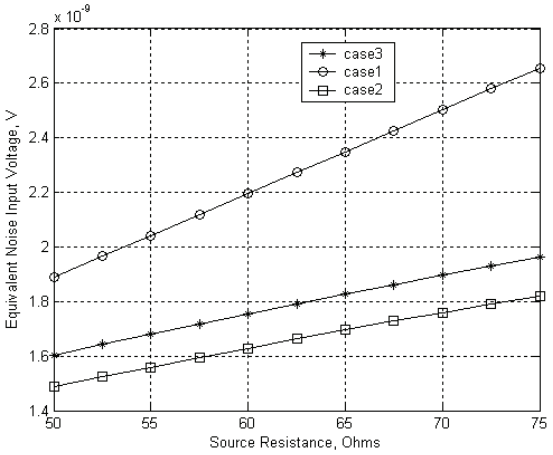


Fig.2. Equivalent noise input voltage versus source resistance.

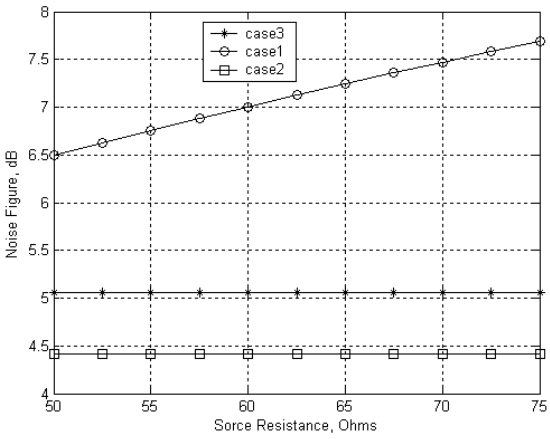


Fig.3. Noise figure dependence on source resistance.

that $NF_{case1} - NF_{min} = 2{,}08$ dB, $NF_{case3} - NF_{min} = 0{,}649$ dB where $NF_{min} = 4{,}413$ corresponds to case2.

Knowing the signal power delivered to the amplifier input, the power gain decrease in percents can be expressed as

$$P_{decrease}\% = \left(P_i / P_{i\,max}\right)100\,\% . \qquad (31)$$

Because the maximum power gain occurs with a conjugate impedance match, in case1 the power gain drops compared to $P_{i\,max} = P_{icase3}$ by the factor P_{icase1} / P_{icase3}, and in case2 it drops by the factor P_{icase2} / P_{icase3}, respectively. The power gain decrease is presented in Fig. 4. It can be concluded that for the amplifier without a matching network the percent decrease in power gain is a linear function of the source resistance, while for a matching network with optimum source impedance, the decrease is a constant. If the source resistance is $R_S = 50\,\Omega$, the decrease in power gain in case1 is 11,11%, or 0,512 dB. In case2 it is 12,53%, or 0,581 dB, respectively.

The results, obtained with different values of the correlation coefficient, show that the correlation coefficient effects on the input power if and only if a matching network with optimum source resistance is used.

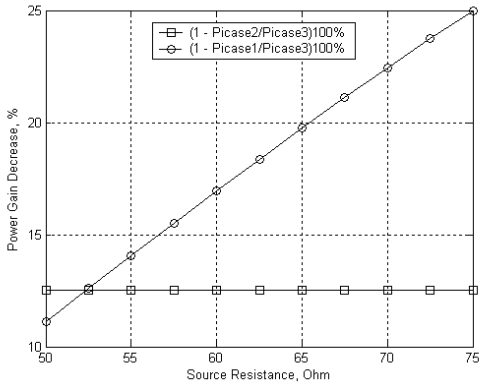


Fig.4. Power gain decrease as a function of source resistance.

IV. Conclusion

An effective approach for analysis of a matching network effect on noise is presented in this paper. The approach is based on the use of an amplifier noise model with a Thévenin input source and a lossless matching network. The analysis is applied to the amplifier without an input matching network, as well as, to a conjugate impedance matching network and to a matching network with optimum source impedance. It is seen from the simulation results that the addition of a conjugate matching network at the amplifier input maximizes the power gain, and the addition of a matching network with optimum source impedance minimizes the noise factor.

It follows from the analysis that because of the dependence of v_{ni}^2 and F of Z_i, it is difficult to predict from the Eqs. (27) and (29) how changes in Z_i affect the noise. This is because V_n, I_n, and c, in the model are, in general, related to Z_i. For example, V_n, I_n, c, and Z_i may all be functions of the bias current in the amplifier input stage. A change in the bias current to vary Z_i can cause a change in V_n and c. Thus the effects cannot be examined in detail unless the relations between the variables are known.

References

[1] S. Maas, "*Noise in Linear and Nonlinear Circuits*", Boston, Artech House, 2004.

[2] Jieh - Tsorng Wu, "*Noise Analysis and Modeling*", National Chiao - Tung University, 2001.

[3] A. Demir and A. Vincentelli, "*Analysis and Simulation of Noise in Nonlinear Electronic Circuits and Systems*", Kluwer Academic Publishers, Boston/Dordrecht/London, 1998.

[4] A. Mehrotra and A. Vincentelli, "*Noise Analysis of Radio Frequency Circuits*", Kluwer Academic Publishers, MA, USA, Hardbound, 2003.

[5] P. Petrova, "Amplifier Noise Model with Thévenin Input Source", ICEST 2005, Paper Proceedings, vol. 1, pp.105-108, Niŝ, Serbia and Montenegro, 2005.

SESSION RCMTA I

Radio Communications, Microwave Technique and Antennas I

Design and Analysis of Wideband Antenna with An Application to Ground Penetrating System

G. E. Atteia[1] , K. F. A. Hussein[2] and A. A. Shaalan[1]

Abstract- **In this paper, a wide band antenna is proposed for Ground-Penetrating Radar (GPR) system. This antenna consists of a dipole antenna housed in a rectangular conducting reflector whose inner walls are coated with an absorbing material. The capability of buried target detection by the proposed antenna is examined by investigating the coupling between the transmitting and receiving antennas in the presence and absence of buried targets. The operating bandwidth of the antenna is shown to be about 50%. The GPR system with the proposed antenna is shown to be capable of detecting the existence of buried targets.**

Keywords- **Ground-Penetrating Radar, GPR, FDTD, Wideband antenna**

I. INTRODUCTION

Ground-Penetrating Radar (GPR) systems are used for the subsurface investigation of earth. They are used in the detection of objects buried beneath the earth surface such as pipes, cables, land mines, and hidden tunnels. GPR system consists of transmitting and receiving antennas. The efficiency of a certain GPR system depends on its capability of the true detection of buried objects. This capability depends mainly on the characteristics of the signal used for the detection. In order to enhance these characteristics, it is required to develop efficient GPR antennas to satisfy a number of demands. A GPR system should have low and short coupling between transmitting and receiving antennas to avoid false detection. Since it operates very close to ground, its characteristics should not be affected strongly with ground properties. In addition it should have light weight and should be economic. Due to the great importance of GPR antenna performance, various types of GPR antennas have received considerable attention in the literature. For example, modeling of GPR antennas with shields and simulated absorbers is discussed in [1]. In [2], a separated-aperture sensor which consists of two dipoles housed in corner reflectors that are separated by a metallic septum is discussed and simulated using the FDTD method. In [3], an efficient bow-tie antenna for GPR system is developed to exhibit good efficiency in ultra-wide band using a combination of a tapered capacitive and resistive loading [4].

[1]G. E Attia and A. A. Shaalan are with the Faculty of Eng., Electronics and Communications Eng. Dept,, Zagazig Univ., Zagazig, Egypt

[2] K. F. A. Hussein is Microwave Eng. Dept., Electronics Research Inst.(ERI), Cairo, Egypt.

This paper proposes a design for new GPR antenna composed of a linear dipole housed in a conducting rectangular reflector whose inner walls are coated with absorbing layers to optimize the antenna performance. The GPR system uses two units of such an antenna; one for transmission and the other for reception.

The FDTD method is used for electromagnetic simulation of the complete GPR system to study the characteristics of the proposed antenna and to evaluate the coupling between the transmitting and receiving antennas during the operation of target detection. FDTD method has been widely used to model wave propagation, scattering, radiation, and simulations of GPR problems [5] since it was first introduced by Yee in 1966 [6]. A major reason for using the FDTD method is that it proved its efficiency in simulating complex geometries and its capability of modeling wide range of realistic mediums and soils [7-8].

The remaining of the paper falls into three parts, the first of which is concerned with describing the construction of the proposed antenna as well as the complete GPR system, the second part describes the electromagnetic simulation of the GPR system using FDTD and the third part presents the results showing the antenna characteristics and the system performance.

II. DESCRIPTIONS OF THE DESIGN OF THE PROPOSED GPR SYSTEM

As shown in Fig.1, two units of the proposed antenna are used to construct the GPR system. This antenna is composed of a linear dipole housed inside a rectangular conducting reflector. The purpose of the reflector is to eliminate direct coupling between the transmitting and receiving antennas. To improve the antenna characteristics and to diminish the internal resonance of the rectangular cavity, the inner walls of the reflector are coated with a lossy absorbing material.

During the operation of buried target detection, and when the GPR system is placed over the ground with the apertures of the reflectors close to the ground surface, the antenna characteristics are significantly affected. The impact of the ground on the antenna characteristics depends on the electric properties of the ground soil. Thus the design of the antenna should take into account the electric properties of a practical soil.

The coating on the inner walls of the conducting reflectors is composed of many layers of lossy materials, which are chosen to achieve a conductivity profile that optimizes the antenna performances for the common types of ground soils.

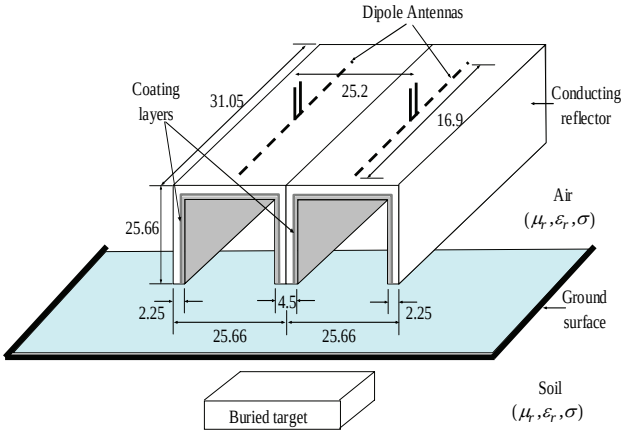


Fig. 1. Geometry of GPR antenna above ground surface (all dimensions in cm) .

III. ELECTROMAGNETIC MODELING OF THE GPR SYSTEM USING FDTD

The FDTD method as described in [9] is applied here to provide electromagnetic simulation for the complete three-dimensional model of the GPR system including the system antenna, soil, and buried target. The entire space is divided into a number of cubic cells with the appropriate resolution. The boundaries of the FDTD volume are terminated using the uniaxial perfectly matched layer (UPML) absorbing boundary conditions. The PML is composed of a number of cells with a specific profile of the electric and magnetic properties and is backed by perfectly conducting walls [9].

A thin-wire model is used to present the transmitting dipole antenna, which is excited with a sinusoidal voltage source applied across an infinitesimal gap at its center [10]. The receiving dipole antenna is also modeled as a thin wire with an infinitesimal gap at its center. The received voltage signal is measured across this gap.

The input impedance of the dipole antenna is a frequency-domain quantity that is measured by, first calculating the current flowing in the antenna arms at the feed point and then dividing the applied voltage on the calculated current, both in the frequency domain. The current is calculated using Ampere's law as described in [9].

The transmission coefficient S_{21} is another frequency-domain quantity whose magnitude expresses the amount of coupling between the transmitting and receiving antennas. This coefficient is defined as the ratio of the voltage measured at the receiving dipole antenna port to the voltage applied at the transmitting dipole antenna port.

IV. RESULTS AND DISCUSSION

In this section, the results for the input impedance, VSWR of the GPR antenna as well as the coefficient S_{21} are presented in the entire frequency band of the GPR system operation.

The type of soil selected for the present study is the most practical one and has properties : $\varepsilon_r = 2.9$, $\mu_r = 1.0$ and $\sigma = 0.02\ S/m$ [2]. The ground is assumed to be homogenous.

The simulation space is a volume of size ($40.05\times59.85\times58.95$ cm), which is divided into cubic cells, each of dimensions 0.45×0.45×0.45 cm. The field components are updated every $\Delta t = 8.66$ ps. This time step satisfies the condition of numerical stability [9]. The dipole is modeled as a perfect electric conductor. The dipole length is 16.9 cm. Its radius is 0.002 of its length, and the thin-wire approximation described in [10] is used. The transmitting dipole antenna is excited with a sinusoidal voltage source of the form : $V(t) = A_0 \sin(2\pi ft)$, where $A_0 = 100$ V.

The coating on the inner walls of the reflectors has a thickness of 2.25 cm and is composed of 6 layers with the conductivity profile: $\sigma = 0.0001\times10^{2n}$, where *n*=0,1,2,3,4,5 is the layer number and *n*=0 is number of the layer just touching the air region of the reflector cavity. The walls of the reflectors are modeled as a perfect conductor.

4.a. Frequency Response of Dipole Antenna inside Rectangular Reflector.

To investigate the proposed antenna performance as regards the optimal band width of operation without being limited by the properties of the ground, the GPR system here is placed in free space. However, the effect of the ground is taken into consideration later on.

Figures (2) and (3) show the frequency response of the antenna impedance and VSWR respectively, while the inner walls of the reflectors are left without coating. The source impedance is taken as 180Ω. It is clear that the antenna is suitable for operation over the band 700-1000 MHz, i.e. about 35% bandwidth.

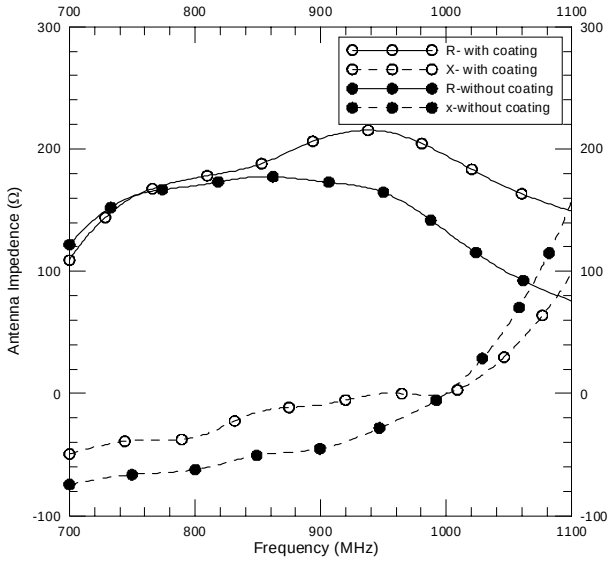


Fig. 2. Input impedance for the antenna with coating and without coating.

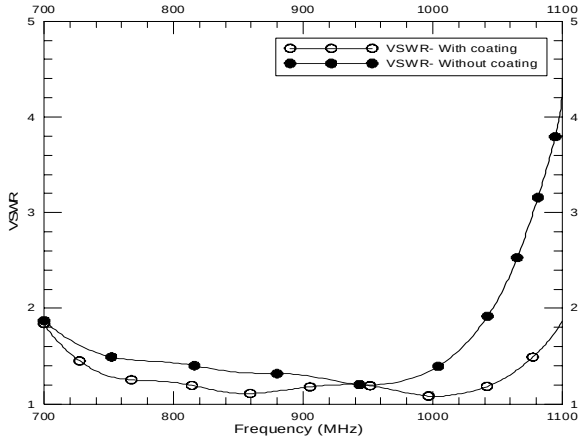


Fig. 3. VSWR for the antenna with coating and without coating.

4.b. Improvement of Antenna Characteristics by Coating the Internal Walls of the Rectangular Reflectors.

To improve the performance of the antenna as regards the proper band width of operation the inner walls of the reflectors are covered by the 6-layer coating with the conductivity profile described above. As shown in Figures (4) and (5), and by comparison with antenna characteristics in the absence of the coating absorber, the addition of this coating results in a significant reduction of reactive part of the antenna impedance and, hence, the VSWR, with respect to 180Ω source impedance, is maintained below 2.0 in the entire band of operation (700-1100 MHz), i.e. the band-width is improved to be about 50%.

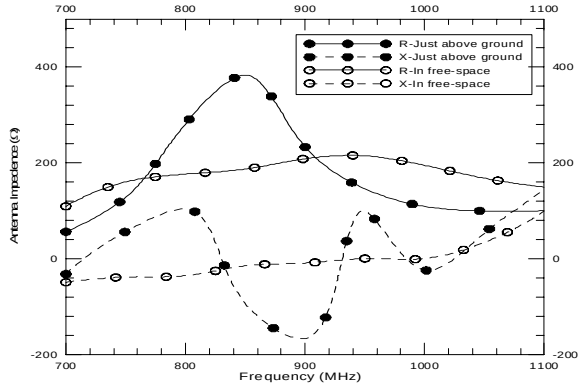


Fig. 4. Input impedance of the GPR antenna against the frequency

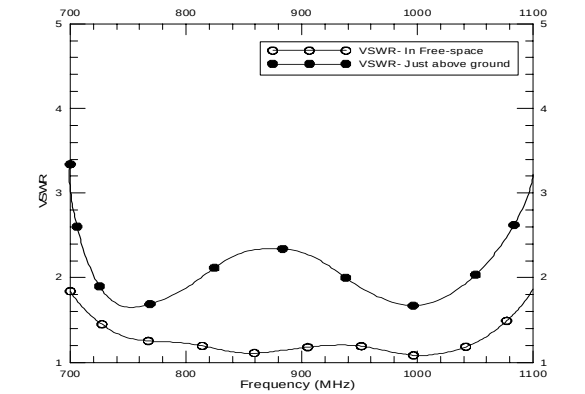


Fig..5. VSWR of the GPR antenna against the frequency.

4.c. Effect of the Ground on the GPR Antenna Characteristics.

The effect of the ground on the characteristics of the GPR antenna should be studied as the antenna is placed very close to the ground during the buried target detection operation. Figures (6) and (7) show the frequency response of the antenna impedance and the corresponding VSWR with respect to 180Ω source impedance when the antenna is placed at different heights over the ground surface. Compared with the characteristics of the same antenna in free-space, the ground soil has a bad effect on the antenna impedance and hence, on the VSWR. The higher the antenna position is above the ground, the less the effect of the ground on the antenna performance. As clear in Fig.(7), the VSWR of the GPR antenna can still be accepted even when the reflector aperture is just touching the ground surface.

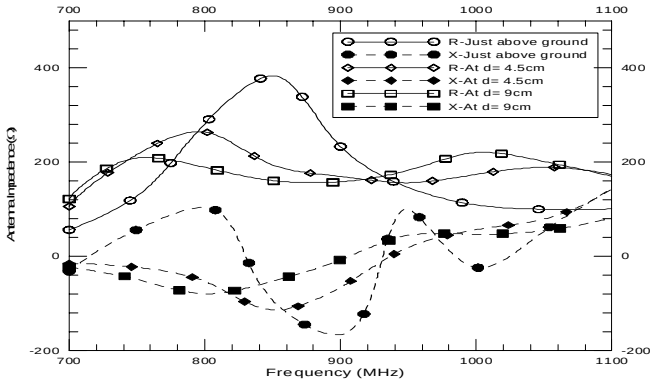


Fig. 6. Input impedance of the GPR antenna against the frequency when it is placed over the ground.

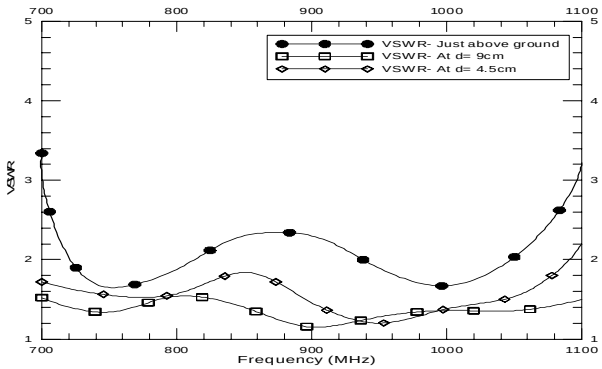


Fig. 7. VSWR of the GPR antenna against the frequency when it is placed over the ground.

4.d. Capability of Buried Target Detection using the Coated-Reflector Antenna.

The capability of target detection by such antenna as GPR system can be measured by change of the electromagnetic coupling ($|S_{21}|$) between the transmitting and receiving antennas due to the presence of a buried target.

In the present investigation, the used target is a dielectric block ($31.05\times49.5\times9$) cm with the properties: $\varepsilon_r = 8$, $\mu_r = 1.0$, $\sigma = 0.0$ S/m. The target is buried in ground at a depth of 6.75 cm.

Figures (8), (9) and (10) show the coupling coefficient $|S_{21}|$ between the transmitting and the receiving antennas in the presence and absence of a buried dielectric target when the GPR system is placed at different heights over the ground. It is clear that the maximum detection is obtained when the GPR system is placed with the reflector aperture touching the ground surface.

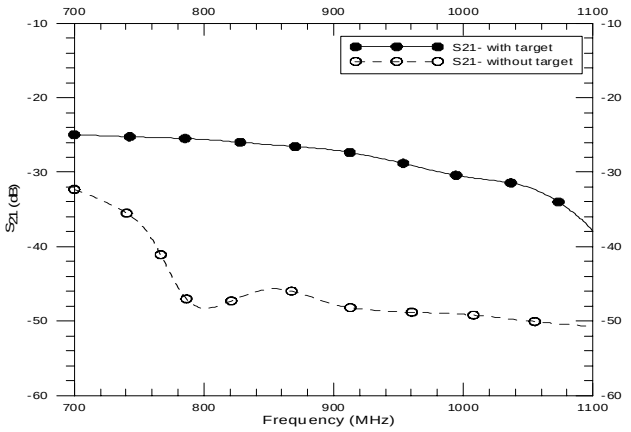


Fig. 8. The coupling coefficient $0|S_{21}|$ for the GPR system when placed touching the ground surface in the absence and presence of a buried target.

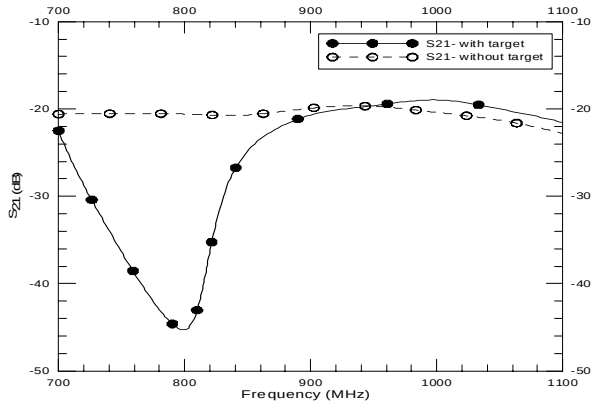


Fig. 9. The coupling coefficient $|S_{21}|$ for the GPR system when placed at a height of 4.5 cm over the ground in the absence and presence of a buried target.

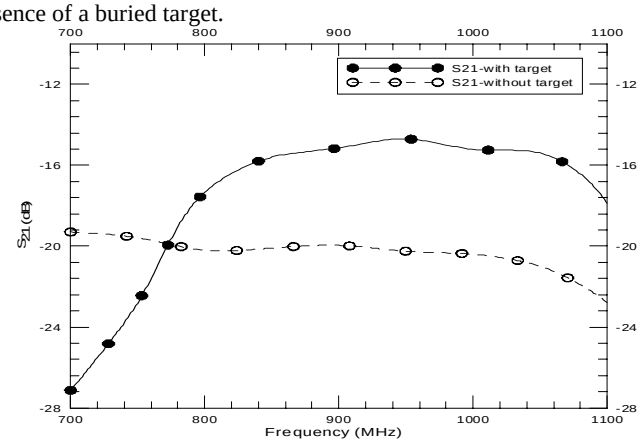


Fig. 10. The coupling coefficient $|S_{21}|$ for the GPR system when placed at a height of 9 cm over the ground in the absence and presence of a buried target

V. CONCLUSION

A wide band antenna designed for Ground-Penetrating Radar (GPR) system is introduced and examined. This antenna consists of dipole antenna housed in a rectangular conducting reflector whose inner walls are coated by an absorbing material. The coating is composed of a number of layers with a conductivity profile designed to achieve the minimum voltage standing wave ratio (VSWR) of the dipole antenna over the frequency band of operation. The antenna impedance and VSWR are calculated using the Finite-Difference Time-Domain (FDTD) method. The antenna impedance and VSWR are presented over a wide-band of frequency. The capability of buried target detection by the proposed antenna is confirmed by investigating the coupling between the transmitting and receiving antennas in the presence and absence of buried targets. The operating bandwidth of the antenna is shown to be about 50%. The effect of the ground soil on the antenna characteristics is studied when the GPR system is placed at different heights above such a soil. The GPR system with the proposed antenna is shown to be capable of detecting the existence of buried targets even when it is placed at different heights above the ground.

REFERENCES

[1] U. Oguz, L. Gurel, "Modeling of Ground-Penetrating Radar antennas with shields and simulated absorbers", IEEE Trans. Antennas Propagat., Vol. 49, No.11, November, 2001.

[2] J. M. Bourgeois, G. S, Smith, "A Complete Electromagnetic Simulation of the Separated-Aperture Sensor for Detecting Buried Land Mines ", IEEE Trans. Antennas Propagat., Vol. 46, No.10, October 1998.

[3] A. A. Lestari, A. G. Yarovoy, L. P. Ligthart, "An efficient ultra-wideband bow-tie antenna", http://www.tudelft.nl/live/binaries/33dc3ad6-3e8a-4bdd-8456-lf3dl0b39c8c/ doc/Leatari_london.PDF.

[4] Y. Nishioka, O. Maeshima, T. Uno, S. Adachi, "FDTD Analysis of Resistor-Loaded Bow-Tie Antenna Covered with Ferrite-Coated Conducting Cavity for Subsurface Radar ", IEEE Trans. Antennas Propagat., Vol. 47, No.6, June, 1999.

[5] J. M. Bourgeois, G. S, Smith, "A Fully Three-Dimensional Simulation of Ground- Penetrating Radar: FDTD Theory Compared with Experiment", IEEE Trans. on Geosciences and Remote sensing, Vol.34, No.4, January, 1996.

[6] K. S. Yee, "Numerical Solution of Initial Boundary Value Problem using Maxwell's Equations in Isotropic Media", IEEE Trans. Antennas Propagat., Vol. AP-14, May, 1966.

[7] F. L. Teixeira, W. Chow chew, "Finite-Difference-Time-Domain Method Simulation of Ground-Penetrating Radar on Dispersive In Homogeneous and Conductive Soils ", IEEE Trans. on Geosciences and Remote sensing, Vol.36, No.6, November, 1998.

[8] A. A. Lestari, A. G. Yarovoy, L. P. Lightar, "Ground Influence on the Input Impedance of Transient Dipole and Bow-Tie Antennas "", IEEE Trans. Antennas Propagat., Vol. 52, No.8, August, 2004.

[9] A. Taflove, "Computational Electrodynamics: The Finite-Difference-Time-Domain Method", Boston, MA: Artech House, 1995.

[10] S. Watanabe, M. Taki, "An Improved FDTD Model for the Feeding Gap of Thin-Wire Antenna", IEEE Microwave and guided wave letters, Vol.8, No.4, April, 1998.

Influence of the Feeding Position Over the Polarization Parameters of Patch Antennas with Different Geometric Shape

Milko V. Stefanov[1]

***Abstract* – Circularly polarized microstrip patch antennas with three different shapes are designed. The influence of the feeding position by a coaxial probe over the polarization parameters – Axial Ratio and Crosspolarization is investigated. Depending on these parameters working area over the patch is defined.**

Keywords – circular polarization, crosspolarization, Axial Ratio, Cross-Polarization Discrimination

I. INTRODUCTION

Microstrip patch antennas are increasing in popularity for use in wireless applications due to their low-profile structure. Therefore they are extremely compatible for embedded antennas in handheld wireless devices such as cellular phones, pagers, GPS receivers etc. Some of their principal advantages are: light weight and low volume, low profile planar configuration, low fabrication cost, hence can be manufactured in large quantities, supports both, linear as well as circular polarization, can be easily integrated with microwave integrated circuits, capable of dual and triple frequency operations, mechanically robust when mounted on rigid surfaces. Microstrip patch antennas suffer from a number of disadvantages as compared to conventional antennas: narrow bandwidth, low efficiency, low Gain, extraneous radiation from feeds and junctions, low power handling capacity, surface wave excitation [1].

A circularly polarized wave has two orthogonal field components that are in phase quadrature. Microstrip antennas can be designed to radiate circular polarization with one feed. It is possible to excite the two modes with one feed by introducing a small perturbation in the patch shape [2]. Many types of perturbations are used. One common configuration is a square patch with a small difference in length of the two opposite sides (asymmetry). This asymmetry can be also realized by adding two rectangular segments in the center of two opposite sides of the square patch. Therefore the orthogonal surface current path lengths are changed, of each additional segment. Therefore the resonant frequencies of the orthogonal modes are different. This difference depends on the segment size, which is designed to achieve phase shift of 90 degrees between the radiated fields.

[1] Milko V. Stefanov is with the Faculty of Communications and Communication Technologies, Technical University of Sofia, 8, Kliment Ohridski Str, 1756 Sofia, Bulgaria,
E-mail: mvs@mail.bg

The next structure is designed by cutting of four rectangular segments near each corner of a square patch. Two opposite rectangles have little different length with the length of the other two rectangles [3]. In such a way the requirements of circularly polarization radiation are fulfilled. In the field of the circular polarized microstrip antennas two bandwidths are defined: impedance bandwidth and axial ratio bandwidth. The Axial Ratio Bandwidth is defined between the two frequencies where Axial Ratio equals 3 dB. In this bandwidth Minimal Axial Ratio is defined. The other important polarization parameter is the Crosspolarization Level. It may be measured with Cross-Polarization Discrimination (XPD) - the difference between the level of Radiation Pattern for working polarization and the opposite one.

All the polarization parameters depend on the shape of the patch, feeding system and the feeding position

II. DESIGN OF THE ANTENNAS

Dimensions design of patch antennas includes calculation of the resonant length, for central working frequency $f = 2.4GHz$. There are not simple analytical expressions for synthesis and analysis of the circularly polarized microstrip antennas because of the complex character of the field in the antenna. That's why numerical methods for analysis as Moments Method, Finite Difference Method etc. have to be used. In the current analysis models of microstrip antennas with different shape in the medium of a software package, working with Moments Method has been designed.

The antennas consist of dielectric substrate with $\varepsilon_r = 2.2$ and thickness h=0.7 mm. They are fed by a coaxial probe.

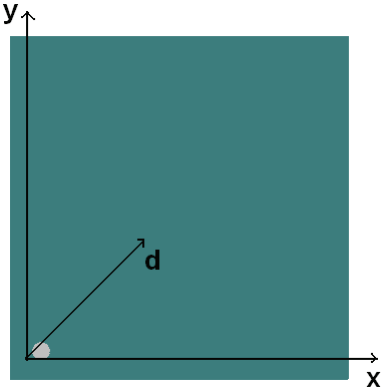


Fig.1. Nearly square microstrip patch antenna (geometry A).

III. Results from the Analysis

In the current article the relations between Axial Ratio (AR) and crosspolarization, measured with the difference between maximal levels of Radiation Pattern for working polarization and the opposite one (XPD), versus feeding position of the coaxial probe are investigated. Three structures of microstrip resonator on the dielectric substrate are designed – nearly square patch [1] - geometry A (Fig. 1), square patch with two rectangular segments in the center of two opposite sides - geometry B (Fig. 2), square patch with four cut rectangular segments near each corner [3], where the two opposite rectangles have little different length with the length of the other two rectangles geometry C (Fig. 3). The coaxial probe is moved along three directions – bisector of a corner (d-direction), horizontally (x-direction), vertically (y-direction). The initial point is x=y=2 mm.

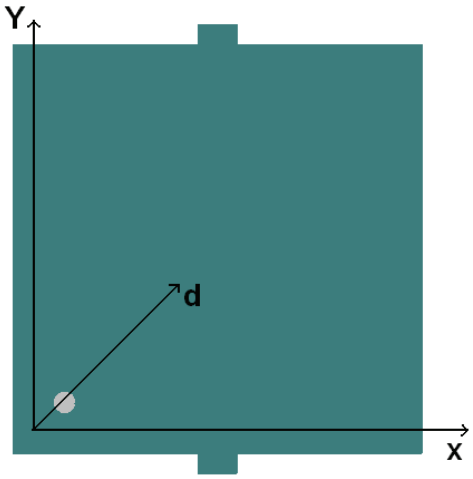


Fig.2. Square microstrip patch antenna with additional rectangular segments (geometry B).

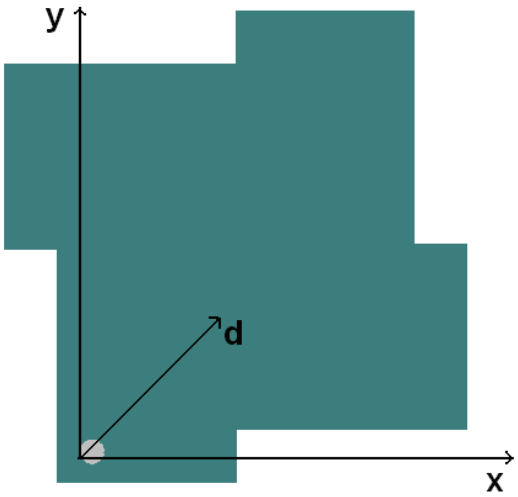


Fig.3. Square microstrip patch antenna with four cut rectangular segments near each corner (geometry C).

First important parameter is the minimal Axial Ratio (AR).

Fig. 4 shoes min AR versus probe position along x, y, d for nearly square patch witch is common used in the practice. min AR is almost constant along bisector, but strongly increasing along horizontal and vertical directions when x, y > 6. That separates an area around the diagonal of the patch.

Fig. 5 shoes the same relation for geometry B. min AR is constant along d, also along y, but for y < 7.5. Along y min AR strongly increases. Here working area between d and y directions is defined.

For geometry C (Fig. 6) the lowest level of min AR is between d and x-directions.

Geometry B has the lowest value of AR and it weakly depends on the feeding point.

The other important parameter is the crosspolarization measured with the Polarization Pattern level difference between working polarization (Right Hand Circular Polarization) and crosspolarization (Left Hand Circular Polarization). This difference is denoted by XPD (Cross-Polarization Discrimination). XPD should be grater than 20 dB.

Fig. 7 shoes XPD versus the probe position for geometry A. XPD is almost constant along x, y and d for x, y < 6 mm. After that XPD increases along d, but decreases under 20 dB along x and y, which degrades the antenna operation. This result along with the area with minimal AR determines the working area that is along d for x and y grater than 6 mm.

For geometry B the greatest values of XPD are between y and d directions when y < 7.5 mm (Fig. 8). This area coincides with that of minimal AR.

For geometry C maximal XPD is again between d and x.

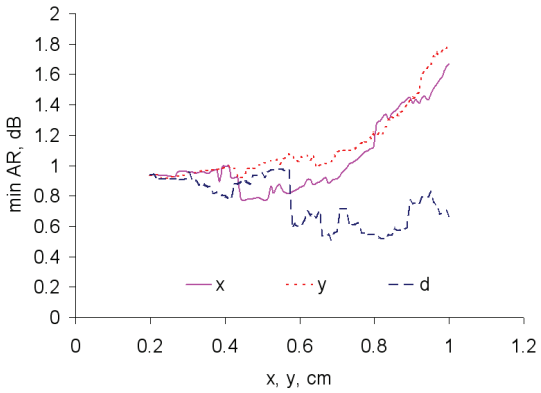


Fig.4. Min. AR versus x, y, d for geometry A

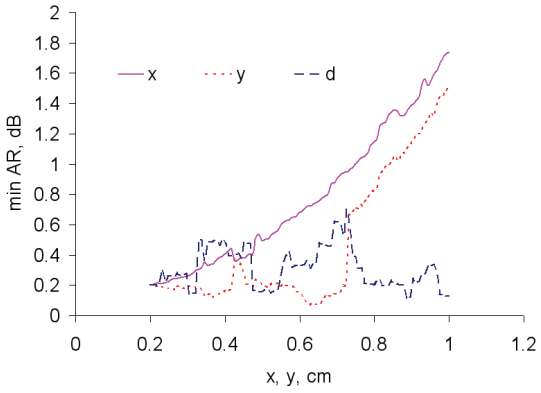


Fig.5. Min. AR versus x, y, d for geometry B

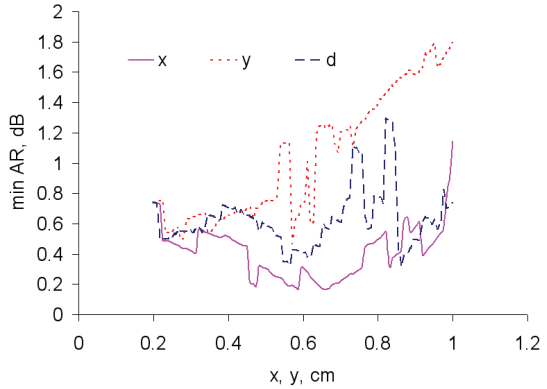


Fig.6. Min. AR versus x, y, d for geometry C

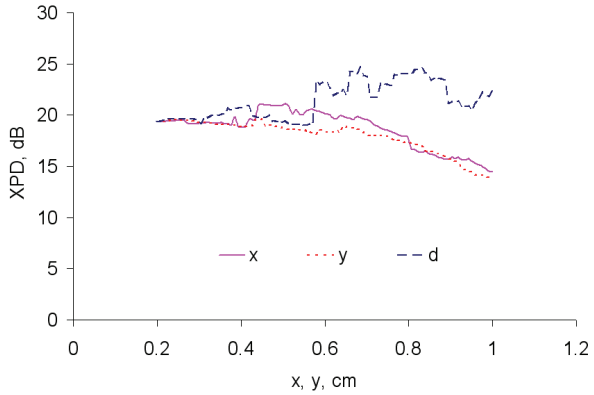


Fig.7. XPD versus x, y, d for geometry A

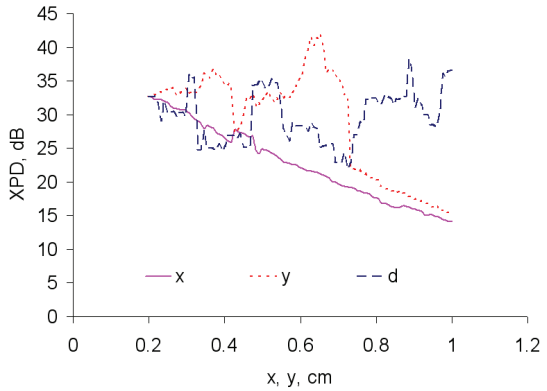


Fig.8. XPD versus x, y, d for geometry B

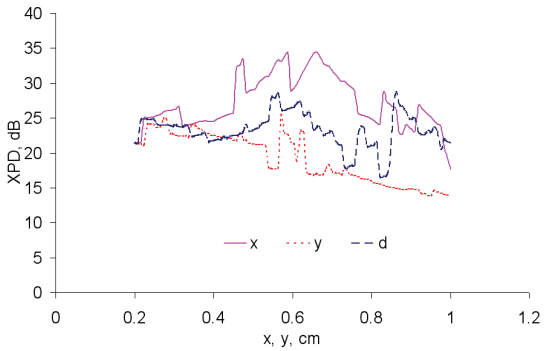


Fig.9. XPD versus x, y, d for geometry C

IV. Conclusion

From the results above some concussions may be drawn. When designing of circularly polarized microstrip antennas the choice of proper feeding position is important from the point of view of impedance matching, crosspolarization and minimal Axial Ratio. A wider working area with minimal crosspolarization leads to better impedance matching because an optimal feeding point with minimal VSWR, AR and maximal XPD may be chosen. Working area is strongly individual for one geometry and may be determined mainly by experiments and simulations. The geometry also determines the Axial Ratio Bandwidth. That's why patches with various shapes are designed.

References

[1]. C. A. Balanis, *Antenna Theory: Analysis and Design*, second edition, John Wiley & Sons, New York, 1997.

[2]. R. A. Sainati, *CAD of Microstrip Antennas for Wireless Applications*, Artech House Inc., 1996.

[3]. H. M. Chen, K. L. Wong, "On circular polarization design of annular-ring microstrip antennas," *IEEE Trans. Antennas Propagat.* vol. 47, pp. 1289–1292, Aug. 1999

This page intentionally left blank.

Novel Methods for Analysis of Smart Antennas for Wireless Communications

Sava V. Savov and Viara Y. Vassileva[1]

Abstract – **In this paper, the principles of uniform circular array (UCA) adaptive beamforming (ABF) and direction-of-arrival (DOA) estimation based on narrowband radio-frequency signals are introduced. An UCA is composed of a number of uniformly distributed identical omnidirectional antenna elements. The problem of DOA estimation of incoming signals (plane waves) incident on a UCA is considered using UCA-ESPRIT. It is a novel closed-form algorithm that provides removal of computationally intensive spectral search procedures and iterative solutions. Improved least mean squares (ILMS) algorithm is applied to antenna beamforming in UCA. It is a method that exploits the array correlation matrix structure and uses previous samples when estimating the gradient at the nth iteration. Limited numerical examples and simulation results are presented to illustrate the novelty of the proposed methods.**

Keywords – **smart antennas, direction of arrival, adaptive beamforming, wireless communications.**

I. INTRODUCTION

Smart antennas (adaptive antennas) have become popular among researchers during recent years. Wireless operators are currently looking for new technologies that would be implemented into the existing wireless communications infrastructures to provide broader bandwidth per user channel, better quality, and new value-added services [1]. These research efforts will enable wireless carriers to maximize the spectral efficiency of their networks so as to meet the explosive growth of the wireless communication industry, and so to take advantage of the huge market opportunity.

Deployed at the base station of the existing infrastructure, smart antennas are capable of increasing the capacity (very important in urban and densely populated areas) in the frequency-resource-limited radio-communication system by an efficient frequency-reuse scheme.

Until now, the investigation of smart antennas suitable for wireless communication systems has involved primary uniform linear arrays (ULA) and uniform rectangular arrays (URA). It is obvious that one can take advantage from the symmetry of *uniform circular array* (UCA).

[1] The authors are with Department of Electrical Engineering, Technical University of Varna, Varna 9010, Bulgaria, E-mail: svsavov@ms.ieee.bg

II. SMART ANTENNA WITH UNIFORM CIRCULAR ARRAY STRUCTURE

The UCA with radius *a* consisting N equally distributed identical omnidirectional antenna elements, as illustrated in Fig. 1 is located on *x-y* plane.

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

As demonstrated in Fig. 1, the array factor *(AF)* of UCA is given by [2]

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

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

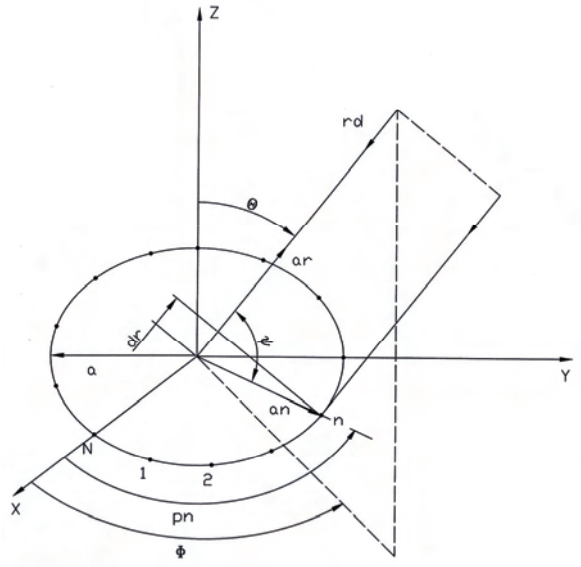


Fig. 1. Geometry of *N*-element UCA, along with an incoming plane wave.

III. DIRECTION OF ARRIVAL ESTIMATION

After that the UCA receives all incoming signals from directions of arrival, the DOA algorithm determines the directions of these signals based on the time delays. Let us assume that a narrowband plane wave impinges at an angle

(θ,ϕ) on the UCA. It produces time delays relative to the other array elements. These time delays depend on array geometry, number of elements, and interelement spacing. For the UCA of Fig. 1, the time delay of the narrowband signal at the *n*th element with respect to the origin, is written as [3]

$$\tau_n = (a/c)\sin\theta\cos(\phi-\phi_n),\quad n=1,2,...,N \tag{2}$$

where c is speed of light in free space.

Several subspace-based methods have been considered in the literature for the significant improvement in smart antenna resolution. For the azimuth and elevation estimation, two algorithms that fall into this category are MUSIC (Multiple Signal Classification) and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Technique). The latter one is used in the paper.

A. Classical ESPRIT – advantages in comparison with MUSIC

Classical ESPRIT is a robust method of DOA estimation. It uses two identical arrays in sense that elements need to form matched pairs with an identical displacement vector (the second element of each pair ought to be displaced by the same distance and the same direction relative to the first element). Furthermore, this does not mean that one has to have two separate antenna arrays but this method exploits this subarray structure for DOA estimation [4].

ESPRIT has become the method of choice because it has ability to offer a number of *advantages* over MUSIC, such as: a) computationally less intensive and more efficient; b) does not require calibration of the antenna array; c) does not involve search through all possible steering vectors to estimate the DOA.

B. The UCA-ESPRIT algorithm for DOA estimation

The UCA-ESPRIT algorithm is unique different from the classical ESPRIT, first of them provides closed-form automatically paired two dimensional estimation as long as the elevation and azimuth of each narrowband signal arrives at the UCA. Applying this novel method under the conditions of a UCA structure, the eigenvalues of each correlation array matrix have the form [3]

$$\lambda_i = \sin\theta_i e^{j\phi_i} \tag{3}$$

where (θ_i,ϕ_i) are respectively elevation and azimuth angles of incoming plane wave of *i*th signal source $(i=1,2,...,M)$, where M is the number of the narrowband sources. In terms of UCA, the three basic steps of real valued estimation are now briefly described [5]:

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

This novel algorithm gives several *advantages* in comparison with classical ESPRIT, such as: a) reduced computational complexity; b) lower SNR (signal-to-noise ratio) resolution thresholds, and c) very accurate finds simultaneously both the elevation and azimuth angles of arrival for impinging signals at the antenna array [6].

IV. METHODS FOR ADAPTIVE BEAMFORMIG ESTIMATION

In literature are represented two classes of adaptive beamforming (ABF) algorithms: 1) DOA-based adaptive beamformig algorithms that utilizes information for angles of arrival of incoming signals to ideally steer the maximum of the antenna radiation pattern toward the desired signal and place nulls toward the unwanted signals or interferences, as depicted in Fig. 2; 2) reference-based ABF algorithms does not need DOA information but instead uses the reference signal to adjust weights of correlation array matrix to match the created time delays. In this section, we consider one of the most popular *reference-based* ABF algorithms applicable to UCA: LMS (least mean squares) algorithm. After that, we considered one version of the LMS algorithm namely ILMS (improved least mean squares).

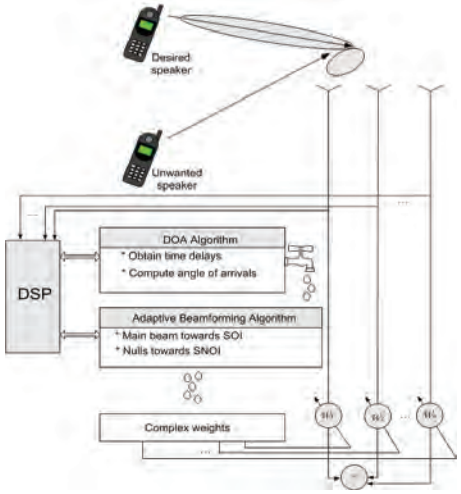


Fig. 2. A block scheme of a smart antenna.

A. LMS algorithm

The LMS algorithm is one of the simplest methods applicable to estimate optimal weights of an antenna array. It is applicable mainly when weights are updated utilizing reference signal. This algorithm uses an estimator of the gradient instead of the real value of the gradient because the real value estimation requires knowledge of incoming signals (DOA information).

The expression of optimal weights is given by [7]

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\mathbf{g}(\mathbf{w}(n)) \tag{4}$$

where $\mathbf{w}(n+1)$ denotes a new computed weights vector at the (n+1)th iteration, μ is the gradient step size, and the array output is given by

$$y(\mathbf{w}(n)) = \mathbf{w}^H(n)\mathbf{x}(n+1) \tag{5}$$

where $\mathbf{x}(n+1)$ is array signal vector computed at the (n+1)th iteration.

In its standard form it uses an estimate of the gradient by replacing array correlation matrix **R** and correlation between array signals and reference signal **z** by their noisy estimates at the (n+1)th iteration [7]

$$\mathbf{g}(\mathbf{w}(n)) = 2\mathbf{x}(n+1)\mathbf{x}^H(n+1)\mathbf{w}(n) - 2\mathbf{x}(n+1)r^*(n+1) \quad (6)$$

where **g** is the gradient vector.

The error between array output and the reference signal is given by [7]

$$\varepsilon(\mathbf{w}(n)) = r(n+1) - \mathbf{w}^H(n)\mathbf{x}(n+1) \quad (7)$$

and

$$\mathbf{g}(\mathbf{w}(n)) = -2\mathbf{x}(n+1)\varepsilon^*(\mathbf{w}(n)) \quad (8)$$

The estimated gradient is a product of the error between the reference signal and the output of the array and the signals after the *n*th iteration.

Apparently, this algorithm provides several advantages: the low complexity, the gradient estimate is unbiased, but the main disadvantage of the LMS is that it tends to *convergent slowly.*

B. ILMS algorithm

In terms of the UCA, the ILMS algorithm is described because it has some advantages in comparison with standard LMS.

It utilizes previous samples and exploits the structure of the array correlation matrix. The ILMS performs better because it has *low sensitivity* to the look direction signal.

V. NUMERICAL EXAMPLES AND SIMULATION RESULTS

Simulation results, utilizing the UCA-ESPRIT algorithm gave precise results when estimate simultaneously both the elevation and azimuth angles of arrival for incoming narrowband signals toward the UCA. In Table 1 are presented results from simulations. The UCA with *N=8* elements, and radius $a = 0.8\lambda$ is examined about two scenarios: a) when the number of incoming signals *M=2* in the presence of the interfering signal (SNOI), and Additive White Gaussian Noise (AWGN) with the mean, and variance 0.1; b) in these terms when *M=4.* The results demonstrate its great performance, accurate estimation ability, and robustness.

To illustrate the ABF algorithm applicability for UCA, we considered the two cases where ILMS algorithm is used: a) the UCA with *N=8* elements, and radius $a = 0.8\lambda$ when angles of arrival of the SOI are $\theta = 45^\circ, \phi = 90^\circ$; b) the UCA with *N=12* elements, and radius $a = 0.8\lambda$ when angles of arrival of the SOI are $\theta = 30^\circ, \phi = 90^\circ$. Fig. 3 and Fig. 4 illustrate the resulting radiation pattern with respect to $\phi_0 = 0^\circ$.

TABLE I
THE DOA ESTIMATIONS OBTAINED UTILIZING UCA-ESPRIT

	Case 1	Case 2
Number of elements	N=8	N=8
Radius of the UCA	a=0.8λ	a=0.8λ
Number of incoming signals	M=2	M=2
Number of data samples	2000	2000
Actual		
SOI 1	$\theta_1=30^0$, $\varphi_1=60^0$	$\theta_1=10^0$, $\varphi_1=150^0$
SOI 2	$\theta_2=40^0$, $\varphi_2=150^0$	$\theta_2=15^0$, $\varphi_2=180^0$
SOI 3		$\theta_3=20^0$, $\varphi_3=210^0$
SOI 4		$\theta_4=70^0$, $\varphi_4=350^0$
SNOI	$\theta=45^0$, $\varphi=120^0$	$\theta=25^0$, $\varphi=200^0$
DOA Estimations		
SOI 1	$\theta_1=29.984^0$, $\varphi_1=59.975^0$	$\theta_1=9.995^0$, $\varphi_1=150.006^0$
SOI 2	$\theta_2=45.032^0$, $\varphi_2=119.995^0$	$\theta_2=15.002^0$, $\varphi_2=179.994^0$
SOI 3		$\theta_3=19.993^0$, $\varphi_3=210.004^0$
SOI 4		$\theta_4=70.035^0$, $\varphi_4=349.995^0$
SNOI	$\theta=44.988^0$, $\varphi=120.004^0$	$\theta=25.024^0$, $\varphi=349.987^0$

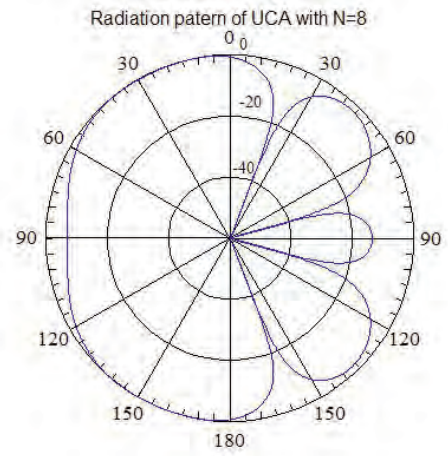


Fig. 3. The radiation pattern of the UCA with N=8 elements.

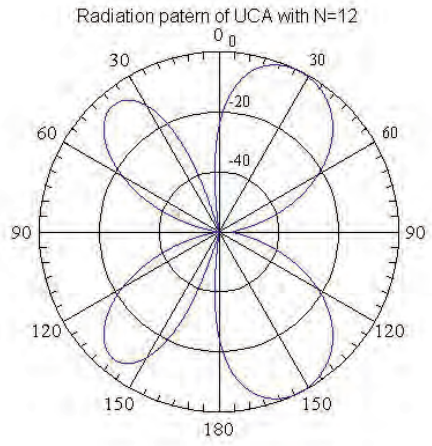


Fig. 4. The radiation pattern of the UCA with N=12 elements.

VI. Conclusion

This paper investigated uniform circular arrays (UCA) for smart antennas. Two main issues: estimation of direction of arrival (DOA) and adaptive beamforming (ABF) were examined. The main approach to DOA here was the algorithm ESPRIT. The main technique for ABF was an improved version of the classical LMS algorithm named ILMS. The symmetry of the UCA antennas was exploited in order to obtain more efficient method for a calculation of accurate eigenvalues. The ESPRIT is a novel algorithm that provides closed-form automatically-paired source azimuth and elevation estimates. These results are proved to be accurate enough.

Concerning beamforming the UCA has shown to be very good regarding both: SOI (maximum) and SNOI (deep nulls). Some deficiency with beamforming along elevation was observed. This can be overcome by using three dimensional arrays (like uniform spherical array, for example).

References

[1] C. Sun et al., “Fast Beamforming of Electronically Steerable parasitic Array Radiator Antennas: Theory & Experiment”, vol. 52, no. 7, pp. 1819 – 1832, 2004.

[2] Z. Huang et al., “Mutual Coupling Compensation in UCAs: Simulations and Experiment”, IEEE Transactions on Antennas and Propagation, vol. 54, no. 11, pp. 3082-3086, 1996.

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

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

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

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

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

Investigation of Periodic EBG Structures with Increased Bandwidth

Nikolay M. Stoyanov[1] and Chavdar Levchev[2]

Abstract –This paper represents investigation of methods for mitigation of mutual coupling between the elements of planar antenna arrays by using EBG substrates. Closed-form expressions for fast calculating of the band-gap frequencies based on simplified equivalent circuits are presented. The results can be used for reducing of mutual coupling between the elements in planar antenna arrays and other RF devices.

***Keywords* – Patch antennas, surface wave, EBG, high surface impedance.**

I. INTRODUCTION

In the last few years a new class of periodic structures, which have frequency band-gap, are designed. This class of structures is known as EBG (electromagnetic band-gap structure) structures. They have a large number of applications as: compact waveguide, suppressing of surface wave, commutation noise in the digital circuits and so on. [1], [2]. The most used periodic structures are two dimensional (2D). 2D metal-periodic structures are mainly investigated in the paper. It is accepted that there are conditions for quasi-stationary.

If the dimension ***a*** of each element is smaller than wavelength, electromagnetic property can be described with equivalent circuits – capacitances and inductances (fig. 1.). In the frequency band with high impedance, tangential magnetic field is low even if there is high electric field in the surface. For this reason sometimes such structures can be described as "magnetic conductor" [4]. The most elementary periodic structure of this type is the periodic structure with grounded microstrip patches. In the paper it is taken notice of such structures with increased frequency band.

II. EBG SUBSTRATE DESIGN

The basic characteristic of periodic structures prototype is the usage of additional equivalent resonance circuits in each element [5]. If each microstrip line is examined like an equivalent inductance and the distance between them is examined like a capacity, it is not hard to be seen that parallel resonance circuits are formed. On the other hand, using the above-mentioned conception, equivalent transmission line between all elements is formed (1).

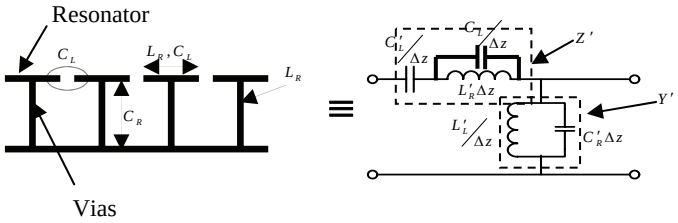


Fig. 1. Equivalent circuits of the meta-materials with increased bandwidth by using transmission line theory

The formula for the self capacitance can be presented:

$$C_1 = \xi \frac{a\varepsilon_0(1+\varepsilon_r)}{\pi} \cosh^{-1}\left(\frac{a+g}{g}\right) \tag{1}$$

The filling ratio ξ f is defined as the area of the metal surface of a spiral-shaped element with sizes $d \times d$ divided by the area of a square patch d^2.

Equivalent inductance is formed on the surface of the dielectric substrate in the area of each of these elements [3]:

$$L_{str}(nH) = L(nH) \pm M(nH) \tag{2}$$

where L is the self-inductance of the microstrip line, M is the mutual inductance. They can be expressed with the following equations:

$$L(nH) = 2.10^{-4} l \left[\ln\left(\frac{l}{W+t}\right) + 1.193 + \frac{W+t}{3l} \right] K_g \tag{3}$$

$$R_s(\Omega) = \frac{KR_{sh}l}{2(W+t)} \tag{4}$$

$$C_{R_str}(pF) = 16{,}67.10^{-4} l \frac{\sqrt{\varepsilon_r}}{Z_0} \tag{5}$$

where all dimensions are in microns and

$$K_g = 0.57 - 0.146 \ln\frac{W}{h}, \qquad \frac{W}{h} > 0.05 \tag{6}$$

$$K = 1.4 + 0.217 \ln\left(\frac{W}{5t}\right) \qquad 5 < \frac{W}{t} < 100 \tag{7}$$

The term K_g accounts for the presence of a ground plane and decreases as the ground plane is brought nearer. Another term

[1]Nikolay M. Stoyanov is with the Ministry of Interior, Sofia, Bulgaria, E-mail: nmstoyanov@gmail.com

[2]Chavdar P. Levchev is with the Ministry of Interior, Sofia, Bulgaria. E-mail: ch_levchev@ncc.mvr.bg

K is a correction factor that takes into account the crowding of the current at the corners of the conductor. The terms *W*, *t* , *h*, *l* and R_{sh} are the line width, line thickness, substrate thickness, length of the section, and sheet resistance per square of the conductor, respectively. The mutual inductance is:

$$M_{ij} = 2.10^{-4} l_e \left[\ln\left\{ \frac{l_e}{d} + \sqrt{1 + \frac{l_e^2}{d^2}} \right\} - \sqrt{1 + \frac{d^2}{l_e^2}} + \frac{d}{l_e} \right] \qquad (7)$$

where, l_e is the effective length of both adjacent sections, *d* is the distance between the sections.

III. Experimental investigations of metalo-dielectric structures with grounded vias.

The analytical expressions for the design of the structures with grounded vias are used. For that purpose the basic configuration from fig. 1. is used. Instead of microstrip resonators microstrip lines with definite width are used (fig. 2). The main objective is bandwidth increase by a group of adjacent resonances. The computation of this structure is performed approximately by the expressions (1) - (8).

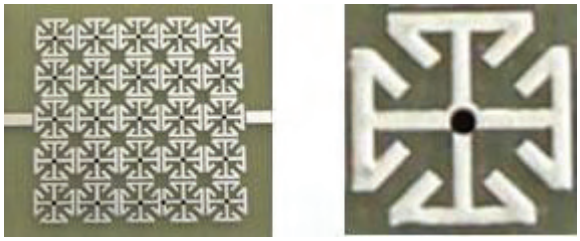
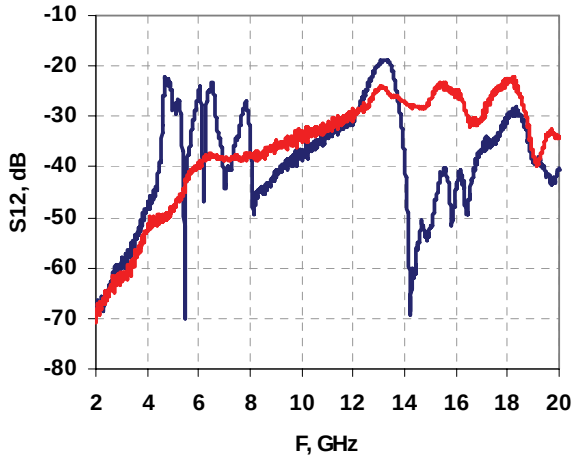


Fig. 2. EBG structure with inductive elements and grounded vias

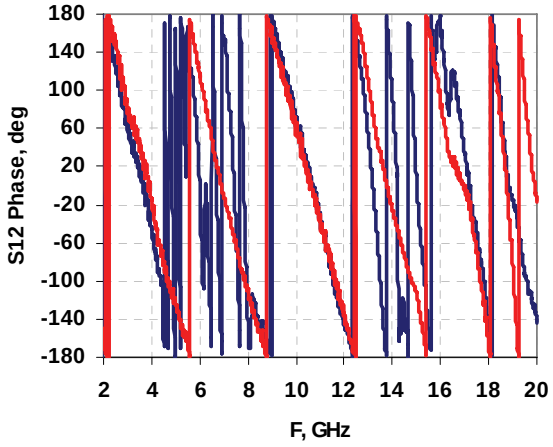
The parameters of this structure are as follows: material of the dielectric substrate *FR4* with relative dielectric permittivity $\varepsilon_r \approx 4{,}4$; height of the dielectric substrate *h = 1,5 mm*; width of microstrip line *W=0,3 mm*; gap between the elements of the structure *g = 0,3mm*; period of the structure *p = 6mm*; radius of the holes *r = 0,4 mm*.

Determination of the band-gap is realized by microstrip lines with characteristic impedance $Z_c = 50\ \Omega$. Self-calibration of the vector network analyzer is performed in the range of *2 GHz to 20 GHz* where the frequency points are *N=1601* ($\Delta f \approx$ *11 MHz)* in number.

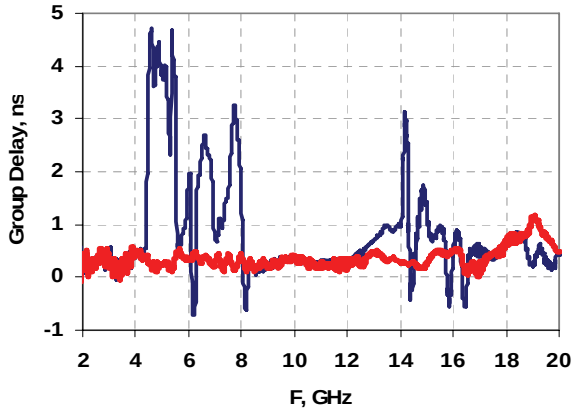
The measurement results are presented on fig. 3.



a)



b)



c)

Fig. 3. The measurement results of the structure are shown on fig. 6.16. a) Magnitude of S_{12} vs. frequency, dB b) Phase of S_{12} vs. frequency c) Group delay τ, *ns* vs. frequency

The plots for S_{12} [dB] show relatively wide band with low mutual coupling that is formed by several narrow band

resonances. A great mitigation of the mutual coupling at 14,2 GHz that is 40 dB and also 16,8 GHz that is 20 dB can be seen. The frequency band with 10 dB mitigation is 3 GHz and above 3 dB is greater than 4 GHz. In practice, the mitigation of the mutual coupling can be also achieved in the frequency band 8-9 GHz.

Fig. 3.b shows multiple phase variation for the regions where increasing or decreasing of the mutual coupling exists. The sign of these variations is shown on the group delay plot. It must be mentioned that in the frequency band between 5 and 8 GHz the sign of the group delay coincides with the sign of magnitude of S_{12} [dB]. This property is inherent to the electrical filters. In this frequency band the structure is alike the band pass filter. For the frequency region about 5 GHz the mutual coupling is 25 dB greater than in the case without this structure.

In the band from 9 to 12 GHz the variations of the magnitude and the phase of S_{12} are setting up in steady state. This results in the phase of S_{12} that is equal to the phase for grounded dielectric substrate without any metal structures. At frequency f = 14,1 GHz the group delay grows very fast but the mutual coupling is reduced due to the increasing of the surface impedance. At frequency f = 14.2 GHz the group delay is decreased and changes its sign. After reaching the minimum value of the group delay it can be seen maximum of the attenuation again (fig. 4).

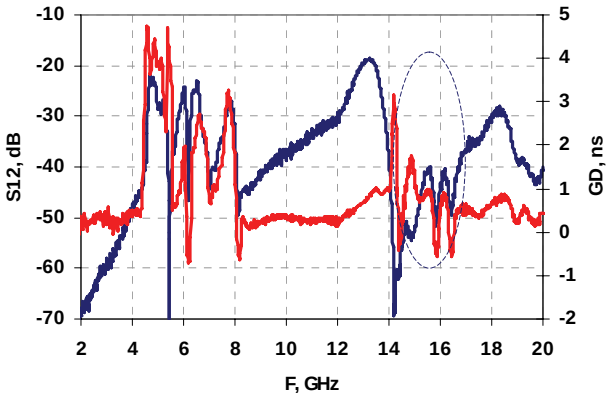


Fig. 4. The magnitude of S_{12}, dB and the group delay τ, ns vs. frequency

The magnitude of S_{12} and the group delay are shown on the same plot in order to represent the variations of these terms in the frequency band with high surface impedance. In fact, in this frequency band the structure changes the signs of the effective permettivity and permeability. At the points where the impedance reaches its maximum the constant of propagation is zero and the phase alters its sign.

Another topology of the periodical structure with high surface impedance that contains vias is shown on fig. 5. Despite of the previous topology, in this structure all inductive elements have common starting point and a capacitive coupling is formed in the opposite ends.

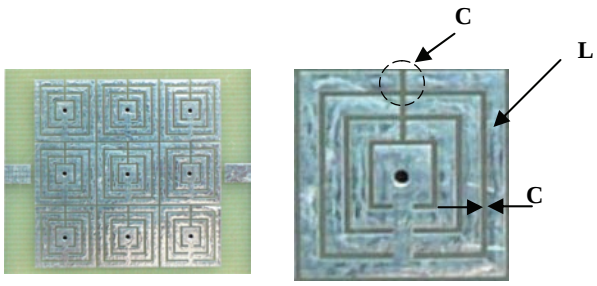
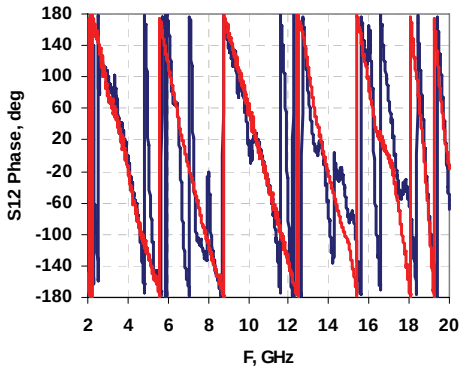
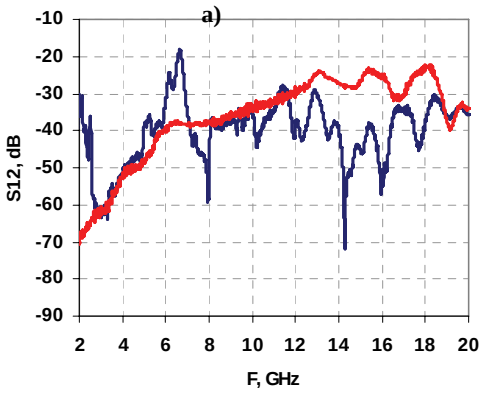
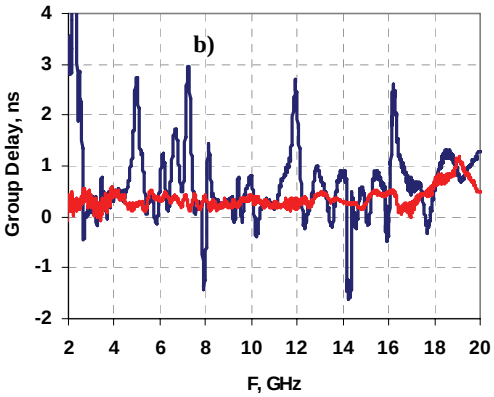


Fig. 5. Combined metal-dielectric periodical structure with modified inductive elements and vias







c)

Fig. 6.The measurement results for the structure a) Magnitude of S12 vs. frequency, dB b) Phase of S vs. frequency c) Group delay τ, ns vs. frequency

These plots show that the variation of the transmission coefficient is very much alike the structure on fig. 6. The main advantage of new structure is the presence of resonators that results in high surface impedance. Consequently, it decreases of the surface wave propagation in sufficiently wide band by appropriate selection of the inductors and capacitors can be achieved. The vias usage is the main disadvantage of the above- mentioned structures.

IV. CONCLUSION

In this paper are made investigations of two kinds of periodic metal-dielectric structures, which are used to increase the surface impedance of the dielectric substrate between the elements. Contemporary methods for design of periodic frequency selective structures with high surface impedance for reducing of mutual coupling between the elements in the planar antenna arrays are used.

The results can be used for development and design of antenna arrays with a low mutual coupling between the elements, for reducing of switching noise in the active antennas and other RF devices.

REFERENCE

[1] C. Balanis "*Antenna Theory (Analysis and Design)*", John Willey and Sons, Inc. N.Y. 1998

[2] Helszajn, J., "*Microwave planar passive circuits, and filters*", JohnWiley&Sons, 1994

[3] I. Bahl, "*Lumped Elements for RF and Microwave Circuits*", Artech House, London, 2003

[4] D. Sievenpiper, Lijun Zhang; R.F.J. Broas, N.G.Alexopolous, E.Yablonovitch, "High-impedance electromagnetic surfaces with a forbidden frequency band*" IEEE Transactions on Microwave Theory and Techniques, , vol.47, No.11, pp.2059-2076, November 2003*

[5] D. F. Sievenpiper, "High-Impedance Electromagnetic Surfaces", PHD dissertation, 1999.

Automated System for Antenna Parameters Measurements in Anechoic Chamber

Chavdar P. Levchev[1] and Nikolay M. Stoyanov[2]

***Abstract* – In this paper presents the block scheme of automated system for measurements control of the antennas with parallel computations in anechoic chamber is presented. Measurement methods and measurement methodology are developed. The functions of each measurement device are explained. Special features related to near-field and far-field measurements are presented. Fragments, figures and diagrams obtained by specially developed software for automated measurements are shown.**

***Keywords* – Anechoic Chamber, Automated System for Measurement, Antenna Parameters**

I. INTRODUCTION

Measurement of antenna parameters often takes a long time. The antenna pattern, gain, direction, efficiency and voltage state wave-ratio (VSWR) are the parameters which require special measurement antenna test areas and precisions equipment ensuring accurate measurements [1]. Antenna test areas have to provide measurements in the far-field zone. They can be realised as open area test site (OATS) or as closed area – Anechoic chambers. The measurement equipment include positioners for the receiving and transmitting sides, providing precise adjustment in azimuth and elevation of receiving antenna and transmitting antenna and device under test (DUT). Measurements conducted by the contemporary measurement equipment are fully automated. Antennas, RF devices, amplifiers and cables are calibrated and their parameters are under permanent control.

In the present work, an automated system for measurements in the anechoic chambers is presented. The explained method and system can be used for measurement of open area test site.

II. AUTOMATED SYSTEM FOR CONTROL OF ANTENNA PARAMETER MEASUREMENTS IN ANECHOIC CHAMBER

The block scheme of an automated measurement system in anechoic chamber is shown on fig. 1.

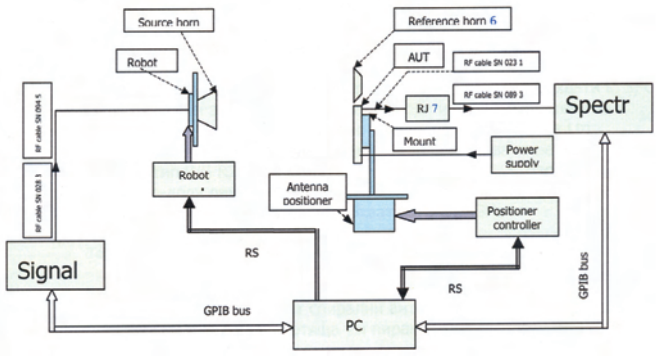


Fig.1. Automated system for measurements of the antenna parameters in anechoic chamber

The above presented system is designed for measurement of antenna gain, antenna pattern of antennas with linear polarization and axial ratio of antennas with circular polarization. The signal-generator, spectral analyzer and the computer are located outside of the anechoic chamber. All the other equipment such as robots, positioners, etalon and measurement antennas and accessories necessary for the measurements is installed in the anechoic chamber. The shielding of the chamber together with the RF absorbing material [2] ensure environment suitable for purpose of the measurements. This kind of measurement chamber simulates free space environment. The shielding reduces the noise level from the surround area and other external influences. The microwave absorbers minimize unwanted waves reflected from the walls, which could influence the measurements. In practice, it is relatively easy high levels of attenuation of the interferences in the surrounding area (from 80 *dB* to 140 *dB)* to be reached by screening, which usually makes these interferences negligible.

The possibility for measurements on the anechoic chamber is limited from the physical dimentions (for aperture antennas), weith, frequency band and equipment which is availability. The physical dimensions are related to the distance R between the transmitter and received antennas. Antennas, measured the anechoic chamber, must be fulfill the following condition:

$$R > 2D_1^2 / \lambda ; \quad (1)$$

where:

$$R > D_1 D_2 / 0.32\lambda \quad (2)$$

R – distance in meters between receiving and transmitting antennas;

[1]Chavdar P. Levchev is with the Ministry of Interior, Bulgaria, E-mail: ch_levchev@ncc.mvr.bg

[2]Nikolay M. Stoyanov is with the Ministry of Interior, Sofia, Bulgaria, E-mail: nmstoyanov@gmail.com

D_1 - the largest dimensions of the aperture of antenna under test in meters;

D_2 - the largest dimensions of the aperture of the transmitting antenna in meters;;

λ - wavelength of the radiated signal in meters.

On the other hand, the distance *R* must be smaller than the largest distance *A* between two antennas, which depends on geometrical dimensions of the anechoic chamber

$$2D_1^2 / < R < A, m .$$

The weight of the antenna under test cannot exceed the maximum weight allowed for each positioner. For example, with the positioner antennas can be manually or automatically orientated to the opposite direction through electrical motor. When automated system with electrical motor is used, should be paid attention to the maximum allowed conditions defined from the nominal start and stop moment of the motor. These parameters are usually marked on the motor in kg/m.

The frequency band of measurements of the antenna under test is related to the geometrical dimensions of the anechoic chamber, to the pyramidal absorbers and their price and to the price of the measurement equipment. With the increasing of the frequency band, the bandwidth becomes wider and the price of the anechoic chamber increases.

III. Antenna Parameters Measurement in Anechoic Chamber

The following main antennas parameters can be automatically measured in the anechoic chamber:

1. Antenna radiation pattern;
2. Antenna gain (G).
3. Polarization characteristics of the antennas:
   - Axial ratio (AR);
   - Cross polarization discrimination (XPD).

Values of antenna directivity (D) and antenna efficiency η, based on the measured diagram pattern, measured and computed antenna gain could be calculated.

$$G = D.\eta \qquad (3)$$

A. Measurement of the antenna radiation pattern

The antenna diagram pattern describes the power or voltage space function. Usually this function describes in spherical coordinate system for fixed distance from the antenna (far field) and fixed frequency. Normally, the diagram pattern consists of main beam, which defines the direction of the maximum radiation and side lobes. Space diagram pattern or antenna pattern cuts, depending on the used method of measurement, could be defined.

Antenna co-polarization and cross-polarization pattern cuts can be taken down, when the wave polarization of the transmitting antenna is opposite to the polarization of the antenna under test. After capturing of these cuts, they may be presented in logarithmic scale.

In order to conduct above-mentioned measurements, special software is developed. The software provides remote control of the positioners, which are situated in the anechoic chamber, and it is necessary to be rotated on azimuth or elevation. In the same way the antenna rotation, height and polarization is controlled. In the most frequently used in practice measurements test set for the antenna pattern, the positioner rotates the AUT and the reference antenna stays fixed at one point from surface of the radiation sphere in the far field, in a way that starting position for axis of the main beam for two antennas is common. Short fragment from the software, which controls of the positioner and orientate the two antennas against each other is shown on fig. 2.

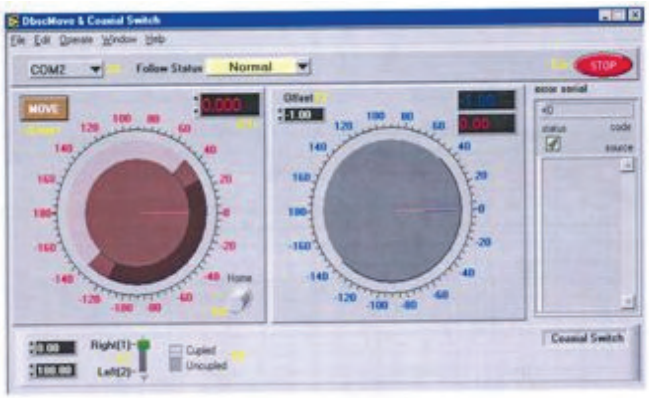


Fig.2. Fragment from the software, which controls of the positioner and orientate the two antennas against each other

The software provides visualization of the measurement procedures. The measured pattern cuts for the co-polarization and cross-polarization are shown on fig. 3.

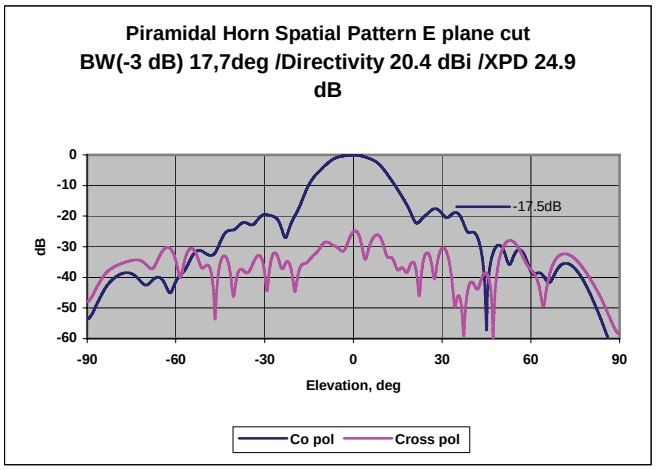


Fig.3. Measured pattern cuts for co-polarization and cross-polarization

B. Measurements of the coefficient of axial ratio of the antenna pattern

It is possible polarization pattern of the anechoic chambers to be taken. The polarization pattern of an antenna is the amplitude response of an antenna, as the antenna is rotated around its axis when illuminated by a linearly polarized plane waves radiated from the transmitting antenna, represented in a polar coordinate system. The axial ratio of an antenna is the ratio of the maximum and minimum of the measured polarization pattern, which is shown on fig.4.

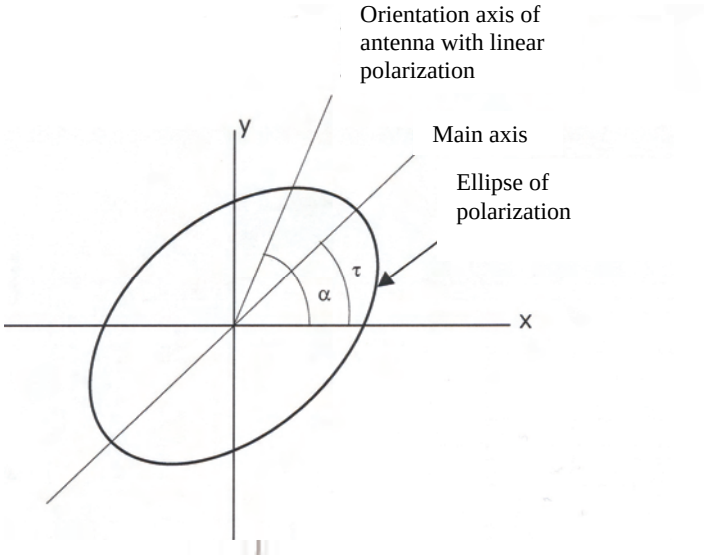


Fig.4. Measurements of the coefficient of axial ratio of the antenna patter

$$AR = P_{R,MAX} - P_{R,MIN}, dB$$

The whole measurement process is controlled by specially developed software allowing precise tuning of the positioners and the antennas. Fragments from the software are depicted on fig. 5 and fig. 6.

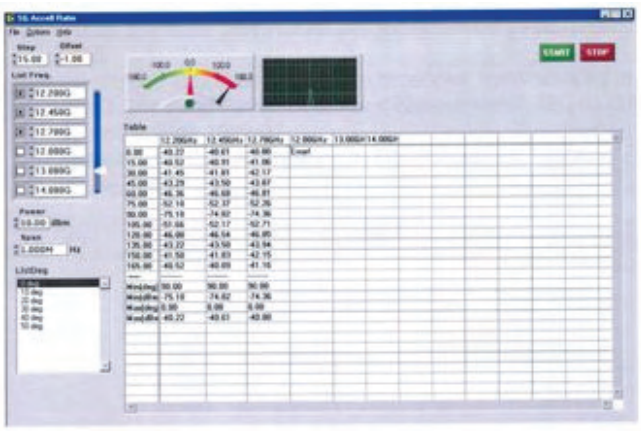


Fig.5. Fragment from the software, which controls the whole measurement process

Fig.6. Fragment from the software, which making tuning of the positioners and the antennas

C. Cross-polarization discrimination of an antenna

This parameter characterizes the ability of an antenna to divide waves with orthogonal polarization, i.e. the main polarization for the AUT and its cross-polarization.

$$XPD = P_{R,CO-POL} - P_{R,CROSS-POL}, dB \quad (4)$$

where $P_{R,CO-POL}, P_{R,CROSS-POL}$ are the received power from AUT for a plane waves with main and cross-polarization respectively.

In order to conduct this kind of measurements, computer sends different commands to the positioner for polarization or antenna rotation, depending on the type of antenna and polarization used.

D. Antenna gain measurements in the anechoic chamber.

There are several different methods for an antenna measurement in an anechoic chamber. The gain comparison method is one of them. The measurements on this method are managed by specially developed measurement software. The computer transmits synchro-pulses for device synchronization through a GPIB buss to the signal generator and spectrum analyzer in the beginning of each measurement procedure.

The two antennas have to be co-axial directly aligned opposite each other. The positioner with control realizes the alignment of the antennas, by tuning the position of the AUT (height, azimuth and elevation) for maximum received power from the transmit antenna (fig. 7).

The signal of the standard antenna comes from the sweep-generator. The level of received signal at the measurement antenna is measured with a spectral analyzer. After the antenna, is measured, as a replacement of measurement antenna is put a standard antenna and measure her level. The levels could be measured by a spectral analyzer and by a vector network analyzer/ VNA (fig.8).

The Gain is calculated as follows:

$$Gaut = L_1 - L_2 + C$$

where Gaut is the gain of the measurement antenna (dBi);

L_1 – measured level of the received signal from the AUT (dB);

L_2 – measured level of the received signal from etalon antenna (dB);

C –gain of the etalon antenna (dBi).

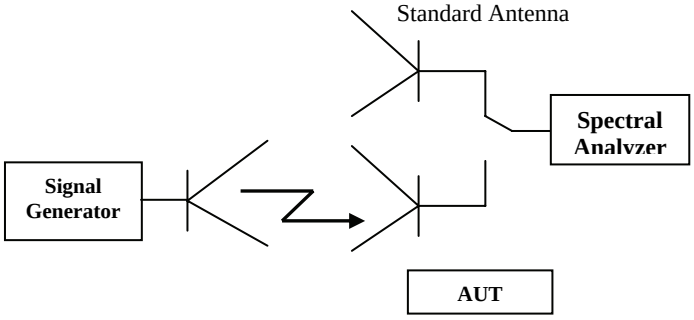


Fig.7 Gain comparison method

The VSWR of the cables and amplifiers used during the antenna measurements are measured by VNA /vector network analyzer/, because of the mishmash loses in them.

Cables and amplifiers are measured for attenuation loses and gain coefficient. All these measurements are made by means of VNA /vector network analyzer/. The VNA can measure the magnitude and phase of S-parameters of the DUT in desired frequency range of measurement. The measurement is automated and together with gain could be done by the computer (fig. 8).

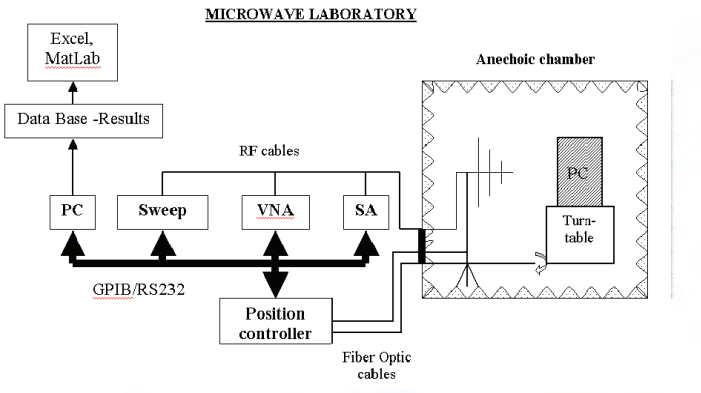


Fig.8 for automatic antenna measurements and radiated emissions from electronic devices

IV. CONCLUSION

The represented automated system for measurements of the antenna parameters in anechoic chamber increases very much the speed and accuracy of the measurements. All space coordinates of the receiving and transmitting antenna are defined by software and optimal tuning levels are automatically obtained. Synchronization between the transmitting and receiving part accelerate data receiving and processing, likewise data visualization and results from the measurements.

As a disadvantage of the system, it could be mentioned the high cost of the equipment, but as practice shows the money investments quickly returns. When the anechoic chamber and software are used for antennas measurements and EMC measurements for different type of devices and commercial technical equipment the profit becomes bigger.

The anechoic chamber on fig.8 can be used for automatic antenna measurements and radiated emissions from electronic devices.

The developed automated system for control of the measurements, parallel data processing and methodology for antenna radiation pattern measurements and surface wave radiation leads to incessantly improvements of the measurements procedures.

REFERENCES

[1] P. Watson, K. C. Gupta, “EM-ANN Models for Microstrip Vias and Interconnects”, IEEE Trans., Microwave Theory Tech.,vol. 44, no. 12, pp. 2395-2503, 1996.

[2] B. Milovanovic, Z. Stankovic, S. Ivkovic and V. Stankovic, "Loaded Cylindrical Metallic Cavities Modeling using Neural Networks", TELSIKS'99, Conference Proceedings, pp.214-217, Nis, Yugoslavia, 1999.

[3] S. Haykin, *Neural Networks*, New York, IEEE Press, 1994.

SDARS Antenna

Peter Petkov[1] , Nikola Dodov[2] and Nikoleta Nikolova[3]

***Abstract* – This paper examines the possibility for the design of a market-oriented antenna with circular polarization. The advantages of the suggested structure are cost-efficiency and simplicity of production compared to traditional ones (thick, high-epsilon antennas). The possibilities for its integration in GPS systems are also discussed.**

***Keywords* – Fractal Antenna, Circular polarization, SDARS, GPS**

I. BASIC PRINCIPLES

SDARS (Satellite Digital Audio Radio Service) is currently operated by Sirius and XM radio. The system operates at 2.320-2.345 GHz (S-band) on Left Hand Circular Polarization (LHCP). The main engineering methods to achieve circular polarization are excitation of 2 modes in the printed resonator (H10 и H01). Some of the configurations shown on fig (1) are used for that purpose. In case 1a, the length of the patch is equal to its width. In the far field a linear polarization will be formed, sloped at 45 degrees. To ensure a circular polarization it is essential to provide a phase shift of 90 deg between these two orthogonal modes. Using non-exact resonant length of the sidewalls ensures this shift, so the L size is a bit longer than the resonant length and the W size is a bit shorter. This nearly square geometry gives +45 and -45 deg phase shift of the resonant modes and guarantees circular polarization in the far field zone of the antenna [1].

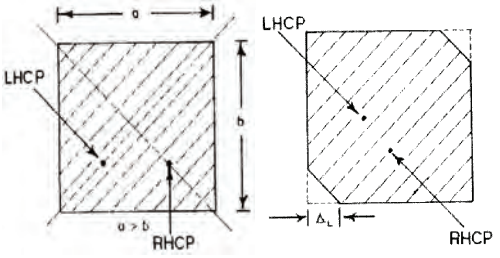


Fig. 1. Figure example

In case 1b, the circular polarization is created by two orthogonal modes, excited in the diagonals of the square. This effect (and 90 deg phase shift) can be explained with the reduction of distributed capacity on both sides of radiating edges. The tapered diagonal is characterized by much more inductive impedance. With variation of the area of the cut ΔS, the ratio E_x/E_y can be adjusted [3].

II. FRACTAL SHAPES

The use of a fractal structure is a basic way for reducing antenna size. Given the same frequency fractal antenna, size can be reduced up to less than 45% of the original size. On Fig2. a fractal Koch island form 0-th , fist and second order is presented. It is quite easy to see the change in the resonator size, with no change in the frequency. A reduction of 75% appears in fractals of second and third order. The second conclusion is that there is no size advantage from using higher order fractals (Higher than first order).[2].

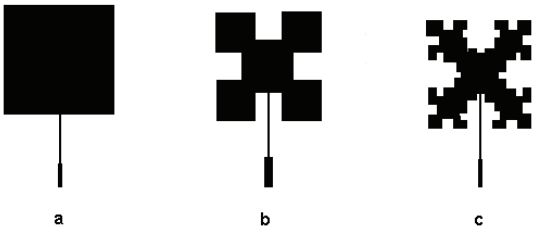


Fig. 2. Generation process of Koch island. (b, c – first and second order)

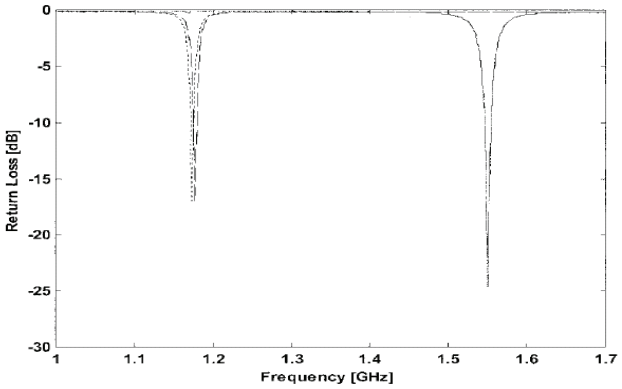


Fig. 3. Return loss of Koch island. (Solid line - 0 order, dashed lines – first and second order)

III. RESULTS

To form a circular polarization in a fractal antenna, a higher order fractal is used for changing corner capacity. Actually, in that case the capacity increases, in spite of the trend described in the case on Fig. 1b. By changing the fractal order and fractal dimension a proper axial ratio can be adjusted

[1]Peter Zh. Petkov is with the Technical University of Sofia, Dept. of Radiotechnics, Faculty of Technical Sciences, 8, Kl. Ohridsky blv, 1000 Sofia, Bulgaria, E-mail: pjpetkov@tu-sofia.bg

[2]Nikola I. Dodov is with the Technical University of Sofia, Dept. of Radiotechnics, Faculty of Technical Sciences, 8, Kl. Ohridsky blv, 1000 Sofia, Bulgaria, E-mail: ndodov@tu-sofia.bg

[2]Nikoleta Nikolova is with the Technical University of Sofia, FCTT, 8, Kl. Ohridsky blv, 1000 Sofia, Bulgaria, E-mail: ladylazar@abv.bg

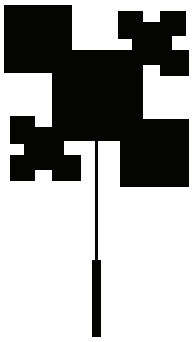


Fig. 3. Proposed design of Circular polarized fractal antenna

The main rule of thumb for producing circular polarization is the area affected of fractalization should be in some definite ratio to the area of initial patch. (Eq.1). In that ratio Q_0 is the quality factor of the patch antenna before fractal changes to be applied. Fractal dimension could be not only a integer number, but also a fraction. This will affect the fractal geometry and gives opportunity for "fine tuning" the shape on the desired frequency.In other words changing both - the fractal size and fractal dimension could be achieved the necessary distributed capacity with shape on proper size and resonating on proper frequency. Such procedure is presented on Fig. 4.

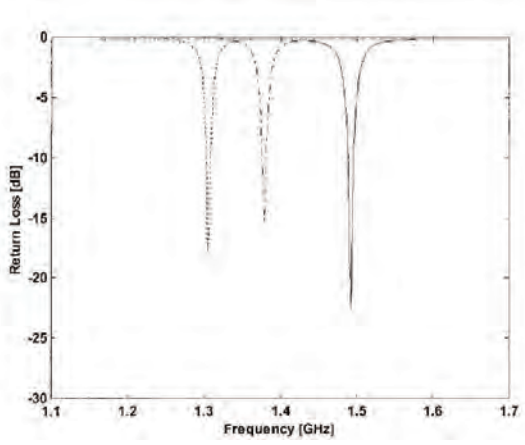


Fig. 4. Tuning of fractal shape by change of fractal dimension. (Solid line – 0.6, dashed lines 0.77 and 2.0)

The use of fractal shapes has one major disadvantage – the bandwidth became narrower with the increase of fractal order. In that exact application this will not be a problem, because SDARS and GPS signals are narrowband enough and antenna properties will not affect the system performance. Because of that reason matching with quarter wave transformer has been used also.

$$\frac{\Delta S}{S} = \frac{1}{2Q_0} \qquad (1)$$

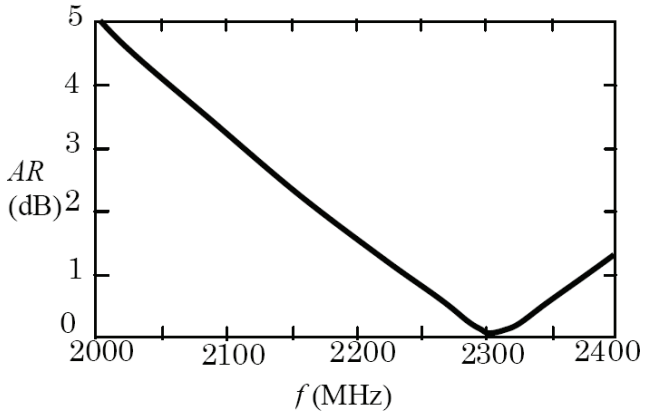


Fig. 4. Axial ratio of the Fractal CP antenna

The simulated axial ratio diagram is presented on Fig.4. Cross-polarization discrimination is maintained in sufficient levels in wide frequency range. Anyway this parameter depends on receiving direction and polarization purity became worse for direction different than main one. For proper functioning of the system it is essential no services to operate on the opposite polarization.

IV. Conclusion

In this paper, the Koch island microstrip patch antenna is proposed. Fractal shape is used in order to reduce antenna size and maintain dielectric permittivity in reasonable values. This patch antenna has a lower resonant frequency compared to the zero-th iteration patch, and this property contributes to the reduction in antenna size. Main advantage of proposed design is significant lower cost than the ordinary high-epsilon antennas and possibility for mass production from inexpensive substrates that can be found on the market. Additional advantage is the increase of antenna efficiency because of lowering the dielectric permittivity. There is also decrease in fringing fields and edge leaking of surface waves. This antenna is intended mainly for automotive industry.

References

[1] Randy Bancroft "Microstrip and Printed Antenna Design", Noble – SciTech, 2004 .

[2] Il-Kwon Kim, Jong-Gwan Yook, and Han-Kyu Park "Fractal shape small size microstrip patch antenna", Microwave and optical technology letters, Vol. 34, No. 1, pp 15-17, July 5 2002

[3] R. Garg, P. Bhartia, I. Bahl and A. Ittipiboon, "Microstrip design antenna handbook". Norwood MA, Artech House, 2001

Investigation of a Wideband Foursquare Microstrip Antenna

Slavi R. Baev[1]

***Abstract* – In this paper a very wideband planar antenna – the Foursquare antenna is investigated. The influence of the feeding system configuration on the antenna characteristics is analyzed. The results obtained for some parameters of the radiating structure reveal in more detail its operation.**

***Keywords* – planar antenna, bandwidth, foursquare element.**

I. INTRODUCTION

The narrow bandwidth of the microstrip antennas is a major constraint. As a result of the research efforts in this direction many radiating structures with improved parameters are created. Some of the antennas with widest impedance bandwidth (BW) are: archimedean spiral (BW≈10:1), equiangular spiral (BW≈8:1), sinuous antenna (BW≈9:1). A relatively new structure in this area is the so-called Foursquare antenna [1], [2]. Its impedance bandwidth is narrower than the above-mentioned antenna types – it is approximately BW≈1,8:1. These values are valid for a single element structure.

The main factors determining the planar antenna application are bandwidth, polarization, physical size. The archimedean and the equiangular spirals work with one sense of circular polarization. This makes them unsuitable for systems with polarization diversity. The sinuous and the Foursquare antenna allow operation with circular or dual linear polarization. The dimension of the radiator for the lowest operating frequency determines the element-to-element spacing in an array structure. To avoid scan angle limitations on higher frequencies this spacing must be as little as possible. Compared to the other wideband elements the Foursquare has a minimum size, which is the reason for the better bandwidth properties in an array environment.

The Foursquare element is suitable for application in systems, which require low profile, multiple frequency or wideband operation, circular, linear or dual linear polarization, array structure.

Only a few publications concerning the Foursquare element are available [1-4]. The information contained is limited and doesn't reveal completely the element behavior. In the present paper the effect of the feeding system on some antenna parameters is analyzed with the purpose to clarify the structure's operation.

[1]Slavi R. Baev is with the Department of Radiotechnics, Faculty of Communications and Communications Technologies, Technical University – Sofia, 8 "Sv. Kliment Ohridski " Blvd., Sofia 1000, Bulgaria, E-mail: sbaev@tu-sofia.bg

II. ANTENNA STRUCTURE

The top and the side view of a Foursquare antenna are shown in Fig. 1.

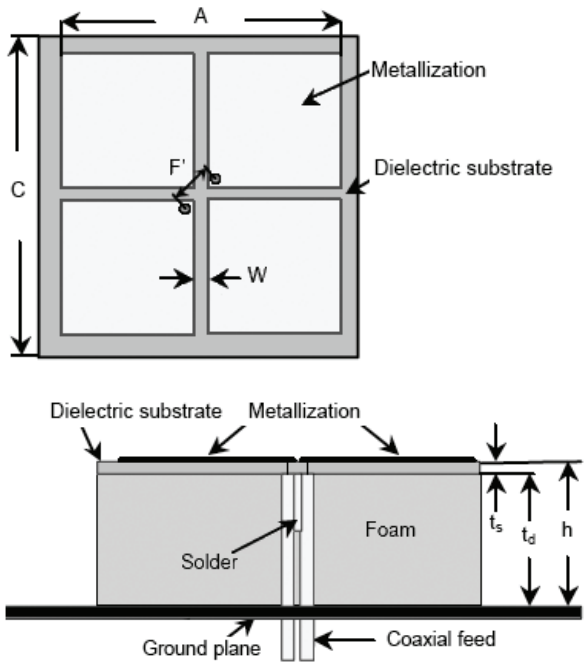


Fig. 1. Construction of a Foursquare antenna element

The antenna consists of four identical metallic patches etched on a dielectric substrate Duroid 5870 (ε_r=2,33). The substrate is positioned over a ground plane and is separated from it by a thick foam layer (ε_r=1,07).

Two of the opposing squares are fed with equal amplitude and opposite phase. The feeding system consists of two coaxial lines placed near the inner corners of the elements. The coaxial lines are embedded into the foam layer and the outer conductors are soldered together near the radiators to ensure balanced feeding. The inner conductors of the coaxial cables pass through the substrate and connect to the metallic squares above it.

The structure described above radiates linear polarization in such a way that the E-field vector is oriented along the main diagonal of the antenna containing the two probe feeds.

Table 1 contains the geometrical parameters of a Foursquare antenna working in a band with a center frequency of about 6 GHz.

The wideband behavior of the Foursquare is due to the change in the current excitation on the four elements with the change of the frequency. As radiators appear not only the two

directly fed elements but also the other two parasitic patches which are fed through a capacitive coupling to the main elements. At the lowest operating frequency f_L the diagonal of the Foursquare $D=\sqrt{2}A$ is about $\lambda_L/2$. Then the resonant length of the antenna is approximately equal to D. When the frequency is increased the condition for resonance is changed and the highest frequency of the working band f_U is determined from the condition $D\approx\lambda_U$ [1], [3], [4].

TABLE I
CONSTRUCTION PARAMETERS OF A FOURSQUARE ANTENNA

Element length	A	21,3 mm
Gap between squares	W	0,25 mm
Distance between probes	F'	4,3 mm
Substrate side length	C	21,8 mm
Substrate thickness	t_s	0,7 mm
Foam thickness	t_d	6,4 mm

As mentioned before the E-field vector is parallel to the antenna diagonal containing the two probes. The Foursquare geometry is such that its long dimension is gradually decreasing away from this diagonal. This ensures the gradual change in the resonant size, which is a prerequisite for a wideband operation.

III. PARAMETRIC ANALYSIS

The main goal of this paper is to present the effect of the feeding line configuration on the antenna excitation ant the corresponding changes in some characteristics – return loss S_{11}, input impedance, radiation pattern, gain G. The simulation results are obtained through the use of a software based on the Finite Element Method – Ansoft HFSS.

The antenna under analysis has the parameters listed in Table 1 except F' – its value is changed through the study. These dimensions are the same as those presented in [1] and [3]. This facilitates the comparison of the results.

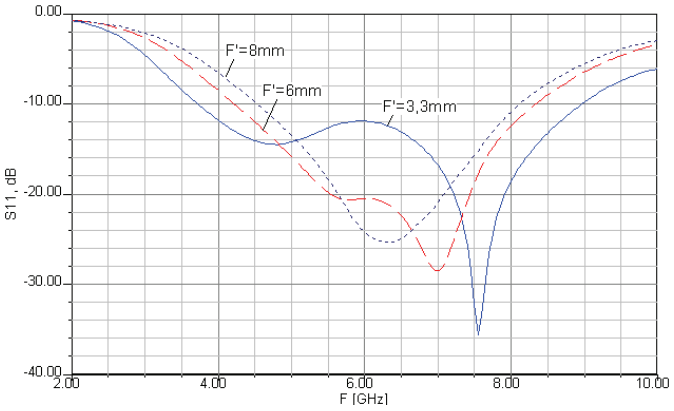


Fig. 2. Return loss S_{11} vs. frequency f for three values of F'

The first investigation shows the dependence of the antenna characteristics from the position of the two probe feeds and particularly from the distance F' between them. It is mentioned in the literature that the optimum position of the probes is as close as possible to the inner corners of the two directly fed squares. The analysis presented here includes three values for F': 3,3mm, 6mm and 8mm. The 50 Ohm coaxial cables used have inner conductor diameter d=0,9mm and outer conductor diameter D=3,1mm.

The return loss for the three values of F' is shown in Fig. 2. It is seen that the widest impedance bandwidth BW (defined for $S_{11} \leq$ -10dB) appears for F'=3,3mm (BW=84%). There are two distinct resonances in the graph. A drawback is the high value of the return loss between the resonances (S_{11} = -12dB), which may be insufficient for some applications. For F'=6mm the bandwidth is approximately 65% and for F'=8mm it is 57%. Increasing F' causes the two resonances of S_{11} to get closer and for F'=8mm there is only one resonance in the middle of the bandwidth. As the distance between the probes increases the matching for the center region of the operating band improves but the minimum achieved values of S_{11} become worse.

The shift of the resonant frequencies in S_{11} is explained with the different surface current excitation on the radiating elements when F' is changed. Although the Foursquare antenna has a complicated resonance behavior some relations can be found. For example when F' is increased the maximum resonant length of the radiator is shorten. The surface currents on the elements flowing in opposite direction to the main currents become stronger. This decrease of the maximum resonant length increases the lower operating frequency f_L and therefore the lower resonance shifts towards the higher frequencies.

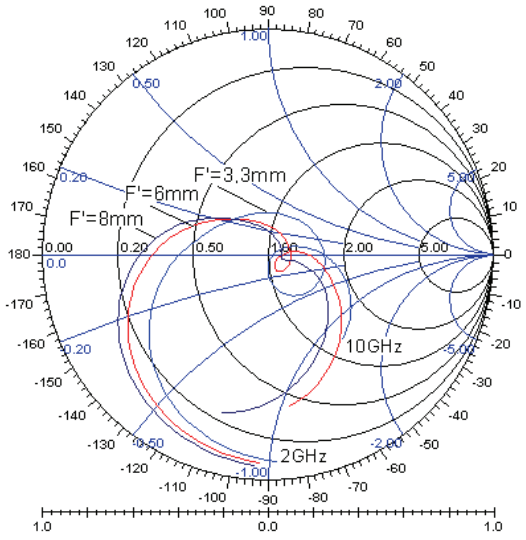


Fig. 3. Smith chart of a Foursquare for three values of F'

Fig. 3 also shows that the two resonances are getting closer. For F'=3,3mm the loop in the Smith chart is wide, which is typical for a two resonance system. When F' is increased the loop is getting smaller and for F'=8mm these is no loop, or it is very tight, which is typical for a single resonance behavior.

The following results present the influence of the distance between the probes on the radiation of the antenna. In Fig. 4 the gain G at boresight for the three configurations is shown.

The maximum achieved values of G are as follows: for F'=3,3mm G_{max}=11,2dB; for F'=6mm G_{max}=9,7dB; for F'=8mm G_{max}=9dB.

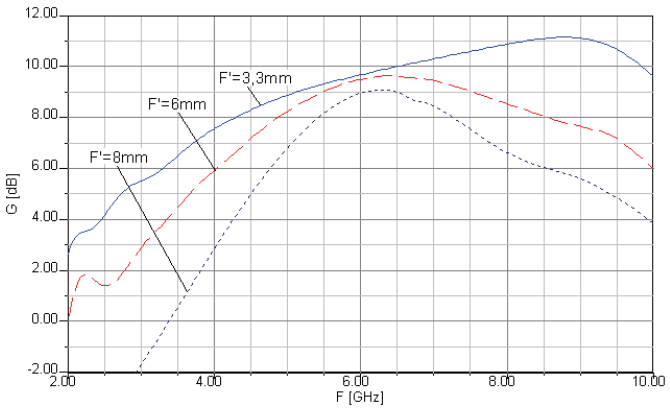


Fig. 4. Gain G vs. frequency f for three values of F'

The radiation pattern bandwidth (determined for a decrease of G with no more than 1dB from G_{max}) for the three cases are: for F'=3,3mm BW=39%; for F'=6mm BW=39%; for F'=8mm BW=29%. It is therefore apparent that the operating bandwidth of the Foursquare is limited by the radiation pattern bandwidth. Nevertheless this band remains very broad for a planar antenna.

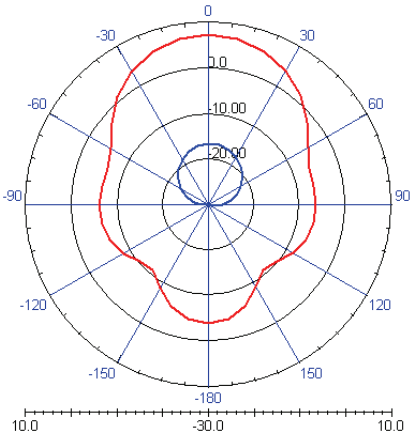
Important result is the drop of the maximum value of G when F' is increased. This is due to the stronger excitation of surface currents flowing opposite to the main currents when the probes are shifted away from the antenna center. The cross-polarization radiation is also increased in this manner. These two effects cause gain degradation.

More detailed description of the Foursquare element gives the E-plane radiation pattern shown in Fig. 5. The E-plane contains the two probes and is perpendicular to the antenna surface. It should be noted that the results are not normalized and the ground plane size is 21,8x21,8mm – the same as the substrate dimensions. In practice it is recommended to use a ground plane at least two times the antenna size [2].

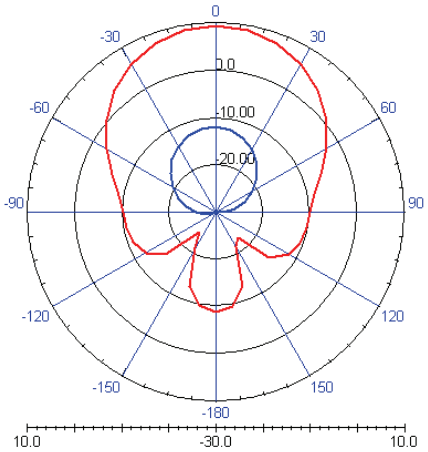
Generally the radiation patterns of the three configurations are similar to the presented in [1-4]. For brevity only the pattern for F'=6mm is shown here.

The main effect that appears when the frequency is increased is that the E-plane pattern becomes wider. This is explained with the change of the main radiating part of the antenna aperture when the frequency is shifted. At lower frequencies the main radiators are the two directly fed patches. For the midband region all the four elements radiate and for higher frequencies the two parasitic patches are the main radiators. The analysis showed that this alteration of the radiation pattern is more significant when F' is larger. This means that for F'=3,3mm the change in the radiation pattern shape with frequency is the weakest and therefore it has the best wideband behavior.

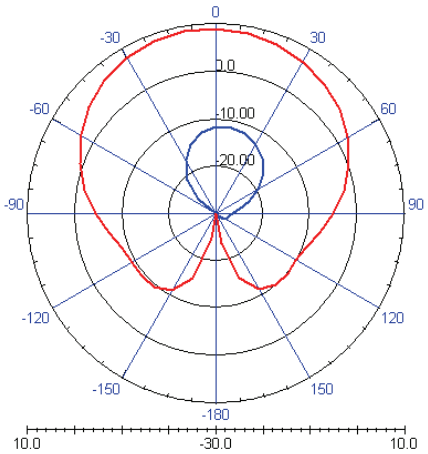
In H-plane there is just a slight alteration of the radiation pattern for the upper half space when the frequency is changed. For different values of F' the H-plane pattern remains almost the same. The backward radiation decreases with f for both E- and H- planes. It is expected that larger ground plane will improve the radiation properties of the antenna.



(a) 4,5GHz



(b) 6 GHz



(c) 7,5 GHz

Fig. 5. Co-polar and cross-polar components of the radiated field in E-plane as a function of Theta for three frequencies. F'=6mm. Values are in dB.

The second experiment resulted from the need for a better impedance matching of the foursquare. It includes an investigation of the input impedance as a function of the type of coaxial line used. Three types of 50 Ohm coaxial cables are

examined: 1. d=0,5mm, D=1,7mm; 2. d=0,9mm, D=3,1mm; 3. d=1mm, D=3,5mm. The antenna under analysis has the dimensions in Table 1 except that F'=8mm. Fig. 6 shows the Smith charts for the three cases.

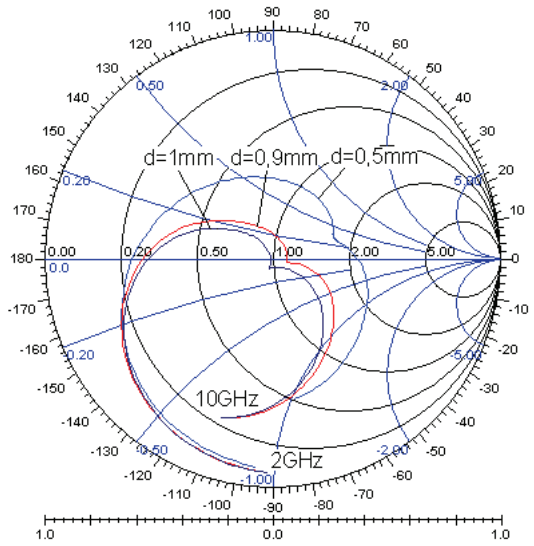


Fig. 6. Smith plot of a Foursquare for three types of coaxial lines: d=0,5mm D=1,7mm; d=0,9mm D=3,1mm; d=1mm D=3,5mm

The impedance curve is shifted to the left towards the lower resistance region as the diameter of the cable is increased. This is caused by the lower input resistance of a probe with larger conductor diameter. This result shows that the size of the coaxial line can be used as an additional instrument for impedance tuning of the Foursquare antenna.

IV. Conclusion

The publication presents the relation between the parameters of a linear polarized Foursquare antenna and the feeding line configuration. The conducted investigation shows that values between 57% and 84% for the impedance bandwidth can be achieved with the Foursquare depending on the feed line position. The operating band is limited by the radiation pattern bandwidth and still the reported values are very good – 29% to 39%.

In contrast to the expectations not all antenna characteristics are optimal when the feed probes are closest to the antenna center (F'=3,3mm). The radiation pattern is the best for this case, but the return loss is not balanced through the whole bandwidth and there is also a mismatch between the impedance and radiation pattern bandwidths.

The results of the present analysis show that the Foursquare antenna can be optimized to a great extent with the use of proper feeding system and its precise placement.

References

[1] J. R. Nealy, “Foursquare Antenna Radiating Element”, *United States Patent № 5,926,137*, 1999.

[2] S.-Y. Suh, and W. L. Stutzman, “Planar Wideband Antennas”, *United States Patent № 7,027,002 B2*, 2006.

[3] S.-Y. Suh, “A Comprehensive Investigation of New Planar Wideband Antennas”, *Ph.D. Dissertation.*, 2002.

[4] C. G. Buxton, “Design of a Broadband Array Using the Foursquare Radiating Element”, *Ph.D. Dissertation.*, 2001.

Investigation of Probe Correction Technique in Planar Near-Field Antenna Measurements

Alexander G. Toshev[1]

Abstract — **Probe correction techniques in near-field theory are investigated theoretically. Simplified probe correction is proposed for the case of good crosspolarization isolation of the probe pattern and angles not far from the boresight of the antenna pattern, which leads to significant reduction of computational effort for obtaining of the far-field radiation pattern of the antenna under test. For the proposed simplified probe correction phase data of the far-field radiation pattern of the probe is not required, which additionally makes it attractive for implementation.**

Keywords — **Near-field measurements, probe correction, planar wave spectrum**

I. Introduction

Modern planar scanning techniques in near-field measurements of antennas are based on planar-wave spectrum (PWS) representation of the fields which could be found in [3], [4]. There is a simple relationship between PWS of the antenna and its far field pattern [3], [4]. This relationship adds significant practical importance of such representation of the fields and particularly to the planar near-field antenna measurements.

During the last quarter century, the advent of the fast Fourier transform (FFT) technique [2] , as well as the rapid progress in electronics, signal processing, and computers, made the planar-scanning near-field measurement increasingly fast, accurate, and economical. As a result, activities in near-field measurement on planar surfaces proliferated [3]-[4], beginning with the probe-correction technique proposed by Kerns [1]. A comprehensive review of this subject was recently conducted by Yaghjian [3].

The problem with the probe correction is as follows. If the pattern of the measuring probe were isotropic the probe wouldn't have influenced the measurement result. Because there are not isotropic probes the last always influence the measurement and has to be considered in the computation. Consideration of the directional properties of the probes used in planar near-field antenna measurements leads to significant complication of computational effort.

This paper emphasizes on PWS representation of the fields and probe correction that has to be applied in planar near-field antenna measurements. Simplified probe correction procedure is proposed, which could save significant amount in computational effort in near-far field transformation.

[1]Alexander G. Toshev is with the faculty of Communications and Communications Technologies, TU Sofia, "Kl. Ohridsky" blvd. N8, 1756 – Sofia, Bulgaria, E-mail: toshev_alex@yahoo.com

The proposed simplified probe correction is applicable in the case of good cross polarization isolation of the probe pattern (example is a open ended rectangular waveguide used as a probe) and for the tilt angles around 30 degrees far from boresight of the antenna pattern.

II. Formulation of Antenna Near fields and Probe Correction Technique

In a source-free free-space region, in which near fields are measured, the time-harmonic Maxwell equations can be transformed into the following vector wave equations:

$$\begin{aligned} \nabla^2 \vec{E} + k^2 . \vec{E} = 0 \\ \nabla^2 \vec{H} + k^2 . \vec{H} = 0 \\ \vec{\nabla} . \vec{E} = \vec{\nabla} . \vec{H} = 0 \end{aligned} \qquad (1)$$

It can be shown that following expressions constitute a solution of the equations (1) for $z \geq 0$ and satisfy the prescribed boundary conditions on the plane $z = 0$:

$$\begin{aligned} \vec{E}(x,y,z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \vec{A}(k_x, k_y) e^{-j \vec{k} \vec{r}} dk_x dk_y \\ \vec{H}(x,y,z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \vec{k} \times \vec{A}(k_x, k_y) e^{-j \vec{k} \vec{r}} dk_x dk_y \end{aligned} \qquad (2)$$

In the equation (2) $\vec{k}$ is the wave vector with the components along the three coordinate axes and $\vec{r}$ is the position radius-vector as follows:

$$\begin{aligned} \vec{k} &= k_x . \vec{i}_x + k_y . \vec{i}_y + k_z . \vec{i}_z \\ k^2 &= \vec{k} . \vec{k} \end{aligned} \qquad (3)$$

$$\vec{r} = x . \vec{i}_x + y . \vec{i}_y + z . \vec{i}_z \qquad (4)$$

Vector $\vec{A}$ represents planar wave spectrum of the field because the expression

$$\vec{A}(k_x,k_y).e^{j\vec{k}.\vec{r}} \qquad (5)$$

represents a uniform plane wave propagating in the direction $\vec{k}$. There is a relationship between the tree components of the PWS so that knowing A_x and A_y is enough for determining all the tree components of the PWS:

$$k_xA_x(k_x,k_y)+k_yA_y(k_x,k_y)+k_zA_z(k_x,k_y)=0 \quad (6)$$

In planar near-field measurements, the antenna is placed in the plane $z=0$, as shown on Figure 1. Planar scanning is conducted in the plane $z=z_t$, near the antenna.

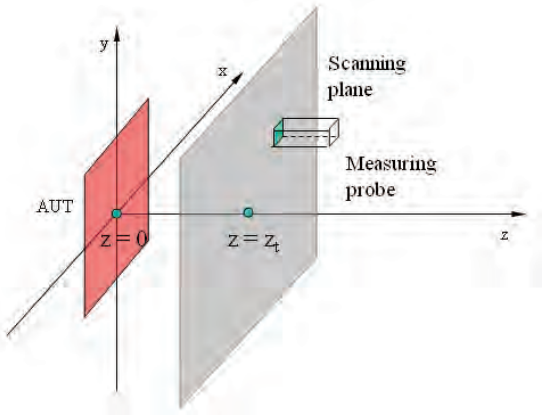


Figure 1: Setup of planar near-field measurement

The radiation condition requires that for $z\geq 0$ the following relations between components of the wave vector should be satisfied [4]:

$$k_z=\begin{cases}\sqrt{k^2-k_x^2-k_y^2}; k_x^2+k_y^2\leq k^2\\ -j\sqrt{k_x^2+k_y^2-k^2}; k_x^2+k_y^2\geq k^2\end{cases} \qquad (7)$$

An imaginary k corresponds to an evanescent PWs, which is rapidly attenuated away from the plane $z=0$.

In the plane $z=z_t$ the x, y components of the field using (2) can be expressed as:

$$E_x(x,y,z_t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A_x(k_x,k_y)e^{-jk_zz_t}.e^{-j(k_xx+k_yy)}.dk_xdk_y$$
$$E_y(x,y,z_t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A_y(k_x,k_y)e^{-jk_zz_t}.e^{-j(k_xx+k_yy)}.dk_xdk_y \qquad (8)$$

we concentrate only in x, y components because z component is easily determined from (1) and (6). Because of the influence of the measuring probe direct measurement of $E_x(x,y,z_t)$ and $E_y(x,y,z_t)$ is not possible. What is actually measured in the plane $z=z_t$ is the reaction of the probe:

$$b_1(x,y,z_t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\vec{R^{b_1}}(k_x,k_y).\vec{A}(k_x,k_y)e^{-jk_zz_t}.e^{-j(k_xx+k_yy)}.dk_xdk_y \qquad (9)$$

Where $\vec{R^{b_1}}(k_x,k_y)$ is the "in advance" known pattern of the probe with representation in spherical and cartesian coordinate system respectively:

$$\vec{R^{b_1}}(k_x,k_y)=R_\theta^{b_1}(k_x,k_y).\vec{i}_\theta+R_\varphi^{b_1}(k_x,k_y).\vec{i}_\varphi \qquad (10)$$

$$\vec{R^{b_1}}(k_x,k_y)=R_x^{b_1}(k_x,k_y).\vec{i}_x+R_y^{b_1}(k_x,k_y).\vec{i}_y+R_z^{b_1}(k_x,k_y).\vec{i}_z \qquad (11)$$

Because of the planar scanning we may choose Cartesian representation and rewrite equation (9) in the form:

$$b_1(x,y,z_t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(R_x^{b_1}.A_x+R_y^{b_1}.A_y+R_z^{b_1}.A_z\right).e^{-jk_zz_t}.e^{-j(k_xx+k_yy)}.dk_xdk_y \qquad (12)$$

Equation (12) is inverse Fourier transform of the scalar product with respect to k_x and k_y. We could obtain the scalar product applying Fourier transform of the measured reaction of the probe and multiplying with the constant $e^{jk_zz_t}$:

$$R_x^{b_1}.A_x+R_y^{b_1}.A_y+R_z^{b_1}.A_z=e^{jk_zz_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}b_1(x,y,z_t).e^{j(k_xx+k_yy)}.dk_xdk_y \qquad (13)$$

Now taking into account equation (6) we can obtain A_z knowing A_x, A_y and rewrite equation (13) in the form:

$$A_z=-\frac{k_x}{k_z}A_x-\frac{k_y}{k_z}A_y \quad (14)$$

$$\left(R_x^{b_1}-\frac{k_x}{k_z}R_z^{b_1}\right).A_x+\left(R_y^{b_1}-\frac{k_y}{k_z}R_z^{b_1}\right).A_y=e^{jk_zz_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}b_1(x,y,z_t).e^{j(k_xx+k_yy)}.dk_xdk_y \qquad (15)$$

Equation (15) is used for obtaining A_x, A_y of the PWS of the antenna under test. Since there is two unknowns another equation is needed, which could be obtained in the same way as equation (15), but using second probe (with different

pattern) or the same probe but 90 degrees rotated. If $\vec{R^{b_2}}(k_x, k_y)$ denotes pattern of the second probe the same equation as (15) is valid:

$$\left(R_x^{b_2} - \frac{k_x}{k_z}R_z^{b_2}\right).A_x + \left(R_y^{b_2} - \frac{k_y}{k_z}R_z^{b_2}\right).A_y = e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_2(x,y,z_t).e^{j(k_x x + k_y y)}.dk_x dk_y \quad (16)$$

Equations (15) and (16) are used for determining of A_x, A_y components of the PWS. The solution is given by the following equations:

$$P_{b_1} = e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_1(x,y,z_t).e^{j(k_x x + k_y y)}.dk_x dk_y$$
$$P_{b_2} = e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_2(x,y,z_t).e^{j(k_x x + k_y y)}.dk_x dk_y \quad (17)$$

$$\Delta_x = \begin{vmatrix} P_{b_1}; \left(R_y^{b_1} - \frac{k_y}{k_z}R_z^{b_1}\right) \\ P_{b_2}; \left(R_y^{b_2} - \frac{k_y}{k_z}R_z^{b_2}\right) \end{vmatrix}$$

$$\Delta_y = \begin{vmatrix} \left(R_x^{b_1} - \frac{k_x}{k_z}R_z^{b_1}\right); P_{b_1} \\ \left(R_x^{b_2} - \frac{k_x}{k_z}R_z^{b_2}\right); P_{b_2} \end{vmatrix} \quad (18)$$

$$\Delta = \begin{vmatrix} \left(R_x^{b_1} - \frac{k_x}{k_z}R_z^{b_1}\right); \left(R_y^{b_1} - \frac{k_y}{k_z}R_z^{b_1}\right) \\ \left(R_x^{b_2} - \frac{k_x}{k_z}R_z^{b_2}\right); \left(R_y^{b_2} - \frac{k_y}{k_z}R_z^{b_2}\right) \end{vmatrix}$$

$$A_x = \frac{\Delta_x}{\Delta}; A_y = \frac{\Delta_y}{\Delta}$$

III. Simplified Probe Correction

Procedure for obtaining of the components A_x, A_y of the PWS of the antenna under test described with equations (17), (18) is common one. In actual measurement discretization of the plane in front of the antenna under test (AUT) is performed and integrals (17) are solved with the help of Discrete Fourier Transform (DFT). Sometimes procedure given with (17) and (18) is too complicated, because for each point of the discretized plane three determinants should be solved (Δ_x, Δ_y and Δ). This makes the speed of the computation very slow and if processing of hundreds of antenna beams is needed the computation time may take hours even on good speed PCs.

The procedure given with equations (17) and (18) could be further simplified if crosspolarized component of the probe pattern is negligible. For linear polarized probes this constitutes the following:

$$R_y^{b_1} \approx 0$$
$$R_x^{b_2} \approx 0 \quad (19)$$

Equations (19) says that $\vec{R^{b_1}}$ is polarized along x axes and $\vec{R^{b_2}}$ is polarized along y axes. Also for angles not very far from the AUT boresight z components of the probe patterns could be considered zero:

$$R_z^{b_1} \approx 0$$
$$R_z^{b_2} \approx 0 \quad (20)$$

Taking into account (19) and (20) equations (15) and (16) could be expressed in a simplified form:

$$R_x^{b_1}.A_x = e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_1(x,y,z_t).e^{j(k_x x + k_y y)}.dk_x dk_y$$
$$R_y^{b_2}.A_y = e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_2(x,y,z_t).e^{j(k_x x + k_y y)}.dk_x dk_y \quad (21)$$

If the same probe is used for the measurement but 90 degrees rotated for each scan the simplified solution for the PWS of the antenna could be write using (21):

$$A_x = \frac{1}{R}.e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_1(x,y,z_t).e^{j(k_x x + k_y y)}.dk_x dk_y$$
$$A_y = \frac{1}{R}.e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_2(x,y,z_t).e^{j(k_x x + k_y y)}.dk_x dk_y \quad (22)$$

where R is the pattern of the probe for the dominant polarization. If circular polarized probes are used equation (22) takes similar form for the two orthogonal circular polarized components of the PWS A_{LHCP} and A_{RHCP}:

$$A_{LHCP} = \frac{1}{R_{LHCP}}.e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_1(x,y,z_t).e^{j(k_x x+k_y y)}.dk_x dk_y \quad (23)$$

$$A_{RHCP} = \frac{1}{R_{RHCP}}.e^{jk_z z_t}.\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b_2(x,y,z_t).e^{j(k_x x+k_y y)}.dk_x dk_y$$

where R_{LHCP} and R_{RHCP} are the patterns of the measuring probe for the two orthogonal linear polarizations.

IV. Discussion and conclusion

In this paper common case of probe correction in planar near field antenna measurements was investigated and revised. A simplified probe correction was proposed for the case of good crosspolarization of the probe pattern and angles not far from the boresight of the antenna under test. The proposed simplified procedure for probe correction leads to significant reduction of the computational effort for calculation of the planar wave spectrum of the antenna, which makes it attractive for implementation.

References

[1] D. M. Kerns, “Analytical techniques for the correction of near-field antenna measurements made with an arbitrary but known measuring antenna,” in Abstracts URSI-IRE Meeting. Washington, DC, Apr-May 1963, pp. 6-7.

[2] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex Fourier series,’’ Math. of Comput., vol. 19, pp. 297-301. Apr. 1965

[3] A. D. Yaghjian, "An overview of near-field antenna measurements," IEEE Trans. Antennas Propagat., vol. 34, pp. 30 - 45, January 1986.

[4] Johnson J. H. Wang, "An Examination of the Theory and Practices of Planar Near-Field Measurements," IEEE Trans. Antennas Propagat., vol. 36, pp. 746 - 753, June 1988.

Coupled Linear Grounding Systems: Quasistationary Antenna Model

Nenad N. Cvetkovic[1] and Predrag D. Rancic[2]

Abstract – Characteristics of the linear grounding systems (LGSs) made of two coupled independently supplied linear ground electrodes, are analyzed in this paper. The unknown current distributions (UCDs) along the conductors are determined using quasistationary antenna model. Ground is assumed as linear, isotropic and homogeneous media having known parameters. UCDs are determined numerically solving system of integral equations (SIE) using point matching method and polynomial approximation for currents. Afterward, the self and mutual impedances (as the measure of mutual LGSs influence) as well as total impedance of connected grounding systems are calculated. For these calculations, so-called symmetric and anti-symmetric regimes of feeding are applied.

Keywords – Quasistationary EM field, grounding systems, linear electrode, antenna model.

I. INTRODUCTION

During last 15 years on the Faculty of Electronic Engineering in Nis, few number of models for analyze of grounding systems treating EM field in the ground as stationary ([1], [2]) and quasistationary ([3]) has been developed. By that, real inhomogeneous ground is modeled as homogeneous, stratified or sectoral structure ([4], [5]). For analyze of grounding systems, especially those ones having massive electrodes, one simple and efficient method named by author as equivalent electrodes method (EEM, [6], [7]) has been also developed. Before 90-ies, two PhD works are done at the Faculty under mentorship of prof. Velickovic in this area (Dr Jelenko Vlajic, Banja Luka and dr Panco Vrangalov, Skopje).

Quasistationary EM field structure in the LGS surrounding is analyzed in the paper, introducing model of longitudinal current along conductor axis, having unknown distribution (UCD). The LGS is fed with very low frequency currents, and domain where LGS is placed provides that surface of connected LGS conductors can be assumed as equipotential. This condition results with integral equation system of Rabenn's type (SIE-R) having UCD as unknown function. SIE-R is numerically solved using Point Matching Method (PMM, [9]) and polynomial approximation for UCD ([10]). By that, inhomogeneous ground is analyzed using various known models, and for each of them, Green function for Hertz vector and electrical scalar potential (ESP) is derived ([3], [8]). In this way is avoided rigorous and very complicated so-called Sommerfeld formulation. The described model corresponds to so-called quasistationary antenna model.

More rigorous quasistationary antenna model demands including of Hertz vector distribution in the way that result tangential electrical field component at the LGS surface is zero. This makes the mathematical model more complicated, and it result with so-called system of integral equation of two potentials (SIE-TP, [11]). If problem is analyzed for very low frequencies, SIE-TP approximately gets the form of SIE-R.

After introductory part, the paper is organized in three sections, followed by the conclusion and reference list.

In the second part of the paper theoretical basics are given, i.e. problem definition as well as expressions for electric scalar-potential (ESP) in the electrode system surrounding and general formulations of applied numerical procedures. The results obtained applying described numerical procedures are presented in the third part of the paper.

II. THEORETICAL BASIS OF THE MODEL

A. Description of the system geometry

System made of total $N = N_1 + N_2$ linear conductors from which N_1 conductors are fed by common isolated earthing conductor with current I_{g1}, and the rest (N_2) of the conductors are fed with common supplying current I_{g2} is observed and illustrated in Fig. 1.

The linear ground conductors of length l_{ik} and circular cross-section radius a_{ik}, $a_{ik} << l_{ik}$, $k = 1,2,...N_i$, $i = 1,2$ are placed in homogeneous ground.

The ground is treated as isotropic semi-conducting media having known electrical parameters σ_1, ε_1 and $\mu_1 = \mu_0$ (σ_1 - conductivity, ε_1 -permittivity, μ_0 - permeability). The following labels are also used in the paper: $\underline{\sigma}_i = \sigma_1 + j\omega\varepsilon_i$, $i = 0,1$ - complex conductivities; $\omega = 2\pi f$ - low angular frequency; and $R_{z10}, T_{z10}, R_{z10} = T_{z10} - 1 = (\underline{\sigma}_1 - \underline{\sigma}_0)/(\underline{\sigma}_1 + \underline{\sigma}_0)$ are quasistationary reflection and transmission coefficients, respectively.

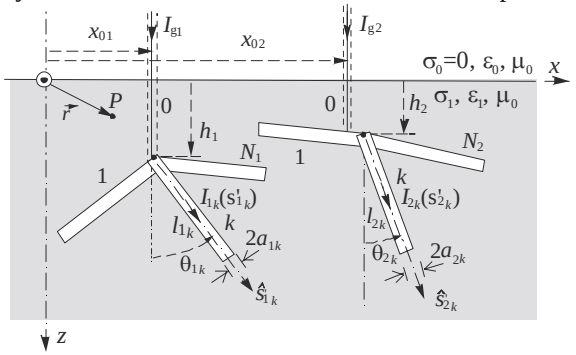


Fig. 1. Illustration of two coupled linear grounding systems (LGS).

[1]Nenad N. Cvetkovic is with the Faculty of Electronic Engineering, P.O.Box 73, 18000 Nis, Serbia, E-mail: csone@elfak.ni.ac.yu.

[2]Predrag D. Rancic is with the Faculty of Electronic Engineering, P.O.Box 73, 18000 Nis, Serbia, E-mail: prancic@elfak.ni.ac.yu.

ESP of the system in the ground point P, having position defined with vector $\vec{r}$, is

$$\varphi_1(\vec{r}) = \varphi_{10}(\vec{r}) + \varphi_{11}(\vec{r}) = \frac{1}{4\pi\underline{\sigma}_1}\left[\sum_{i=1}^{2}\frac{T_{z10}I_{gi}}{|\vec{r} - x_{0i}\vec{x}|} + \right.$$

$$\left. + \sum_{i=1}^{2}\sum_{k=0}^{N_i}\int_{s'_{ki}=0}^{l_{ki}} I_{ik}(s'_{ik})\frac{\partial}{\partial s'_{ik}}K(\vec{r}, s'_{ik})\,\mathrm{d}\,s'_{ik}\right], \quad (1)$$

where x_{0i}, $i = 1,2$ determine the positions of the supplying points at the ground surface, $I_{ik}(s'_{ik})$ -longitudinal current distribution and $K(\vec{r}, s'_{ik})$ is kernel of the form,

$$K(\vec{r}, s'_{ik}) = \frac{1}{r_{1k}} + \frac{R_{z10}}{r_{2k}}. \quad (2)$$

In Eq, (2), r_{1k} and r_{2k} are distances between the field point P and conductor current element, and its image in modified flat mirror, respectively.

B. Numerical procedure evaluation

Matching the ESP value (1) in the conductors surface points P_n, $\varphi_1 \cong U_n$, the SIE-Rs are obtained

$$U_n = \frac{1}{4\pi\underline{\sigma}_1}\left\{\sum_{i=1}^{2}\sum_{k=1}^{N_i}\left[I_{ik}(l_{ik})K(s_n, l_{ik}) - \right.\right.$$

$$\left.\left. - \int_{s'_k=0}^{l_k} I'_{ik}(s'_{ik})K(s_n, s'_{ik})\,\mathrm{d}\,s'_{ik}\right]\right\},\ n = 1,2 \quad (3)$$

where: $I_{ik}(l_{ik})$ are currents values at free ends of ground electrodes, $I'_{ik}(s'_{ik}) = -\partial I_{ik}/\partial s'_{ik}$ are leakage currents density per unit length and s_n denotes distance between the matching point P_n at the n-th conductor surface and its origin.

The solution of the SIE-Rs given with Eq. (3) is current distribution along the conductors, $I_{ik}(s'_{ik})$ $k = 1,2,...N_i$, $i = 1,2$, where feeding currents satisfy conditions

$$\sum_{k=1}^{N_i} I_{ik}(0) = I_{gi},\ i = 1,2. \quad (4)$$

Assuming the current distribution in polynomial form,

$$I_{ik}(s'_{ik}) = \sum_{m=0}^{M_k} I_{ikm}\left(\frac{s'_{ik}}{l_{ik}}\right)^m,\ k = 1,2,...N_i,\ i = 1,2, \quad (5)$$

where I_{km}, $k = 1,2,...N_i$, $m = 0,1,...M_k$ are unknown current coefficients, the SIE-Rs (3) got the form

$$U_n = \frac{1}{4\pi\underline{\sigma}_1}\left\{\sum_{i=1}^{2}\sum_{k=1}^{N_i}\sum_{m=0}^{M_k} I_{ikm}\left[K(s_n, l_{ik}) - \right.\right.$$

$$\left.\left. - \frac{m}{l_{ik}}\int_{s'_{ik}=0}^{l_{ik}}\left(\frac{s'_{ik}}{l_{ik}}\right)^{m-1} K(s_n, s'_{ik})\,\mathrm{d}\,s'_{ik}\right]\right\},\ n = 1,2. \quad (6)$$

The total number of unknown coefficients in (6) is $N_u = \sum_{i=1}^{2}\sum_{k=1}^{N_i}(M_k + 1)$ and there are $N_1 + N_2$ conditions for currents having form [3],

$$I_{ik}(s'_{ik} = l_{ik}) = -a_{ik}I'_{ik}(s'_{ik} = l_{ik}),\ k = 1,2,...N_i,\ i = 1,2. \quad (7)$$

The rest of the linear equations is formed matching (6) at the conductors surfaces points defined with

$$s_{nj} = \frac{j}{M_k + 1}l_{ik},\ j = 1,2,...M_k,\ k = 1,2,...N_i,\ i = 1,2. \quad (8)$$

After solving the system given with Eqs. (6) and (7), the current coefficients in (5), i.e. current distribution along the conductors are obtained.

The SIE-Rs (6) with conditions (7) are solved for two regimes of feeding ([12]) so-called: symmetric ($U_1 = U_2 = U_s = 1V$) and anti-symmetric ($U_1 = -U_2 = U_a = 1V$). After determining coefficients in (5) the feeding currents using (4) are obtained, i.e. $I_{gk}^{s/a}$, $k = 1,2$. Labels "*s/a*" denote solution for the currents corresponding to the supplying potential U_s - symmetric, i.e. U_a - anti-symmetric. Using the same procedure as in [12], after determining "*Y*" parameters, corresponding "*Z*" parameters of the LGSs are obtained.

In the general case the known following relations between electrodes voltage and their feeding currents are,

$$U_1 = Z_{11}I_{g1} + Z_{12}I_{g2},$$

$$U_2 = Z_{21}I_{g1} + Z_{22}I_{g2}. \quad (9)$$

In (9), Z_{11} and Z_{22} are self-impedances of the LGS-1, i.e. LGS-2, while $Z_{12} = Z_{21}$ are mutual-impedances between two systems. "*Z*" parameters represent integral grounding system characteristics and measure of mutual influence level between coupled LGSs.

If two LGSs are connected, i.e. formed the unique grounding system, substituting $U_1 = U_2 = U_s(=1V)$ in (9), grounding impedance can be determined as

$$Z_g = R_g + \mathrm{j}X_g = 1/(I_{g1}^s + I_{g2}^s). \quad (10)$$

III. NUMERICAL RESULTS

In order to illustrate describe general model for coupled LGSs system analyze, two linear grounding electrode having length l_1 and l_2 and circular cross section of radius a_1 and a_2 ($N_1 = 1$, $N_2 = 1$), arbitrary placed in homogeneous ground in the plane $y = 0$, are observed. Electrodes are fed over isolated earthing conductors of vertical length in the ground h_1 and h_2 with currents I_{g1} and I_{g2}, respectively. The distance between feeding point is $d = x_{02} - x_{01}$.

In the following text, angles θ_1 and θ_2 are defined as

$$\theta_k = \pm \arccos\left(\hat{s}'_k \hat{z}\right),\ k = 1,2, \qquad (11)$$

where $\hat{s}'_k$ is the ort defining direction of k-th conductor, $k = 1,2$. The sign "-" is used when angle between $\hat{x}$ and $\hat{s}'_k$-ort is less than 90^0.

The convergence of "Z" parameters (real parts values) with increasing polynomial degrees of the approximation (5) are presented in Table I. System geometry parameters are: $h_1 = h_2 = 0.5\,\text{m}$, $l_1 = 1.5\,\text{m}$, $l_2 = 2\,\text{m}$, $a_1 = a_2 = 25\,\text{mm}$, $d = 4\,\text{m}$, $\theta_1 = 30^0$, $\theta_2 = 45^0$, $\sigma_1 = 0.01\text{S/m}$ and $\varepsilon_{r1} = 10$.

TABLE I

REAL PARTS OF "Z" PARAMETERS, $Z..$ IN [Ω], VERSUS POLYNOMIAL DEGREES

$M_1 = M_2$	$\text{Re}\{Z_{11}\}$	$\text{Re}\{Z_{12}\}$	$\text{Re}\{Z_{21}\}$	$\text{Re}\{Z_{22}\}$
1	46.4306	3.4715	3.4445	37.8645
2	45.8434	3.4651	3.4402	37.4187
3	44.5614	3.4592	3.4579	36.4543
4	44.3778	3.4594	3.4569	36.3150
5	43.9354	3.4588	3.4586	35.9826

Real parts of self impedance Z_{11} and mutual impedance Z_{12} versus angle θ_2, with angle θ_1 as parameter, are shown in Fig. 2 and Fig. 3, respectively. The parameters values are: $h_1 = h_2 = 0.5\,\text{m}$, $l_1 = 1.5\,\text{m}$, $l_2 = 2\,\text{m}$, $a_1 = a_2 = 25\,\text{mm}$, $d = 4\,\text{m}$, $\sigma_1 = 0.01\text{S/m}$ and $\varepsilon_{r1} = 10$. The polynomial degrees are $M_1 = M_2 = 3$.

Real parts of self Z_{22} and mutual impedance Z_{21}, versus angle θ_2 are shown in Fig. 4. The parameter values are: $h_1 = h_2 = 0.5\,\text{m}$, $l_1 = 1.5\,\text{m}$, $l_2 = 2\,\text{m}$, $a_1 = a_2 = 25\,\text{mm}$, $d = 2\,\text{m}$, $\theta_1 = 0$, $\sigma_1 = 0.01\text{S/m}$ and $\varepsilon_{r1} = 10$. The polynomial degrees are $M_1 = M_2 = 3$.

Real parts of self Z_{22} and mutual impedance Z_{21} versus distance d are presented in Fig. 5. The parameter values are: $h_1 = h_2 = 0.5\,\text{m}$, $l_1 = 1.5\,\text{m}$, $l_2 = 2\,\text{m}$, $a_1 = a_2 = 25\,\text{mm}$, $\theta_1 = 30^0$, $\theta_2 = 45^0$, $\sigma_1 = 0.01\text{S/m}$ and $\varepsilon_{r1} = 10$. The polynomial degrees are $M_1 = M_2 = 3$.

The real part of the impedance of the unique grounding system formed by two conductors versus angle θ_2 for $M_1 = M_2 = 3$ is shown in Fig. 6. The parameter values are: $h_1 = h_2 = 0.5\,\text{m}$, $l_1 = 1.5\,\text{m}$, $l_2 = 2\,\text{m}$, $a_1 = a_2 = 25\,\text{mm}$, $d = 2\,\text{m}$, $\theta_1 = 0$, $\sigma_1 = 0.01\text{S/m}$ and $\varepsilon_{r1} = 10$.

Finally, ESP distribution at the ground surface in the conductors surrounding for grounding geometry parameters from Fig. 6 is presented in Fig. 7, for $\theta_2 = 45^0$.

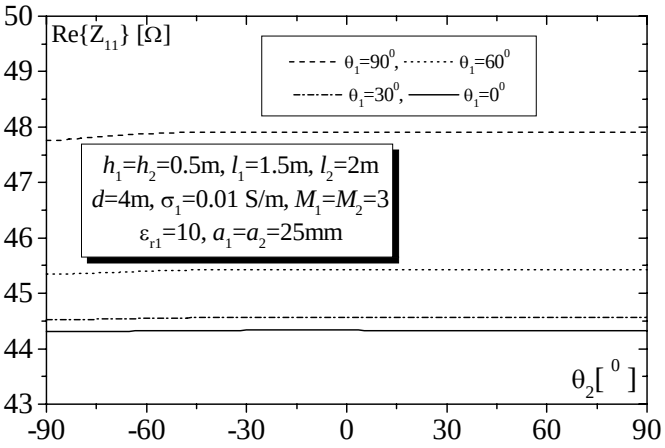


Fig. 2. The real part of self impedance Z_{11} of two coupled ground wire conductors, versus angle θ_2 and angle θ_1 as parameter.

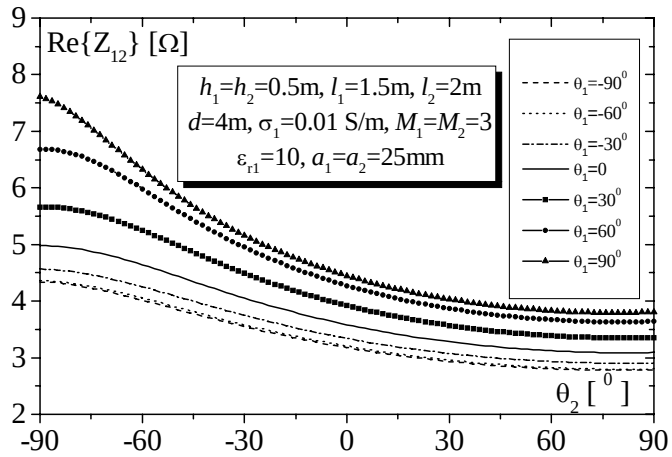


Fig. 3. The real part of mutual impedance Z_{12} of two coupled ground wire conductors, versus angle θ_2 and angle θ_1 as parameter.

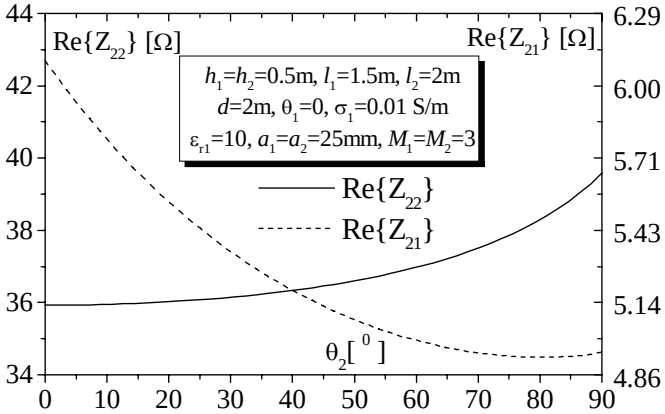


Fig. 4. The real part of self Z_{22} and mutual impedance Z_{21} of two coupled ground wire conductorss versus angle θ_2.

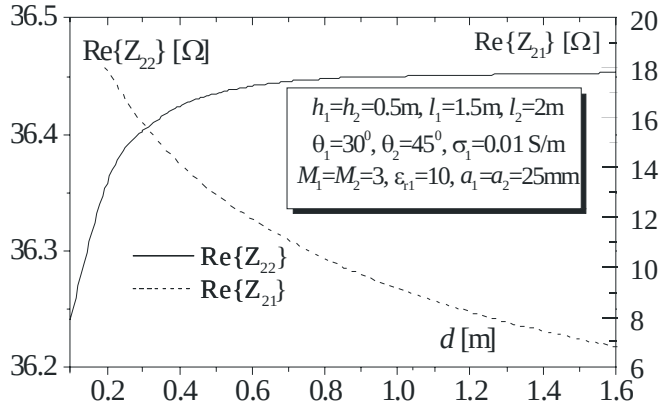


Fig. 5. The real part of self Z_{22} and mutual impedance Z_{21} of two coupled wire ground conductors, versus distance d .

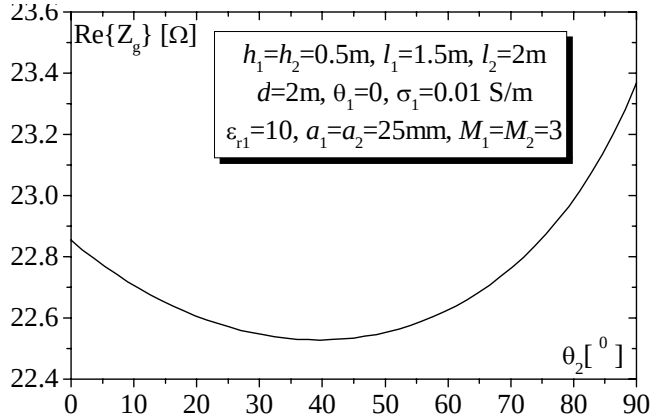


Fig. 6. The real part of grounding impedance of the unique grounding system formed by two conductors versus angle θ_2 .

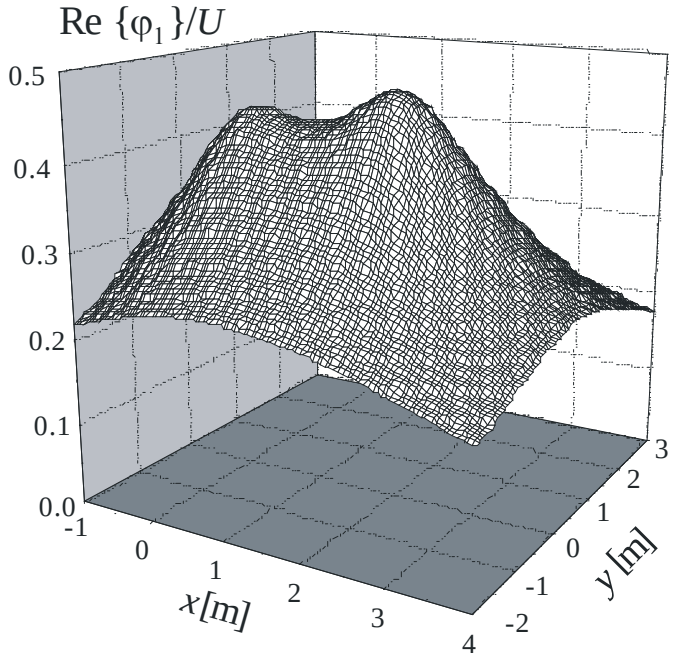


Fig. 7. Potential distribution at the ground surface.

IV. Conclusion

Unknown distribution of longitudinal current along the LGS conductors axis are numerically determined in the paper. It is carried out by numerical solving SIE-R using PMM and polynomial approximation for current distribution.

The described procedure is illustrated applying it on the problem of two coupled linear grounding electrode in symmetric and anti-symmetric regime. For this case, self and mutual impedance (determined from current distribution) are numerically calculated. ESP distribution at the ground surface is also determined.

V. Aknowledgement

This paper is partly financially supported by Ministry of Science and Environmental protection of Republic of Serbia.

References

[1] D. M. Veličković, "Calculation and optimization of earthing grid", The third Symposium of applied electrostatics ΠΕС ‘90, Nis, Yugoslavia, October 23-26, 1990, Proceedings, pp. 70-74 (in Serbian)

[2] D. M. Veličković, N. N. Cvetković, "Grounding on the hill", Engineering of Modern Electric Systems-EMES 97, Oradea, Romania, 30. May-1. Jun 1997. (ANALELE UNIVERSITATII DIN ORADEA, ISSN-1223-2106, pp. 64-69.)

[3] P.D. Rančić, "A new concept for linear grounding system analysis", Fourth Int. Symposium of Applied Electrostatics, ΠΕС ‘96, Nis, Yugoslavia, May 22-24, 1996, Proc. of Papers, pp. 103-116. (Invited paper, in Serbian)

[4] P.D. Rančić, L.V. Stefanović and Đ. R. Đorđević, "An Improved Linear Grounding System Analysis in Two-Layer Earth", IEEE Trans. on Magnetics., Vol. 32, No. 5, Sep. 1995, pp. 5179-5187.

[5] P.D. Rančić, S.Z. Đokić and Z.P. Stajić, "Vertical Ground Rod (VGR) in Inhomogeneous Earth of Sectoral Type", El. Letters, Vol. 30, No. 19, Sep. 1994, pp. 1562-1564.

[6] D.M. Veličković, "Equivalent Electrodes Method", Scientific Review, pp. 207-248, Belgrade, 1996.

[7] D.M. Veličković., "Equivalent Electrodes Method Application for Grounding Problems Solving", Elektrotehnika, 32, Vol. 3-4, May-August 1989, pp. 149-160. (in Serbian)

[8] Nenad N. Cvetkovi}: "Linear grounding system: Quasi-stationary structure of Hertz's vector and electric scalar potential", International PhD Seminar NUMERICAL FIELD COMPUTATION AND OPTIMIZATION IN ELECTRICAL ENGINEERING, Ohrid, Macedonia, September 20-25, 2005, Proc. of Full Papers, pp 55-60.

[9] R.F. Harrington, *Field computation by Moment Methods*, The Macmillan Company, New York, 1969.

[10] B. Popović, "Polynomial Approximation of Current Along Thin Symmetrical Cylindrical Dipole", proc. IEE, Vol. 117, No 5, 1970, pp. 873-878.

[11] P. D. Rančić, "Contribution to linear antennas analysis by new forms of integral equations of two potentials", 10-*th* Conference on the Computation of Electromagnetic Fields, COMPUMAG '95, Berlin, 1995, pp. 328-329,

[12] D. Mitić, *One new proceedure for analyse of arbitrary oriented coupled anntenas*, M.Sc. Thesis, Faculty of Electronic Engineering, University of Nis, Serbia, 1976.

SESSION MS

Measurement Systems

Metrological Support of the Distributed Power Quality Monitoring

Božidar R. Dimitrijević[1] and Milan M. Simić[2]

***Abstract* – This paper describes basic attributes and specific examples of software based solution for distributed measurement and information system for monitoring and analysis of relevant power quality parameters, at distant locations of power delivery network, in accordance with the European quality standards (EN 50160 and IEC 61000). The presented solution of distributed measurement and information system is based on GPRS network communication between computer system in power plant control center and local remote measurement stations. The metrological support block, inside each local measurement station, contains application of advanced methods for measurement and recording of critical voltage quality parameters, with the LabVIEW based virtual instrumentation software support.**

***Keywords* – Distributed measurement and information system, Power quality parameters, Virtual instrumentation**

I. INTRODUCTION

Planning and development of electrical power distribution system is very important and complex task, that includes strong interactions with planning procedures in many other economic and industrial areas. The electrical power supplier's primary goal is to meet the power demands of its customers under all possible conditions. Consistently in order to avoid the high cost of equipment failures, all customers have to make sure that they obtain an electrical power supply of satisfactory quality, and that their electrical devices and equipment are capable of functioning as required, when small level network disturbances occur.

This can only be guaranteed if the limits within which power quality parameters may vary can be specified. These limits can be defined by specific demands of a customer in a power quality contract with supplier, by national regulation documents, or by international power quality standards, such as European Union standard EN 50160 [1,2] and International Electrotechnical Commission electromagnetic compatibility standard IEC 61000 [3]. On the one hand, previous documents define the main characteristics in low and medium voltage networks under normal operating conditions, and on the other hand it gives customers new rights as regards the quality of electrical power supply. For example, according to EN 50160, a customer is entitled to complain, if the voltage quality at the coupling point does not conform to the mentioned standard demands. Therefore in a number of countries, suppliers have adjusted their pricing in accordance with accomplished quality. As a consequence, electrical power quality parameters should be constantly monitored, analyzed and assessed, and these procedures must be performed by means of advanced measuring and recording devices and equipment. Instruments for measurement and analyzing of power quality parameters, such as Fluke 433/434 or Chauvin Arnoux C.A.8332/8334 power analyzers [4], based on complex digital processing functions, are capable to allow continuous voltage quality monitoring and control. It has been designed to measure relevant power quality parameters, and to perform computer based statistical and diagnostic activities, on single phase or three phase low voltage networks, in order to verify its compliance with relevant quality standards. These devices are using as the single units at selected points in the network, or alternatively, several separated units can be combined into a complete distributed data acquisition system, for monitoring and analyzing of an entire electrical power delivery network.

Authors are with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia. E-mail's: dimitrijevic@elfak.ni.ac.yu; milansm@elfak.ni.ac.yu

II. DISTRIBUTED POWER QUALITY MONITORING BASED ON GPRS NETWORK COMMUNICATION

The possibility of fast and reliable data transmission by using GPRS devices, inside the mobile GSM networks, can be used for designing and realisation of distributed measurement and data acquisition systems, typically consisting of a number of remote terminal units, that perform real-time measurements and send this data over a computer network, to system control center, for further proccessing and presentation. Each terminal measurement unit is supplied with single or a number of digital instruments and background computer software. The applied instrument is remotely configurable and controllable, able to control other devices and to temporarily retain acquired data, in order to avoid loss of information in the case that communication breakdown occurs [5].

Permanent monitoring of power system quality, including automated recording and classification of power disturbances, and calculations of relevant statistical parameters, above the rest, act as preventive activity. In addition it allows power suppliers to confirm their system's compliance to relevant quality standards. Electrical power delivery processs to be monitored and controlled is partitioned into geographically distributed cells, that can be dealt with by a single processing unit or a number of locally connected units. Internet based TCP/IP protocols, Ethernet technology or GSM/GPRS based data transfer protocols are using for designing the networking infrastructure, between single terminal units, namely remote measurement stations, and power plant control center [6,7]. General Packet Radio Service is designed as an extension to mobile GSM networks. It greatly improves and simplifies the wireless access to packet data networks, such as the Internet, allowing data rate of up to 100 kbps.

Simplified block configuration of distributed power quality monitoring system, based on data transmitting from remote measurement stations, at different network sites, by using a

GPRS transfer protocols, to delivery control center inside the power plant,. is presented on Figure 1. Presented measurement and information system consists of three functional connected parts; the measurement station at distant location of delivery network, GPRS network communication and electrical power utility control center. Sensors and measurement transducers, inside each local mesurement station, provide informations in relation to real values of voltage and current parameters, and send them to metrological support block based on application of virtual instrumentation software and analog to digital PCI acquisition card. Metrological support software, developed by using graphical programming software package LabVIEW [8], provides measurement data in respect to monitored power quality parameter values, and records it into local date base.

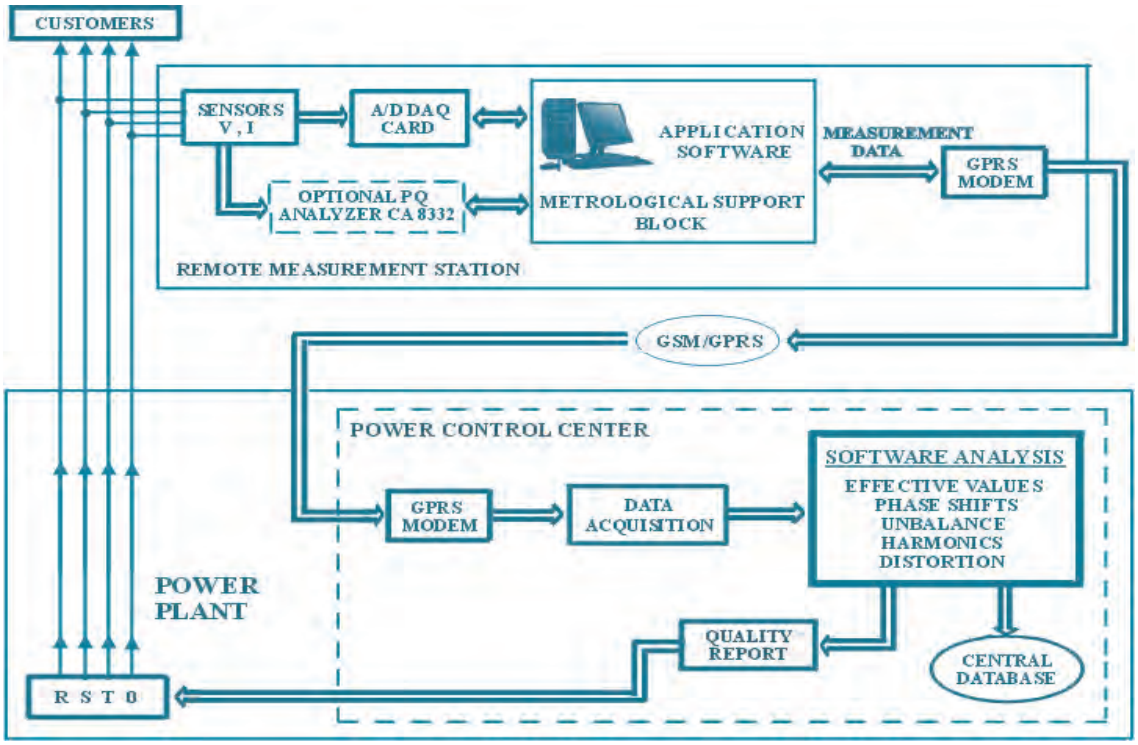


Fig. 1. Distributed power quality monitoring system based on GPRS network communication

An optional solution in local measurement station considers using of one or more power quality analyzers, such as CA 8332/34, including corresponding PC background analyzing software Dataviewer Pro. The second segment of presented monitoring system refers to the sending of measurement data, by serial port RS 232 and GPRS modem through a mobile GSM network, towards the power control center, to second receiving GPRS modem and data acquisition block, for further processing and analyzing. The major activities in the third segment of the system, power plant control center, are data acquisition, analysis of measured effective voltage values, phase shifts, voltage unbalance, level of harmonic distortion, sags and dips, transient voltages, and presentation of final power quality report, including and recording of processed measurement data into central system database.

By comparing detected unwanted events, chronologically recorded in central system database, with existing libraries of typical power quality variations, the main causes for specific network disturbance can be determined and eventually eliminated. Also, recorded data are valuable for characterizing sensitivity of end use equipment to power quality variations, that could be helpful in selecting of proper equipment, and providing specifications for protection that must be installed. Publishing the final quality report and measurement and analyzing results from central database on power utility Internet site, allows reliable and reference data to be viewed by individual power customers on remote home computers.

III. METROLOGICAL SUPPORT BLOCK WITH LABVIEW APPLICATION SOFTWARE

Applications of virtual instrumentation software allow users to implement a device that will best suit to his actual demands and requirements. These graphical-programming applications contain libraries of predesigned graphical components, that enable fast and easy implementation of the user interfaces and instrument control panels. Graphical user interface of the virtual instrumentation is a software replacement for the front panel of a classical type of instruments. It enables interaction between a program user and a software application [7].

A front panel of the virtual instrument, developed by using a graphical programming software package LabVIEW, that monitors power spectrum and three-phase voltage waveforms, performing measurement of relevant electrical power quality parameter values, is presented on Figure 2. Presented virtual instrument is developed within the scope of the project carried out at Measurement department of the Faculty of Electronic Engineering in Nis, with a support of the Ministry of Science and Environmental Protection of Republic of

Serbia. Parallel with monitoring of waveforms and effective values graphs of phase voltages V_1, V_2 and V_3, this system provides power spectrum and measurement data, regarding effective voltage values, phase shifts and total harmonic distortion level, for each of three displayed signals, showing exact date and time of recording for six successive measured values of controlled power quality parameters. Combining with additional data acquisition hardware components described virtual instrument is using as a part of the metrological support block, inside the local mesurement station, for monitoring and measurement of critical quality parameters, and for recording measurement information into local datebase. Block diagram or LabVIEW programming sequence, corresponding to shown front panel of developed virtual instrument, is presented on Fig. 3 below. A substantial number of data obtained from measurement of different power quality parameters, creates the need to analyze and evaluate practical usability of these data, with reference to measurement uncertainty of the final results. Actually, it is necessary to determine and specify the uncertainty level that is associated to the measurement results, in accordance with the recommendations of paper "Guidelines for evaluating and expressing the uncertainty of NIST measurement results" [9].

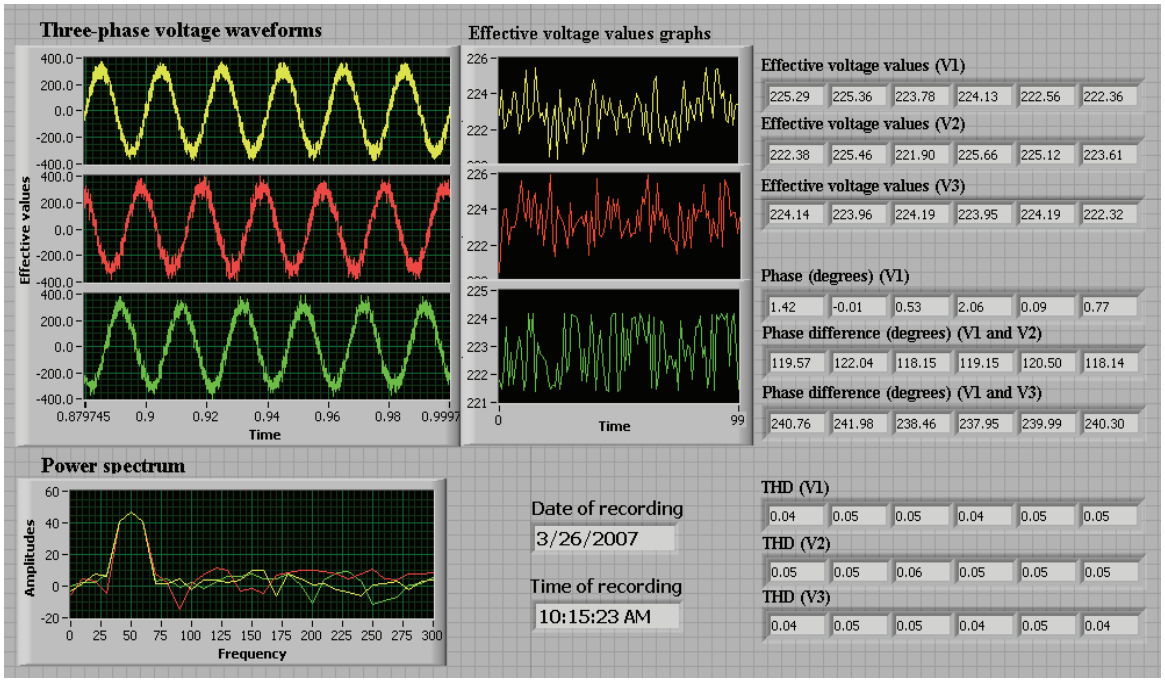


Fig. 2. Front panel of the virtual instrument developed in a graphical programming language LabVIEW

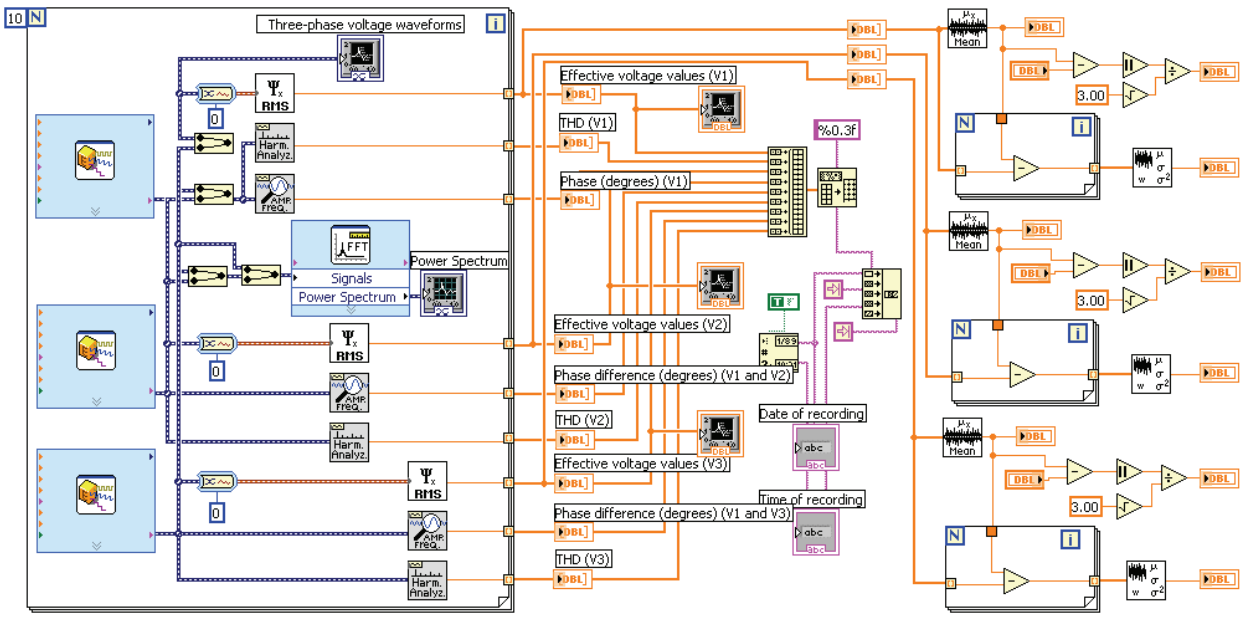


Fig. 3. LabVIEW program sequence (block diagram) of developed virtual instrument

The information concerning uncertainty level are providing by means of either statistical methods applied to experimental measurement results (uncetainty type A), or by using accuracy specifications for devices included in the system, provided by manufacturer (uncetainty type B). The automatic procedure for both types of measurement uncertainty calculation is developed and associated to the virtual instrument for metrological support, presented on Figure 2 above. A part of LabVIEW virtual instrumentation program for type A and type B measurement uncertainty calculation, concerning effective values of three-phase voltages V_1, V_2 and V_3, measured and presented on Figure 2, is presented on Figure 4.

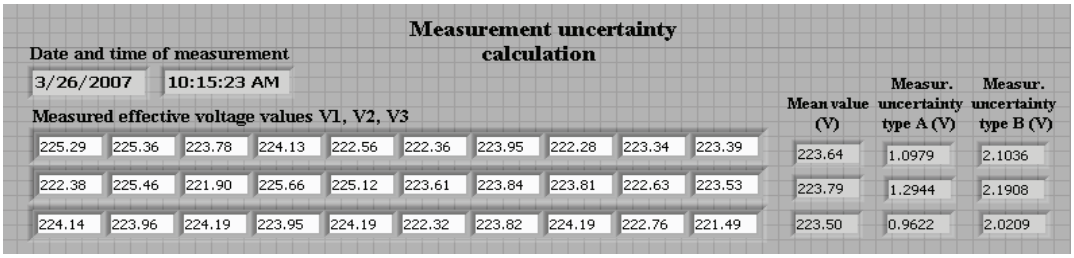


Fig. 4. LabVIEW virtual instrument for automatic type A and type B measurement uncertainty calculation

A type A evaluation of standard uncertainty is based on calculation of standard deviation of the mean of a series of independent observations. An input quantity is V_i whose mean value is estimated from n independent observations $V_{i,k}$, in this case, ten measured effective voltage values, obtained under the identical measurement conditions. The estimated value V_i is usually the sample mean:

$$\overline{V_i} = \frac{1}{n}\sum_{k=1}^{n} V_{i,k}\,, \tag{1}$$

and the type A standard uncertainty to be associated with V_i is:

$$u(V_i) = \sqrt{\frac{1}{n(n-1)}\sum_{k=1}^{n}\left(V_{i,k} - \overline{V_i}\right)^2}\,. \tag{2}$$

Calcualtion of the type B standard uncertainty considers an input quantity V_i whose value is estimated from an assumed rectangular probability distribution of the lower limit V_{min} and the upper limit V_{max}. In this case, the estimated value V_i is the expectation of the distribution:

$$\overline{V_i} = \frac{V_{max} + V_{min}}{2}\,, \tag{3}$$

and the type B standard uncertainty to be associated with V_i is the positive square root of the variance of the distribution:

$$u(V_i) = \frac{V_{max} - V_{min}}{\sqrt{3}}\,. \tag{4}$$

IV. Conclusion

According to configuration of the distributed power quality monitoring system presented in this paper, based on GPRS communication between remote measurement stations and power plant control center, the key components for providing satisfactory measurement results uncertainty level, are placed in metrological support block inside each local measurement station, that is functionally based on graphical programming application software. Presented virtual instrument, developed by using LabVIEW programming environment, continuously monitors waveforms, effective values graphs and signal power
spectrum of three-phase voltages, including measurement and recording of effective values, phase shifts and total harmonic distortion level, for each of three displayed signals, indicating precise date and time of measurement. Finally, is presented an automatic LabVIEW based procedure for type A and type B uncertainty calculation of measured effective voltage values, functionally implemented into described virtual instrument. Chronological recording of previously processed and analyzed measurement quality parameter values in power plant central database and publishing these recorded data over the Internet, allows correct and reliable measurement data to be viewed by relevant individual customers on local home computers.

This paper is result of research activities within the project "Applying of modern methods and devices for monitoring and analyzing of electrical power quality according to relevant EU standards", no. 223010, supported by the Ministry of Science and Environmental Protection of Republic of Serbia, within the National program of Energy efficiency.

References

[1] Standard BS EN 50160:2000 Voltage characteristics of electricity supplied by public distribution systems, Reo, UK (www.reo.co.uk).
[2] EN 50160 Power quality standard, Power quality access meters and EN50160, Siemens, May, 2003.
[3] IET Guidance document on Electromagnetic compatibility and Functional safety, Power quality (Annex B1), The Institution of engineering and technology, (www.theiet.org/factfiles).
[4] C.A. 8332/34 Power quality analyzers specifications, Chauvin Arnoux Group (www.chauvin-arnoux.com).
[5] W. Nawrocki, *Measurement Systems and Sensors, Chapter 7, Wireless Measurement Systems*, ARTECH HOUSE INC, 2005.
[6] L. Cristaldi, A. Ferrero and S. Salicone, "A Distributed System for Electric Power Measurement", IEEE Transactions on Instrumentation and Measurement, Volume 51, Number 4, 2002, pp. 776-781.
[7] B. Dimitrijević and M. Simić, "Virtual Instrumentation Applied to Electrical Power Quality Measurement", ICEST 2006, Conference proceedings, pp. 395-398, 29 June – 1 July 2006, Sofia, Bulgaria.
[8] LabVIEW User manual, National Corporation, USA, 2005 (www.ni.com).
[9] Guidelines for evaluating and expressing the uncertainty of NIST measurement results, NIST Technical Note 1297, 1994. (www.nist.gov).

Power Resistors With Low Parasitic Inductance

Pavlik Rahnev[1] and Silvija Letskovska[1]

Abstract – In this work three types of power non inductive resistors are described. The design and technology for every one is shown. The subject of the experiments are bulk B_4C, thin film and resistors from parallel connected wires. The comparison of main parameters such as resistance (R_n), TCR, accuracy and prices is done.

Key words – **power resistors, non inductive resistors, sintered resistors, low ohm resistors.**

I. INTRODUCTION

The need of high power (high amperage) resistors arrases simultaneously, especially for AC applications. In conjunction with the base resistor parameters $R_{nominal}$, TCR and tolerance $(\varepsilon\%)$, now the parasitic capacitance and inductance are very important. There are a lot of producers in wide range of resistors from parts of ohms to kohms and power up to $1000\,W$ [1÷5].

Almost all technologies and construction are presented such as bulk (sintered), wired, film resistors, grid and ribbon est.

The goal of this work is to describe three possible type of power resistors with low parasitic inductance.

This includes sintered (B_4C) resistors, thin Nichrom and special designed resistors from Cantal or Nichrom wire. The experiments are provided to produce small numbers, accurate and in wide range of parameters resistors. They are used for AC power application in laboratory conditions.

II. DESIGN AND TECHNOLOGY OF EXPERIMENTAL SAMPLES

II.1. Resistor from boron carbide (B_4C).

The boron carbide is well known with its hardness, chemical resistance and electrical parameters: $\rho = 0.1 \div 10\ \Omega cm$ and $TCR \approx -0.5\ \%/^0C$.

The TCR is strongly negative but with addition of Co, Mo, it is closed to $0.01\ \%/^0C$. The preparation of these resistors include the next steps:

- mixing the powder of boron carbide (with $1 \div 5$ % Co or Mo);
- hot pressing (sintering) - $T \approx 2200\ ^0C$;
- ohm contacts made with Ag paste (low $T = 100\ ^0C$) or $AgPd$ ($T=850\ ^0C$).

The shape of thee resistors is as a nozzle (Fig. 1).

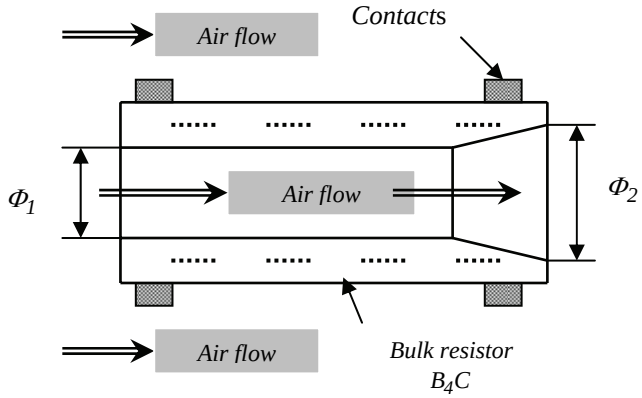


Fig. 1. The shape of the samples.

This design allows well air cooling and measuring the temperature in the centre of the resistors.

The main parameters of boron carbide (B_4C) resistors are:

- nominal value $R_n = 1 \div 10\ \Omega$);
- $TCR = 0.5 \div 10.01\ \%/^0C$;
- production tolerance $\varepsilon = \pm 10\ \%$.

II.2. The film resistors.

Next, very used technology for power low ohm resistors is $NiCr$ thin resistive films on insulating substrate (Fig. 2).

The technology steps are well know as:

- deposition of film combination $NiCr - Ni - Ag$;
- photolithography for pattering resistor $NiCr$ and contacts Au;
- making external connections and attachment to the heat sink;
- trimming and surface protection.

For this type of resistors is important the very accurate resistance could be obtained using electro spark trimming (L - cut) (Fig. 3). With this technique the accuracy of $\pm 0.1\%$ could be done and it depends on method of measurements.

[1]Pavlik Rahnev is with of Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, rahnev@bfu.bg

[1]Silviya Letskovsta is with of Burgas Free University, San Stefano 62, 8000 Burgas, Bulgaria, silvia@bfu.bg

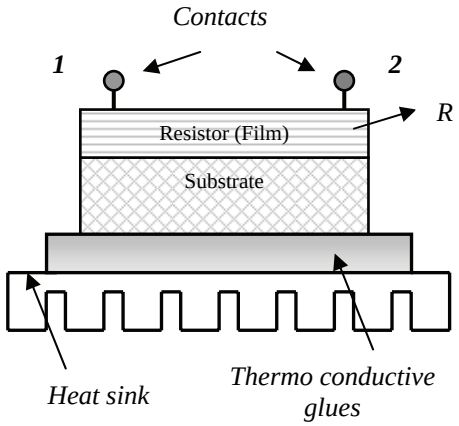


Fig. 2. Thin film $(NiCr)$ resistor.

These thin film power resistors have $R_{nominal}$ from 10 to $100\ ohms$. The disadvantage is very high parasitic capacitance $(10 \div 20\ pF/cm^2)$ depending to the dielectric substrates (Sital, Al_2O_3) and surface.

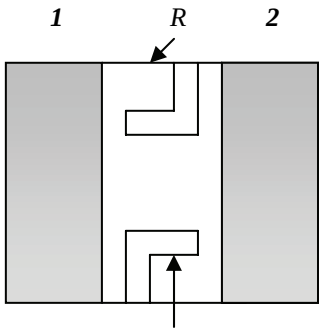


Fig.3. Trimming with L - cut.

II.3. Resistors based on resistive wires.

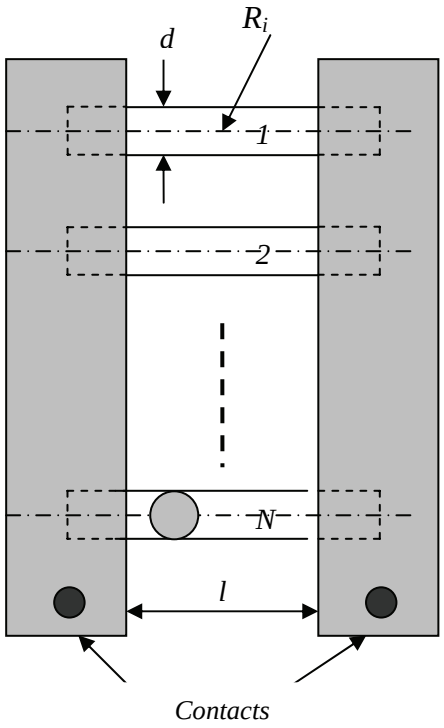
The construction of these resistors is shown on Fig. 4.

If consist a number of resistive wires connected in parallel between $Cu\,(Cu + Ag)$ external contacts. As a resistive material the wires from Nichrom or Cantal is used with different diameters. From the Fig. 4 it could be seen that they are several possibility for reaching dose tolerance:

- resistivity of ρ_W;
- length material of wires L;
- diameter of wires d;
- number of wire N.

For example, of the need is power shunt resistor $R = 0.006\ \Omega$, the design could be calculated as fallows:

- $\rho_{WIRE} = 20 \div 30 Mohm/cm^2$;
- $L = 1\ cm$;
- $R_{nominal} = 0.006\ \Omega$;
- $R_n = \rho_W / N$;
- $N = (20 \div 30)/0.006 \approx 4 \div 5$.



N – number of wires

Fig. 4. Low ohm, low inductive resistors.

These type of resistors are really low ohm with low parasitic capacitance and inductance, low TCR, high accuracy and without special protection.

CONCLUSION

The described three type of power resistors have advantage and disadvantages on the base of construction, technology and materials. The best parameters are:

- boron carbide (B_4C). – high stability (temperature and chemical);
- thin film $(NiCr)$ - best accuracy (tolerance $\pm 0.1\%$);
- Cantal and Nichrom parallel wires – easiest technology and low price.

For laboratory use (small numbers - $1 \div 10\ psc$) the last type of resistors is a good solution.

REFERENCES

[1] www. rftechniques.com
[2] www. str-restech.de
[3] www.powerohm.com
[4] www. heiresistors.com
[5] www.digikey.com

PC Based RLC Meter

Mare Srbinovska[1], Vladimir Dimcev[1], Cvetan Gavrovski[1] and Zivko Kokolanski[1]

Abstract –This paper presents method that convert resistance, capacitance or inductance values into corresponding time intervals - frequency. Virtual RLC meter, based on a PC with a DAQ and external relaxation oscillator is used for measurements. The fair accuracy of the Virtual RLC meter is verified by experimental results for resistance and capacitance measurements.

Keywords – Virtual RLC meter, DAQ, relaxation oscillator

I. INTRODUCTION

Most of the methods and techniques for measuring resistance, capacitance and inductance are very complicated. RLC meters are specialized laboratory equipment, dedicated in measuring the component's resistance, inductance and capacitance, and are typically expensive. In recent years is used one popular method that converts resistance, capacitance or inductance into corresponding voltage, time intervals or frequency. Two basic techniques for R, L and C measurement are called "Step response" and "Frequency domain" methods. Both methods are based on a resistor – capacitor circuit and a resistor – inductor circuit.

By using a PC with a DAQ board, and developing an interface with relatively reduced complexity, we were able to build a medium performance RLC meter at an exceptionally reduced cost.

This paper describes how we developed a Virtual RLC meter, based on a PC with a DAQ board (National Instruments PCI 6221) and an external relaxation oscillator, using Lab VIEW as the development software, and our results and conclusions are presented.

II. STEP RESPONSE

The RC circuit and the RL circuit have some very similar properties that can be used to determine the value of either the inductor or capacitor in the circuit. If the input (Ui) of the circuit is abruptly stepped from 0 volts to some positive value, the output voltage (U0) rises in an exponential manner (Fig. 1).

[1]Authors are with the Faculty of Electrical Engineering and Information Technology, Karpos 2 bb, 1000 Skopje, Macedonia, E-mail: mares@etf.ukim.edu.mk, vladim@etf.ukim.edu.mk

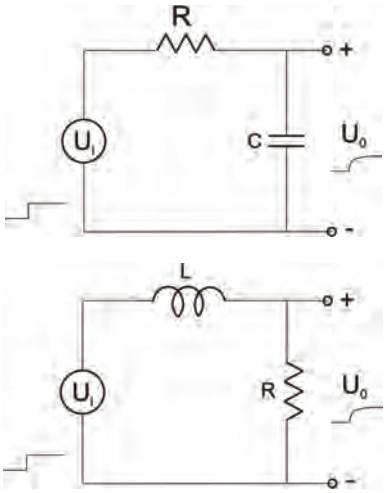


Fig. 1. RC and RL circuits

These circuits can be used to determine C and L values by measuring the time constant of the circuit.

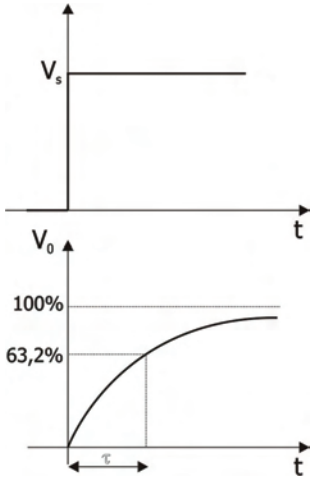


Fig. 2. The step response input (U_i) and output (U_0) waveforms for the RC and RL circuits shown in Fig. 1.

The output voltage that results from a voltage step at the input is called the step response. The mathematical expression for the step response of the circuit shown is

$$U_0(t) = U_i(1 - \ell^{-t/\tau}),\ t \geq 0 \quad (1)$$

In one time constant, the step response reaches 63,2% of its final value.

$$U_0(\tau) = U_i(1 - \ell^{-1}) = 0{,}632 \cdot U_i \quad (2)$$

The time constant can be computed from the circuit values.

$\tau = RC$ for the RC circuit

$\tau = L / R$ for the RL circuit

An unknown capacitor or inductor whose value is to be determined can be connected to a known resistor in the appropriate circuit. The time constant of the circuit can be measured and the unknown component value computed.

$C = \tau / R$ for the RC circuit

$L = \tau \cdot R$ for the RL circuit

III. SINUSOIDAL RESPONSE

These same circuits, with voltage sine wave as excitation, can be used in the frequency domain to measure the value of a capacitor or inductor (Fig. 3). The frequency response of these two circuits is the same.

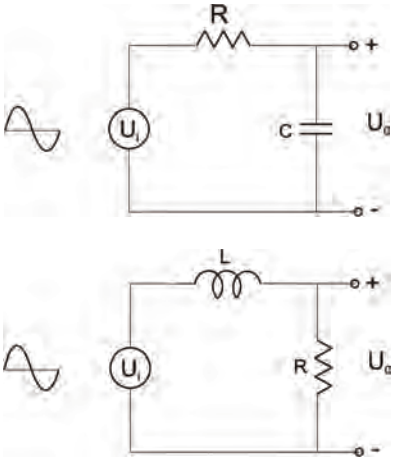


Fig. 3. The frequency response of the RC and RL circuits can be used to measure a capacitor or inductor

Low frequency signal are passed from the input to the output with a little or no attenuation, while high- frequency signals are attenuated significantly, resulting in a low – pass filter. The point at which the response is under 3dB is used to define the filter bandwidth. A loss of 3dB corresponds to a reduction in output voltage to 70,7% of the original value (Fig.4).

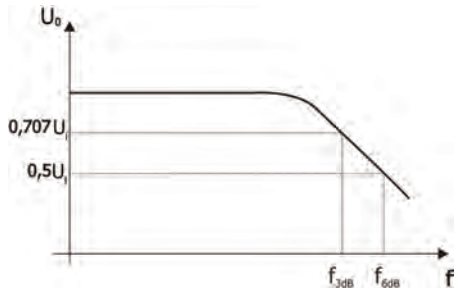


Fig. 4. The frequency response of the RC and RL circuits

$$f_{3dB} = \frac{1}{2\pi\tau} \qquad (3)$$

Where τ is the time constant of the circuit.
A 6-dB reduction in voltage corresponds to a 50% reduction in voltage.

$$f_{6dB} = \frac{\sqrt{3}}{2\pi\tau} \qquad (4)$$

The 3dB or 6dB frequency can usually be determined from the frequency control of the source, or for more accuracy, a frequency counter can be used to measure the source frequency. After the f_{3dB} or f_{6dB} frequency is measured, the value of the capacitance or inductance is calculated.

IV. RELAXATION OSCILLATOR

A relaxation oscillator is an oscillator in which a capacitor is charged gradually and then discharged rapidly. Relaxation oscillator is a circuit that repeatedly alternates between two states with a period that depends on the charging of a capacitor. The capacitor voltage may change exponentially when charged or discharged through a resistor from a constant voltage. With exponential charging, the timing is expressed in terms of time constants RC. In a basic circuit, the capacitor will charge with time constant τ_1 from a starting voltage U_1 towards a supply voltage U_u for a time t_1. When it reaches some voltage U_2, it will then discharge with time constant τ_2 towards a lower voltage U_d for a time t_2, when the cycle will repeat. The period of the oscillator is $T = t_1 + t_2$, and its frequency $f = 1/T$. In terms of these parameters,

$$t_1 = \tau_1 \ln\left[(U_2 - U_1)/(U_u - U_2)\right] \qquad (5)$$

$$t_2 = \tau_2 \ln\left[(U_2 - U_d)/(U_1 - U_d)\right]. \qquad (6)$$

A classic op-amp relaxation oscillator is shown on Fig. 5. The noninverting input is biased at $\lambda \cdot U_0$, where U_0 is the output voltage. The inverting input chases this value and when it reaches it, the output voltage changes from one saturation level to the opposite one from $+Vcc$ to $-Vcc$.

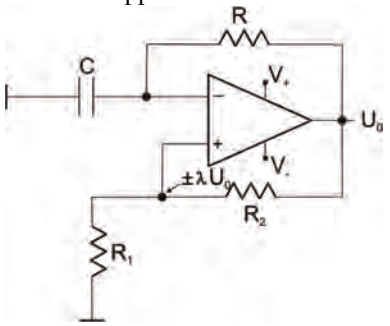


Fig. 5. Relaxation oscillator

This square wave generator is like the Schmitt trigger circuit where the reference voltage for the comparator action depends on the output voltage.

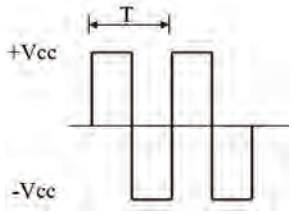


Fig. 6. Square waveform

$$T = 2RC\ln\left[\frac{1+\lambda}{1-\lambda}\right],\ \lambda = \frac{R_1}{R_1+R_2}$$

$$\text{If } R_1 = R_2 \Rightarrow \lambda = 1/2$$

$$f_{out} = \frac{1}{kRC} \qquad (7)$$

The output voltage is a square wave (Fig.6), with a frequency of f_{out} =1/1.22RC, and duty cycle of 50% if the saturation levels are symmetrical.
Using expressions for R-L circuit it is possible to achieve that if R is replaced by L and C by R the output frequency will be:

$$f_{out} = \frac{R}{kL} \qquad (8)$$

For measurement of R, C or L the unknown component is connected in an R-C or R-L circuit as appropriate. For example for measurement of C, the resistor is known or can be accurately measured and vise versa.

V. HARDWARE AND SOFTWARE REALIZATION

The designed virtual RLC meter consists of the following blocks: the relaxation oscillator, switching block, data acquisition system, personal computer and Labview software.

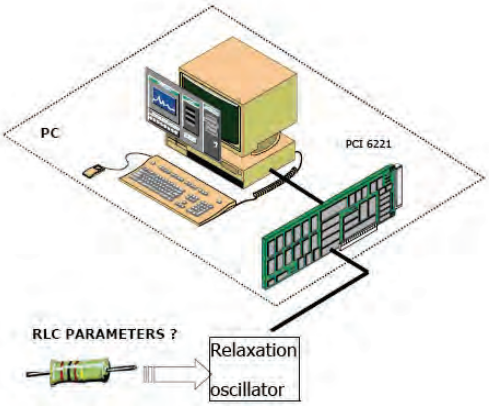


Fig. 7. The block diagram of virtual RLC meter

The designed virtual RLC meter has tree operation modes for R, L and C measurement respectively. The oscillator circuit consists of a comparator with a positive feedback loop, which converts the time – varying signal into a square wave with a varying time period. This signal can be counted and digitized into a value that is proportional to the oscillator frequency and in turn proportional to R, C or L measured.

The software of the RLC meter was entirely developed using Lab VIEW. Programming an application in Lab VIEW is very different from programming in a text based language such as C or Basic. Lab VIEW uses graphical symbols (icons) to describe programming actions. Data flow is "wire" into a *block diagram*. This software provides an extensive library of functions and tools for data analysis, report generation, data acquisition and file input/output.

The choice of this development software was mainly due to its easy-of-use in communicating with the DAQ board for data generation and acquisition, and the extreme simplification in the processing and visualization of the data.

The created programming code made in Lab VIEW is shown on the Fig. 8.

For frequency measurement is used one counter on a Counter Input Channel. The edge, minimum value and maximum value are all configurable (Fig.9).

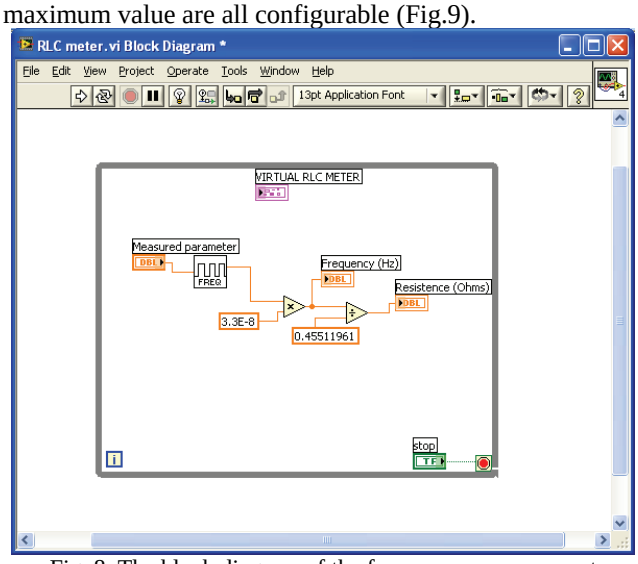


Fig. 8. The block diagram of the frequency measurement

The measurement edge is used to specify the edge on which the counter has started measuring.

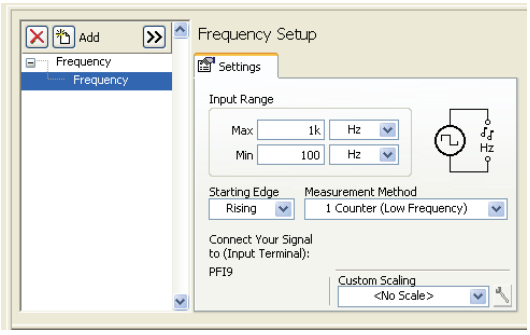


Fig. 9. Frequency setup

The starting edge parameter is used to determine if the counter will begin measuring on a rising or falling edge. It is important to set the maximum and minimum frequency values as accurately as possible so the best internal timebase can be chosen to minimize measurement error. The default values specify a range that can be measured by the counter using the 20 MHz Timebase.

The user interface of the created virtual system is shown on Fig. 10.

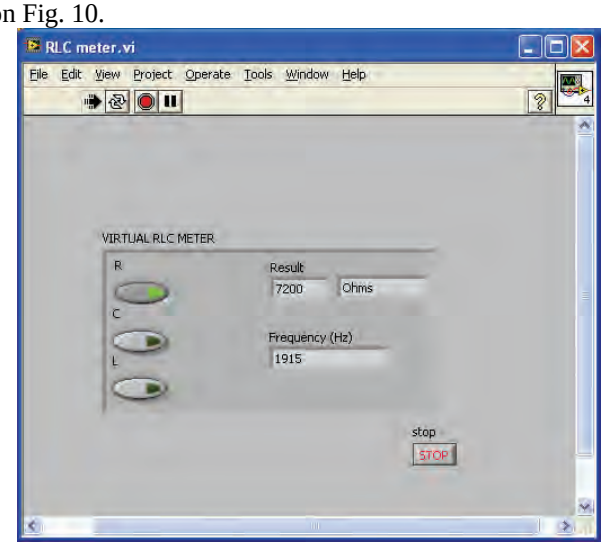


Fig. 10. The block diagram of the frequency measurement

The front panel consists of basic parts. The first one is placed in the left corner and is related to the selection of measurement parameter. In the example shown above, the button that is switch on means that the measurement parameter is resistance. The numeric indicators show the measured value and the current frequency.

VI. EXPERIMENTAL RESULTS

In order to define the measurement accuracy of the realized virtual system a set of arbitrary selected resistors and capacitors are measured. The measured results are summarized into Table I (for resistance), Table II (for capacitance).

	$R_{ref}(\Omega)$	$R_{cal}(\Omega)$	$C_{ref}(nF)$	$(\delta_1\%)$	$(\delta_2\%)$
1	1M	1,23M	0,133	22,87	12
2	47k	40k	10	0,6	-14,8
3	33k	35,3k	10	1	2,7
4	22k	20,7k	10	1,8	-6
5	10k	10,4k	10	4	-2
6	4,7k	5,36k	10	14,28	10
7	2,2k	2,45k	100	4,54	9
8	1k	1,084	100	12,5	8,4
9	680	765	100	17,6	12,5
10	470	604,2	100	28,78	15

Table I Resistor measurement results

	$C_{ref}(F)$	$C_{cal}(F)$	$R_{ref}(k\Omega)$	$(\delta_1\%)$	$(\delta_2\%)$
1	680p	660p	105	10	-11,7
2	1n	1,067n	105	6,7	-5
3	3,3n	3,35n	105	1,5	-4
4	6,8n	7,1n	10	4,4	1,5
5	10n	10,3n	10	3	-2
6	22n	22,7n	10	3,2	-4,5
7	33n	33,9n	10	2,7	-6
8	100n	112n	1	12	8
9	470n	530n	1	12,7	-1
10	1μ	$1,3\mu$	0,1	30	-8

Table II Capacitor measurement results

The calculated errors are shown in the Tables I and II. The first error δ_1 is calculated using the operational amplifier TL, while δ_2 is calculated by using higher performance comparator OP27. As it can be seen errors can be lower using OP27.

VII. CONCLUSION

In this paper, a design and development of virtual RLC meter is presented using a low-cost general-propose DAQ board from National Instruments (PCI 6221, 16 bit, 250kSamples/s) and simple external hardware.

This approach offers a simple solution for an application that needs the low cost measurement of resistors, capacitors and inductors. The realized virtual system can be used to any applications where frequency-time parameters have to be measured with highest resolution and programmable accuracy.

Using a relaxation oscillator with higher performance comparator would significantly improve the performance and precision of this RLC meter.

REFERENCES

[1] Georgi Todorov Nikolov, Ivan Iliev Stoyanov, Boyanka Marinova Nikolova, "RLC Measurement using Virtual Counting Tecniques", Electronics' 06 Conference Proceedings, book 4, 69-74, Sozopol, Bulgaria, 2006.

[2] Writte R. A. Electronic Test Instruments, Prentice Hall PTR, 2003.

[3] National Instruments, DAQ – STC Technical Reference Mannual, Part Number 373427A-01, August 2005.

[4] Henrique Silva, Habil Silva and Jorge Martins, PC- Based RLC Meter , DataFlow, Setubal, Portugal

[5] Sensor Platform Inc. SSP1492 Sensor Signal Processor Chip, Product Datasheet, V2.0

Method for Calculating Residual Strains and Visual Diagnostics by a Specialized Software System

Nikolay T. Kostov[1], Slava M. Yordanova[2] and Ginka K. Marinova[3]

Abstract – **Abstract - In this article the authors suggest specialized software system, which revises roentgenografic method for calculation of residual strains in metals with poly-crystal grating, by applying a roentgenostructural analysis on a complex strained condition. The character of this strained condition has a significant influence on the conduct of the metal article in the conditions of exploitation. The definition of the macro tensions is brought to a precise definition of the periods of the crystal grating and recalculating the received values deformations and tensions.**

Keyword – **specialized software system, method for calculation.**

I. INTRODUCTION

The roentgenographical method for definition of the macro tensions is based on the fact that by the presence of a macro-strained condition the intra-surface distances are changing, which has effect on the situation of the diffraction lines. Consequently The definition of the macro tensions is brought to a precise definition of the periods of the crystal grating and recalculating the received values deformations and tensions.[1,2,3]

II. EXPERIMENTAL PART

The deformation of the crystal grating can be expressed through the intra-surface distance:

$$\varepsilon = (d - d_0)/d_0 \quad (1)$$

Hook`s law,

where: d – intra-surface distance of a material with tension; d_0 – intra-surface distance of a material in non-strained condition.

Dependencies by a different strained condition:

$$\Delta\theta = \Delta d/d \cdot tg(\theta_0) \quad (2)$$

Equation of Woolf-Brek;

$$\sigma = E/\mu \cdot \cot g(\theta_0) \cdot \Delta\theta; \quad (3)$$

μ – Plank constant;

Δθ – changing of the angle of the reflected roentgen wave;

θ_0 – base angle.

From the equation (3) it is clear, that in order to define the tensions by a roentgenographical method we have to find θ_0 и Δθ.

For carrying out this experiment it is needed the space angles, between the roentgen ray and the surface of the model to be defined:

φ – placed in a plain, parallel of the surface of the model and connects the main strain σ_1 and projection of the roentgen ray on the surface of the model.

Ψ – This is an angle between σ_3 and the direction of the roentgen ray.For definition of the tensor of the deformations and the strains man has to do the following:

- Roentgenograms by azimuthally angles φ = 0° (in the direction of the grinding) and several values for ψ (0°> ψ>90°) and φ = 45° и 90° are being recorded;
- The angles θ(φψ) are being defined and calculated d(φψ);
- Graphical dependencies $d\theta_{\varphi\psi}$) - sin2ψ by φ = 0°, 45° and 90°, which are straight functions, are being build.

The components ε_{ij} of the deformations are defined through the tangencies of the inclination according to the method Dolle-Hauk.

$$\begin{aligned} tg(1) &= d_0(\varepsilon_{11} - \varepsilon_{33}) \\ tg(2) &= d_0(1/2\varepsilon_{11} + 1/2\varepsilon_{22} - \varepsilon_{33} + \varepsilon_{12}) \\ tg(3) &= d_0(\varepsilon_{22} - \varepsilon_{33}) \\ tg(4) &= d_0\varepsilon_{33} \\ tg(5) &= d_0 \frac{1}{\sqrt{2}}(\varepsilon_{13} - \varepsilon_{23}) \\ tg(6) &= d_0\varepsilon_{23} \end{aligned} \quad (4)$$

The value of a_1,φ = 0 does not depend on φ and it is equal to $d_0*\varepsilon_{33}$. Through ε_ij we can get the tensor of the deformations:

[1]Nikolay T. Kostov is with the Technical University of Varna, Department of Radio engineering, Studentska Street 1, Varna 9010, Bulgaria, E-mail: n_kostov@mail.bg

[2]Slava M. Yordanova is with the Technical University of Varna, Department of Computer Science and Technologies, Studentska Street 1, Varna 9010, Bulgaria, E-mail:slava_yordanova@mail.bg

[3]Ginka K. Marinova is with the Technical University of Varna, Department of Computer Science and Technologies, Studentska Street 1, Varna 9010, Bulgaria, E-mail: gin_kaleva@abv.bg

$$\begin{matrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{matrix} \quad (5)$$

and by the general Hook`s Law the components σij of the tensor of the strain can be calculated:

$$\sigma_{ij} = 2/(S_2)\left[\varepsilon_{ij} - \delta_{ij} S_1(\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33})/(S_2/(2)+3S_1)\right]$$

$$\begin{matrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{matrix} \quad (6)$$

If the components σ_{13} have values above the mistake of the experiment, this means that the surface of the main strains σ_1, σ_2 is inclined towards the surface of the model in the following angle:

$$\Delta\Psi = \frac{1}{2} arctg\left(-\frac{2\sigma_{13}}{\sigma_{13}-\sigma_{31}}\right) \quad (7)$$

Calculation of the main strains σ_1, σ_2, σ_3 in this case is done by the following equations:

$$\begin{aligned} \sigma_1 &= (\sigma_{11}+\sigma_{33}) + r_\sigma \\ \sigma_2 &= \sigma_{22} \\ \sigma_3 &= (\sigma_{11}+\sigma_{33}) - r_\sigma \end{aligned} \quad (8)$$

where:

$$r_\sigma = \sqrt{\frac{2(\sigma_{33}-\sigma_{11})}{4+\sigma_{13}}}$$
$$tg(7) = d_0(\varepsilon_{11}-\varepsilon_{33})$$
$$tg(8) = d_0(\varepsilon_{22}-\varepsilon_{33})$$

From here

$$\begin{aligned} \varepsilon_{11} &= tg(7)/d_0 + \varepsilon_{33} \\ \varepsilon_{22} &= tg(8)/d_0 + \varepsilon_{33} \\ \varepsilon_{33} &= (d_\varphi = 0,\ \Psi = 0 - d_0)/d_0 \end{aligned} \quad (9)$$

$$\begin{matrix} \varepsilon_{11} & 0 & 0 \\ 0 & \varepsilon_{22} & 0 \\ 0 & 0 & \varepsilon_{33} \end{matrix}$$

The tensor of the strain is being calculated by the formula:

$$\sigma_{ii} = \frac{2}{S_2}\left[\varepsilon_{ii} - \frac{S_1(\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33})}{\frac{S_2}{2}+3S_1}\right] \quad (10)$$

Using the center of the weight we can receive values of the diffraction angles.

$$a_{\varphi\psi} = (d_{\varphi\psi} > 0 + d_{\varphi\psi} < 0)/2$$
$$a_{\varphi\psi} = (d_{\varphi\psi} > 0 - d_{\varphi\psi} < 0)/2$$

With the already calculated values for a_1 and a_2, for all the combinations φ(ψ) the grafics for every φ are beeing build:

$$\begin{aligned} a_1\varphi &= 0(\sin 2\psi) \\ a_1\varphi &= 45(\sin 2\psi) \\ a_1\varphi &= 90(\sin 2\psi) \\ a_2\varphi &= 0(\sin|2\psi|) \\ a_2\varphi &= 45(\sin|2\psi|) \\ a_2\varphi &= 90(\sin|2\psi|) \end{aligned} \quad (11)$$

To define the exact function of the separate graphics the method of the least squares is being used. [1, 2]

$$f(x) = a + b\cdot x \quad (12)$$

– equation of interpolation function

$$aM + b\Sigma_1 Mx_i = \Sigma_1 My_i$$
$$a\Sigma_1 Mx_i + b\Sigma_1 Mx_{i2} = \Sigma_1 Mx_i y_i$$

where:

(x_i, y_i) – coordinates of the experimental points;

M – number of points.

$$\begin{aligned} a &= \frac{\Sigma_1 My_i \cdot \Sigma_1 Mx_{i2} \cdot \Sigma_1 Mx_i y_i}{M\Sigma_1 Mx_{i2} - 2(\Sigma_1 Mx_i)} \\ b &= \frac{M\Sigma_1 Mx_i y_i - \Sigma_1 Mx_i \cdot \Sigma_1 My_i}{M\Sigma_1 Mx_{i2} - 2(\Sigma_1 Mx_i)} \\ b &= tg\alpha \end{aligned} \quad (13)$$

where: α – slope angle of the straight line.

Into the methods, which are used it is necessary the angle factors of all the graphics to be found (14). Than through them the deformations are being found: [1, 2, 3]

$$\begin{aligned} \varepsilon_{33} &= (d\varphi = 0,\ \psi = 0 - d_0)/d_0 \\ \varepsilon_{11} &= tg(1)/d_0 + \varepsilon_{33} \\ \varepsilon_{13} &= tg(4)/d_0 \\ \varepsilon_{12} &= tg(2)/d_0 - \varepsilon_{11}/2 - \varepsilon_{22}/2 + \varepsilon_{33} \\ \varepsilon_{22} &= tg(3)/d_0 + \varepsilon_{33} \\ \varepsilon_{23} &= tg(6)/d_0 \\ \varepsilon_{21} &= \varepsilon_{12} \\ \varepsilon_{31} &= \varepsilon_{13} \\ \varepsilon_{32} &= \varepsilon_{23} \end{aligned} \quad (14)$$

$$\sigma_{ij} = \frac{2}{S_2}\left[\varepsilon_{ij} + \frac{S_1(\varepsilon_{11}+\varepsilon_{22}+\varepsilon_{33})}{S_2/2+3S_1}\right] \quad (15)$$

$i = 1, 2, 3;\ j = 1, 2, 3$

tensor of the deformations:

$$[\varepsilon] = \begin{matrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{matrix} \quad (16)$$

tensor of the strain:

$$[\sigma] = \begin{matrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{matrix}$$

Calculation the slope of the surfaces of the main tensions:

$$\Delta\Psi\sigma = \frac{1}{2} arctg\left[-\frac{2\sigma_{13}}{\sigma_{11}-\sigma_{33}}\right] \quad (17)$$

Definition of the main deformations and tensions:

$$r_\varepsilon = \sqrt{\frac{2(\varepsilon_{33}-\varepsilon_{11})}{(4)+\varepsilon_{13}2}}$$
$$r_\sigma = \sqrt{\frac{2(\sigma_{33}-\sigma_{11})}{(4)+\sigma_{13}2}} \quad (18)$$

$$\begin{aligned} \varepsilon_1 &= (\varepsilon_{11}+\varepsilon_{33})/2 + r_\varepsilon \\ \varepsilon_2 &= \varepsilon_{22} \\ \varepsilon_3 &= (\varepsilon_{11}+\varepsilon_{33})/2 - r_\varepsilon \\ \sigma_1 &= (\sigma_{11}+\sigma_{33})/2 + r_\varepsilon \\ \sigma_2 &= \sigma_{22} \\ \sigma_3 &= (\sigma_{11}+\sigma_{33})/2 - r_\varepsilon \end{aligned} \quad (19)$$

$$[\varepsilon] = \begin{matrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{matrix} \quad (20)$$

$$[\sigma] = \begin{matrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{matrix}$$

III. CHOICE OF SOFTWARE PRODUCTS FOR REALIZATION.

A. Visual Basic .NET 2003

The used language for projecting and programming of the current application is Visual Basic .NET. The user`s interface is developed in Visual Basic.

B. Visual Basic .NET Power Pack

Visual Basic.NET Power Pack consists of seven controls, written on Visual Basic .NET 2003.

In the current application is used ImageButton, there you can insert picture and text. It implements IButtonControl and in this way it can be AcceptButton or CancelButton of the form. [4,5]

In order to use VB Power Pack, in the project has been made a link to VbPowerPack.dll.

In the current computer system InnoSetup is used for creating of the installator of the application. [3, 4, 5]

C. Macromedia RoboHelp

Macromedia RoboHelp is an application for creating Help systems and documents for desktop and web based applications, including.NET.

Some main advantages of RoboHelp are:

- Work in preferred from the user editor – we use the built in RoboHelp HTML editors, as Dreamweaver, Frontpage or Microsoft Word;
- Pasting of content into an HTML documents, Microsoft Word, Adobe PageMaker, PDF documents, XML documents or existing HELP projects.
- Easy crating of context-sensible assistant;
- Generating of every popular help format - FlashHelp, WebHelp, Microsoft HTML Help, WinHelp, JavaHelp, Oracle Help for Java and XML.
- Creating of HELP systems, working with every platform and browser;
- Creating a great number of versions for the existing HELP system.

In the current development RoboHelp is used for creating the Assistant to the application.[5]

D. Macromedia Captivate

Macromedia Captivate is a product for creating an interactive simulations and software demonstrations in Macromedia Flash format. The advantages of Captivate are:[3,4,5]

- It takes hold of everything, which is being shown on the screen even web based applications with or without parallel audio;
- Automatically adding interactivity during recording;
- Possibility for publishing in a great number of formats for distribution;
- Possibility for changing or deleting the trajectory of movement of the mouse;
- Automatically creating titles for every step.

For testing the present program system there were used external and structure tests aiming finding and correcting of the made mistakes. The structural method for testing is based on the text of the program module. For that purpose the following is needed:

- Program separation of the module and definition of its information connection to the system. If other modules are being evoked it is necessary they to be structurally tested or simulators to be put on their place;
- In the program text the structure defining operators need to be found and numbered;
- Depending on the type and the semantics of the operators, they have to be represented into a conditional IF – structures;
- The operators between two adjoining IF are representing a straight line section. It can be without operators. If a feedback connection is present, causing multiple implementation of the straight line section, then a cyclic section is found in the program module;
- IF operators are interpreted as peaks of the structural graph, and the linear sections are interpreted as its arcs;
- The structural graph is represented with its matrix of the connections as a list structure;
- The structural graph of the program module is decomposing as a sequence of linear and deposed cyclic sections.

In order to fulfill this test the debugger of Visual Studio NET and its Watch module are used. The breaking points are placed. The values accepted by a given variable are traced and in which branch of the graph it goes. The testing was performed in the "up-down" direction and all the arcs were enveloped.

There were no mistakes found during the conduction of the test. The graphical results visualize the distribution of the residual strains in depth of the examined sample.

V. CONCLUSION

In this stage the experiment was successfully concluded.

The method for calculating residual strains through computer roentgen-structural analysis in complex strain condition allows:

1. Automated calculation of the parameters, needed for calculation of the residue strains in depths;
2. Presentation of the experimental results in digital and graphical mode – convenient for analyze;
3. The method of „ψ – splitting" allows the special features of the deformed and strained condition of the construction materials after grinding to be determined and analyzed;
4. The usage of the method Dolle-Hauk allows the following to be determined: the full tensors of the deformations and strains, the values of the main deformations and strains in the grinded layer, as well as the inclination of the main axes of the ellipsoid of the deformations and respectively of the strains, and as a whole the distribution of the residual strains in depth;
5. The results of the experimental examination in near to the practice conditions are showing high exactness and reliability of the system;
6. In economical aspect, this method is saving the usage of specialized mathematical software as MathCad (by Mathsoft) и Mathlab (by Mathworks) and it is many times cheaper comparing to this software.

REFERENCES

[1] E. Ruseva, R. Rusev – "Physical Metal science", VMEI – Varna, 1988 ;
[2] E. Ruseva, R.Rusev – "Physical metal science– handbook manual for laboratory studies", VMEI – Varna, 1986 ;
[3] John Conel „Programming by Microsoft Visual Basic.NET", Softpress, 2002;
[4] Professional projects Microsoft –"Visual Basic NET"
[5] http://d.linuxbg.org/download/docs/translate/

Prepar Measuring the Main Parameters of Optical Transfer Media – Optical Fibers and Cables

Ivelina Balabanova

***Abstract* – Attenuation measurement is performed at a definite wavelength λ. The main part of such a device is the monochromator. Attenuation is measured using an optical time domain reflectometer (OTDR). It is based on back scattering of the signal due to Rayleigh scattering. A series of short optical pulses is input into the fiber and the reflected power is measured. Optical reflectometers are used to determine and measure distances to interruptions and cracks in optical fibers, as well as distances to splitters and connectors.**

***Keywords* – Optical reflectometer, Monochromator, Numerical aperture, Fiber absorption, Pulse dispersion**

I. Prepar Measuring the Main Parameters of Optical Transfer Media – Optical Fibers and Cables

Attenuation constant of the signal in the optical cable - α

$$\alpha = \frac{1}{\ell}.10.\log\frac{P_I}{P_O} \qquad (1.1)$$

l – cable length; P_I – input power; P_O – output power.

In the case of a cable of length l = 1 km; $n_1 = 1{,}48$; $n_2 = 1{,}46$; dc = 50 μm a beam falling at an angle of 80,6° to the optical axis will travel not 1000 m, but 1014 m, i.e. 14 meters more than a beam parallel to the optical axis.

If $V_c \approx 2.10^8\ m/s$ in a fiber of $n_1 \approx 1{,}5$ the beam delay will be:

$$td = \frac{14m}{2.10^8\ m/s} = 69\,ns \qquad (1.2)$$

The optical power is measured in watts (W), to be turned in decibels (dB)

$$dB = 10.\log\frac{P_{изм}}{P_E} \qquad (1.3)$$

$P_{изм}$ - the power measured; P_E - reference power.

For 1mW reference power

$$dB_m = 10.\log\left(\frac{P_{изм}}{1mW}\right) \text{-(absolute power).} \qquad (1.4)$$

For 1 μW reference power

$$dB_\mu = 10.\log\left(\frac{P_{изм}}{1\mu W}\right) \text{- (absolute power).} \qquad (1.5)$$

The light power losses in an element of the communication systems are:

$$L(dB) = 10.\log\frac{P_O}{P_I} \qquad (1.6)$$

The loss factor in the cable is:

$$CLF(dB/km) = \frac{P_I(dB) - P_O(dB)}{l(km)} \qquad (1.7)$$

Optical devices

Optical multimeter – measures: power in dBμ or dBm; losses in optical fibers, connectors, splitters; reflection losses, etc.

Optical power meter – measures optical power; radiation angles of laser diodes and LEDs; losses in fibers, connectors; numerical aperture, etc.

Optical ohmmeter – measures losses in the fibers.

Optical reflectometer – determines the length of optical fibers; identifies the fault location and the crack location; measures attenuation in fibers, splitters and connectors.

Optical attenuator – measures the speed of optical systems and reflection losses.

Optical spectrum analyzer – analyzes spectra of various types of light source.

Monochromator – operates in circuits for measuring the spectral dependence of the attenuation constant.

- Quantities measured in optical communications: signal attenuation in optical fibers and cables – dB/km;
- constant α of signal attenuation in the fiber – dB/km;
- radiation power of the source (absolute value);
- numerical aperture;
- pulse dispersion – ns;
- fiber absorption;
- losses in fiber connections – dB;
- losses in fiber scattering – dB;
- sensitivity of the optical receiver – A/W;
- radiation spectrum of laser diode and LED – nm, μm;
- physical and design features of fibers:
  - profile measurement;
  - diameter of the fiber core;
  - diameter of the cover.

Measurement of the signal attenuation in fibers and cables

Methods:

- comparison method for input and output signal
  - insertion loss method
  - cut – back method
- pulse method

[1]Ivelina Balabanova Technical university of Gabrovo, department: "KTT", 8 Hadji Dimitar, Gabrovo 5300, Bulgaria, e-mail:ivstoeva@yahoo.com

♦ back scattering method by means of a reflectometer

Attenuation measurement is performed at a definite wavelength λ. The main part of such a device is the monochromator.

Fig.1.1a,b show a circuit for measurement of attenuation using the *insertion loss method*.

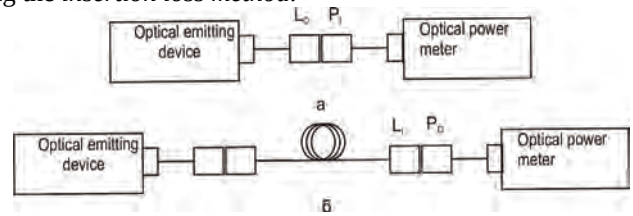


Fig.1.1

If P_I designates the optical power at the beginning of an optical fiber of length L, and P_O - at the end of the fiber, then the attenuation is:

$$a = 10.\log\frac{P_I}{P_O}, dB \qquad (1.8)$$

The attenuation constant α is:

$$\alpha = \frac{a}{L}, dB/km \quad ; \quad L \approx L_1,\ L_0 << L_1. \qquad (1.9)$$

When applying this method, a small standard piece of optical fiber of length L_0 is used (for example, up to 3 m), and the optical power meter is connected to it – Fig.1.1a. Then the meter is connected to the cable being measured of length $L_1 >> L_0$ (circuit in Fig.1.1b).

Fig.1.2 shows measurement of attenuation in optical fibers using the *cut-back method*.

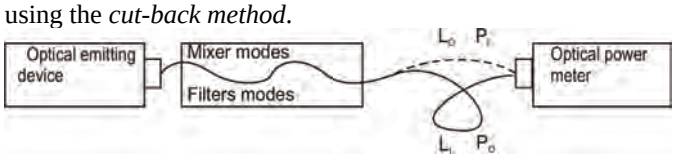


Fig.1.2

First the entire fiber of length L_1 is measured and a reading is taken of power P_O, after that a short fiber of length L_0 (1÷2 m) is cut and a reading is taken of power P_I. The attenuation constant is:

$$\alpha = 10.\log\frac{P_I}{P_O}\Big/(L_1 - L_0), dB/km \qquad (1.10)$$

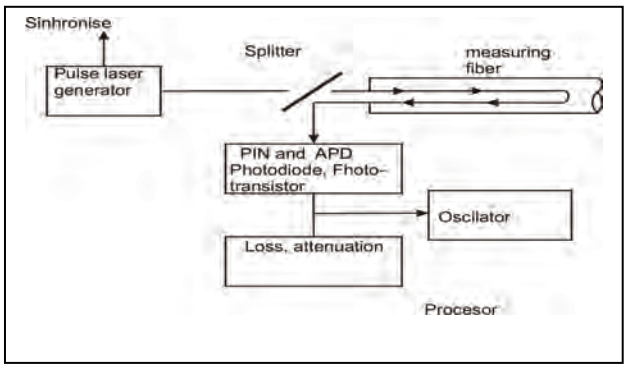


Fig.1.3

Attenuation is measured using an *optical time domain reflectometer* (OTDR).

It is based on back scattering of the signal due to Rayleigh scattering.

A series of short optical pulses is input into the fiber and the reflected power is measured.

Pulses are radiated with t_I = (10- 100) ns, and pulse power of 0,5 W at the respective λ.

The circuit of measurement using an optical reflectometer is shown in Fig. 1.3.

The optical coupler is located at an angle of 45°.

Optical reflectometers are used to determine and measure distances to interruptions and cracks in optical fibers, as well as distances to splitters and connectors. It becomes clear that access to both ends of the fiber is not necessary.

When a reflectometer is used for measurement, the results are presented graphically, by means of reflectograms.

Measuring the dependence of the optical fiber attenuation constant α on the wavelength λ. The measurement circuit is shown in Fig. 1.4.

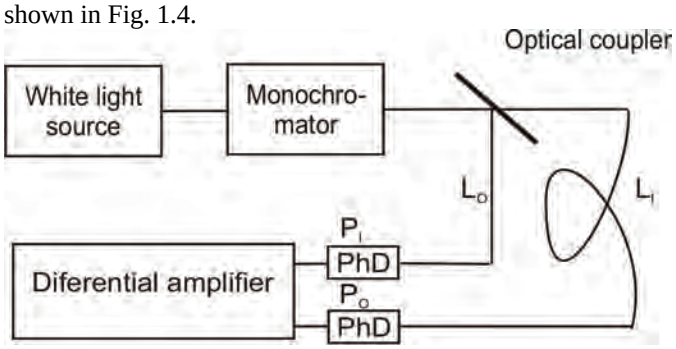


Fig.1.4

The monochromator with a diffraction grating transmits only beams of wavelength:

$$\Delta\lambda = \lambda_1 - \lambda_2 \qquad (1.11)$$

An incandescent lamp is used as a colourless light source.

Measuring the sensitivity of the receiver (R) relative to the digital error (BER) probability – Fig. 1.5

In the circuit T designates a transmitter, E/O – electro-optical, for example laser LED, and R designates a receiver.

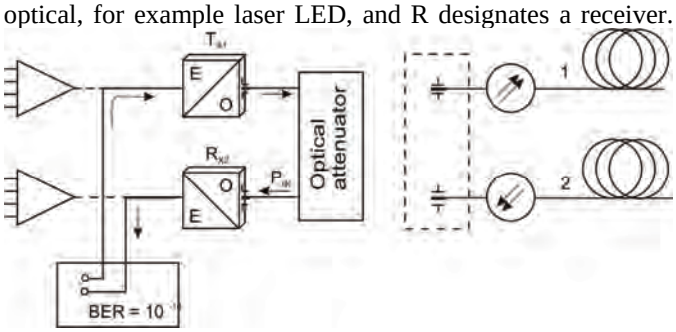


Fig.1.5

REFERENCES

[1]. OSRAM. Opto semiconductors,2001

[2]. Personal Computer Word, 2003, май.

[3]. Wohlrabe, Frank. Infrarot-Datenubertragung. Elektor-Verlag, Aachen. (I Aaufgabe), 2002.

[4]. Велчев, Н. Б. Основи на оптоелектрониката., Пловдивски университет, Изд. "Паисий Хилендарски", 2002.

Estimation of the Safety Vehicle Speed in Curve Using Inertial Sensors

Emil Iv. Iontchev[1] and Rosen G. Miletiev[2]

Abstract - **This paper presents a method for calculation of the maximum authorized vehicle speed in the curve using inertial sensors to accomplish an active safety tracking system. The curve radius and road tilt are calculated by the roll and pitch vehicle axis accelerations. The system performance and the method functionality is tested and analyzed.**

Keywords - **land vehicle rollover, inertial sensors, tilt**

I. INTRODUCTION

MEMS sensors allow the implementation of lots of different functions, as free-fall detection, car navigation, map browsing, gaming, menu scrolling, motion control, vibration monitoring, antitheft and many others. With the introduction of the low-cost controllers and low-cost MEMS technology, much more economical inertial sensors can now be targeted toward automotive steering/traction control. Steering/traction control systems require a three variable inertial sensor system. In this example, the system must detect motion in the X- and Y-axis and rotation on the Z-axis (yaw) [1]. This system is able to detect the direction of actual vehicle movement in comparison to the desired movement and provide feedback to the overall system controller [2]. Incidents with vehicles as a result driving with a high speed and consequence turnover are a serious problem for the transportation companies [3].

Therefore, developing systems of warning of the driver of decrease to the speed accordingly the geometry a route, the state of the vehicle and the atmospheric conditions is particularly actual question.

As an example for such system witch warning the driver ahead is explain in [4]. The Estimation of the safety curve speed is based on the information for the lateral acceleration. The term that is being used for estimation the hazard of rollover is:

$$RSA_{score} = \frac{a_n(actual)}{a_n(critical)} 100\% \qquad (1)$$

where a_n is the vehicle lateral acceleration.

The critical lateral acceleration is inversely proportional to the mass of the vehicle and was determined experimentally with a tilt table.

The system computes the RSA score every half second, and once it generates a high score, it delivers a warning message to the driver. This system require combination of precise placement, digital maps with the accuracy data about the road, accurate calculate the dynamic parameters to the vehicle, as well as predict it the speed. They would be able to deliver a warning signal to the driver some seconds before the dangerous section of the road.

The other kind of systems are these that calculate the current state, they are registering and can be exploited for subsequent analysis on the causes that had led the incident.

In the last years the MEMS sensors are integrated with GPS receivers, with which the measured parameters are increased, as well as the accuracy of measurement. There are developed methods of measurement of some of the above parameters in [5, 6].

This paper explains a method of estimation the critical vehicle speed according to the inertial sensors data to accomplish an active safety tracking system.

II. ROAD AND VEHICLE KINEMATICS

The high safe speed in a curve depends on the geometry of the curve, the state of road surface, the stability of the vehicle against rollover and an experience of the driver.

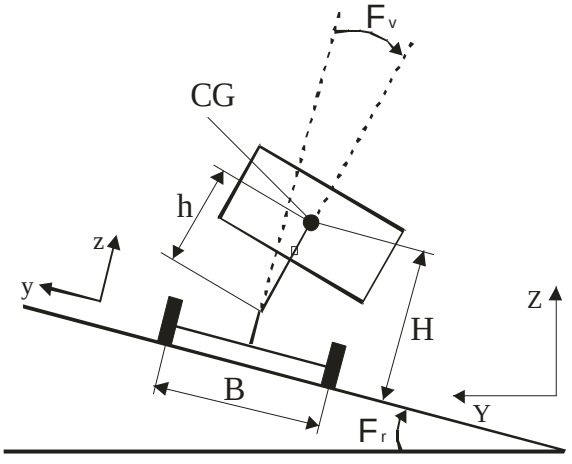
A vehicle roll model with road bank angle is shown on Fig. 1.

The highest speed which a vehicle can pass in a curve without a side-slip can be estimated by the expression [7]:

$$V_{\max} = \sqrt{gR\mu_{lat}}, \qquad (2)$$

and the highest vehicle speed in a curve without a rollover can be estimated by the expression [7]:

$$V_{\max} = \sqrt{gR\tan\phi_{\lim}}. \qquad (3)$$



[1]Emil Iv. Iontchev is with the Departament of Material Handling and Construction Machines, Higher Scool Of Transport "T. Kableshkov" Sofia, Bulgaria E-mail: e_iontchev@yahoo.com

[2]Rosen G. Miletiev is with the The Faculty of Communications and Communications Technologies, Technical University of Sofia Bulgaria E-mail: miletiev@tu-sofia.bg

Fig.1 Vehicle kinematics

It is seen that V_{max} value is determined by the curve radius R, Earth gravity acceleration g, lateral coefficient of coherence μ_{lat}, which is function of the state of the surface on the road, the vehicle tires, atmospheric conditions and the static stability angle $tan\phi_{lim}$. The last parameter is determined by the base of the vehicle B and the position of vehicle gravity center H according to the equation [7]:

$$\tan\phi_{\lim} = \frac{B}{2h} \qquad (4)$$

When the curve there is a bank as a whole the stability against rollover and side-slip is increased. The expression (2) in this case is transformed to the following equation:

$$V_{\max} = \sqrt{gR\left(\frac{\mu_{lat} + \tan\phi}{1 - \mu_{lat}\tan\phi}\right)}, \qquad (5)$$

and the expression (3) is transformed to the next equation:

$$V_{\max} = \sqrt{gR\tan\left(\phi_{\lim} + \phi_p\right)} \qquad (6)$$

The total tilt of the vehicle in a curve is a sum from the route bank ϕ_r and the suspension tilt ϕ_v in dependence of the parameters μ_{lat} and $tan\phi_{lim}$ on a side-slip before rollover and vice versa.

All parameters needed for calculation of the speeds can be measured carefully by the inertial sensors. Their values are measured with respect to a vehicle frame that is shown of Fig.2.

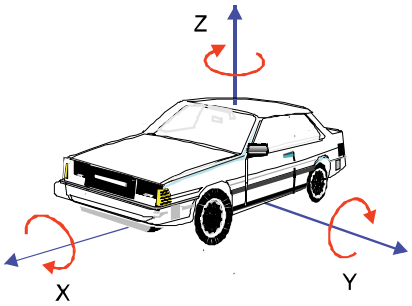


Fig.2 Vehicle frame

III. Inertial Measured System

The acceleration data for calculation of the critical vehicle speed are obtained from a system, which block diagram is shown on Fig.3. The inertial sensor measures the linear accelerations on the three axes. The tilt of vehicle measures by the same accelerometer, whose axis's x and y match the same axis's of co-ordinate system of the vehicle.

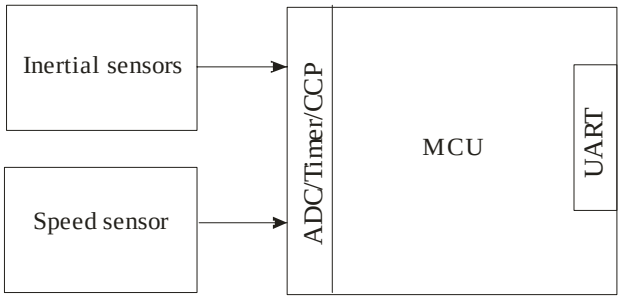


Fig.3 Block diagram of the measured system

The block for gaining the speed can use the sensors rigged with the gear, a GPS receiver or it can be calculated by mean of the data of accelerometer about the x axle.

A very precise analog to digital converter AD7739 [8] was used in this design to minimize the errors from conversion of analog outputs of the sensors to digital data. The MCU takes care of the synchronization of measurements and transfer of gathered inertial data to the other block. The inertial block algorithm has been developed to make measurements with frequency of 10Hz which is quite enough for land-vehicle applications.

IV. Performance of the System

The ingredient in the signal which is being changed from bank is changing more slowly that ingredient is being changed from tilt of vehicle therefore is made filtration to take more reality data. In our case it is thought that the tilt of vehicle due to the suspension is little, therefore the received data is mainly from the bank of route. The relationship between the measured values from accelerometer about y and x axes and the angles are:

In case that the vehicle is laterally tilted:

$$\phi = \arcsin\frac{a_y}{1g}, \qquad (7)$$

and when it is longitude tilted:

$$\theta = \arcsin\frac{a_x}{1g}. \qquad (8)$$

The radius of curve is estimated from the linear vehicle speed v and the normal acceleration a_y:

$$R = \frac{v^2}{a_y} \qquad (9)$$

The estimation of the vehicle speed in a real time experiments have been carried out. The circular shape of trajectory is known. Thus a stadium (Fig.4) is selected as reference trajectory because its shape is very close to the circular one. The selected place for experiments has very low level of slope.

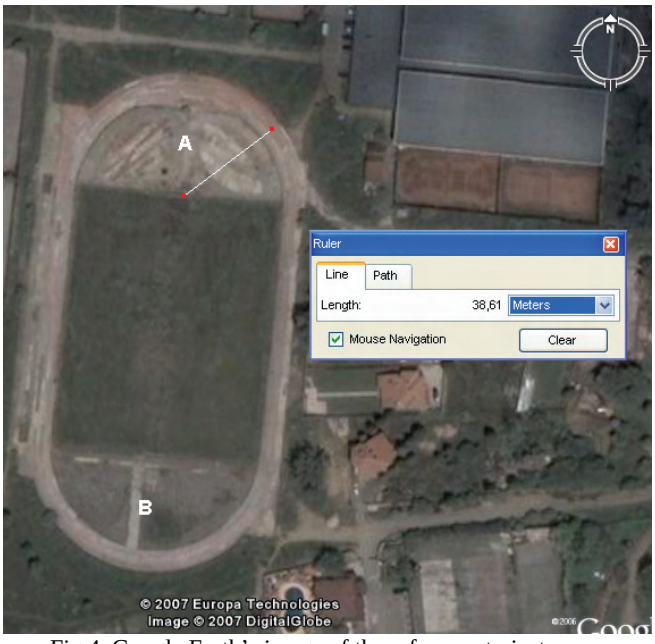
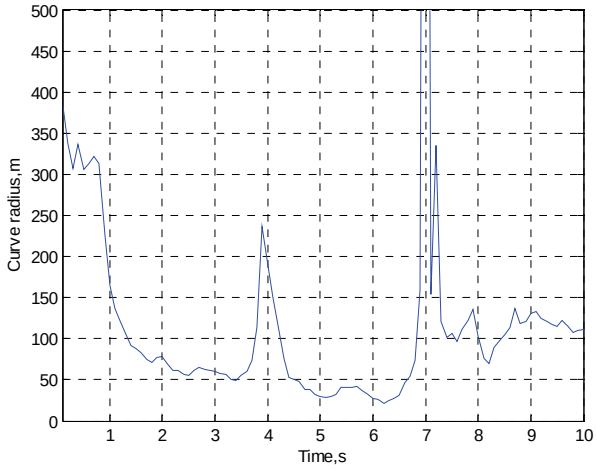
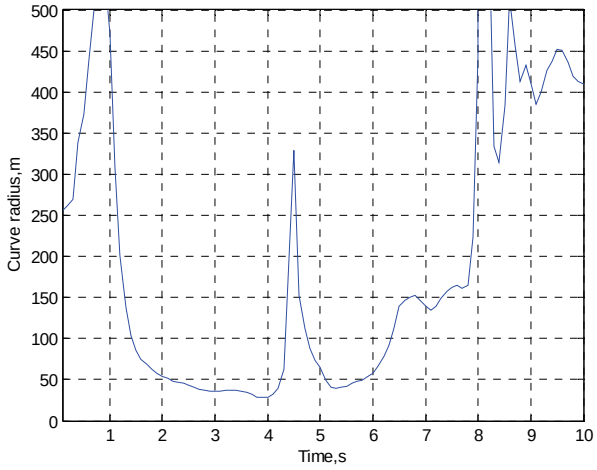


Fig.4. Google Earth's image of the reference trajectory

The results of the system performance experiments are shown at Fig.5.

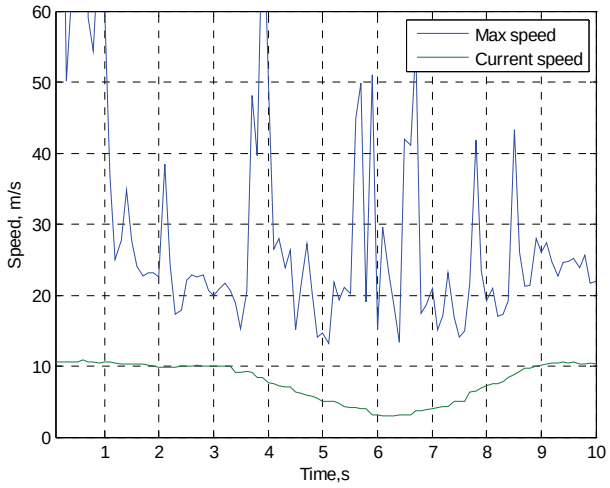


(a)

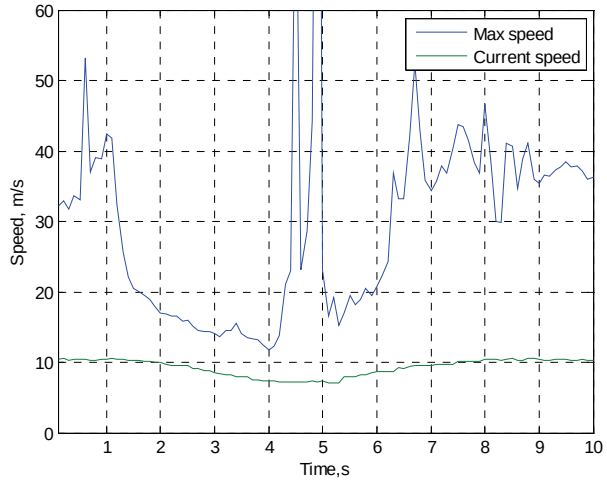


(b)

Fig.5. Estimated curve radius *R*



(a)



(b)

Fig.6. Estimated safety curve speed

Figure 5a and Figure 5b represents the estimated curve radius *R* for A and B area (Fig.4) respectively according to the equation (9). The estimated radius value is close to the real curve radius. The both figures contain a time period (approximately 0,5s) when the estimated curve radius is much higher then the real one. This time period corresponds to the middle curve point when the driver restores the wheel angle and put out to the next curve.

The safety curve speed is also calculated according to the equation (6). It is compared with the driving speed and the estimated results are represented at Figure 6. Figure 6a and Figure 6b represents the estimated safety curve speed and the current driving speed for A and B area (Fig.4) respectively.

The represented results show that the driving speed is lower than the safety curve speed. If the estimated maximum curve speed is exceeded, then the traction control system reduces the driving speed and retains the car stability.

The system performance can be improved if the low-cost angular rate sensor (gyroscope) is added. In this case the car rotation rate can be estimated directly from the gyroscope data.

V. CONCLUSION

The represented method allows determining the curve radius and maximum safety speed by using only one three dimensional linear accelerometer. This system is able to detect the direction of actual vehicle movement in comparison to the desired movement and provide feedback to the overall system controller to prevent the road accidents. It also accomplishes the real-time calculations of the critical speed and is able to deliver a warning signal to the driver up to few seconds before the road dangerous section or can directly access to the car traction control systems to reduce the current speed to the safety one.

VI. REFERENCES

[1] Inertial Sensor for Steering/Traction Control, www.freescale.com

[2] Richard Dixon, Prospects for MST sensors in automotive applications, WTC document, 2006

[3] Ryu, J. and Gerdes, J. C., "Estimation of vehicle roll and road bank angle". American control conference, 2004, Boston, US

[4] Seth Rogers, Wenbing Zhang, "Development and Evaluation of a Curve Rollover Warning System for Trucks", DaimlerChrysler Research & Technology North America

[5] Rumen Arnaudov, Yordan Hristov, Yasen Angelov, „Application of inertial navigation systems in control of military subdivisions", National Millitary Forum, V. Tyrnovo, 2004

[6] Dr.-Ing. Edgar von Hinüber, "Inertial Measuring Systems with Fiber-Optic Gyroscopes", ATZ magazine, 2002, vol.6

[7] Jordan Dimitrov, "Theory of the automobile, tractor and the forklift trucks". TU, 1991, Sofia

[8] Analog Devices, AD7739 8-Channel, High Throughput, 24-Bit ∑-Δ ADC, product datasheet, www.analog.com

Application of Ecological Monitoring in SMEs in Bulgaria

Ina Nikolova-Jahn[1]

***Abstract* – In the paper are considered aims and instruments of the ecological monitoring. In the last years the increasing ecological awareness of the consumers, the continuous resources shortage as well as the presence of ecological standards, led to the fact that ecological monitoring is essential for successful auditing and management. The first part of paper clarifies the definition of and differences between monitoring and auditing, and outlines their potentially important role in ecological impact assessment. The second part draws briefly instruments of monitoring and their application in SMEs in Bulgaria.**

***Keywords* – Ecological Monitoring, Auditing, SME, ISO 14001:2004, ISO 9001:2000, Ecological Indicator.**

I. INTRODUCTION

Today the prosperity of each enterprise depends on the effectiveness of its economic activity and competitive power. The conditions of the open market are based on extensive analysis and application of technical, ecological and economic information. Monitoring can provide essential information to improve ecological impact assessment and project management.

II. THEORETICAL ASPECTS OF THE ECOLOGICAL MONITORING

As a function of the activity of an enterprise the ecological requirements will be differently pronounced.

Monitoring involves the measuring and recording of physical, social and economic variables associated with development impacts.

Rafees [4] represents monitoring as factor to provide information for decision making and also includes broad public participation and stringent administrative review procedures.

Sadlers [5] writes of the need to introduce feedback in order to learn from experience; we must avoid the constant "reinventing of the wheel" in environmental project management.

Davies [6] have examined three types of ecological monitoring which might be associated with the life cycle of an undertaking, these are a) baseline monitoring, b) effects or impact monitoring and c) compliance monitoring.

"Baseline monitoring" refers to the measurement of ecological variables during a representative project period to determine existing conditions, ranges of variation and process of change.

" Effects monitoring " involves the measurement during project.

"Compliance monitoring" takes the form of periodic sampling and continuous measurement of levels of waste discharge, noise or other similar emissions to ensure that conditions are observed and standards are made.

The terms of the *ecological monitoring* contain an extension of the monitoring and auditing concept in the sense of all ecologically relevant facts of an enterprise. According to this goal, definitions of monitoring system have to be developed, which consider both economic and ecological and as well social aspects in the long run of preparing a decision. According to ecological monitorings the comprehensive provision of information, information collection and information retrieval, as planning, auditing and a check of all ecological relevant enterprise activities,are thus understood. Monitoring can also provide an accepted date base, which can be useful between interested parties : customers, public, trade unions, owners, outside capital givers, public, coworkers, status of competition, special syndicates, parties etc.(see in Fig.1)

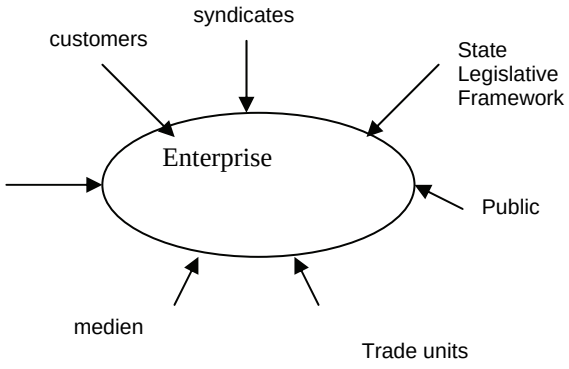


Fig.1. Parties interested in the process of ecological monitoring

The enterprise can exclude itself from the requirement pressure of these groups only if it integrates *social values* antizipativ into the target system and enterprises, the politics result, which consider equally ecological and economic facts.

Above all the ecological monitoring is the impact on the ecological production, which seize and systematically dispose

[1]Ina Nikolova-Jahn,PhD, Faculty of Management, Department Economics, Industrial Engineering & Management
Technical University, Sofia, E-mail: jahn_ina@yahoo.com

consumption, in order to make it moderate and controllable[3].

Purposes of the ecology-oriented monitoring The definition of the aims of ecological monitoring and auditing is very important for the enterprises and their activities.

In table 1 there is an example for a systematization of the aims of the ecological monitoring, based on the study of the SME Agency in Bulgaria[2].

TABLE I
GOALS OF THE ECOLOGICAL MONITORING

Intern goals	Reduction of procurement and production costs High identification of the Personal with the enterprise
External goals	Satisfaction of the demand for ecologically perfect products Ecological image Easier entrance to financial markets Legal fulfilment of ecological protection demands

Ecological Monitoring provides information which can be used to evaluate the effectiveness of innovation activities.

Furthermore it manages to recognize ecological risks connected with the *Ecological monitoring* eliminates them as soon as possible, in order to convert ecological challenges into operational success factors.

II. INSTRUMENTS OF THE ECOLOGY-ORIENTED MONITORING AND AUDITING

A. ***Ecological information system*** for the systematic collection of all relevant information, existing planning and control system are to be supplemented around an ecology-oriented ecological information system. Above all such system should fulfill the following functions:

- Supply of information about the enterprise-internal material and energy flows (also regarding the product life cycle;
- Supply from the information to the analysis of economic-ecological restrictions
- Situation analysis of the enterprise;
- Development of ecological programs;
- Supply of the ecological specified condition of the enterprise regarding products and processes representations;
- Supply of control system for the integration of the ecological questions into the master planning of the enterprise
- Supply of all relevant data, facts and information both for enterprise-internal decision makers and for all external information target groups by accordingly arranged information politics.

Above all the intensified meaning of ecological/ecology information in the past turned to external communication. The preferred medium is usually an annually ecological report, which is also required explicitly in connection with Certifying for the Ecological system ISO 14001:2004 or implementation of EMAS II (Eco-Management and Audit Scheme).

B. Reduction of costs reserves in operational ecological protection can lie considerable reduction of costs. Depending on enterprises and industry the cost can be in the range of 5 to 15 per cent of the operational total expenditure[2].

Sociological investigations show to some extent a similar typology in the behavior of the small and medium enterprises in Bulgaria while they have to apply special measures for ecological protection. The enterprises become passive participants in the process of assimilation of the ecological requirements and they almost never undertake voluntary measures for this aim.

C. ***Ecological cost***

There are two possibilities to set up an ecological cost calculation. In the first case the cost calculation system, which already exists, remains complete[1].

The ecological costs have to be calculated:

- Ecological costs of the piece/total costs;
- Ecological costs of the material/personnel expenditure and writings-off;
- Ecological cost characteristic numbers in the time comparison.

In the other case the ecological costs are assumed for each cost category. These are then assigned on an extended cost apportionment sheet, cost centres and cost objectives.

The expenses for investments are not a priority for the enterprises in Bulgaria. Being asked what part of the investments in last two years are used for protection of the environment, over 73% of the enterprises answer that the investments are not connected with the environment and there is no special control above the ecological expenses. The explanation may be, that enterprises see their competitive priority only cost reduction and may be technological innovations. At the same time SMEs are an important part of the solution of Bulgaria's ecological issues and are therefore the object special support and government institutions (see Table II).

TABLE II
Ecological projects

Areas	Number of projects financed	Total value of projects, BGN
Waste management	10	4485000
Air protection	8	5513000
Waste water	4	4494756
Eco-Agriculture	4	280729
Total	26	14493036

Source : EMEPA and MOEW

D: Ecological Indicators

Numbers in the overview for the setting up an effective ecological monitoring are needed in the operational everyday life information about the effects of operational acting on the natural environment.

This information can be determined by characteristic numbers, whereby characteristic numbers comparisons (time

comparison, comparison of nominal and actual values, operating comparison) are from special interest.

Following ecologically important kinds of characteristic numbers and/or characteristic numbers can be differentiated (see Table 2).

TABLE III
Ecological Indocators

Energie	Total energy consumption Source of energy Energy utilization
Air pollutants	Emission quantity of air pollutants Emission ratios
Water and waste characteristic	Quantities required of individual kinds of water (for example drinking water, groundwater, industrial water, rain water etc.) Quantities of the kinds of waste water (for example process water, cooling water, sanitary and kitchen water etc.) Water utilization
Material usage	Raw material utilization Fuel employment and fuel utilization Secondary raw material utilization
Waste	Waste characteristic numbers arranged according to valuable and special refuse Waste ratio

F. ***Ecological audit***

A goal of the EMAS II is specified in such way, that enters into force in all countries of the European union, is the continuous promotion of operational ecological protection. The enterprises participating in the system receive the authorization, after an external examination, to advertise with the *EMAS II Logo* in order to have an advantage secure itself so a market lead over not participating competition. Since the beginning of the 90's the ISO 14001 series is available an international standard for ecological management systems, which today emphasises a process orientation, contain many references to the implementation of ecological management systems. For this reason is Implementation of ISO 14001:2004 frequently compared with the development of the quality management and the standard series ISO 9000 ff.

Monitoring is also essential for successful ecological impact auditing. And can to be one of most effective guarantees of commitment to undertakings and to mitigation measures.

The following individual measures can be differentiated during the introduction of ecological monitoring and Auditing:

- Development of an ecological goal catalog;
- Development of ecological enterprise guidelines;
- Development and definition of ecological company targets;
- Definition of ecological valuation criterion;
- Development and/or improvement of an ecological cost accounting procedure;
- Periodic collection, documentation and evaluation of the material and energy flows in all parts of the enterprise;
- Regular evaluation of all operational activities under ecological aspects of risk;
- Regular evaluation more again and continuing monitoring of old means of production, production procedures and products regarding their ecological characteristics;
- Elaboration and documentation of ecological product guidelines.
- Evaluation and documentation of incidents and elaboration of measures for incident reduction and/or minimization of incident damages;
- Current control of the ecological measure catalogs as well as constant development and improvement of ecology-oriented control indicators;
- Open communication and documentation of all control results in the enterprise

During investigation of the SMA Agency in Bulgarian is determined that the enterprises in Bulgaria have a positive attitude on the ecological monitoring and legislative framework in relation with their competitive abilities. 76,8% consider that the legislation is strict but useful for Bulgarian enterprises and have ecological program. On other hand only 15,5% think that the legislation is too hard and it negatively affects the firm's activity. In addition, these enterprises look take care for their results regarding process effectiveness auditing ecological pollution.

III. CONCLUSION

The relationship between a project and its environment is needed throughout the life of a project. Environmental assessment is meant to establish the terms and conditions for project implementation. Such ecological monitoring can improve project management and contribute to the auditing.

In the paper aims and instruments of the ecological monitoring and their possibilities for application in small and medium enterprises in Bulgaria are discussed.

REFERENCES

[1] Meffert, D., K, Kirchgeorg (1998) Marktorientiertes Umweltmanagement, 3rd. Edition, Stuttgart
[2] MOEW, Research survey for SME (2006)
[3] Nikolova, I. (2005) Anwendung einer system von Ecological-Indikatoren bei der Einführung des UMS, V Konferenz Management und Engeneering, Sofia
[4] Raffee, H., (1991)Marketing und Umweltschutz, Stuttgart, p. 46 – 54
[5] Sadler, B (1988). The Evauation of assessment: post –EIS Research and Process. In Environemtal Impact Assessment. P. Wathern (ed.) London: Udwin Hyman
[6] Winter, G. (2003)Das umweltbewusste Unternehmen, Verlag Vahlen, S. 785

This page intentionally left blank.

Uncertainties Estimation of Measurement Devices by Using Finite Element Method

Marija M. Cundeva-Blajer[1] and Ljupco S. Arsov[2]

***Abstract* – In this paper an advanced CAD methodology for metrological uncertainties estimation of two electromagnetical measurement devices will be presented. The metrological characteristics of a 20 kV combined current-voltage instrument transformer, as well as of an electrical steel sheet testing device-Epstein frame will be determined by using an original program package developed at the Faculty of Electrical Engineering and Information Technologies-Skopje, based on the finite element method, FEM-3D. The FEM-3D outputs will be input data for further metrological improvement, of the test devices.**

***Keywords* – Finite element method, Epstein frame, instrument transformer.**

I. INTRODUCTION

This paper will provide a study of several inaccuracy sources that appear in the electromagnetic measurement devices. Two metrological objects of analysis: 20 kV combined current-voltage instrument transformer (CCVIT) and electrical steel sheet testing apparatus-Epstein frame (EF) will be studied. The measurement devices must comply with the rigorous metrological standards such as IEC 60404-2, [1] and IEC 60044-2, [2]. The metrological uncertainty of the electromagnetic testing devices must be estimated as exactly as possible. These devices are complex electromagnetic systems, which are highly non-linear and their analysis by using the analytical methods introduces a lot of approximations and their design will be with very high metrological uncertainty. So the application of the numerical methods, like the finite element method, for the analysis of the electromagnetic fields phenomena is indispensable.

II. ELECTROMAGNETIC FIELD ANALYSIS

A. FEM-3D

The magnetic field distribution in closed and bounded systems (such as the EF and the CCVIT) is described by the system of Maxwell's equations. After the magnetic vector potential $\vec{A}$ as an auxiliary quantity is introduced, the magnetic field distribution is expressed by the Poison's non-linear partial differential equation:

$$\frac{\partial}{\partial x}\left(\vec{\nu}\left(\vec{B}\right)\frac{\partial \vec{A}}{\partial x}\right)+\frac{\partial}{\partial y}\left(\vec{\nu}\left(\vec{B}\right)\frac{\partial \vec{A}}{\partial y}\right)+\frac{\partial}{\partial z}\left(\vec{\nu}\left(\vec{B}\right)\frac{\partial \vec{A}}{\partial z}\right)=-\vec{j}(x,y,z) \quad (1)$$

where $\vec{B}$ is the magnetic flux density, $\vec{\nu}$ the magnetic reluctivity and $\vec{j}(x,y,z)$ the volume current density. The analyzed devices are heterogeneous and non-linear with prescribed boundary conditions and eq. 1 can be solved by numerical methods, only. The magnetic field analysis is done by an original and universal program package FEM-3D developed at the Faculty of Electrical Engineering and Information Technologies-Skopje (FEIT), [3].

The FEM-3D program comprises five main modules: G1-automatic mesh generator; G2-flux plot; G3-input definition of boundary conditions, material magnetic properties, magnetizing current sources; G4-magnetic field distribution calculator; G5-electromagnetic characteristics calculator. The 3D iterative finite element method calculations are done by taking into account the magnetic anisotropy and the different magnetic reluctivities along the co-ordinate axes, due to the lamination of the magnetic cores of the devices.

B. Case Study No. 1-Epstein Frame

The Epstein frame forms an unloaded transformer comprising a primary magnetising winding, a secondary voltage winding and electrical steel sheet as test specimen, as given Fig. 1.

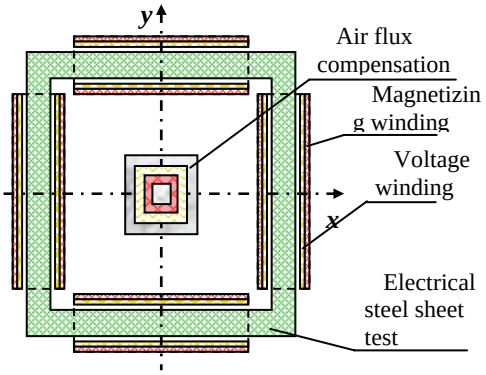


Fig. 1. The electromagnetic system of Epstein frame

Systematic errors which exist at the Epstein frame must be eliminated or reduced. However, the Epstein frame is a highly non-linear electromagnetic system, therefore for its proper design and analysis the numerical methods such as the finite element method, are indispensable, [3].

The initial analysis of the Epstein frame electromagnetic

[1]Marija M. Cundeva-Blajer is with the Faculty of Electrical Engineering and Information Technologies, Karpos II bb, POBox 574, 1000 Skopje, Macedonia, E-mail: mcundeva@etf.ukim.edu.mk

[2]Ljupco S. Arsov is with the Faculty of Electrical Engineering and Information Technologies, Karpos II bb, POBox 574, 1000 Skopje, Macedonia, E-mail: ljarsov@etf.ukim.edu.mk

system is done by application of the classical analytical transformer theory. However, that is an approximate calculation because of the numerous neglections that are included. Some of these approximations which introduce systematic errors are:

- constant magnetic path length l_m=0,94 m, like in [1];
- approximate magnetic field distribution;
- not exactly calculated leakage fluxes in the air.

For appropriate geometrical and mathematical modeling of the EF, its compound configuration is divided into 7 layers along the axial axis of the frame, as follows: the magnetizing winding (1. and 7. layer), the voltage winding (2. and 6. layer), the isolation+air (3. and 5. layer), the magnetic core of the specimen examined by the frame (4. layer). Computations are carried out for various magnetic polarization and for different grades of reference specimen of electrical steel sheet. The reference specimen parameters are given in Table I.

TABLE I
MAIN PARAMETERS OF THE THREE REFERENCE SPECIMEN GRADES

	Grade A	Grade B	Grade C
Grade according to IEC 60404-8-7, 1998	M95-23P 5	M110-27P 5	M140-30S 5
Mass *m* [kg]	0,5551	0,5973	0,7100
Number of strips *n*	40	36	40
Cross-section surface per branch $A_{j,br}$ [cm^2]	0,6479	0,6971	0,8287
Stacking factor k_{fe}	0,945	0,950	0,955
Active mass m_a [kg]	0,4659	0,5013	0,5959
Thickness *d* [mm]	0,216	0,258	0,276
Height *h* [mm]	4,320	4,644	5,520

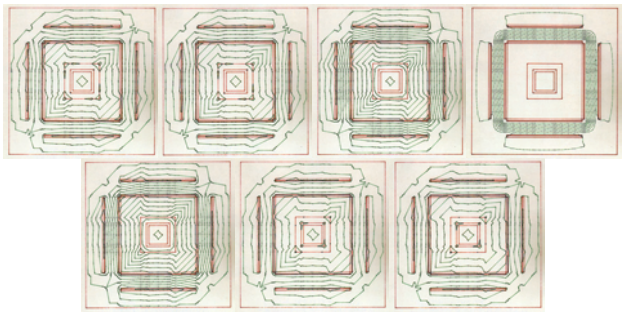


Fig. 2. Magnetic flux plots in the 7 cross-sectional layers along the z-axis of the EF with electrical steel sheet (magnetic core) at J_m = 0,8 T

After the non-linear iterative FEM-3D calculations of the magnetic vector potential are accomplished, the flux plots of the magnetic field distribution in the seven cross-sections along the z-axis of the Epstein frame three-dimensional domain are derived, as displayed in Fig. 2.

C. Case Study No. 2-Combined Instrument Transformer

The combined current-voltage instrument transformer (CCVIT) comprises four windings and two magnetic cores in one housing. The initial CCVIT (voltage measurement core VMC ratio $20000\,V/\sqrt{3} : 100\,V/\sqrt{3}$ and current measurement core CMC ratio 100 A : 5 A) analysis is done by application of the analytical transformer theory. Still, this is a very complex electromagnetic system in highly non-linear regime and with two cores mutually magnetically coupled, as shown in Fig. 3.

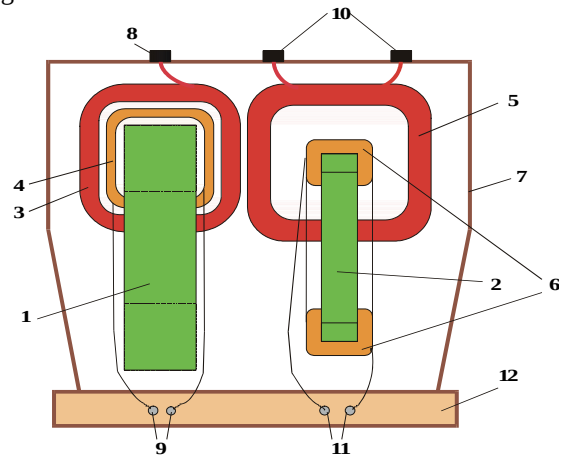


Fig. 3. Electromagnetic system of the 20 kV CCVIT

(1-VMC magnetic core; 2-CMC magnetic core; 3-VMC high voltage primary winding; 4-VMC low voltage secondary winding; 5-CMC high current primary winding; 6-CMC low current secondary winding; 7-common isolation housing; 8-high voltage socket lit; 9-low voltage socket lits; 10-primary current socket lits; 11-secondary current socket lits; 12-isolation base)

The CCVIT electromagnetic phenomena is analysed by FEM-3D, [3]. The objective of the FEM-3D analysis is the estimation of the four CCVIT metrological parameters:

- p_u VMC voltage error
- p_i CMC current error
- δ_u VMC phase displacement error
- δ_i CMC phase displacement error.

They are calculated through the following procedure:

No load voltage error:

$$p_{0u} = -100 \cdot \left(r_{1eu} \cdot g_{0eu} + x_{\sigma 1eu} \cdot b_{0eu}\right) \quad (2)$$

No load phase displacement error:

$$\delta_{0u} = 3440 \cdot \left(r_{1eu} \cdot b_{0eu} - x_{\sigma 1eu} \cdot g_{0eu}\right) \quad (3)$$

Voltage error introduced by the load:

$$p_{tu} = -100 \cdot \left((r_{1eu} + r_{2eu}) \cdot g_{eu} + (x_{\sigma 1eu} + x_{\sigma 2eu}) \cdot b_{eu}\right) \quad (4)$$

Phase displacement error introduced by the load:

$$\delta_{tu} = 3440 \cdot \left((r_{1eu} + r_{2eu}) \cdot b_{eu} - (x_{\sigma 1eu} + x_{\sigma 2eu}) \cdot g_{eu}\right) \quad (5)$$

VMC total voltage error:

$$p_u = p_{0u} + p_{tu} \quad (6)$$

VMC total phase displacement error:

$$\delta_u = \delta_{0u} + \delta_{tu} \quad (7)$$

$\vec{Z}_{2i}$ is total CMC secondary circuit impedance:

$$Z_i'' = \sqrt{(R_{ti} + R_{2i})^2 + (X_{ti} + X_{\sigma 2i})^2} = \sqrt{R_i''^2 + X_i''^2} \quad (8)$$

$$tg\boldsymbol{\beta}_{2i} = \frac{X_{ti} + X_{\boldsymbol{\sigma}2i}}{R_{ti} + R_{2i}} = \frac{X''_i}{R''_i} \quad (9)$$

CMC current error:

$$p_i \approx -\frac{I''_{0i}}{I''_{1i}} \cdot cos(\boldsymbol{\varphi}_{fei} - \boldsymbol{\beta}_{2i}) \cdot 100\% \quad (10)$$

CMC phase displacement error:

$$\boldsymbol{\delta}_i \approx 3440 \cdot \frac{I''_{0i}}{I''_{1i}} \cdot sin(\boldsymbol{\varphi}_{fei} - \boldsymbol{\beta}_{2i})[\text{min}] \quad (11)$$

where: $cos\boldsymbol{\varphi}_{fei} = \frac{P_{fei}}{S_{fei}}$ and

- r_{1eu} and $x_{\sigma 1eu}$ are the resistance and the leakage reactance of the VMC primary winding (per turn);
- r_{2eu} and $x_{\sigma 2eu}$ are the resistance and the leakage reactance of the VMC secondary winding (per turn);
- R_{1i} and $X_{\sigma 1i}$ are the resistance and the leakage reactance of the CMC primary winding (total);
- R_{2i} and $X_{\sigma 2i}$ are the resistance and the leakage reactance of the CMC secondary winding (total).

TABLE II
FEM-3D CALCULATED VMC PRIMARY WINDING LEAKAGE REACTANCES PER TURN

U_u/U_{ru}	Only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
I_i/I_{ri}	$x_{\sigma 1eu}$ [μΩ]							
0,2	91,61	91,22	83,48	69,41	58,30	47,10	36,16	25,3
0,4	92,12	92,09	86,56	81,01	75,46	69,84	64,38	59,0
0,6	91,28	91,29	87,61	83,88	80,19	76,48	72,78	69,2
0,8	91,51	91,49	88,73	85,97	83,17	80,40	77,62	74,9
1,0	91,61	91,64	89,40	88,20	84,97	82,74	80,54	78,4
1,2	91,81	91,79	89,95	88,11	86,26	84,40	82,54	80,8

TABLE III
FEM-3D CALCULATED VMC SECONDARY WINDING LEAKAGE REACTANCES PER TURN

U_u/U_{ru}	Only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
I_i/I_{ri}	$x_{\sigma 2eu}$ [μΩ]							
0,2	63,58	63,31	57,94	48,17	40,46	32,69	25,10	17,56
0,4	63,93	63,91	60,07	56,22	52,37	48,47	44,68	40,95
0,6	63,35	63,35	60,80	58,21	55,66	53,08	50,51	48,04
0,8	63,51	63,49	61,58	59,66	57,72	55,80	53,87	52,02
1,0	63,58	63,60	62,04	61,21	58,97	57,42	55,89	54,40
1,2	63,72	63,71	62,43	61,15	59,87	58,58	57,28	56,06

TABLE IV
FEM-3D CALCULATED CMC PRIMARY WINDING LEAKAGE REACTANCES PER TURN

I_i/I_{ri}	Only CMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	$x_{\sigma 1ei}$ [μΩ]							
0,2	28,09	27,75	27,47	27,21	26,91	26,68	26,39	26,18
0,4	28,10	27,76	27,62	27,50	27,35	27,21	27,09	26,96
0,6	28,09	27,75	27,67	27,58	27,48	27,39	27,30	27,22
0,8	28,11	27,76	27,70	27,63	27,56	27,49	27,43	27,37
1,0	28,10	27,75	27,69	27,64	27,58	27,52	27,48	27,42
1,2	27,99	27,66	27,59	27,58	27,51	27,45	27,43	27,38

TABLE V
FEM-3D CALCULATED CMC SECONDARY WINDING LEAKAGE REACTANCES PER TURN

I_i/I_{ri}	Only CMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	$x_{\sigma 2ei}$ [μΩ]							
0,2	8,910	8,801	8,713	8,630	8,537	8,462	8,370	8,305
0,4	8,913	8,806	8,760	8,723	8,676	8,630	8,592	8,551
0,6	8,911	8,803	8,775	8,747	8,716	8,688	8,661	8,633
0,8	8,915	8,806	8,785	8,764	8,741	8,720	8,699	8,681
1,0	8,914	8,801	8,784	8,767	8,749	8,730	8,715	8,698
1,2	8,877	8,772	8,752	8,747	8,726	8,705	8,701	8,685

The leakage reactances of the four windings can be calculated by numerical methods only, because they include the non-linearity of the magnetic cores and their mutual electromagnetic coupling. By using the program package FEM-3D they have been iteratively derived and are given in Tables II-V.

III. POSTPROCESSING OF THE FINITE ELEMENT METHOD ANALYSIS RESULTS

A. Case Study No. 1-Epstein Frame

After FEM-3D calculations of the magnetic vector potential are accomplished, by using numerical integration, and the exact 3D magnetic field distribution in the frame is derived. In the IEC standard 60404-2, [1] the effective magnetic path length l_m is taken as constant and equal to 0,94 m. Still, in [4] it has been shown that it depends on the material as well as on the magnetic polarisation. The l_m has been derived for three reference specimen of electrical steel sheet as shown in Fig. 4.

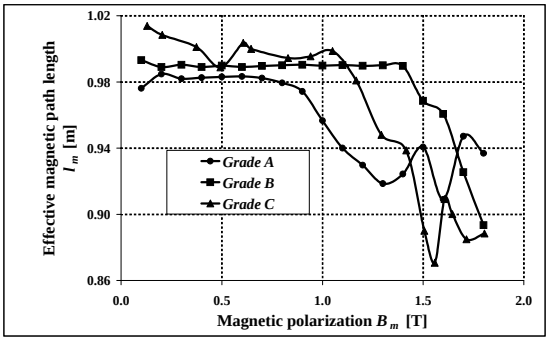


Fig. 4. FEM-3D calculated effective magnetic path length l_m via the magnetic polarization

The precise finite element estimation of the magnetic path length l_m of the EF enables reduction of the measurement uncertainties during the process of experimental testing of the electrical steel sheet properties (e. g. the specific power losses P_s=f(J_m) or the magnetic characteristics H=f(J_m)).

B. Case Study No. 2-Combined Instrument Transformer

The FEM-3D derived values of the CCVIT leakage reactances are introduced in the CCVIT metrological parameters given in Tables VI-XI.

TABLE VI
VMC NO LOAD VOLTAGE ERROR

I_i/I_{ri}	only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	p_{u0} [%]							
0,2	-0,34	-0,34	-0,31	-0,26	-0,22	-0,18	-0,15	-0,11
0,4	-0,28	-0,28	-0,26	-0,25	-0,23	-0,22	-0,20	-0,19
0,6	-0,24	-0,24	-0,23	-0,23	-0,22	-0,21	-0,20	-0,19
0,8	-0,22	-0,22	-0,22	-0,21	-0,21	-0,20	-0,19	-0,19
1,0	-0,21	-0,21	-0,20	-0,20	-0,20	-0,19	-0,19	-0,18
1,2	-0,21	-0,21	-0,20	-0,20	-0,19	-0,19	-0,19	-0,18

TABLE VII
VMC VOLTAGE ERROR INTRODUCED BY THE LOAD

I_i/I_{ri}	only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	p_{ul} [%]							
0,2	-12,35	-12,3	-11,5	-9,91	-8,69	-7,46	-6,26	-5,07
0,4	-6,20	-6,20	-5,90	-5,59	-5,29	-4,98	-4,68	-4,38
0,6	-4,10	-4,10	-3,97	-3,83	-3,70	-3,56	-3,43	-3,30
0,8	-3,08	-3,08	-3,01	-2,93	-2,85	-2,78	-2,70	-2,63
1,0	-2,47	-2,47	-2,42	-2,39	-2,32	-2,27	-2,23	-2,18
1,2	-2,06	-2,06	-2,03	-1,99	-1,96	-1,93	-1,89	-1,86

TABLE VIII
VMC VOLTAGE ERROR

I_i/I_{ri}	only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	p_u [%]							
0,2	-12,69	-12,6	-11,8	-10,2	-8,92	-7,65	-6,41	-5,18
0,4	-6,48	-6,48	-6,16	-5,84	-5,52	-5,20	-4,88	-4,57
0,6	-4,35	-4,35	-4,21	-4,06	-3,92	-3,77	-3,63	-3,49
0,8	-3,31	-3,31	-3,23	-3,15	-3,07	-2,98	-2,90	-2,82
1,0	-2,68	-2,68	-2,63	-2,60	-2,52	-2,47	-2,42	-2,37
1,2	-2,27	-2,27	-2,23	-2,20	-2,16	-2,12	-2,08	-2,05

TABLE IX
VMC NO LOAD PHASE DISPLACEMENT ERROR

I_i/I_{ri}	only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	δ_{u0} [min]							
0,2	-4,99	-4,96	-4,41	-3,39	-2,59	-1,79	-1,00	-0,22
0,4	-4,77	-4,77	-4,41	-4,04	-3,67	-3,33	-2,94	-2,59
0,6	-4,95	-4,95	-4,70	-4,45	-4,21	-3,90	-3,71	-3,48
0,8	-4,72	-4,72	-4,54	-4,37	-4,19	-4,02	-3,84	-3,68
1,0	-4,48	-4,48	-4,34	-4,27	-4,08	-3,95	-3,82	-3,69
1,2	-4,43	-4,43	-4,32	-4,21	-4,11	-3,99	-3,88	-3,78

TABLE X
VMC PHASE DISPLACEMENT ERROR INTRODUCED BY THE LOAD

I_i/I_{ri}	only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	δ_{ul} [min]							
0,2	-407	-405	-366	-295	-240	-183	-128	-73,9
0,4	-205	-205	-191	-177	-163	-149	-135	-121
0,6	-135	-135	-129	-122	-116	-110	-104	-98,3
0,8	-101	-101	-98,3	-94,8	-91,3	-87,8	-84,3	-80,9
1,0	-81,5	-81,5	-79,3	-78,1	-74,8	-72,6	-70,4	-68,2
1,2	-68,1	-68,1	-66,5	-65,0	-63,4	-61,9	-60,3	-58,8

TABLE XI
VMC PHASE DISPLACEMENT ERROR

I_i/I_{ri}	only VMC	0	0,2	0,4	0,6	0,8	1,0	1,2
U_u/U_{ru}	δ_u [min]							
0,2	-412	-410	-371	-299	-242	-185	-129	-74,1
0,4	-209	-209	-195	-181	-166	-152	-138	-124
0,6	-140	-140	-133	-127	-120	-114	-108	-101
0,8	-106	-106	-102	-99,2	-95,5	-91,8	-88,1	-84,6
1,0	-86,0	-86,0	-83,6	-82,4	-78,9	-76,5	-74,2	-71,9
1,2	-72,5	-72,5	-70,8	-69,2	-67,5	-65,9	-64,2	-62,6

TABLE XI
CMC CURRENT ERROR

U_u/U_{ru}	only CMC	0	0,2	0,4	0,6	0,8	1,0	1,2
I_i/I_{ri}	p_i [%]							
0,2	-1,16	-1,16	-1,16	-1,16	-1,16	-1,16	-1,16	-1,16
0,4	-0,89	-0,89	-0,89	-0,89	-0,89	-0,89	-0,89	-0,89
0,6	-0,75	-0,75	-0,75	-0,75	-0,75	-0,75	-0,75	-0,75
0,8	-0,68	-0,68	-0,68	-0,68	-0,68	-0,68	-0,68	-0,68
1,0	-0,68	-0,68	-0,68	-0,68	-0,68	-0,68	-0,68	-0,68
1,2	-0,64	-0,64	-0,64	-0,64	-0,64	-0,64	-0,64	-0,64

TABLE XI
CMC PHASE DISPLACEMENT ERROR

U_u/U_{ru}	only CMC	0	0,2	0,4	0,6	0,8	1,0	1,2
I_i/I_{ri}	δ_i [min]							
0,2	8,47	8,54	8,59	8,64	8,69	8,74	8,79	8,83
0,4	3,82	3,87	3,89	3,90	3,92	3,95	3,96	3,98
0,6	2,59	2,63	2,64	2,65	2,66	2,67	2,68	2,69
0,8	0,84	0,87	0,88	0,89	0,90	0,90	0,91	0,92
1,0	0,50	0,54	0,54	0,55	0,55	0,56	0,57	0,57
1,2	0,17	0,20	0,21	0,21	0,22	0,23	0,23	0,23

IV. CONCLUSION

In this paper an original methodology for measurement devices uncertainty estimation, by using the finite element method, has been given. The derived results can be further used for metrologically improved design of the Epstein frame and the combined instrument transformer.

REFERENCES

[1] IEC (International Electrotechnical Commission) 60404-2, 1996: Magnetic Materials, Part 2: Methods of measurement of magnetic properties of electrical steel sheet and strip by means of an Epstein frame, Geneve, 1996.

[2] IEC (International Electrotechnical Commission) 60044-2, 1980: Instrument transformers, Part 3: Combined transformers, Geneve, 1980.

[3] M. Cundev, L. Petkovska, The Weighted Residuals Method for Electromagnetic Field Problems in Electrical Machines, Proc. of the 32nd Universities Power Engineering Conference UPEC'97, Vol. 2, UMIST, Manchester, UK, pp.934-937, 1997.

[4] M. Cundeva, L. Arsov, Experimental Verification of the Results from the FEM-3D Magnetic Field Analysis of Epstein Frame, Studies in Applied Electromagnetics and Mechanics, vol. 22, IOS Press Ohmsha, Amsterdam, pp.379-384, 2002.

First Year of Wind Parameters Measurements in Macedonia

Vladimir Dimcev, Krste Najdenkoski and Vlatko Stoilkov

Abstract — In Macedonia there is no accurate knowledge of country's wind resources and it is a major barrier for any possible development of the utilization of wind power. The paper explains developments in last couple of years considering exploration of Macedonian wind resources. First of all a Wind Atlas of Macedonia was created in 2005. According the Wind Atlas the sites with the most promising wind potential were chosen. On four of this sites complete measurement stations were installed in June, 2006. The measurement campaign is lasting for 12 -15 months and this year the first results are available.

Keywords — Wind energy, wind atlas, measurement of wind parameters.

I. Introduction

In last 15-20 years the wind energy exploitations faced dramatic development world-wide. The target set by EU commission for 10% wind generation penetration in overall electricity production in European Union by the 2010 now it seems feasible, and this goal will be reached even earlier. In Macedonia few dedicated professors from Faculty of Electrical Engineering (FEE) in Skopje started to push the idea that investigation of country wind resources is something that should be done as soon as possible. In fact the first step for wind farms development is to determine accurately the wind resources and potential wind energy production of a future wind farms in selected candidate sites. Unfortunately in Macedonia there is no accurate knowledge of country's wind resources and it is a major barrier for any possible development of the utilization of wind power. The available wind speed information in Macedonia originates from the national network of meteorological stations, which - as in the rest of the world – used alone are not sufficient for accurate wind resource assessment.

This situation was inspiration to make strategic plan for investigation of wind resources and potential and possible development of wind farms in the near future. The plan is consisting of three main phases: preparation of wind atlas which is numerical modeling based on geophysical and meteorological inputs, conducting of measurement campaign on the most promising sites defined from the atlas and preparation of feasibility studies as basis for possible erection of wind farms.

Authors are with the Faculty of Electrical Engineering, Skopje, Republic of Macedonia, e-mail: vladim@etf.ukim.edu.mk;: krste@etf.ukim.edu.mk; stoilkov@etf.ukim.edu.mk

The first phase was conducted by ESM (the former Electric Power Company of Macedonia) and Wind Atlas was prepared by AWS Truewind Company from U.S.A in May and June, 2005. The second phase, now underway, is measurement campaign of wind parameters on selected sites. This phase is sponsored by Norwegian Ministry of Foreign Affairs with active participation of FEE and ELEM.

The four complete measurement stations with sensors, data loggers and musts were provided in April, 2006 and installed in June, 2006. The height of towers is 50 meters.

II. Wind Energy Atlas

The Wind Atlases are based on numerical modeling of the large-scale climatology of the atmosphere. The inputs for modeling are typically taken from global databases: the wind field could be taken from the NCEP/NCAR database, the land-use from the GLCC data base of USGS (United States Geological Survey) and the height information of the country from the SRTM30 database by NGA (National Geospatial-Intelligence Agency) and NASA. This information is validated with all other available information (e.g. maps, satellite images, etc.).

In our case the medium scale modeling has been made to resolve atmospheric phenomenon on scales down to the order of 10's of kilometers. The model enables an understanding of the overall wind resource to such an extend that informed choices can be made for setting up measuring campaigns on selected sites. This wind atlas represents the mean of winds modeled in the medium scale class simulations after adjustment to specific standard surface conditions, uniform roughness and level terrain.

The Atlas output are two main products: (1) color maps of mean wind speed and power density at various heights above ground and (2) data files containing wind speed and direction frequency distribution parameters. The maps and data can be compared with actual wind measurements, if any are available, and adjustments to the wind maps can be made.

A. Wind maps

The Atlas are consisting of maps for wind speed at four heights from the ground (40m, 60m, 80m and 100m, hard copy) together with the ArcReader software which allows user to obtain the "exact" wind speed and wind direction values at any point. The wind maps are showing that the best wind

resources in Macedonia is generally found along high mountain ridges, while lowlands and valleys are likely to have much lower average wind speeds. The predicted mean wind speed at 80m height on the ridge tops varies from 6,5m/s to 8,5m/s. Unfortunately the windiest areas are at elevations above than 2000 m, which may be expensive for wind projects.

Sites of moderate to good wind resources exist at lower elevation, however. Perhaps the most important examples are the hills on either sides of the Vardar River between Demir Kapija and Gevgelija, where the predicted mean wind speed reaches 7 – 7,5m/s at 500 -800m above see level. The river valley forms a gap in the mountain range, through which the wind flow is concentrated under some weather conditions.

The surface roughness surrounding a site also plays a big role, open fields and bodies of water experience greater wind speeds than urban or forested areas. It should be clear that the predicted wind speed at any particular location by the Atlas may depart substantially from real values due to errors on surface roughness, land cover, etc.

B. Site selection

The analysis identified 15 candidate sites for wind farms, which were selected and ranked on the base of predicted mean wind speed, wind plant size, turbine output, cost of energy, interconnection cost and other factors. A brief summary of parameters for each site is provided in Table 1. The predicted wind resource at each point on the map was combined with elevation and temperature data to estimate the net annual output of 1,5 MW turbine with a 77 m rotor diameter and 80 m hub height.

The GIS-based approach considered the factors like: the costs for new transmission lines for connection to electricity grid, the costs for new access roads, and terrains with slope greater than 20% based on SRTM digital elevation model, the gross turbine output was reduced by 12% to account planned maintenance and outages. These factors were combined with the minimum output for wind farm of 25 MW to predict the cost of energy (COE).

The amount of land area required for a 25 MW wind farm project depends on the type of terrain, direction of the wind, and other factors. In the program were assumed two extreme cases: a broad valley or plain with a multi-directional wind resources and a narrow ridge with a nearly one-directional wind resource. The program interpolated between these two extremes depending on the shape of the site.

The selected 15 candidate sites have a potential plant capacity between 25 MW and 33 MW. The cost of energy (COE) for the sites range from $0,07/kWh to $0,094/kWh, assuming a 25MW plant capacity, 15% fixed charge rate and no subsides or other incentives for wind projects. The predicted mean wind speed ranges from 6,7 m/s to 8,4 m/s. These figures are within the normal range for commercial wind projects being developed today.

An AWS Truewind engineer together with ESM engineers carried out a field inspection of selected sites between May 29 and June 3, 2005. The purpose of the trip was to assess the accuracy of the data used in the site selection, to assess the suitability of mountain ridge for installation of meteorological masts and wind turbines, to check the quality of roads, any restricted areas such as military facilities or telecommunication towers and microwave links.

After the site screening was completed and careful consideration of all obtained data by the Atlas were made, four sites were chosen for wind measurement campaign.

Site 7 - Kozuf Mountain has greatest potential for development of wind farm. It is large open grassland and the top of the ridgeline consists of gently rolling hills. The site elevation between 1300m to 1760m may produce problems during installation phase and maybe for maintenances of turbines.

Site 10 – Ranovec Mountain, Bogdanci, the ridge has 450-500 m elevation and an east-west orientation in an area where the wind is predominantly from the northwest-southeast. The vegetation on the site is grass and low shrubs and the site is located between three 110kV lines linked into triangle. Maybe this site is the most promising.

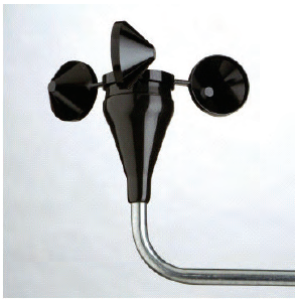
TABLE 1. The predicted parameters of 15 chosen sites from Wind Atlas

General site data				25 MW Capacity				
ID	LON	LAT	Avg Elevation (m)	SPD 80m (m/s)	PWR 80m (W/m2)	AREA (km2)	DENS (MW/km2)	MW
1	21,02	41,25	1896	8,41	700	2,1	12	25
2	20,7	41,41	2079	7,97	502	2,1	12	25
3	22,39	41,35	566	7,35	482	2,2	11,5	24,9
4	20,82	41,54	1994	7,63	482	2,1	12	25
5	20,52	41,26	2088	7,85	690	2,1	12	25
6	21,95	42,31	1159	7,53	518	2,1	12	25
7	22,3	41,25	1453	7,45	581	2,1	12	25,4
8	22,27	41,4	641	6,96	370	2,2	12	26,4
9	20,81	42	2511	8,06	640	2,1	12	25,4
10	22,56	41,23	408	7,04	502	2,2	11,4	25
11	22,46	42,13	2003	7,3	488	2,1	12	25
12	22,18	41,15	1998	7,43	666	2,2	12	25,9
13	20,72	41,8	2134	7,13	413	2,1	12	25
14	21,38	41,72	2319	7,29	566	2,3	12	27,4
15	22,45	41,78	1577	6,68	384	2,2	12	25,9

Site 16 – Sasavarlija, Stip the site was identified by ESM people. It is located around 20 kilometers to the southeast from the town of Stip on highland with several dispersed hills and a maximum elevation of 996 m. The predicted mean wind speed at 80 meters is 5,81 m/s with capacity factor of 22,44%.
Site 20 – Bogoslovec, Sveti Nikole this site was also identified by ESM people. The site is on short ridge of up to 750 m elevation. The site has moderate winds of 6,0 – 7,0 m/s at 80 m according to the wind map and is estimated to have the potential to accommodate wind farm to 10 MW

III. MEASUREMENT CAMPAIGN

The objective of measurement campaign is to measure and to analyze the wind parameters on four locations with promising potential of wind energy. The final goal is to obtain valid results for feasibility studies of wind energy potential on selected locations and final decision for building wind farms. This is the next phase, after the preparation of Wind Atlas, in overall scientific determination of regions and locations with promising wind energy potential. This initiative is perfectly fitted with governmental strategy for developing of renewable resources. The possible development of wind power plants in the near future will also have positive implication on the local economy.
The project is running by FEE (Faculty of Electrical Engineering) – Skopje together with ELEM (Electric Power Plants of Macedonia).
For the realization of the measurement campaign four complete measurement stations are provided and installed on four previously mentioned sites. The height of the towers is 50m, this is a trade off between being as close possible to hub heights of today's wind turbines (80 m or higher) and reducing cost of measurements towers.



The measurements system is consisting of four three cups anemometers, two wind direction vanes and thermometer (fig. 1.a and 1.b). All sensors are by NRG Systems Company, USA and high quality anemometer on the top of the towers are made by RISO, Denmark. The wind speed sensors are calibrated according to the MEASNET standard with accuracy class of 0,1m/s, which is used by the wind industry in Europe. Two of the anemometers are on the top of the mast at 50 m height, the third is on 30 m height and the last is on 10 m height above the ground. The wind direction is measured in two different heights for redundancy. Temperature is measured at 2 m, which is a meteorological standard.
In addition to sensors there are data loggers (fig. 2) inside weather proof cabinets with communication capabilities over GSM line. The data loggers is an internet ready, ultra-low power microprocessor-controlled data logging system, specially designed for the wind energy industry. Together with the data logger there is iPack system, which enables transfer of data via GSM to the local Internet provider than emails are sent to the final users. The iPack device is connected to the back of the logger and its battery is charged with solar panel.
The logger has a 12 channel, 6 counter inputs for wind sensors or other frequency signals and 6 analog inputs for wind direction, temperature, barometric pressure, etc. The storage medium is non-volatile FLASH 16 MB multimedia card, which provides 2 years time of data storage.

The sensors are sampled every 2 seconds, but the average values, minimum and maximum values, standard deviation are calculated for time interval of every 10 minutes, according IEC 61400-12. The data loggers are sending data once per day to dedicated computers.
The data are processing and archiving in a well-defined, systematic and consistent manner. All raw data received via remote communications are archiving monthly into a complete database. The raw files are converting to engineering values (wind speed, wind direction, temperature) using appropriate software.
Incoming data are checking regularly for any signs of equipment damage and/or malfunction. The completeness of the received data files are analyzing weekly and the data are analyzing on a bi-weekly basis for errors and equipment failures. The need for accuracy is usually higher for wind energy projects than for weather forecasting. The analysis of data will be performed in close cooperation with the local meteorological office.

IV. OBTAINED RESULTS

The obtained resilts are shown on the figures 3 to 7. Figure 3 presents the average values of wind speed during one month

for one of the locations, the obtained value is 6.3 m/s. Next figure 4 presents wind rose for the same location and same month. Dominant direction of the wind is north-northwest. Relative frequency distribution of wind speed for the same location and month is shown on fig. 5.

The short analysis of obtained results from the four anemometers mounted at one measurement station for 7 months is presented in fig. 6.

Next phase of the project is to analyze all results obtained from the measurements and to prepare feasibility study of the chosen locations.

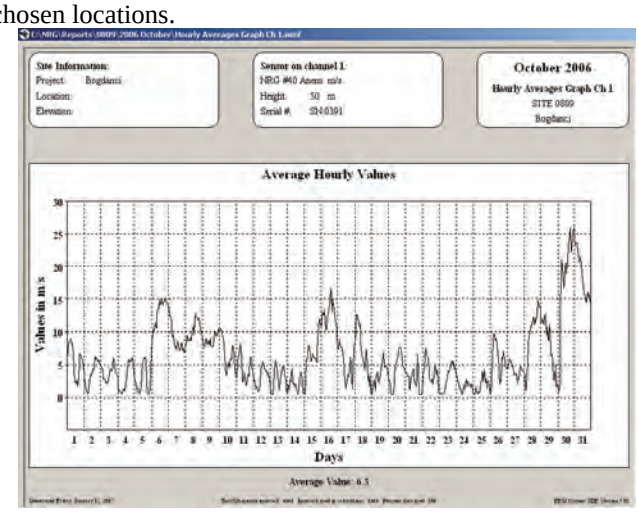


Fig. 3 Average 10 minutes wind speed for one month

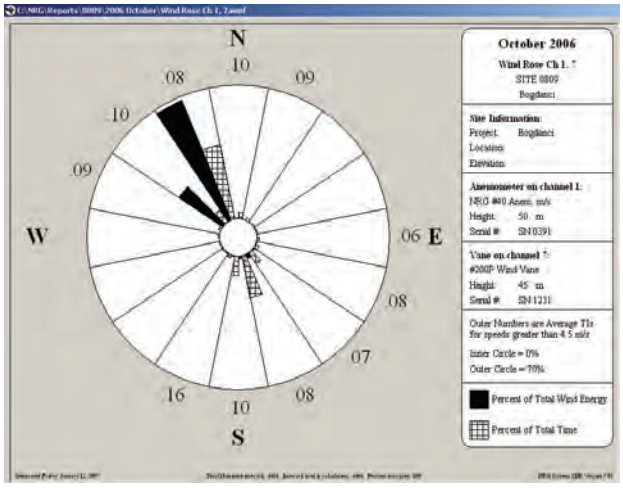


Fig. 4 Wind rose for one month

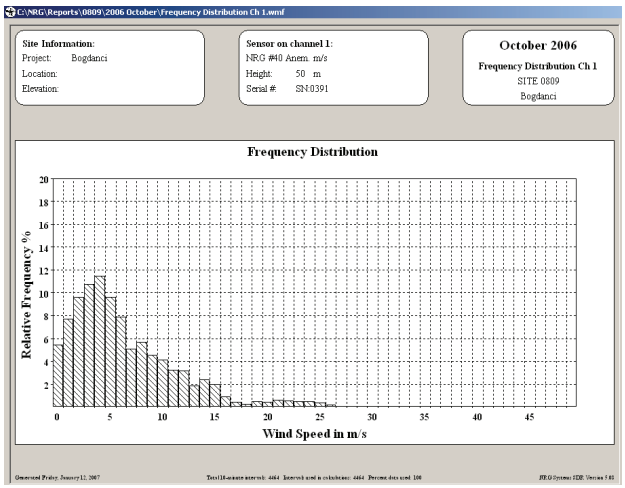


Fig. 5 Relative frequency distribution of wind speed

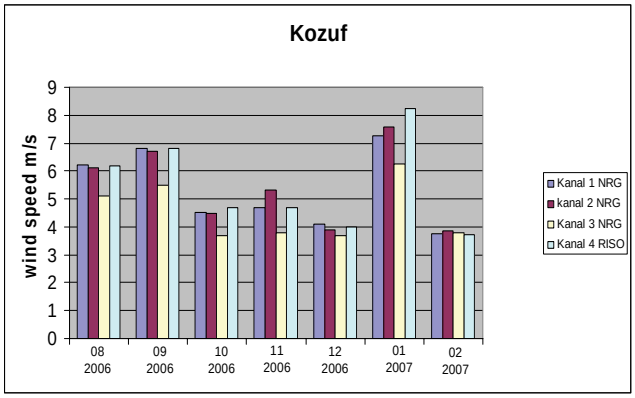


Fig. 6 Average wind speeds from 4 sensors for 7 months

V. Conclusion

The paper explains latest developments considering exploration of Macedonian wind resources. A Wind Atlas of Macedonia was created in 2005. The maps formed the basis for a GIS-based selection of 15 prospective sites where measurements should be done. Another five additional sites were proposed by ESM (now ELEM). From those 20 sites, after on sites visits four most promising sites were chosen for the measurement campaign.

The four complete measurement stations were installed in June, 2006. The planned measurement campaign is going for ten months and will last another 3 - 6 months. The final goal is to provide reliable results for feasibility studies of wind energy potential on selected locations and final decision for building wind farms.

VI. References

[1] Wind Energy Resource Atlas and Site Screening of the Republic of Macedonia, AWS Truewind, June, 2005.

[2] Monitoring Program of Macedonian Wind Resources, Project Funded by Norwegian Ministry of Foreign Affairs, November, 2005.

[3] V. Dimcev, K. Najdenkoski, V. Stoilkov: Exploration of wind energy potential in Republic of Macedonia, Balkan Power Conference, Ohrid, 2006.

[4] http://www2.jpl.nasa.gov/srtm/

[5] http://edcdaac.usgs.gov/modis/mod12q1.asp

SESSION CSIT I

Computer Systems and Internet Technologies I

Low-Power Multihop Routing for Wireless Sensor Networks

Zdravko Karakehayov

***Abstract* – This paper presents ALS-MAC, a medium access control protocol for wireless sensor networks. The protocol employs beacons to synchronize internode communications. Scalability and collision avoidance are achieved via contention-based advertising slots mapped to scheduled-based transmission slots. ALS-MAC has two dedicated modes of operation for data intensive traffic which assign temporary priority to nodes with many or long buffered packets. We study collision-free concurrent data transmissions and multihop optimization schemes under ALS-MAC.**

Keywords – Wireless sensor networks, MAC protocol, Low-power routing

I. INTRODUCTION

Distributed sensor networks are made up of a large number of small sensing nodes which cooperatively perform complex tasks. Environmental monitoring, healthcare, building automation, surveillance and rescue missions are applications where wireless ad-hoc networks provide benefits that we would not otherwise be able to obtain.

The interaction between the nodes is based on wireless communication. Sensor-actuator networks employ sensors to gather information and actuators to perform appropriate actions in a area of interest. Since the energy is a scarce and usually non-renewable resource, the functionality of distributed sensor-actuator networks must be viewed from low-power perspective. Normally, the nodes have a limited radio footprint and packets are forwarded in a multihop manner. When a node receives a packet it applies a routing algorithm to select a neighbor for forwarding. Different criteria can guide the local decision. One approach is to choose the closest to the destination neighbor.

The greater than linear relationship between transmit energy and distance promises to reduce the energy cost when the radio link is partitioned. Nodes calculate the distance and tune their transmit power accordingly. Consequently, it would be beneficial to forward the packet via several hops instead a single link.

Medium access control (MAC) mechanism has a significant impact on the energy efficiency. Currently available MAC protocols for wireless sensor networks can be broken down into two major types: contention based and scheduling based. While under contention-based protocols nodes compete among each other for channel access, scheduling-based schemes rely on prearranged collision-free links between nodes. There are different methods to assign collision-free links to each node. Links may be assigned as time slots (TDMA), frequency bands (FDMA), or spread spectrum codes (CDMA). However, size and cost constrains may not permit allocating complex radio subsystems for the node architecture. Logically, TDMA scheduling is the most common scheme for the domain of wireless sensor networks. The limited communication range of network nodes provides an extra opportunity for collision-free interaction, space division access (SDMA). Two major cases of redundant energy consumption are associated with contention-based communication. Collision occurs when two or more nodes transfer data to a single node at the same time. Overhearing is the situation when a node receives a packet which is not directed to it. Location-aware nodes and radios with variable power levels would help alleviate both collision and overhearing.

In this paper we present a MAC protocol for wireless sensor networks, ALS-MAC. The protocol achieves good scalability and collision avoidance by utilizing contention-based advertising slots mapped to scheduled-based transmission slots. ALS-MAC provides extra opportunities for collision-free data transmissions and low-power multihop routing which we study for non-regular topologies.

II. RELATED WORK

Previous approaches have studied MAC protocols and multihop low-power routing separately. Span, a distributed, randomized algorithm declines the power consumption via short advertising periods and long beacon cycles [1]. S-MAC medium access control protocol establishes a low duty cycle operation in nodes [2], [3]. TRAMA, another MAC protocol, reduces energy consumption by avoiding collisions via a distributed election scheme [4]. ExOR (Extremely Opportunistic Routing) is a routing method developed to reduce the total number of transmissions taking into account the actual packet propagation [5]. DTA, data transmission algebra, has been developed to generate complex transmission schedules based on collision-free concurrent data transmissions [6]. Multihop optimization has already been studied for simple linear settings [7], [8]. In related research we studied multihop optimization for general topologies under a simplified 802.11 MAC protocol [9].

Zdravko Karakehayov is with the MCI-CPD, University of Southern Denmark, Grundtvigs Alle 150, DK-6400, Sønderborg, Denmark, E-mail: zdravko@mci.sdu.dk

III. NOTATION

Assume that the nodes of a wireless sensor network are members of the following set: $N = \{N_1, N_2, N_3, ... N_{n(N)}\}$. Nodes are placed in a rectangular region of *X* by *Y*. The distance between node i and node j is *d*(i,j). The distance between node k and the halfway point between node i and node j is $d(k, m_{i,j})$. Routing algorithms are employed to determine the next hop node of N_i, N_i^{+1}. The distance between N_i and its next hop N_i^{+1} is *d*(i,+1). Likewise, the distance between N_k and the halfway point between N_i and N_i^{+1} is $d(k, m_{i,+1})$. R is the radio communication range.

A statement **power**(*d*(i,j)) in the pseudocode listing adjusts the transmit power according to the distance *d*(i,j) to save energy. A statement **send**$(N_i \rightarrow N_j)$ indicates a packet forwarding from node i toward node j.

Algorithm 1 $N_i^{+1} \leftarrow \text{NextHop}(N_i, N_D, N_i^R)$

```
if N_D ∈ N_i^R
    return N_D
end if
s = (X^2 + Y^2)^{1/2}
for 1 ≤ j ≤ n(N), j ≠ i do
    if N_j ∈ N_i^R and d(j,D) < s
        N_i^{+1} = N_j, s = d(j,D)
    end if
end for
```

Algorithm 1 describes a single-hop selection procedure used to find N_i^{+1}, the next hop of N_i. N_D is the final destination. N_i^{+1} is the closest to the destination neighbor of N_i.

IV. HARDWARE MODEL

A typical node is built around a low-power microcontroller. Wireless transceivers create physical links between nodes. Hardware provides the following low-power mechanisms.

- Each node is capable of determining its coordinates.
- The receiver and transmitter can be individually enabled and disabled.
- The transmit power can be adjusted gradually.

V. ALS-MAC PROTOCOL

We propose a single channel MAC protocol suitable for location-aware sensor networks. We call this procedure Advance Local Scheduling, ALS-MAC. In order to save energy, nodes should stay in a sleeping mode as long as possible. Periodically, nodes must wake up and receive the packets buffered for them. ALS-MAC employs beacons to synchronize internode communications. A beacon period, T_B, includes two major sections. The period begins with a contention-based, time-slotted traffic indication window, T_A. During T_A all nodes are listening and pending packets are advertised. The set of slots

$$S = \{S_0, S_1, S_2, ... S_{n(S)}\} \quad (1)$$

In an IEEE 802.11 style, short request-to-send (RTS) and clear-to-send (CTS) control frames are exchanged. The second section of the beacon period is a collection of transmission slots for scheduled access. The number of the slots is equal to the number of the slots in the indication window. Once a node succeeds in establishing a session via RTS and CTS frames in the traffic indication section, it gets an access to the corresponding slot in the data transmission section. Data transmissions are followed by acknowledgement (ACK) frames to confirm successful reception. The one-to-one mapping between advertising and transmission slots allows all nodes to sleep during the second section and wake up to exchange packets only at specific slots.

Along with this essential mode, ALS-MAC is capable of organizing three extra interaction schemes. L-mode and M-mode help to adapt to different traffic demands. S-mode aims merely to keep nodes well synchronized for future sessions.

L-mode. A node generates a L-beacon when the number of packets it has already buffered is greater than a predefined threshold. The beacon announces a number of slots equal to the number of buffered packets. Addressed nodes send ACK frames in the same order they have been indicated in the beacon. Neighbors that are not involved in the advertised traffic are still allowed to participate. However, ALS-MAC discipline requires they to follow collision-free SDMA with regard to the advertised links. L-mode allows to transmit very long packets when a single data frame utilizes a complete transmission section.

S-mode. S-beacons are broadcast by nodes which have no packets to transmit. The motivation behind S-beacons is to maintain synchronization. When a node does not receive beacons for a predefined number of periods, it generates a S-mode frame. If neighbors have buffered packets, they exchange control and data frames in the same way they do after normal beacons.

M-mode. This interaction scheme allows multihop routing within a single beacon period. The beacon indicates one or more routing paths. All involved nodes are within the communication range. Nodes can advertise buffered packets in the following slots or immediately after the beacon if they do not interfere with the routing paths. The main advantage of M-mode is energy-efficient partitioning of the communication links and collision avoidance. Also, M-beacons can be used to send a single packet to multiple nodes.

Fig. 1 shows an example when a source sends a packet to a destination via an intermediate node. Since the routing path has been indicated in the M-beacon, the destination sends CTS in the first advertising slot. The intermediate node receives this frame and sends back its own CTS indicating also the status of the path. As a result, the source will know

which nodes are available. The interaction goes on with data packets and acknowledgements.

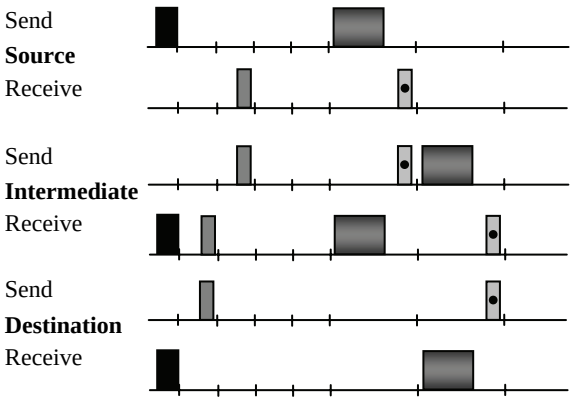


Fig. 1. Forwarding under M-mode

VI. ENERGY MODEL

The energy used to send a bit over a distance d may be written as $E = ad^n$, where a is a proportionality constant [10], [11]. The radio parameter n is a path loss exponent that describes the rate at which the transmitted power decays with increasing distance. Typically, n is between 2 and 4 [10]. $E = ad^n + b$ emerges as a more realistic model. The b constant is associated with specific receivers, CPUs and computational algorithms. The power consumption of a turned on receiver is yet another constant, P_R.

VII. MULTIHOP OPTIMIZATION

Fig. 2 shows a partitioning of the link in the general case, the intermediate node is neither equally spaced from S and D, nor located on the line SD.

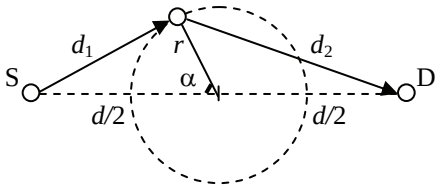


Fig. 2. Partitioning the link in the general case

Theorem 1. Let the energy per bit sent over a distance d scales as $ad^4 + b$. The power consumption of a turned on receiver is P_R. The bit rate is B. Referring to Fig. 2, if the distance

$$d \geq \left((8P_R + 8bB)/7aB\right)^{1/4} \tag{2}$$

and the distance between the intermediate node and the halfway point between S and D

$$r \leq \left(-1.25d^2 + \left(2d^4 - 0.5a^{-1}B^{-1}P_R - 0.5a^{-1}b\right)\right)^{1/2}\Big)^{1/2} \tag{3}$$

the two-hop communication requires less energy than the direct link.

Proof. We must prove when the following inequality holds. Suppose the communication range is R, the packet has p bits and signaling frames are q bits long. The advertising part of the beacon period is T_A and the number of slots is n(S).

$$\begin{aligned}
&q(aR^4 + b) + p(ad_1^4 + b) + T_A P_R + 2(q/B)P_R + 2q(ad_1^4 + b) \\
&+ p(ad_2^4 + b) + (T_A / n(S))P_R + T_A P_R + (p/B)P_R + (q/B)P_R \\
&+ 2q(ad_2^4 + b) + (T_A / n(S))P_R + T_A P_R + (p/B)P_R \\
&+ (n(N_S^R) - 2)((T_A / n(S))P_R + T_A P_R) \leq q(aR^4 + b) + p(ad^4 + b) + T_A P_R \\
&+ (q/B)P_R + 2q(ad^4 + b) + (T_A / n(S))P_R + T_A P_R + (p/B)P_R \\
&+ (n(N_S^R) - 1)((T_A / S)P_R + T_A P_R)
\end{aligned} \tag{4}$$

$$d_1 = \left(d^2/4 - dr\cos\alpha + r^2\right)^{1/2} \tag{5}$$

$$d_2 = \left(d^2/4 + dr\cos\alpha + r^2\right)^{1/2} \tag{6}$$

Thus, we get

$$16aBr^4 + 8ad^2B\left(1 + 4\cos^2\alpha\right)r^2 - 7ad^4B + 8P_R + 8bB \leq 0 \tag{7}$$

Since the threshold value for the distance r will vary with α, we take the worst case, $\cos\alpha = 1$. Then the inequality has solutions if and only if $d \geq \left((8P_R + 8bB)/7aB\right)^{1/4}$. Using the quadratic formula,

$$r \leq \left(-1.25d^2 + \left(2d^4 - 0.5a^{-1}B^{-1}P_R - 0.5a^{-1}b\right)\right)^{1/2}\Big)^{1/2} \tag{8}$$

VIII. SDMA UNDER M-MODE

Calculated SDMA under M-mode allows collision-free parallel data transmissions. Assume that a M-beacon includes a single routing path. The locations of the nodes involved constitute the following set:

$$G = \left\{L_1, L_2, L_3, \ldots L_{n(G)}\right\} \tag{9}$$

Algorithm 2 describes a search procedure for a collision-free slot. Such a slot can be used to send a packet from N_i to N_i^{+1} in parallel with the advertised in the beacon routing path.

Algorithm 2 $S_j \leftarrow \mathrm{SDMA}\left(N_i, N_i^{+1}, G\right)$

```
if n(S) > n(G) − 1
    return j = n(S) - (n(G)-1)
end if
for 0 ≤ j ≤ n(S) -1 do
    if
    MIN[d(L_{j+1}, N_i), d(L_{j+1}, N_i^{+1}), d(L_{j+2}, N_i), d(L_{j+2}, N_i^{+1})] >
    MAX[d(L_{j+1}, L_{j+2}), d(N_i, N_i^{+1})]
        return j
    end if
end for
j = n(S)
```

Initially, in lines 1-3, we check if the number of available slots is higher than the required number for the transmissions advertised in the beacon. If there are not empty slots, the procedure goes on with comparing the distances between the nodes. As soon as SDMA has been found, the name of the current slot is returned. When the network layout does not allow SDMA, the procedure returns a non-existing slot, $S_{n(S)}$.

Algorithm 3 $S_j \leftarrow \mathrm{SDMA}(N_i, N_i^{+1}, G)$

1: **if** $n(S) > n(G) - 1$
2: **return** $j = n(S) - (n(G) - 1)$
3: **end if**
4: $m = n(S)$
5: $s = 0$
6: **for** $0 \le j \le n(S) - 1$ **do**
7: **if**
$$\mathrm{MIN}\left[d(L_{j+1}, N_i), d(L_{j+1}, N_i^{+1}), d(L_{j+2}, N_i), d(L_{j+2}, N_i^{+1})\right] - \mathrm{MAX}\left[d(L_{j+1}, L_{j+2}), d(N_i, N_i^{+1})\right] > s$$
8: $m = j$,
$$s = \mathrm{MIN}\left[d(L_{j+1}, N_i), d(L_{j+1}, N_i^{+1}), d(L_{j+2}, N_i), d(L_{j+2}, N_i^{+1})\right] - \mathrm{MAX}\left[d(L_{j+1}, L_{j+2}), d(N_i, N_i^{+1})\right]$$
9: **end if**
10: **end for**
11: $j = m$

Algorithm 3 is an improvement of Algorithm 2 that aims to decline the impact of radio irregularity on wireless links [12]. Radio irregularity is a common phenomenon which arises from multiple factors, such as uncontrolled variations of the transmit power and change of the path loss. In case of parallel transmissions, Algorithm 3 selects the slot which is characterized with longest distance between communicating nodes.

IX. Conclusion

This paper presents ALS-MAC, a low-power medium access control protocol. The protocol achieves scalability and collision avoidance via variable transmit power levels and a correspondence between contention-based and scheduled-based slots. ALS-MAC has a special mode to support multihop low-power routing. We proved that any node located within a certain area can be successfully employed to partition the communication link. We discussed algorithms for collision-free parallel data transmissions under ALS-MAC.

References

[1] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: An Energy-efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks", ACM Wireless Networks Journal, vol. 8, no. 5, pp. 481-494, 2002.

[2] W. Ye, J. Heidemann, and D. Estrin, "Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks", IEEE/ACM Transactions on Networking, vol. 12, no. 3, pp. 493-506, 2004.

[3] D. Dewasurendra and A. Mishra, "Design Challenges in Energy-Efficient Medium Access Control for Wireless Sensor Networks", in *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, edited by Mohammad Ilyas and Imad Mahgoub, CRC Press LLC, pp. 28-1 - 28-25, 2005.

[4] V. Rajendran, K. Obraczka, and J.J.Garcia-Luna-Aceves, "Energy-efficient, Collision-free Medium Access Control for Wireless Sensor Networks", Proc. ACM SIGMOBILE Int. Conf. Embedded Networked Sensor Systems, pp. 181-192, 2003.

[5] S. Biswas and R. Morris. "Opportunistic Routing in Multi-hop Wireless Networks", ACM SIGCOMM Computer Communication Review, vol. 34, Issue 1, pp. 69-74, January 2004.

[6] V. Zadorozhny, D. Sharma, P. Krishnmurthy, and A. Labrinidis, "Tuning Query Performance in Mobile Sensor Databases", Proc. 6th Int. conference on Mobile data management, Ayia Napa, pp. 247-251, 2005.

[7] I. Stojmenovic and Xu Lin, "Power Aware Localized Routing In Wireless Networks", IEEE Transactions on Parallel and Distributed Systems, vol. 12, no. 11, pp. 1122-1133, November 2001.

[8] Z. Karakehayov, Low-power Design for Smart Dust Networks, in *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, edited by Mohammad Ilyas and Imad Mahgoub, CRC Press LLC, pp. 37-1 - 37-12, 2005.

[9] Z. Karakehayov, "Low-power Communication for Wireless Sensor-Actuator Networks", Proc. of the Fifth IASTED Int. Conference on Communication Systems and Networks, Palma de Mallorca, Spain, ACTA Press, pp. 1-6, 2006.

[10] J. L. Gao, *Energy Efficient Routing for Wireless Sensor Networks*, Ph.D. dissertation, University of California, Los Angeles, 2000.

[11] J. M. Rabaey, M. J. Ammer, J. L. Silva, D. Patel and S. Roundy, "PicoRadio Supports Ad Hoc Ultra-low Power Wireless Networking", IEEE Computer, vol. 33, pp. 42-48, July 2000.

[12] G. Zhou, T. He, S. Krishnamurthy and J. Stankovic, "Models and Solutions for Radio Irregularity in Wireless Sensor Networks", ACM Transactions on Sensor Networks, vol. 2, no. 2, pp. 221-262, 2006.

Low Power Application Specific Processing Element

Tatjana R. Nikolić[1], Goran Lj. Djordjević[1], and Mile K. Stojčev[1]

***Abstract* – Today the power consumption of integrated circuits (ICs) is considered as one of the most important problems for high performance chips. Many low power techniques have been proposed during the last 15 years [1]. In this paper the structure and instruction set of a processing element (PE), as a basic building block of complex ICs, is described. The micro-power PE's model is described writing efficient register transfer logic (RTL) code from a low- power standpoint using clock gating technique. The PE is implemented in FPGA technology. Simulation results show that power reduction of up to 10 times, in respect to standard PE design solutions, can be achieved using micro-power PE.**

***Keywords* – Low power, Clock gating, Processing element.**

I. INTRODUCTION

The ongoing progress in silicon technology fosters the transition from board level integration towards System-on-Chip (SoC) implementations of embedded systems. According to the International Technology Roadmap for Semiconductors by the end of the decade SoCs will grow to 4 billion transistors running at 10 GHz and operating at one volt [2]. In order to achieve high performance embedded systems three primary goals have to be successfully solved. The first one relates to high-speed computational PE capability, the second one is with an efficient data transfer among SoC's building blocks, and the third one is concerned with micro-power consumption. Systolic arrays, as a regular arrangement of PEs, are good candidate accelerator-architectures that are used in many SoC designs with aim to achieve high computational and communication performance. A challenging problem which should be taken into consideration by the SoC designers now relates both to increase performance and improve energy efficiency. The architectural PE design for low-power is in focus of our interest in this paper.

In particular we will consider RTL synthesis of a simple micro-power PE which represents a constituent of a complex SoC design.

The rest of the paper is organized as follows. In Section II we identify the main sources of power consumption in CMOS circuits. Section III describes the basic principles of clock gating technique. In Section IV the instruction set and hardware structure of the micro-power PE are presented. Section V deals with results obtaining during simulation and implementation of standard- and micro-power-PE on Spartan2 and Virtex4 series. Finally, Section VI gives a conclusion.

[1]Tatjana R. Nikolić, Goran Lj. Djordjević, Mile K. Stojčev are with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: {tatjanas,gdjordj,stojcev}@elfak.ni.ac.yu

II. POWER CONSUMPTION COMPONENTS

Total power dissipation of CMOS inverter is composed of static and dynamic consumption

$$P_{total} = P_{dynamic} + P_{static} \quad (1)$$

The static dissipation can be expressed as

$$P_{static} = I_{leakage} \cdot V_{DD} \quad (2)$$

As can be seen from (2) P_{static} depends on leakage current, $I_{leakage}$, and supply voltage V_{DD}. The leakage current depends on technology and the number of used gate.
The dynamic dissipation is given by the fallowing formula

$$P_{dynamic} = P_{cap} + P_{scc} \quad (3)$$

where P_{cap} is caused by the charging and discharging of parasitic capacitances in the circuit, while P_{scc} is short circuit dissipation.

$$P_{cap} = \alpha_{o\to 1} f_{CLK} \int_0^T i_{V_{DD}}(t) V_{DD} dt = \alpha_{0\to 1} f_{CLK} C_l V_{DD}^2 \quad (4)$$

P_{cap} is proportional to the operating frequency, f_{CLK}, switching activity, α, switched capacitor, C_l, and V_{DD} at power of two. The power dissipation when both transistors are switched on is given by

$$P_{scc} = \alpha_{0\to 1} f_{CLK} I_{sc} (\frac{t_r + t_f}{2}) V_{DD} \quad (5)$$

where t_r corresponds to the rising and t_f to falling pulse edges, respectively. According to (5) we see that P_{scc} directly proportional to the short circuit current, I_{scc}. Approximately, 20% of the total power dissipation correspond to P_{scc}, 75% to P_{cap}, and 5% to P_{static} [3]. In submicron technology dominant influence on power dissipation has $P_{dynamic}$. According to this, the main design effort related to power reduction are directed towards optimal choice of parameters α, C_l, V_{DD}, and f_{CLK}. Theoretically, power reduction can be achieved by decreasing each of the mentioned parameters. In practical realization, designers use ICs with fixed technologically defined supply voltage, so really, in this case there is no choice. The parasitic capacitance, C_l, is defined both by circuit fanout and layout. This means that this parameter is hardly to control. If we suppose now that for execution of some task n steps (operations) are needed, then there is not effect if we decrease f_{CLK}, since the number of transitions is identical in both cases. A parameter to which we can have real influence is α. In other words, α determines the number of transitions. Correct

choice of α is determined by the selected code for data presentation, and by elimination of unnecessary switching in parallel and pipeline computation. An approach based on decreasing the number of signal transitions, α, referred as clock gating, is adopted in this paper.

III. CLOCK GATING TECHNIQUE

The clock gating technique is extensively used in the design of low-power circuit [4]. It involves dynamically shutting off the clock to portions of a design that are idle or are not performing useful computation. Fig. 1 depicts the concept of the clock gating using an AND gate. The basic idea is to AND the clock with an enable signal, so that the register receives a clock signal only when the enable is high.

The granularity of the circuit block at which clock gating is applied greatly affects the power savings that can be achived because gating larger blocks results in higher power savings in the "off" clock cycle, but allows fewer number of "off" clock cycles. The following three levels of granularity can be distinguished: 1) modul-level clock gating, 2) register-level clock gating, and 3) cell-level clock gating [4].

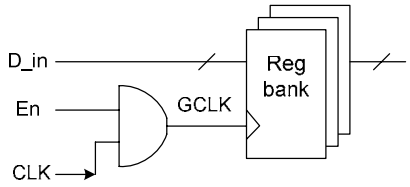


Fig. 1. Clock gating using a latch and AND gate

In this paper we will consider implementation of clock gating technique on simple PE as a constituent of 1D or 2D linear systolic array. We define systolic array as an arrangement of PEs in an array (often rectangular) where data flows synchronously across the array between neighbors, usually with different data flowing in different directions. More details concerning systolic array can be found in [5].

IV. PROCESSING ELEMENT STRUCTURE

The function of a PE, described in this paper, is defined by the following formula:

$$c_{out} = (a_i \otimes b_i) \oplus c_{in}$$

where: a_i, b_i and c_{in} are n-bit input operands; c_{out} is n-bit output operand; symbols $\otimes$ and $\oplus$ correspond to one of the arithmetic/logic operations defined in Table I.

Having in mind that in one instruction cycle the PE have to perform two operations, our design choice was to use two stage pipelining technique (see Fig. 2). The first pipeline stage (PS1) performs the operation $\otimes$ on input operands a_i and b_i, and generates output results denoted as c and time delayed input operands, a' and b'. Select input signals SEL1 define operation of PS1. The second pipeline stage (PS2) performes the operation $\oplus$ on input operands c and c_{in} and generates output results c_{out} and time delayed input operands, a_{i-1} and b_{i-1} (a_{i-1} is obtained by implementing Z^{-1} to a_i). Select input signals SEL2 define operation of PS2.

TABLE I
INSTRUCTION SET

Mnemonic	Description
`OR c,a,b`	c = a ∨ b
`AND c,a,b`	c = a ∧ b
`NOT c,a`	c = $\bar{a}$
`ADD c,a,b`	c = a + b
`SUB c,a,b`	c = a - b
`MIN c,a,b`	c = min{a,b}
`MAX c,a,b`	c = max{a,b}
`CMIN c,a,b`	$c = \begin{cases} a \text{ if } a = b \\ min(rangevalue) \end{cases}$
`CMAX c,a,b`	$c = \begin{cases} a \text{ if } a = b \\ max(rangevalue) \end{cases}$
`NEG c,a`	c = - a
`MUL c,a,b`	c = a ∗ b
`NOP`	no operation
`others operations`	reserved for future use

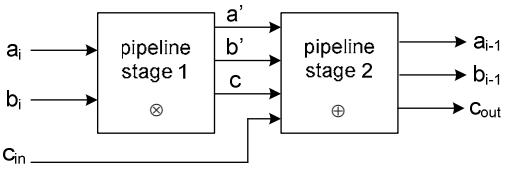


Fig. 2. Two-stage processing element

The internal PE hardware structure, composed of two PSs, is pictured in Fig. 3. As can be seen from Fig. 3, both PS are identical. Each PS is realized as a parallel hardware composed of 16 arithmetic/logic execution units, marked as EX0 to EX15. All execution units operate in parallel. The multiplexer, MUX1/MUX2, according to the control inputs, SEL1/SEL2, selects one of the 16 execution-unit-outputs at its output. Blocks CG1/CG2 marked with dashed lines correspond to the hardware structure which is needed for implementation of the clock gating technique. Let note that clock gating technique is independently realized for each PS. The block CG1/CG2 is composed of: a) 16 clocking circuits, CLKIN0 to CLKIN15; b) time delay stage for input operands, DSZ; c) opcode decoder, DEC1/DEC2; and d) clock distribution tree, CLK_TR1/CLK_TR2. The hardware block COUNT&DEC is common to both CG1 and CG2 blocks and is used as a multiphase clock generator.

In our proposal, the clock gating is realized as a two steps technique. During the first step, in order to select the specified PS's function ($\otimes$/$\oplus$ referred as opcode), the control signal SEL1/SEL2 is decoded and one of DEC1/DEC2 outputs Eni(0) to Eni(15), i=1,2 is activated. At the second step, the

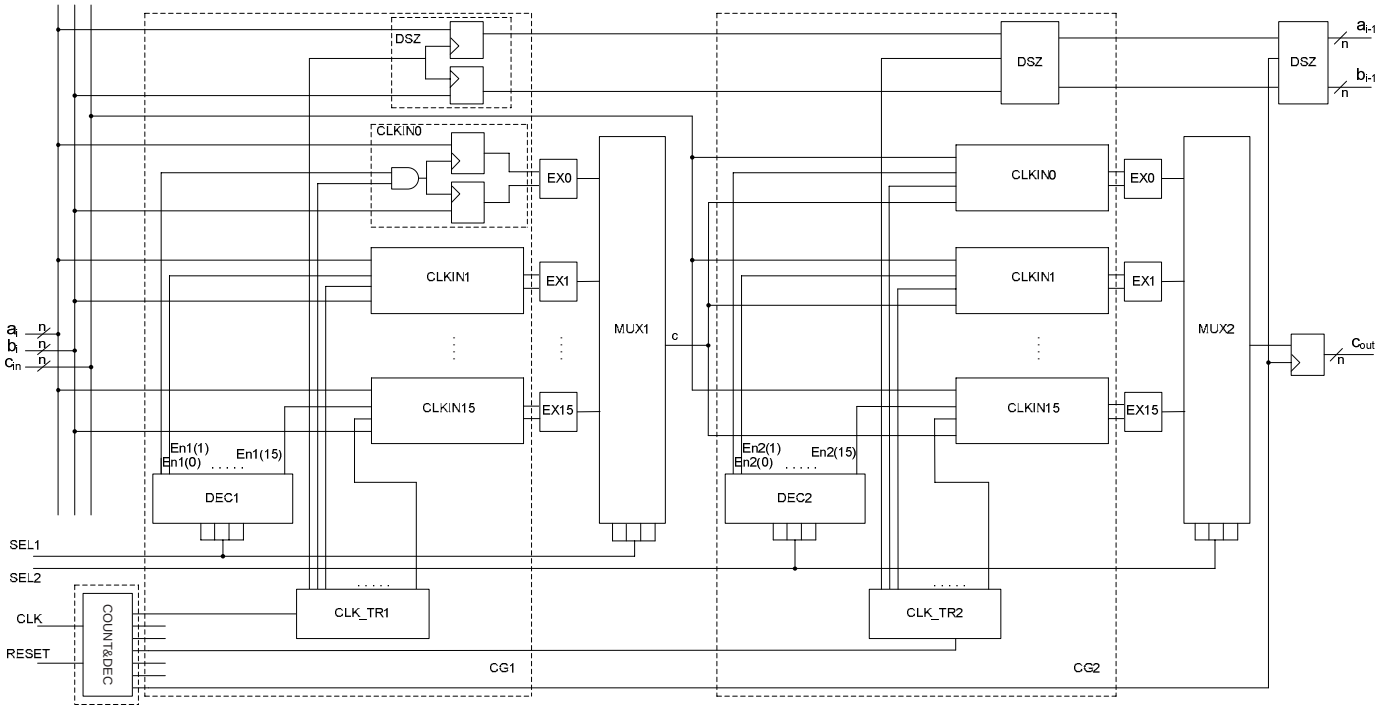


Fig. 3. Micro-power PE hardware structure

latching operation of clocking circuits CLKIN0 to CLKIN15 is enabled or disabled. Only one clocking circuit is enabled, all others are disabled.

Power reduction using this aproach is achieved thanks to the following fact: Instead to switch inputs to all execution units, only inputs to the enabled execution unit are swithed. In this way, we significantly reduce switching activity, α, defined in (4) and directly decrease energy consumption.

The PE for which we have implemented the clock gating technique is referred as a *micro-power PE*. The energy reduction is achieved at cost of decreased PE's operating speed and increased Silicon area, i.e. the number of occupied gates.

Another crucial property of the proposal relates to its possibility to implement the method at the RTL level of abstraction. In this way the proposal can be implemented independently of the target technology. In other words, standard HDL code (VHDL, Verilog) as well as standard CAD tools intended for simulation and synthesis can be used.

V. RESULTS

In order to evaluate the performance of a proposal we have implemented both the standard and the micro-power PE on Xilinx FPGA technology using a software tool ISE 9.1. We have considered and analyzed several *n*-bit PEs which can manipulate with 8-, 12-, 16-, 24- and 32-bits operands. All PEs, at behavioral level, are described using VHDL code. During the phase of HDL code-creation special care was devoted to power consumption optimization by using clock gating technique. Namely, to each execution unit a corresponding clock gating logic is appended. In this manner, significant power reduction is achieved.

The PEs are implemented on two different types of FPGA circuits, the first one to Spartan2, and the second one to Virtex4 series. Parameters that have influence on power consumption were selected to be identical in both cases. For evaluation of power consumption the Xilinx Xpower tool was used. As performance metrics the area overhead, AO, and power reduction factor, PRF, were used. The metrics AO and PRF are defined as

$$AO = \frac{\mu_power_PE_gate_count - standard_PE_gate_count}{standard_PE_gate_count}$$

$$PRF = \frac{standard_PE_power - \mu_power_PE_power}{standard_PE_power}$$

where: μ_power - micro-power.

The obtained results which correspond to standard- and micro-power-PE, for Spartan2 and Virtex4 families, are given in Table II and III, respectively. PE's hardware complexity is presented in columns 2 and 4, and is specified as an equivalent gate count. In column 3 and 5 PE's power consumption in mW is given. The last two columns include AO and PRF metrics specified in %.

By analyzing the obtained results we can conclude the following:

1) The micro-power-PE has significantly lower power consumption in respect to standard-PE.

2) The proposed method is more efficient for larger *n*. Namely, with increasing *n* both metrices improve, i.e. the PRF increases (from 87.6% for 8-bit PE up to 97.4% for 32-bit PE, for Spartan2 serie; and from -9% for 8-bit PE up to 91.6% for 32-bit PE, for Virtex4 serie), while AO decreases (from 99.6% for 8-bit PE down to 36.7% for 32-bit PE, for Spartan2 serie; and from 172.6% for 8-bit PE down to 38.6% for 32-bit PE, for Virtex4 serie).

TABLE II
IMPLEMENTATION RESULTS FOR XILINX SPARTAN2E FPGA

number of input operands bits	standard PE		micro-power PE		area ovrhead [%]	power reduction [%]
	gate count	power [mW]	gate count	power [mW]		
8 bit	3636	104.70	7256	13.02	99.6	87.6
12 bit	7068	290.78	12270	16.22	73.6	94.4
16 bit	11236	1399.28	18032	50.24	60.5	96.4
24 bit	22065	2451.18	32106	125.98	45.5	94.9
32 bit	36240	11729.89	49550	299.85	36.7	97.4

Note: Target Device: xc2s300e-6fg456

TABLE III
IMPLEMENTATION RESULTS FOR XILINX VIRTEX4 FPGA

number of input operands bits	standard PE		micro-power PE		area ovrhead [%]	power reduction [%]
	gate count	power [mW]	gate count	power [mW]		
8 bit	2688	6,98	7328	7.61	172.6	-9.0
12 bit	6966	19,25	12252	5.19	75.9	73.0
16 bit	11020	62,30	18026	8.44	63.6	86.4
24 bit	21702	128,18	32100	15.06	47.9	88.2
32 bit	35730	462,35	49520	28.38	38.6	91.6

Note: Target Device: xc4vlx15-12sf363

3) Area overhead for both series are comparable, while power reduction is significantly larger for Spartan2 FPGA. Lower power reduction factor for Virtex4 is consequence of allready factory implemented advanced low power techniques such as, for example, the reduced power supply voltage, decreased leakage current, decreased load parasitic switching capacitors, etc.

VI. CONCLUSION

Clock gating technique is one of the most successful and widely used techniques for power reduction. In this paper the hardware structure, instruction set, and implementation of a simple micro-power processing element based on FPGA technology are described. For power reduction the clock gating technique was used. The obtained results show that power reduction up to 97.4%, with area overhead of 36.7% can be achieved. The processing element can be used as a basic building block of 1D or 2D linear systolic array.

REFERENCES

[1] C. Piguet, ed, *Low-Power Electronics Design*, Boca Raton, CRC Press, 2005.
[2] T. Kogel, R. Leupers and H. Meyr, *Integrated System-Level Modeling*, Dordercth, Springer, 2006.
[3] L. Benini, A. Bogliolo, G. De Micheli, "A Survey of Design Techniques for System-Level Dynamic Power Management", IEEE Trans. on VLSI, Vol. 8, No. 3, pp. 299-316, 2000.
[4] R. Mehra, B. Pangrle, *Synopsys Low-Power Design Flow*, pp. 40.1-40.20 in C. Piguet, ed, *Low-Power Electronics Design*, Boca Raton, CRC Press, 2005.
[5] M.O. Esonu, A.J. Al-Khalili, S. Hariri, D. Al-Khalili, "Systolic Arrays: How to Choose them", IEE Proceedings-E, Vol. 139, No. 3, pp. 179-188, 1992.

Throughput and Fairness in IEEE 802.16 Broadband Wireless Networks

Radostin A. Pachamanov[1] and Boris P. Tsankov[2]

***Abstract* – The IEEE 802.16 standard defines the air interface specifications for broadband access in wireless metropolitan area networks. These networks support various services that require specific data throughput. In order to maintain appropriate quality of service (QoS), the network should perform an adequate resource allocation. Optimizing the network throughput as a primary objective may lead to unfairness from the users' point of view. In this paper we investigate the "fair" resource allocation problem. We propose an optimization-based approach for resource allocation, which takes into consideration fairness issues.**

***Keywords* – IEEE 802.16, resource allocation, quality of service, throughput, fairness.**

I. INTRODUCTION

The newest generation of Internet connectivity is the broadband wireless access (BWA) technology, based on the IEEE 802.16 standard, also known as WiMAX (Worldwide Interoperability for Microwave Access) [1]. The main advantages of IEEE 802.16 wireless access are high transmission rate, flexibility for the connection, and the ability for pre-defined quality of service (QoS) framework. WiMAX, as a wireless solution, is a promising alternative for "last mile" access in crowded urban or suburban areas where installation of cable-based infrastructure is economically or technically infeasible.

The WiMAX network is able to provide a wide range of services that are varied in their nature and demand different performance levels in order to maintain an appropriate QoS. The mechanisms for providing and granting QoS in WiMAX are through connection admission control (CAC) and resource allocation. CAC is used to limit the number of connections in the network in order to maintain the pre-defined QoS parameters of the applications that are already accepted for service. If the available resources could be allocated among the ongoing and the incoming connections so that the QoS requirements of both types of connections could be fulfilled and maintained at the target level, the new connections are accepted. Through a resource allocation schemes the applications that require better throughput are granted more bandwidth. The resource allocation scheme should take into consideration the fact that in wireless networks the transmission rate is affected by channel conditions and also depends on allocated power [2]. It needs to take into consideration not only the different rates and QoS requirements of the applications, but the SSs' positioning in the cell (the distance to the BS, SINR, modulation used, etc.). Although the IEEE 802.16 standard comprises the signaling for the multiple access, the algorithms for CAC and resource allocation remain open issues.

There are several articles published on the resource allocation [3-5] and CAC problems [6-7]. They investigate different architectures and scheduling algorithms to guarantee QoS. A very interesting approach is the formulation of an optimization problem having as an objective function QoS parameters such as delay, loss, throughput, etc. [5-7].

In this paper, we investigate the resource allocation problem in order to obtain maximum network throughput. But maximizing the network throughput as a primary objective may lead to "unfair" resource allocation. Within a wireless network, two different aspects of fairness can be distinguished. First (and this is typical for every network serving simultaneously traffic flows with different traffic characteristics), the applications with less stringent QoS requirements may be neglected compared to high priority traffic. This problem has been broadly studied in recent years with focus on IP networks, and different schemes for providing fairness have been introduced [8-9]. A new and specific problem with fair resource allocation appears in IEEE 802.16 wireless networks. Since the SSs distant from the BS use different modulation schemes, the throughput achieved for the same amount of bandwidth may differ. Therefore, the base station may refuse service to remote subscribers in order to maximize network throughput. To the best of our knowledge, this problem has not yet been discussed in the literature. Considering these two aspects, we propose an optimization-based joint connection admission control and resource allocation scheme for IEEE 802.16-based wireless networks where a fairness element is implemented. Our scheme takes into account traffic characteristics of the ongoing and incoming connections such as mean data queue length, maximum allowed mean data delay, and mean packet arrival rate. We formulate a corresponding optimization problem where we assume that each user is assigned a throughput-dependent utility function. This function, which may be different for different SSs, represents the utility of the corresponding throughput achieved and implements the fairness concept. Our goal is to optimize the total utility over all users.

Optimization of such a network is a complicated task since it is a complex problem consisting of joint throughput and power allocation optimization. One possible approach is to

[1]Radostin A. Pachamanov is with the Dept. of Telecommunications, Technical University of Sofia, "Kl.Ohridski" Blvd. No8, Sofia 1000, Bulgaria, e-mail: radostin_ap@tu-sofia.bg

[2]Boris P. Tsankov is with the Dept. of Telecommunications, Technical University of Sofia, "Kl.Ohridski" Blvd. No8, Sofia 1000, Bulgaria, e-mail: bpt@tu-sofia.bg

split the problem in two parts [9]. According to [10], the gain obtained from throughput optimization is bigger than the gain from power optimization. Therefore, we assume that the power is equally spread over all slots and solve the slot allocation problem, considering it as a part of the global optimization problem.

The rest of the paper is organized as follows. Section II gives a brief overview of IEEE 802.16 specifics taken in consideration. Section III presents the fairness problem with resource allocation in a cell structure. Section IV presents the system model and the formulated optimization problem. Section V concludes with final remarks.

II. An Overview of the Considered IEEE 802.16 Specifics

The IEEE 802.16 standard, or so called WirelessMAN, is an air interface standard that defines the first two network layers (physical-PHY and data-MAC) [1] of the OSI model. The IEEE 802.16 system architecture consists of two logical entities – a base station (BS) and a subscriber station (SS).

The medium access control layer (MAC) is structured to support multiple PHY specifications, depending on the particular operational environment. The physical layer operates at 10-66 GHz, or 2-11 GHz frequency band, where for the latter propagation without direct line-of-sight is possible. For frequencies below 11 GHz, three alternatives for PHY specifications are provided: Orthogonal Frequency Division Multiplexing (OFDM), Orthogonal Frequency Division Multiple Access (OFDMA) and Single-Carrier (SC). In our investigations we consider OFDMA, since it is proposed for the mobile version IEEE 802.16e of the standard. To maximize the spectral efficiency of the air link, each specification uses multilevel modulation scheme. The modulation is optimized for each SS based on the quality of the radiofrequency channel. If link conditions permit, a more efficient modulation is used to maximize the tradeoff between bandwidth and robustness. The supported modulations are QPSK, 16-QAM and 64-QAM.

The MAC protocol is performed in a way that provides high data rates in both uplink (UL) and downlink (DL) directions, and comprises medium access admission and bandwidth allocation algorithms. The protocol is connection-oriented and all data communications for both transport and control are in the context of a unidirectional connection. BS is responsible for coordinating the access of SSs to the medium. Data transmission is organized in framed format. Separate subframes are used for uplink and downlink directions. There are two ways supported by IEEE 802.16 for separating uplink from downlink transmissions: frequency division duplexing (FDD) and time division duplexing (TDD). In FDD, downlink and uplink are performed at different frequencies, and therefore may overlap in time. TDD divides time into uplink and downlink transmission periods. In both schemes the frame has fixed duration. We consider TDD since it has the advantage that the network can adjust the size of the UL and DL subframes within the frame depending on the traffic load in the corresponding direction.

In order to provide a QoS-based network operation, the IEEE 802.16 standard groups the applications with similar QoS requirements and traffic characteristics into a small number of classes, named *scheduling services*. Each scheduling service is tailored to support specific class of application. Four types of services are defined.

Unsolicited grant service (UGS): Supports constant-bit-rate (CBR) real time traffic with stringent QoS requirements, such as voice over IP. The medium is granted on a periodic basis.

Real-time polling service (rtPS): Supports various-bit-rate real time traffic. The amount of bandwidth required for this type of service is determined based on the required QoS parameters, the channel quality and the traffic arrival rates of the sources. An example for such kind of applications is VoIP with silence suppression.

Non real-time polling service (nrtPS): Supports traffic with less stringent QoS requirements. This is suitable for applications such as file transfer. The bandwidth allocation is also adaptive as in the case of rtPS.

Best-effort service (BE): Supports best-effort traffic with no QoS guarantee.

In our model we focus the latter three types of services. We assume that a certain amount of bandwidth is granted for the UGS, and the BS has to allocate the rest among the other three types.

III. Fairness in IEEE 802.16-Based Networks

In telecommunications the term “fairness” is used to represent a criterion for distributing the available network resources among competing traffic flows – the resources should be “fairly” distributed among the subscribers in such a way that everybody is satisfied with the achieved QoS. Generally, policies for resource sharing that are characterized by low level of fairness provide high average throughput, but low stability in the service quality, meaning that the achieved service quality is varying in time, depending on the behavior of other users. If this instability is a frequent event, it may result in displeased customers that may choose another, more stable communication network.

In a simple vision, fairness may be interpreted as allocating the same share of bandwidth to all. However, in a case of a network supporting a wide range of applications with various QoS requirements, such a simple view does not make sense.

As we mentioned above, two different aspects of fairness following the specifics of the wireless networks can be distinguished. The network allocates the resources depending on the type of *scheduling service* to which the application belongs. The applications from classes with lower priority may suffer. Therefore, a special scheme for resource distribution is needed.

The second important and specific for IEEE 802.16 aspect concerns the throughput that can be achieved with a certain amount of bandwidth (we consider OFDMA where a number of subcarriers are dedicated to a subscriber for a certain time period). The SSs are randomly situated within the cell and the distance between a particular SS and the BS may vary. Depending on the wireless link condition, SINR, the distance, etc., a different modulation scheme may be used. Some

modulation schemes are more efficient than others, and the same amount of bandwidth can deliver better throughput. An example is given in Fig.1.

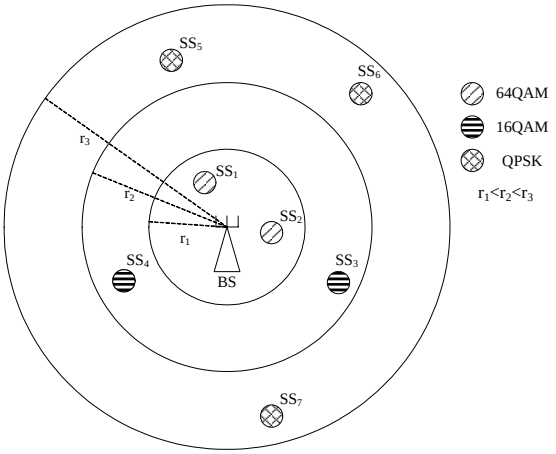


Fig.1. SSs working with different modulation schemes

It is natural for a network operator to focus on maximizing the network throughput, since in that way the system utilization is increased. However, maximizing the throughput without taking into consideration the fairness issues described above may worsen the network operation. A simple way to evaluate the level of fairness achieved is through the Fairness Index (FI), proposed in [12]:

$$FI = \left(\sum_{i=1}^{N} r_i \right)^2 / \left(N \sum_{i=1}^{N} r_i^2 \right), \qquad (1)$$

where N denotes the number of active SSs, and r_i represents the resource portion allocated to, or throughput achieved by, user i. Since the applications have different traffic characteristics, they require different resource portions to achieve certain QoS level. Therefore, we define r_i as a "surplus" resource portion/throughput of user i, and $r_i = a_i - a_i^{\min}$, where $a_i^{\min}$ denotes the minimum required data throughput for user i. If all users get the same surplus throughput, then FI will be 1, and the system will be 100% fair.

Three fairness criteria are presented in the literature – max-min fairness, proportional fairness, and utility fairness [8-9]. Max-min fairness puts emphasis on maintaining high values for the smallest rates. Proportional fairness is an alternative definition and provides fairness on a proportion basis. In the "utility" fairness approach, every user has a utility function that indicates the value to that source to have a certain throughput/rate. In our model, we consider utility fairness, since it is a more general concept, and comprises both max-min and proportional fairness.

IV. THROUGHPUT OPTIMIZATION

Our system model consists of a BS that serves incoming and outgoing traffic from and to a certain number of SSs. We assume that every SS has a corresponding queue in the BS. Let us denote the number of active SSs in the system by *N*. We assume that the BS is responsible for both UL and DL and follows the state of *2N* queues, since each user/application has separate queues for UL and DL directions. Henceforth, by user/application we will mean the traffic flow that enters the corresponding queue. Let $\vec{\lambda} = (\lambda_1,...,\lambda_N)$ represent the vector of arrival rates for data that are entering the queues. Depending on the QoS parameters of the connections, the BS allocates resources to a particular queue, and the data are transmitted.

We consider WirelessMAN-OFDMA as a PHY specification. In OFDMA, a set of transmitter's carriers is divided into subsets, each of which can address a different receiver at any given time depending on the specific throughput requirements of the user. In this paper we assume that the available bandwidth is divided into a number of subchannels (*m*), each of which is made of multiple subcarriers according to OFDMA. The duration of one frame is divided into a given number of time-slots (*n*) (Fig.2).

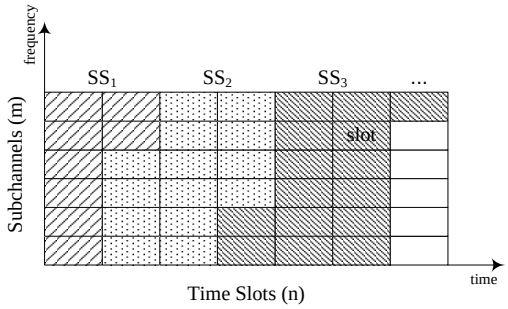


Fig.2. Bandwidth allocation

Each subchannel/time-slot pair will be referred to as a *slot*. A slot can be assigned to at most one user/application and one transmission direction. Let us denote the transmission rate that can be achieved within a single slot with s_k ($k=1,\ldots,nm$). It varies depending on the used modulation and coding scheme, the condition of the wireless link, the allocated power, etc. The number of slots assigned to a user x_i ($i=1,\ldots,nm$) and the specific transmission rates depending on the chosen type of modulation of these slots determine its overall throughput which we will denote by $a(x_i)$.

In our optimization problem for allocating the total available bandwidth to the ongoing and the incoming connections, we associate with each user a utility function of its throughput $u_i(a(x_i))$ We assume that these functions are increasing, strictly concave, continuously differentiable over the range $x_i \geq 0$, and $u_i(0) = 0$. These functions represent the utility of that user/application to the corresponding throughput achieved. The objective is to maximize the overall utility of all users/applications. The main advantage in using such an approach is that these utility functions could be sophisticated and able to provide fair resource allocation. All the fairness issues concerning the type of scheduling service to which the application belongs may be taken into consideration through the specification of the utility functions. Let $u_i(a(x_i))$ indicate the "value" to source i of having throughput $a(x_i)$. We can assume that each "slot" has a cost function $g(x_k), k = 1,..,nm,$

which indicates the cost to the network of supporting x_k slots for connection i, and may represent the second aspect of the fairness problem – the type of modulation used. Then, a "utility-fair" allocation of resources in an allocation which maximizes $U(\vec{x})$, defined by:

$$U(\vec{x}) = \sum_{i=1}^{2N+2} u_i\left(a(x_i)\right) - \sum_{k=1}^{nm} g(x_k), \quad (2)$$

where $\vec{x}$ presents the optimal solution vector. The BS has to estimate if the available resources may be allocated among the ongoing and the new incoming connection in such a way that QoS for each application can be met. In other words, if the optimization problem is feasible, the new request should be accepted for service, and the optimal solution itself gives the optimal resource allocation.

We add a number of constraints to the optimization problem. First, the total number of slots that are allocated should be less than or equal to the total number of available slots (bandwidth) in the system. The second constraint is related to the mean queue length for each user. When an application has a certain packet arrival rate, it is expected that in the corresponding queue there would be a certain number of packets, as that number is related to the arrival rate. Slots should be allocated to this queue only if we expect that there will be enough data to fill them. We denote by $\overline{L_i}$ the mean number of packets in the i^{th} queue and by $l(x_i)$ the amount of data that can be transmitted with x_i allocated slots.

Data delay as an important QoS parameter determines the third constraint. We assume that each user demands a service that has stringent requirements about the maximum allowed average data delay $\hat{d}_i$, and the next constraint in our optimization problem will ensure that the rate achieved through resource allocation should be enough to guarantee it. Let $\overline{w}(\lambda_i, x_i)$ denote the average delay for packet arrival rate λ_i and x_i allocated slots. We assume a Poisson request arrival process.

The optimization problem can be formulated as follows:

$$\begin{aligned} &\max_{\mathbf{x}} \quad U(\vec{x}) \\ &s.t. \quad \sum_{i=1}^{2N+2} x_i \le nm \\ &\qquad l(x_i) \le \overline{L_i}, \quad i = 1,\ldots,2N+2 \\ &\qquad \overline{w}(\lambda_i, x_i) \le \hat{d}_i, \quad i = 1,\ldots,2N+2 \\ &\qquad x_i \ge 0, \quad i = 1,\ldots,2N+2 \end{aligned} \quad (3)$$

If the problem is feasible, the optimal solution will contain the amount of bandwidth allocated to each connection, including the new one. If the solution is infeasible, there is no bandwidth allocation scheme such that the throughput and delay requirements for all users can be satisfied, and the incoming connection is blocked. The optimization problem may be used as a stand-alone scheme for optimal resource allocation that can be performed at the beginning of each frame. The fairness is provided through the specification of the utility functions.

V. Conclusions

In this paper we presented the specifics of the IEEE 802.16 standard, the methods for QoS provisioning, and the corresponding problems with "fair" resource allocation. We formulated an optimization problem where we focused on maximizing a utility-based objective function related to users' throughput, which comprises fairness. In future studies we intend to develop further the concept of utility functions.

Acknowledgements

This research was done under contract BУ-TH-105/2005 between Technical University of Sofia and the Bulgarian Ministry of Education and Science.

References

[1] IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems. June, 2004,

[2] Eklund C., R. Marks, S. Ponnuswamy, K. Stanwood, N. van Waes, *WirelessMAN: Inside the IEEE 802.16TM Standard for Wireless Metropolitan Area Networks*, IEEE press, 2006,

[3] Cicconetti C., C. Eklund, L.Lenzini, E. Mingozzi, "Quality of Service Support in IEEE 802.16 networks", IEEE Network, vol. 20, no. 2, March 2006

[4] Sayenko A., O. Alanen, J. Karhula, T. Hamalainen, "Ensuring the QoS Requirements in 802.16 Scheduling", Proc. of the 9th ACM symposium on Modeling analysis and simulation of wireless and mobile systems, 2006, Pages: 108 – 117,

[5] Iyengar R., K. Kar, B. Sikdar, "Scheduling Algorithms for Point-to-Multipoint Operation in IEEE 802.16 Networks", Proc. Of 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006, pp1-7,

[6] Niyato D., E. Hossain, "Joint Bandwidth Allocation and Connection Admission Control for Polling Services in IEEE 802.16 Broadband Wireless Networks", IEEE ICC Proceedings, 2006,

[7] Hosein P. "On the Optimal Allocation of Downlink Resources in OFDM-Based Wireless Networks". Proc. of WWIC 2006, Bern, Switzerland,

[8] Bertsekas D., R. Gallager, *Data Networks*, Englewood Cliffs, NJ: Prentice Hall, 1987,

[9] Kelly F., A. Mauloo, D. Tan "Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability", J. Operational Research Society, 1998, pp. 237-252,

[10] Bohge M., J. Gross, A. Wolisz, M. Meyer, "Dynamic Resource Allocation in OFDM Systems: An Overview of Cross-Layer Optimization Principles and Techniques", IEEE Network, Jan/Feb 2007,

[11] Song G., Y. Li, L. J. Cimini Jr., H. Zheng, "Joint Channel-Aware Data Scheduling in Multiple Shared Wireless Channels", IEEE WCNC 2004 Proceedings, Vol.3, pp. 1939-1944,

[12] Kim K., Y. Han, "A Proportional Fair Scheduling for Multicarrier Transmission Systems", IEEE Commun. Lett. vol.9, no.3, March 2005, pp. 210-212.

Implementation of TCP/IP Protocols in Wireless Sensor Networks

Mirko R. Kosanovic[1] and Mile K. Stojcev[2]

***Abstract* – Wireless Sensor Networks (WSNs) are large networks composed of small sensor nodes with limited computer resources capable for gathering, data processing and communicating. WSNs usually can not operate in isolation, i.e. they have to be connected to some Wide Area Network (WAN). As the Internet protocol becomes de-facto standard for WANs, it is a challenge now to connect WSNs to WAN through TCP/IP protocol. In this paper we will analyze, first, the current well known solutions, which relate to connecting WSNs to TCP/IP based networks. In order to make WSN transparent to TCP/IP networks, we involve local node addressing within the WSN. By using this approach the traffic density between sensor nodes and the access point is decreased. In order to evaluate performance of the communication we define a traffic reduction factor $\Phi(m,h)$ as a metric. For several messages and hops we point to the effects of the decreased traffic that can be achieved using the proposed method.**

***Keywords* – TCP/IP, Wireless Sensor Networks, Internet**

I. INTRODUCTION

Wireless sensor networks are composed of a large number of radio-equipped sensor devices, that autonomously form a network, through which sensor nodes are capable of sensing, processing and communicating among each other. Some crucial properties of WSNs are the following: a) nodes are densely deployed in a region and are very often prone to failures; b) broadcast communication paradigm, mainly used without global identification (ID), is implemented; c) nodes operate under limited power; and d) computational capacity and memory space of each sensor node is limited, too. In general, WSNs do not work efficiently in full isolation. So, it is imperative to connect them to some other kind of networks, such as Local Area Network (LAN), Metropolis, WAN and Internet. In this way very complex heterogeneous distributed network (HDN) can be configured. Such connection, from one side, provides transparent operation of the WSN for all HDN`s end users, but, from other side, creates new challenges related to the development and research in this field. Having in mind that TCP/IP protocol suite becomes de-facto standard in network connectivity, it is quite reasonable to look at some efficient methods for interconnecting protocol specific WSN to TCP/IP based network, such as for example Internet. This problem is in focus of our interest in this paper. At the start, a short survey related to current researches on this field will be given. Next, in order to evaluate communication performance of the WSN-TCP/IP-middleware we define a traffic reduction factor. After that, we propose a method for traffic reduction between WSN and TCP/IP based network. Finally, for several messages and different number of hops we present results obtained by implementing the proposal.

[1]Mirko R. Kosanović is with the High Technical School, A.Medvedeva 20, 18000 Nis, Serbia, E-mail: vkosanovic@sbb.co.yu
[2]Mile K. Stojcev is with the Faculty of Electronic Engineering, A.Medvedeva 14, 18 000 Nis, Serbia, E-mail: stojcev@elfak.ni.ac.yu

II. APPLICABILITY OF TCP/IP PROTOCOLS IN WSNs

With the rapid development of wireless technology, the increased requirements for implemmentation of TCP/IP based networks become a necessity. To solve this problem efficiently is not an easy task, especially in WSNs, where sensor nodes (SNs) are realized with many limited resources. Without doubth, one of the more pronounced design challenge relates to SN's power consumption. Namelly, since the energy of the battery powered SN is limited, in order to prolong it's live, micro power consumption for SN is of paramaunt importance. Numerous researches conveyed in this field [4], [5], [6], [7] show that the communication building block is the largest energy consumer of the wireless SN. For example, the energy which is consumed to send only one bit of data is less or equivalent than the amount of energy needed to process 100 instructions for Berkely node [1]. The TCP/IP protocol is often perceived to be „heavy-weight" protocol, because, as a first, its implementation requires large amounts of resources both in terms of memory and processing power, and as a second, the size of its header is too large (IPv4-24 byte, IPv6-40 byte, UDP-8 byte, TCP-24 byte). Having this in mind, a direct implementation of TCP/IP protocol in WSN results to an inefficient design solution, i.e. we have to send 30 bytes of message for only 2-3 bytes of useful payload data. To cope with this problem several methods have been proposed [7], [8], [9].

II.1 Communication models

WSNs use the following three types of communication models:

1. Adress-Centring communication – in this case all sensor nodes have an unique ID number and the routing is performed according to IDs. This kind of protocols, is referred as table-driven routing protocols [2].
2. Data-Centring communication – sensor nodes are without ID numbers. Communication is based on broadcast messages. There are two kind of messages: *Interest packet*, which propagates an information interest through the network, and *Advertisement packet* which is replay from sensor nodes on which an interest has been registered [3].
3. Location-Centring – sensor nodes use the location as a primary means for addressing and data routing. Each

sensor node has an unique spatial address which depends of its physical location in the deployed region..

If we analyze the communication possibilties of all three models and try to implement these models in TCP/IP network, we can conclude the following: Data-Centring model is not good candidate for TCP/IP networks. In order to provide consistency between WSN and TCP/IP, the Address-Centric and Location-Centring communication paradigms are better solutions for interconnecting WSNs and TCP/IP networks [4].

II.2 Communication architecture

According to the communication architecture, we divide the communication methods (see Figure 1.) as those that are based on: Proxy architecture, Overlay based architecture and Gateway architecture.

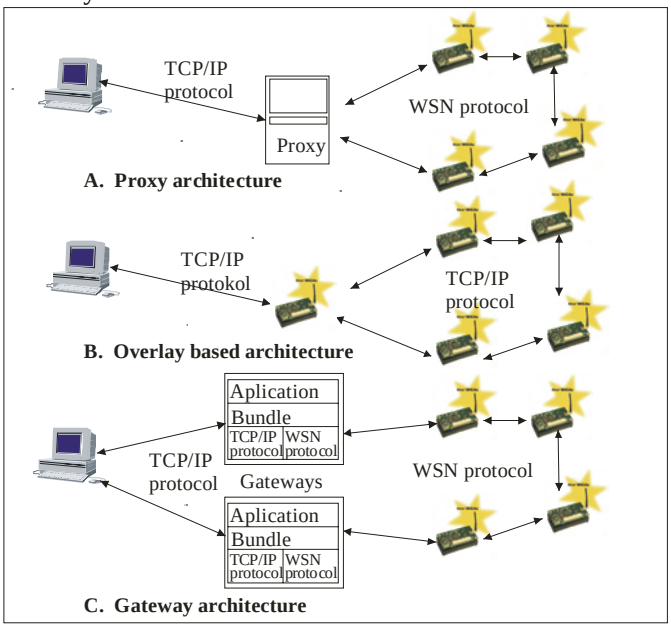


Fig. 1. Three types of communication architectures

II.2.1 Proxy architecture

The comunication between TCP users and sensor nodes is done through the proxy computer. The communication protocol used in the sensor network may be chosen freely. There are two various modes according to which proxy can be operative and can interconect WSN with TCP/IP networks as:

1. *relay* – in this mode all data which are coming from one network are passed to the another network.
2. *front-end* – the proxy pro-actively collects data from sensor nodes and stores this information in its database. The users from TCP/IP networks can query for specific data in a variety ways, such as SQL queries or WEB based interfaces.

Both solutions have some drawbacks, what make these solutions not applicable in general. As a first, a single point of failure exists. When the proxy stop working, all communications to and from the WSN are broken. As a second, the proxy implementation usually depends on the specific task or a particular set of protocols. This means that for each application different proxy is needed.

II.2.2 Overlay based architecture

There are two kinds of overlay based methods: *TCP/IP overlay sensor networks*, and S*ensor networks overlay TCP/IP* [5]. The first approach points that it is possible to implement TCP/IP protocol stack to microcomputer system with very poor resourses: 8-bit microprocesor with only 2kB RAM memory [2]. In *Sensor networks overlay TCP/IP* the protocol stack of WSN is deployed over the TCP/IP stack and each Internet user is considered as a virtual sensor node. The virtual sensor node can interpret WSN packets since it has installed the WSN protocol stack in addition to TCP/IP stack. Numerous problems accompany the implementation of TCP/IP in WSNs. They can be identified as: header overhead, high bit error rates, high energy consumptions for end-to-end multi hop retransmissions, differences in routing protocols and implementation of adressing and routing schemes [4].

II.2.3 Gateway architecture

One of the essential device who provides a connection between wireless and TCP/IP network is a gateway. It performs several tasks such as protocol conversion, message delay,etc. All solutions, that use gateway as an interconecting device, can be grouped into the following two categories: *Aplication gateway* and *Delay Tolerant Network (DTN)*. Aplication gateway is a simple gateway based approach which works in aplication layer [9]. The DTN, is a similar solution. The main difference in respect to *Aplication gateway* is the following: It implements one new layer, both in TCP/IP and WSN networks, referred as *Bundle Layer*. The main function of the bundle layer is to store and forward packets beetwen two network (Figure 1.).

III. A GATEWAY AS A TCP/IP TO WSN ADDRESS TRANSLATOR

Main design challenges concerning the interconection between TCP/IP based networks and WSNs are related to the fact that it is neccessary to provide: a.) access to each SN through the TCP/IP based network; b.) efficient communications from aspect of SN's energy consumption; and c.) transparency in operation between TCP/IP based protocols and WSN protocols.

The method which we propose in this paper is sutiable for applications area such as: health monitoring (diagnostics, telemonitoring), environmental monitoring (fire detection, water pollution, tracing movements of birds, animal or insects, detection of chemical and biological agents), military and security (movements of soldiers and vehicles, monitoring critical terrain), industrial process control, smart buildings, traffic control, etc. In general, both from aspect of toplogical hierachical network organization, from one side, and control, from other side, these systems are heterogeneous in nature. Therefore, in order to design an optimal solution, for a particular case, it is neccessary first to forsee the curcial assumptions and requirement that these kinds of networks have to fulfill:

- WSN is organized as set of clusters.
- Cluster topology is aribritrary.
- Each cluster is organized around one Main Sensor Node, referred as a MSN.

- The MSN acts as a gateway between the WSN and TCP/IP based network.
- To each MSN two addresses are appended. The first one is TCP/IP address, while the second is local WSN's address.
- SNs can accsess to TCP/IP users through MSN, or contrary.
- Within each cluster, it is possible to address 255 single-addressed SNs, 128 double-addressed SNs, 64 quad-addressed SNs, etc.
- The number of addresses appended to each SN determines the number of messages with which the SN can manipulate simultaneosly, i.e. at the time.
- For data transfer, store and forward technique is used.
- Single hop and multi-hop data transfers whitin the cluster are possible.
- During the initialization phase, the MSN assigns different group of addresses (single, double, quad, octal etc.) to each SN, according to the predicted traffic intensity among SNs and MSN.
- To the TCP/IP users each SN is visible through its first group address.
- MSN can transfer data to: a.) SNs located within single cluster area; b.) to MSNs that are constituents of other clusters; and c.) to TCP/IP users.

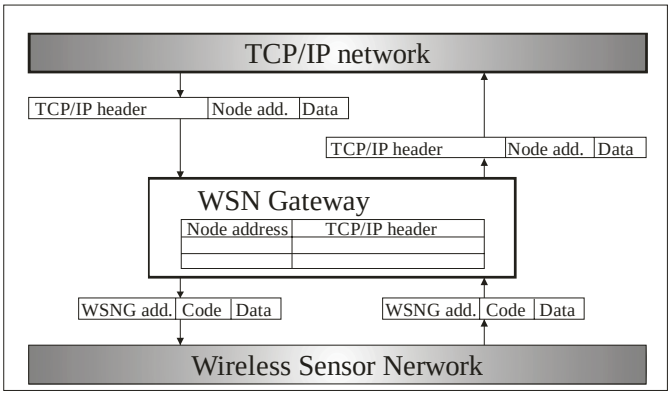


Fig. 2. Translation table

Figure 2. depicts the principle of messages transfer between TCP/IP network and WSN. As can be seen from Fig.2 the MSN acts as a protocol translator between both networks. From software piont of view it maps addresses from one network domain to another, and translates larger TCP/IP header into smaller WSN header. In order to perform this activity it uses data stored in a translation table.

The translation table has 256 entries. Each entry has two fields. The first field points to the SN's local address, while the second to the TCP/IP header. For single addressed SN one table entry is appended, to double addressed SN two table entries are appended, etc. This kind of table organization allow us to direct several TCP/IP messages to the same SN at the time, without extending the WSN header. Smaller header sizes directly implies to lower communications cost, and indirectly to decreased energy consumption of the SN.

IV. A GATEWAY: PRINCIPLE OF OPERATION

In order to evaluate the performance of the proposed solution we have assumed the following: a.) each SN within a cluster is seen as TCP/IP addressible unit; b.) protocols above and bellow the network layer remain unchanged; c.) data transfer between two communication units is store-and-forward type; d.) single hop or multihop are allowed; e.) all data transfers are error free, i.e. without retransmissions. Successible data transfer of one message between two SNs (message transfer and responce) depends of the shortest end-to-end delay. This kind of communication delay includes the following items:

1. T_t (transmission delay) – time for transmition one message It depends on the channel bandwidth, bit rate, message length, and coding technics.
2. T_p (propagation delay) – signal propagation time between two SNs.
3. T_c (processing delay) – the time needed for processing one message.
4. T_q (queueing delay) – an averge time during which message wait in a queue for transmission.

The total communication time for the solution given in reference [4] is defined as:

$$Tref = 2mh(2Tc + Tt + Tp) + 2m(h-1)Tq \qquad (1)$$

while for the solution proposed in this paper is:

$$Tps = 2(m+1)hTc + 2(m+h-1)Tt + 2hTp + 2m(h-1)Tq \qquad (2)$$

where m is a number of transfered messages, and h coresponds to a number of hops. In Figure 3A. the principle of single message data transfer among four SNs is pictured. Figure 3B. corresponds to data transfer of four messages. Let note that overllaping between data transfers in this case exist.

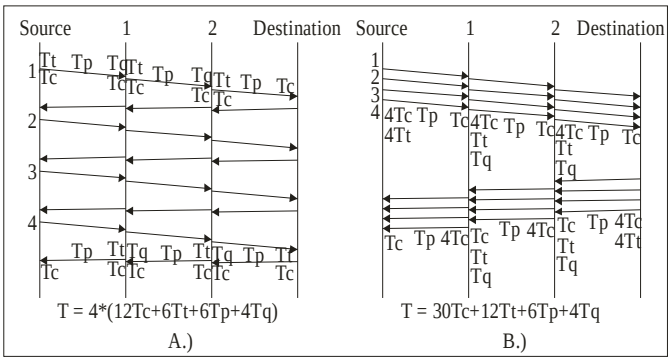


Fig. 3. Time needed for transmission of four messages

In order to evaluate the performance of both proposal we assume the following:

1. the data transfer rate is R = 720 kbps,
2. one WSN message usualy consists of N = 16 bytes,
3. the signal propagation velocity is $v_p = 3*10^8$ m/s,
4. the distance between SNs is uniform, and is within a range d = (10-150) m,
5. CPU clock frequency is f = 12 MHz and the average number of instruction to process one byte is n = 10 instructions with t = 4 clock periods per instruction.

According to the invoved assumptions and by substituting these values into T_c, T_t and T_p we obtain:

$$Tt = N / R = 16*8/720\,kbps = 173{,}6\,\mu s \quad (3)$$

$$Tp = d / v_p = 0{,}1\,\mu s \text{ for } d = 30\,m \qquad (4)$$

$$Tc = N * n * t / f = 16 * 10 * 4 / 12 = 53{,}33\,\mu s \qquad (5)$$

For traffic without retransmissions $Tq = 0$

Having in mind that $T_t >> T_p$ and $T_c >> T_p$ we can ignore T_p and T_q in respect to T_p and T_c, respectively. We will involve now a new metric $\Phi(m,h)$ called trafic reduction factor. The metrics $\Phi(m,h)$ is defined as a ratio between the total communication time defined in our porposal and the total comunication time defined in Ref. [4]. This metrics points to the percentage of decreasing the total commuication timesT_{ps} in respect to T_{ref}, in terms of number of messages *m*, and number of hops *h*, as parameters.

$$\Phi(m,h) = \frac{Tps}{Tref} = \frac{mh + h + \frac{Tt}{Tc}(m + h - 1)}{2mh + \kappa mh} \qquad (5)$$

By substituting the values for T_t=173,6 ms and T_c=53,33 ms we define $T_t/T_c \approx 13/4$. Accordingly (5) we obtain:

$$\Phi(m,h) = \frac{Tps}{Tref} = \frac{4mh + 17h + 13m - 13}{21mh} \qquad (6)$$

Figure 4. sketches the metric Φ in terms of *m*, with *h* as parameter. As can be seen from Fig.4 by increasing *m* and *h*, the metric $\Phi(m,h)$ decreases what means that our proposal has better performance (from 14% for *m*=4 and *h*=1, up to 50% for *m*=4 and *h*=4). Let known that further hop increases (*h*>=5), does not result a linear $\Phi(m,h)$ decrease.

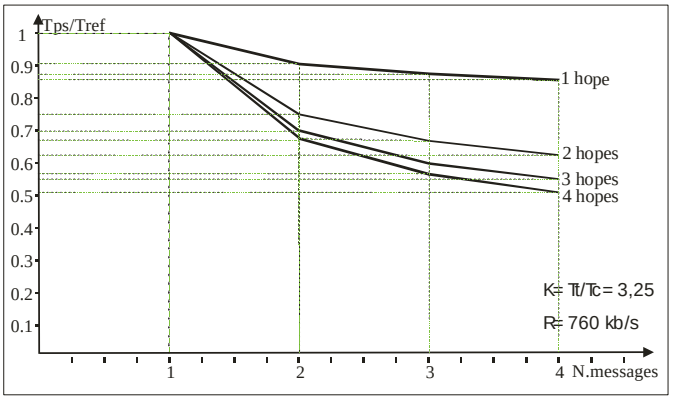


Fig.4 Metric $\Phi(m,h)$ in terms of *m* with *h* as a parameter

Figure 5. presents the metric $\Phi(m,h)$ in terms of *h*, with *m* as parameter. By analyzing Fig.5 we can conclude that for larger messages (m>=2) the metric $\Phi(m,h)$ decreases what implies that our proposal has better performance in respect to the Ref.[4] (from 10% for *m*=2 and *h*=1, up to 49% for *m*=4 and *h*=4).

V. CONCLUSION

The problem of impementation of TCP/IP protocols in WSNs is considered. This possibility allow us to access each sensor node as an TCP/IP addressable unit. In order to decrease the node access time and to pertain standard WSN protocol in a process of addresses mapping, we have involved within the gateway architecture a translation table which can provide multiple node addresses for a single sensor node. In this manner we have decreased the number of bytes in WSN message header and increased data transfer efficiency. In order to evaluate the performance of our proposal we have involved a metrics called traffic reduction factor. The obtained results, that performace improvement during message transfer among sensor nodes in WSN up to 50 %, in respect to the solution proposed in Ref.[4], are obtained.

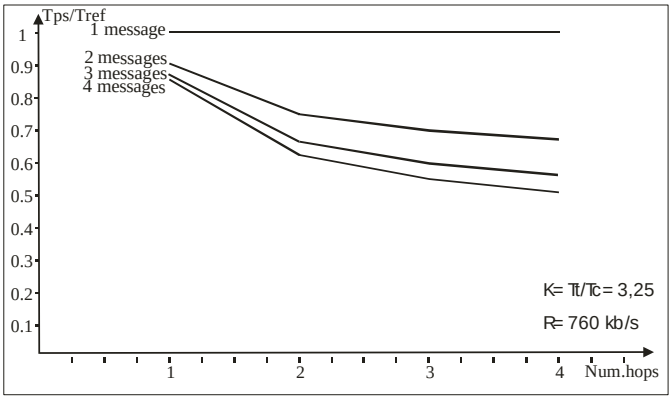


Fig.5 Metric $\Phi(m,h)$ in terms of *h* with *m* as a parameter

REFERENCES

[1] M.Zuniga, B.Krishnamachari, ''Integrating Future Large-scale Wireless Sensor Networks with Internet'', www.cs.usc.edu/Research/techreports/papers/03-792.pdf

[2] A.Dunkels, ''Towards TCP/IP for Wireless Sensor Networks'',Malardalen University Licentiate Thesis No. 45, Swedish Institute of Computer Science, March 2005.

[3] C.Intanagonwiwat, R.Govindan, D.Estrin, ''Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks'', Proc.ACM MobiCom'00, Boston, MA, 2000, pp. 56-67

[4] S.Lei, W.Xiaoling, Xu Hui, Z.Jie, J.Cho, S.Lee, ''Connecting Heterogeneous Sensor Networks with IP BasedWire/WirelessNetworks'',SEUS-WCCIA`06,2006

[5] H.Dai, R.Han, ''Unifying Micro Sensor Networks with Internet via Overlay Networking'', Proc.IEEE Emnets-1, Nov, 2004

[6] K.Mayer, W.Fritsche, ''IP-enabled Wireless Sensor Networks and their integration into the Internet'', http://portal.acm.org/ft_gateway.cfm?id=11426878&type=a5-mayer.pdf

[7] M.Zhang, S.Pack, K.Cho, D.Chang, Y.Choi, T.Kwon, ''An Extensible Interworking Architecture (EIA) for Wireless Sensor Networks and Internet'', www.mmlab. snu.ac.kr/publications/docs/EIA_APNOM2006.pdf

[8] C.Westphal, ''Layered IP Header Compression for IP-enabled Sensor Networks'', www.people.nokia.net/cedri c/Papers/icc06.pdf

[9] Z.Z.Marco, K.Bhaskar, ''Integrating Future Large-scale Wireless Sensor Networks with Internet'', www.cs.usc.edu/Research/techreports/papers/03-792.pdf

Simulation Modelling of Overflow Probabilities in Tandem Queuing System

D. Radev, E. Rashkova and D. Stankovski

***Abstract* – An approach for simulation modelling of overflow probability based on the rare event simulation and limited relative error is suggested. The developed algorithm for tandem queuing system help to speed-up the simulation for estimation of overflow probability, which in tandem queuing system is a rare event. Numerical and simulation results are shown.**

***Keywords* –Simulation Modelling, Markov Chains, Queuing Theory, Overflow Probability, Rare Event.**

I. INTRODUCTION

Queuing systems are convenient reference models for research and analysis of communication systems performance, where very important is to evaluate the overflow probability. In real communication systems it is impossible to receive the values of these probabilities with direct methods, that's why is necessary to involve simulation modeling [1].

The model of tandem queuing system is under study in order to be able to compare received simulation measures with known analytical or numerical results. For example, in tandem queuing system, the probability that the content of the second buffer exceeds some high level L before it becomes empty starting from a given state is a rare event, because it probability most often is less than 10^{-6}.

The estimation of rare event probability using Monte Carlo simulation requires a very long computer processing time, and cannot be implemented [4]. Lately two basic methods of the rare event simulation were developed: Importance splitting of the sample path after reach definite intermediate level between the starting level and the rare event [6]; and the Importance Sampling (IS) generation, where the probability density function is used as the rare event evaluation measure, which can be compared and changed based on the likelihood ratio of the less rare event probability density function [2].

RESTART (Repetitive Simulation Trials After Reaching Thresholds) approach is an accelerated simulation method, which belongs to the so-called importance splitting methods, used to the speed-up the rare event simulation [5,7].

The Limited Relative Error (LRE) measures the complementary distribution function of the queue occupancy and performs the Run Time Control (RTC) of the simulation. The limited relative error performs the run time control with two conditions: first the Large Sample Conditions and second the Relative Error Condition. The first condition assures that the queuing system has reached the steady state. The second condition represents a measure to estimate the relative error at the current state of the simulation.

For simulation modeling of overflow probabilities in tandem queuing system in this paper consider RESTART approach combined with limited relative error.

II. OVERFLOW PROBABILITY OF TENDEM QUEUES

Many models used for evaluation of queuing systems performance can be presented as discrete-time Markov chains (DTMCs) [3]. This is the case if all inter-arrival times and service times have an exponential distribution, and the quantity of interest is the overflow probability. Note that some other performance measures, like delays, cannot be obtained from the discrete-time Markov chain description. A DTMC for a Markov queuing model has a highly regular structure. First of all, the states typically can be arranged conveniently on a grid with as many dimensions as the number of queues, where each of coordinates represents the number of customers in one of the queues. Secondly, every transition in the DTMC corresponds to an elementary event in the queuing model: an arrival or a service completion at one of the queues. These events are known as "transition events" and they are defined independently of the state; i.e., there is only one transition event for a service completion at a given queue, and this single transition event corresponds to a transition out of every state in the DTMC in which this particular queue is non-empty. From the other side not all transition events are "enabled" in every state: e.g., in a state where a particular queue is empty, the service completion event of that particular queue is not possible, i.e., not enabled.

Consider the overflow probability of the total population in a queuing network consisting of two queues in tandem presented on Figure.1, where customers arrive at the first queue according to a Poisson-process with arrival rate λ. Both servers have exponentially-distributed service times with rates μ_1 and μ_2.

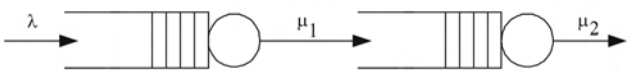


Figure 1. Two-queue tandem network.

The state of the system at any time is given by the two integer values n_1 and n_2, which are the number of customers in the first and second queues, respectively.

[1]D. Radev, Department of Communication Technique and Technologies, University of Rousse, Phone: +359 82 888 841, E-mail: dradev@abv.bg

[2] E.P.Rashkova, Department of Communication Technique and Technologies, University of Rousse; E-mail: elena_pl@gbg.bg

[3] D. Stankovski, Department of Communication Technique and Technologies, University of Rousse; E-mail: draganstankovski@yahoo.com

The difficulty of applying standard simulation techniques arises when the first queue is the bottleneck and the rare set definition is related to the value of n_2. In this case it would be better to apply one of splitting methods for rare event simulation with careful choice of the importance function [3]. On Figure 2 is presented state space of two queues in tandem network.

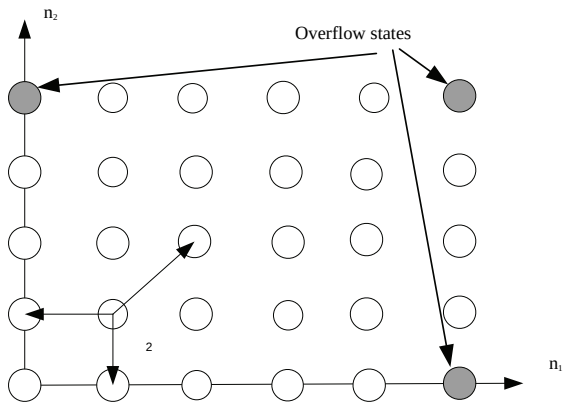


Figure 2. State space of two queues in tandem network

For the considered two-queue tandem network the rare event set A include the following three cases, where L is the level of overflow of the queues (1).

$$\begin{aligned} &n_1 + n_2 \geq L, \\ &n_2 \geq L, \\ &\mathrm{Min}(n_1, n_2) \geq L \end{aligned} \qquad (1)$$

There exist three possible cases for rare event:

A. Rare event defined as $n_1 + n_2 \geq L$

For this definition of the rare event set, the possible states for rare event are $(0,n_1+n_2)$ or $(0,(n_1+n_2-1))$. The importance of these states is different. The higher value of λ_1 (for n_1+n_2), given that customer at n_1 has to be served by both servers before leaving the system, while a customer at n_2 has to be served only at the second one. Here the bottleneck is the first queue, so the states with high value of n_1 and low value of n_2 have the highest probability. μ_1 μ

B. Rare event defined as $n_1 \geq L$

For this definition of the rare event set $n_1 \geq L$, to obtain states with closer importance, it seems that the weight given to n_2, have to be given to n_1.

C. Rare event defined as $Min(n_1,n_2) \geq L$

For $n_1 \leq L$ and $n_2 \leq L$, the rare event set is in the queue with fewer customers.

III. The LRE approach

The limited relative error approach tries to construct independent samples from a given time series by building batches that tend to have lesser autocorrelation than the original series. It leads to reduction of autocorrelation and results in confidence intervals for the mean of the simulation results. The limited relative error approach is implemented with counters to measure the quantities: n - number of trials; h_i - state frequency; l_i - number of times the queue was in state i when a cell arrived; a_i - transition frequency for decreasing number of transitions from j to i between two arrivals, with $j>I$. From these counters the quantities l_i and d_i are (2) and (3).

$$l_i = n - d_i \qquad (2)$$

$$d_i = \sum_{j=1}^{k} h_j, with \quad d_0 = n \qquad (3)$$

If the large sample conditions and relative error condition has to be fulfilled and the results of it are summarized in the end of the algorithm. The large sample conditions are presented as (4).

$$\begin{aligned} &n \geq 10^3, \\ &(l_i, d_i) \geq 10^2, \\ &(a_i, d_i - c_i, l_i - a_i) \geq 1 \end{aligned} \qquad (4)$$

The relative error condition is presented as (5).

$$\mathrm{RE}_{\max i} = \frac{1 - d_i/n}{d_i} \cdot \mathrm{Cor}_i \qquad (5)$$

The LRE approach starts with index i=1, tests the conditions (4) and (5), and if the conditions are fulfilled the LRE advances to the next index i. The simulation is terminated for the last index $i=L_m$, where L_m is the overflow conditional probability.

IV. Algorithm with RESTART

The RESTART method is a simple simulation method for the estimation of small probabilities. Applying the RESTART means to split the possible range of values of λ into regions with different importance. The given set of thresholds is L_i, $i=0,1,2,...,m$. The system reaches the state L_0, while evaluating the first interval $[0, L_0]$ the system states are saved. Once the values for the first interval are established at the time t_0, the simulation restarts from one of the previously saved states at threshold L_0. Each time when the threshold L_0 is crossed a restart is performed with one of the saved states, which is chosen randomly and the system is reload with this state.

The RESTART method is used to the speed-up the rare event simulation for estimation of overflow probability defined as (6).

$$\begin{aligned} &G_l = P\{\lambda \geq L\} = \\ &P\{\lambda \geq L | \lambda \geq L_{m-1}\} P\{\lambda \geq L_{m-1} | \lambda \geq L_{m-2}\} \\ &\ldots P\{\lambda \geq L_1 | \lambda \geq L_0\} P\{\lambda \geq L_0\} \end{aligned} \qquad (6)$$

The relative error $RE_{\max_i}$ is a function of the overall relative error RE_{max} and the number of thresholds m is defined as (7).

$$RE_{\max_i} = (1 + RE_{\max})^{\frac{1}{m+1}} - 1 \qquad (7)$$

The proper choice of the thresholds L_i leads to a maximum speed-up according to (8).

$$\frac{G_i}{G_{i-1}} = e^{-2} \quad (8)$$

Since the complementary cumulative density function is monotonically decreasing, and the thresholds are properly chosen, the RESTART method will speed-up the simulation.

The following algorithm with RESTART/LRE for tandem queuing systems was developed:

Algorithm RESTART/LRE

Step 1 Initialization

Specification of the rare event L, number of levels m, the values of the thresholds $L_0,\ldots,L_{m-1}$, and the maximal relative error $RE_{\max}$;

Definition of generated random variables β and the upper and lower boundary for generated variables $\beta_{\max}$ and $\beta_{\min}$;

Definition of parameters for arrival and service process;

Calculation maximal relative error $RE_{\max i}$ from $RE_{\max}$;

Generate model objects (traffic sources, queues, network nodes);

Initialization of counters, parameters of the system and the matrix of the state space (Fig. 2)

Definition of the rare event set and kind of the model:

Case 1: Rare event set defined as $n_1 + n_2 \geq L$

Case 2: Rare event set defined as $n_2 \geq L$

Case 3: Rare event set defined as $\text{Min}(n_1 + n_2) \geq L$

Initialization of number of runs n:=0 and the simulation time s:=1;

Set the scaling factor U:=1;

Start simulation;

Step 2 Outer loop

If ($n \geq 10^3$ AND (l_i, d_i) $\geq 10^2$ AND (a_i, l_i-a_i, d_i-a) ≥ 10)

Step 3 For i=0

While not (error< $RE_{\max}$ for L_0-1 to L_0)

Simulation continues with generating a new value of random variables β with Poisson distribution.

Estimation of generated random variable β: whether β is in the left side of L_0-1 or in the right side of L_0; Result of estimation: the new value of β is accepted if it is in the right side and the state is stored for RESTART;

Evaluation Counter=Counter+1,

End;

Step 4 Inner loop

For i=1 to m

Set the thresholds $L_0,\ldots,L_m$ for the rare event L;

While not (error< $RE_{\max}$ for L_{i-1} to L_l)

Restore one random state from L_{i-1} ;

Calculation Relative Error for Complementary Cumulative Distribution Function for $G_{L_{i-1}|\beta \geq L_{i-1}}$ to $G_{L|\beta \geq L_{i-1}}$. If the $RE > RE_{\max}$ the simulation stop and the new random state is restored;

Simulation continues with generating data: new value of random variables β with Poisson distribution end estimation whether β is in the left side of L_i-1 or in the right side as L_i. The new value of β is accepted if it is in the right side and the state is stored for RESTART.

Evaluation Counter=Counter+1;

n:=n+1

Set scaling factor U:= $G_{L|\beta \geq L_{i-1}}$;

Multiplication of G-values with U for $L_{i-1},\ldots, L_i$;

Continue simulation till the last value of n;

End;

Step 5 Output results

Calculation of final results:

$$\hat{G}(x) = \hat{G}_i = d_i / n$$

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{k} d_i$$

$$\hat{\rho}(x) = \hat{\rho}_i = 1 - \frac{c_i / d_i}{1 - d_i / n}$$

$$\text{Cor}[x] = \text{Cor}_i = (1 + \hat{\rho}_i)(1 - \hat{\rho}_i)$$

$$\text{RE}[x]^2 = \text{RE}_i = \frac{1 - d_i / n}{d_i} \cdot \text{Cor}_i$$

END.

V. Numerical Example

The simulation results for RESTART/LRE algorithm are received with simulation program on Visual C++ with implemented of Ptolemy as a powerful, object-oriented simulator and Microsoft Visual Studio 3.0. The simulator is available with source code for operating systems Windows 9x, NT, XP. All the simulation examples were executed on Personal computer with properties: Pentium, 1,4 GHz, 312Mb RAM, operating system Windows XP Professional. One example of a part of developed simulation software is shown on Figure 3.

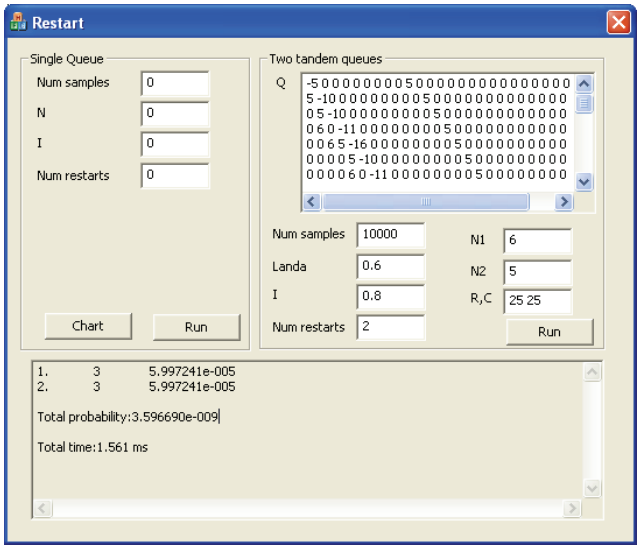


Figure 3. Program realization

Customers with Poisson arrival distribution enter the first queue with buffer size n_1=6 for service completion and enter the second one with buffer size n_2=5 for service completion as is shown on Figure 2. The mean arrival rate at the first queue

is λ and the service time is exponentially distributed in each queue with mean service rates μ_1 and μ_2, respectively. The load at each queue is $\rho_1 = \lambda/\mu_1$ and $\rho_2 = \lambda/\mu_2$. The generated random variables with Poisson arrival distribution are shown on Figure 4.

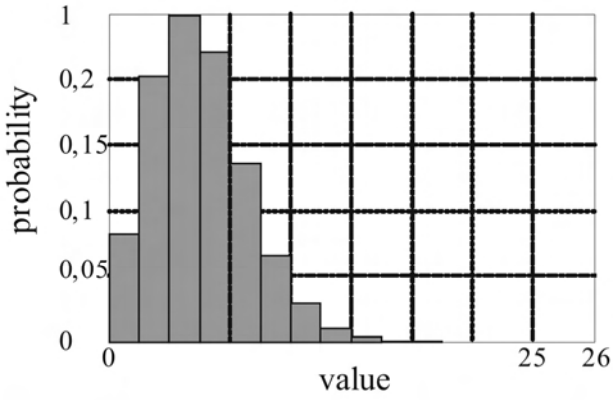


Figure 4. Generation of variables with Poisson distribution

The buffer size at each queue is assumed to be finite. Consider example with arrival rates λ=0,6 and service completion with rates μ_1=0,7 and μ_2=0,8. The number of simulation samples is n=10 000.

In all the runs, the simulation length was adjusted to have a relative half width of the 95% confidence interval (relative error) equal to 10%. The interval width was evaluated using the batch means method [3].

The investigated states defined as rare event L are in the end of the first queue (5,0), the minimum (5,4), and in the end of the second queue (0,4). The importance of each of these states is different. The thresholds were chosen as L=0,8 and L=0,9. The number of restarts in the numerical example is 2. The number of samples is n=10 000.

In Table 1 are shown the received simulation results for L=0,8; relative error RE=5%, λ=0,3, μ_1=0,4 and μ_2=0,9. For comparison are given theoretical results and simulation time.

Table 1 Results for tandem queue network with n=10 000 simulation samples,

Definition of Rare state	G_L theoretical	G_L	Time, *ms*
First case: (5,0)	0,0318598651	9,292000e-006	2,549
Second case: (0,4)	0,0020832578	6,912656e-006	2,346
Third case: (5,4)	0,0000417815	7,849558e-006	2,836

In Table 2 are shown the received simulation results for L=0,9; relative error RE=10%, λ=0,4, μ_1=0,7 and μ_2=0,9. For comparison are given theoretical results and simulation time.

Table 2 Results for tandem queues network with n=10 000 simulation samples,

Definition of Rare state	G_L theoretical	G_L	Time, *ms*
First case: (5,0)	0,0134789964	9,582591e-006	2,836
Second case: (0,4)	0,0078628157	5,983916e-006	2,513
Third case: (5,4)	0,0003996907	3,129537e-006	3,349

VI. Conclusion

The simulation modeling of overflow probability based on the rare event simulation and limited relative error was under consideration. The developed algorithm for tandem queuing system help to the speed-up the simulation for estimation of overflow probability, which in tandem queuing system is a rare event. The received simulation results for the basic reference model, two server queuing system – FIFO with a finite buffer size N, shows that suggested approach and algorithm are suitable for rare event simulation with large number of customers in complex communication networks.

Acknowledgments

This work is a part of the research project ВУ-ТН-105/2005 of Bulgarian Science Fund at Ministry of Education and Science.

References

[1] Boer. P. Analysis and Efficient Simulation of Queuing models of Telecommunication Systems. PhD thesis, Univ. of Twente, 2000.

[2] Georg, C., Schreiber, F. The RESTART/LRE method for rare event simulation. Winter Simulation Conference, Coronado, CA, USA, 1996, 390-397.

[3] Radev, D. Modeling of Rare events in broadband digital networks, Sofia, 2006.

[4] Radev, D., Rashkova, E. Algorithms for traffic flows modelling with Monte Carlo simulation TELEKOM'2006, Varna, 2006.

[5] Radev, D., Iliev, M., Arabadjieva, I. RESTART Simulation in ATM networks with tandem queue. - Proceedings of the International Conference Automatics and Informatics'2004, Sofia, Bulgaria, 37-40.

[6] Radev, D. Splitting method in rare event simulation for high - speed communication networks. - Proceedings of the International Conference Automatics and Informatics'2003, Vol.1, Sofia, 103-107, 2003.

[7] Villen Altamirano, M. Villen Altamirano, J. RESTART: a method accelerating rare event simulations, in 13th International Teletraffic Congress, Copenhagen, North- Holland, 1991, pp. 71-76.

The Effect of Load/Store Queue Size on System Performance

Daniela Curcievska[1], Pece Mitrevski[1] and Marjan Gusev[2]

Abstract. **As processors continue to exploit more instruction level parallelism, greater demands are placed on the performance of the memory system. Load latency is identified as a major bottleneck in modern superscalar processors. In this study we propose load-forwarding via LSQ in issue and execution phases of the pipeline, in order to get shorter execution time. We come to conclusion how changing the LSQ size affects system performances. Namely, on average, a 6-9% decrease in the number of execution cycles with the increase of LSQ size up to 32 entries, which is not the case afterwards. With additional increase in size, gain speedup becomes insignificant, while for some of the test-examples the number of execution cycles slightly increases.**

Keywords. **ILP, memory, memory instructions, load forwarding, pipeline, RUU, LSQ**

I. Introduction

The objective of modern superscalar processors is to maximize the instruction level parallelism (ILP) that can be extracted from programs. The most basic method uses for extracting more ILP from program is out-of-order execution. Unfortunately, out-of-order execution by it self does not provide the desired level of ILP. The program's control flow [1] and data flow [2] impose serious limits on level of parallelism that can be extracted. Therefore, most modern processors employ aggressive branch prediction mechanisms to relax the control flow constraints that limit the ILP. To overcome the data-flow limits, researchers have suggested data speculation [2, 3, 4, 5, 6] to be used.

II. Memory Dependence Problem

Modern processors exploit ILP by executing instructions in an order that is defferent from the sequential program order, which is called out-of-order execution. In other words, independent instructions whose operanda are ready can be scheduled and executed before older instruction that are still waiting for their operands.

Hence, to support out-of-order execution, the hardware needs to be able to precisely determine the dependence among instructions so that sequential program semantics will not be violated. In case of register dependences, determining which intructions are dependent is easy due to the explicit encoding of architecture register names in the instruction format. Memory dependences are much harder to determine, because memory addresses are not explicitly encodedin the instruction format and need to be dynamically generated. However, this dynamic generation of memory addresses is not done in sequential program order. Hence, when a load instruction is ready to be scheduled, it is likely that there are older store instructions in an instruction window whose addresses have not yet been determined. This problem is know as the memory anti-aliasing [7]. A related concept, the process of determining if two memory instuctions access the same memory location is called *memory disambiguation*.

There are several ways to attack the memory anti-aliasing problem. One possible solution is to execute all store and load in total program order. Considering that load and store instructions comprise a large fraction of instructions in most programs, imposing a total order on memory reference instructions would seriously limit the ILP that can be extracted from programs. A slightly less conservative approach is to delay the scheduling of load until all previous store addresses become available. This approach limits the amount of ILP extracted from memory disambiguation, because it is unlikely that a load will conflict with many of the previous stores.

The aggressiveness of memory disambiguation using speculative scheduling also depends on how much more parallism we want to exploit. A less aggressive approach is to predict whether a load will conflict with any older store in the instruction window. In this case, the load can not be scheduled until all older stores execute. A more aggressive scheme is to predict that a load conflicts with a particular earlier store, if any, and delay the scheduling of the load until that particular store executed. However, a load instruction will not be unnecessarily delayed when the predict is corect.

One extreme form of speculative memory operation scheduling is to always assume that the load that is to be scheduled will not conflict with any of the unknown store addresses. This kind of extreme speculation is not the best performing technique due to the cost of recovery as a result of misprediction.

It is real to assume that existing of Load and Store Queue (LSQ) that enables, when a store instruction executes, it writes its data into LSQ, and a later load that is dependent on the store to access and read data from LSQ will increase the amount of ILP. This is called *load forwarding* [8]. In order to

[1]Daniela Curcievska and Pece Mitrevski are with the Faculty of Technical Sciences, University "Sv. Kliment Ohridski", I. L. Ribar bb, 7000 Bitola, Macedonia, E-mail: daniela.curcievska@uklo.edu.mk, pece.mitrevski@uklo.edu.mk

[2]Marjan Gusev is with the Institute of Informatics, Faculty of Science, "Ss. Cyril and Methodius" University, Arhimedova 5, P.O Box 162, 1000 Skopje, Macedonia, Email: marjan@on.net.mk

be able to do load forwarding, a load needs to have its address calculated. In continuation, the Baseline Superscalar Processor Model which supports usage of LSQ data structure will be presented.

III. Baseline Superscalar Processor Model

In order to describe the usage and the management of the LSQ data structure, an overview of Superscalar Processor Model will be given. Baseline Superscalar Processor Model with its five phases of pipelining is given in Figure 1.

Register Update Unit (RUU) is a hardware data structure that is used to resolve data dependencies by keeping track of an instruction's data and execution needs and that commits completed instructions in program order.

Each register in the general purpose RegFile has two associated n-bit counters (n of 3 is typical)

1. NI (number of instances) – the number of instances of a register as a destination register in the RUU
2. LI (latest instance) – the number of the latest instance

When an instruction with destination register address Ri is dispatched to the RUU, both its NI and LI are incremented. Dispatch is blocked if a destination register's NI is 2n -1, so only up to 2n – 1 instances of a register can be present in the RUU at the same time. When an instruction is committed (updates the Ri value) the associated NI is decremented. And finally when NI = 0 the register is "free" (there are no instruction in the RUU that are going to write to that register) and LI is cleared.

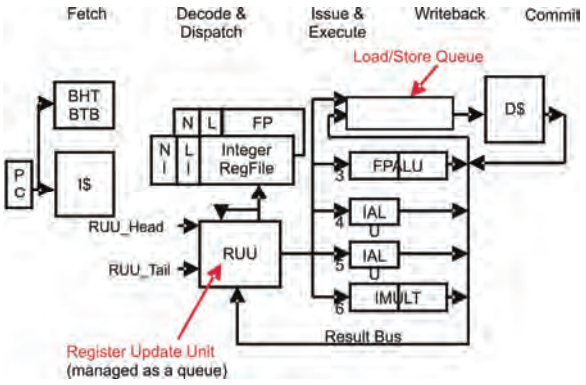


Fig. 1. Baseline Superscalar Processor Model

This processor model support pipeline instruction flow in five phases:

- Fetch (in program order): Fetch multiple instructions in parallel from the I$
- Decode & Dispatch (in program order):
  In parallel, decode the instr's just fetched and schedule them for execution by dispatching them to the RUU; Loads and stores are dispatched as two (micro)instructions – one to compute the effective addr and one to do the memory operation
- Issue & Execute (out of program order): As soon as the RUU has the instr's source data and the FU is free, the instr's are issued to the FU for execution
- Writeback (out of program order): When done the FU puts its results on the Result Bus which allows the RUU and the LSQ to be updated – the instr completes
- Commit (in program order): When appropriate, commit the instr's result data to the state locations (i.e., update D$ and RegFile)

By managing the RUU as a queue, and committing instruction from RUU_Head, instructions are committed (aka retired) in the order they were received from the Decode & Dispatch logic (in program order). Committing the instr's result data to the state locations (RegFile and D$) are buffered in the RUU and LSQ until commit time.

The LI counter allows multiple instances of a specific destination register to exist in the RUU at the *same* time via register renaming, and with that it prevents output dependencies.

The RUU perform the following four tasks in parallel every cycle:

1. Accepts new instructions from the Decode & Dispatch logic
2. Monitors the Result Bus to resolve true dependencies and to do write back of result data to the RUU
3. Determines which instructions are ready for execution, reserves the Result Bus, and issues the instruction to the appropriate FU
4. Determines if an instruction can commit (i.e., change the machine state) and commits the instruction if appropriate

Loads and stores are dispatched to the RUU as two (micro)instructions – one to compute the effective addr and one to do the memory operation:

lw R1,2(R2)	becomes	addi	R0,R2,2
		lw	R1,R0
sw R1,6(R2)	becomes	addi	R0,R2,6
		sw	R1,R0

At the same time with dispatching memory instruction to RUU, LSQ entry for that instr is allocated. Each LSQ entry consists of a Tag field (RegFile addr || LI) and a Content field. The LI counter allows for multiple instances of stores (writes) to a memory address.

When a load completes (the D$ returns the data on the Result Bus) or a store commits (in program order) the LSQ entry is released.

Instruction dispatch is blocked if there is not a free LSQ entry and two free RUU entries.

For load instruction, when the address is computed then the load instruction is ready to issue. The address is copied to the LSQ entry with the (associatively) matching RegFile||LI address and is sent to the D$ for execution. When the load value is returned it is copied into the destination content field of some instruction on the RUU. Advance the RUU_Head and LSQ_Head appropriately.

For store instruction, when the address is computed then the store instruction is read to issue. The computed address is copied to the LSQ entry with the (associatively) matching

RegFile||LI address along with the store data value. When the store instruction reaches the RUU_head, its companion LSQ entry (which by now is at the head of the LSQ) is sent to the D$ (to complete the store operation) and both the RUU_Head pointer and the LSQ_Head pointer are advanced.

Performance can be improved if loads are allowed to bypass previous stores, this is called *load bypassing*. That can be done only if the memory addresses of *all* previous stores dispatched to the RUU are known since must first check for load data dependency on previous uncommitted store instructions. That is facilitated by having a common load and store queue.

Store instructions are not allowed to bypass previous loads, so there are no antidependencies between loads and stores. Also store instructions are committed to the D$ in program order, so there can not be output dependencies between stores.

When a load's address becomes known, the address is compared (associatively) to see if it matches an entry already in the LSQ (i.e., if there is a pending operation to the same memory address).

- If the match in the LSQ is for a load, the current load does not need to be issued (or executed) since the matching pending load will load the data
- If the match in the LSQ is for a store, the current load does not need to be issued (or executed) since the matching pending store can directly supply the destination Content for the current load
- If there is no match, the load is issued to the LSQ and executed when the D$ is next available

When the RUU# of the load instruction appears on the Result Bus along with the memory data, the load completes by updating the RUU and releasing the RUU and LSQ entries for committed load.

When a store's address (and the store data) becomes known, the address is compared (associatively) to see if it matches an entry already in the LSQ (i.e., if there is a pending operation to the same memory address)

- If the match in the LSQ is for a load, the current store is issued to the LSQ
- If the match in the LSQ is for a store, the current store is issued to the LSQ with an incremented LI
- If there is no match, the store is dispatched to the LSQ

Store instructions are held in the LSQ until the store is ready to commit (i.e., until its partner instruction reaches the RUU_Head) at which time the store is executed (i.e., the data and address are sent to the D$) and the RUU and LSQ entries are released.

IV. Out-Of-Order Processor Timing Simulation

In order to show how LSQ size affects system performance, simulations of execution on different test-examples with sim-outorder simulator were made, which is part of The Simple Scalar Tool Set [11].

This simulator support out-of-order issue and execution, based on the RUU. The RUU scheme uses a reorder buffer [10] to automatically rename registers and hold the results of pending instructions. Each cycle the reorder buffer retires completed instructions in program order to the architected register file.

The processor's memory system employs a load/store queue. Store values are placed in the queue if the store is speculative. Loads are dispatched to the memory system when the memory addresses of all previous stores are known. Loads may be satisfied either by the memory system or by an earlier store value residing in the queue, if their addresses match.

The main loop of the simulator, located in sim_main(), is structured as follows:

```
ruu_init();
for ( ; ; ) {
        ruu_commit();
        ruu_writeback();
        lsq_refresh();
        ruu_issue();
        ruu_dispatch();
        ruu_fetch();
}
```

This loop is executed once for each target (simulated) machine cycle. By walking the pipeline in reverse, inter-stage latch synchronization can be handled correctly with only one pass through each stage. When the target program terminates, the simulator has generated the statistics.

The fetch stage of the pipeline is implemented in ruu_fetch(). The fetch unit models the machine instruction bandwidth, and takes the following inputs: program counter, the predictor state and mispericition detection from the branch execution unit(s). Each cycle, it fetches instructions from only one I-cache line. After fetching the instructions, it places them in the dispatch queue, and probes the line predictor to obtain the correct cache line to access in the next cycle.

The code for dispatch stage of pipeline resides in ruu_dispatch(). The routine is where instruction decoding and register renaming is performed. The function uses the instructions in the input queue filled by the fetch stage, a pointer to the active RUU and the rename table. Once per cycle, the dispatch takes as many instructions as possible from the fetch queue and places them in the scheduler queue. The dispatch routine enters and links instructions into the RUU and LSQ.

The issue stage of the pipeline is contained in ruu_issue() and lsq_refresh(). These routines model instruction wakeup and issue to the functional units, tracking register and memory dependences. Each cycle, the scheduling routines locate the instructions for which the register inputs are all ready. The issue of ready loads is stalled if there is an earlier store with unresolved effective address in the LSQ. If the address of the earlier store matches that of the waiting load, the store value is forwarded to the load. Otherwise, the load is sent to the memory system.

The execute stage is also handled in ruu_issue(). Each cycle, the routine gets as many ready instructions as possible from the scheduler queue. The functional units availability is also checked and if they have available access ports, the

instruction are issued. Finally, the routine schedules writeback events using the latency of the function units.

The writeback stage resides in ruu_writeback(). Each cycle it scans the event queue for instruction completions. When it finds a completed instruction, it walks the dependence chain of instruction outputs to mark instructions that are dependent on the completed instruction. If a dependent instruction is waiting only for that completion, the routine marks it as ready to be issued.

Ruu_commit() handles the instructions from the writeback stage that are ready to commit. This routine does in-order committing of instructions, updating of the data caches with store values and data. The routine keeps retiring instructions at the head of the RUU that are ready to commit until the head instruction is one that is not ready. When an instruction is committed, its result is placed into the architected register file and the RUU/LSQ resources devoted to those instructions are reclaimed.

We have simulated the execution of five test-examples that are part of the SimpleScalar Tool Set (math, fmath, llong, lswlr, printf) by changing LSQ size between 4 and 512 entries (Fig. 2).

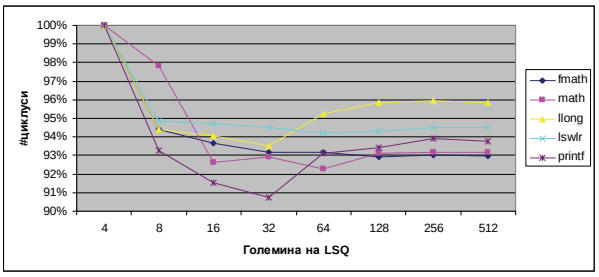


Fig 2. Simulation Results

It is observed that all test-examples follow the same performance trends in the eight different simulations. There is, on average, a 6-9% decrease in the number of execution cycles with the increase of LSQ size up to 32 entries, which is not the case afterwards. The additional speedup becomes insignificant, while for some of the test-examples (llong, printf) the number of execution cycles slightly increased.

V. Conclusion

The LSQ structure enables memory instructions that address the same memory locations to bypass data values among them during Issue and/or Execute phases in the pipeline. However, it is only possible if all memory dependences have already been resolved. As memory instructions get their values from the LSQ (not further down from the memory hierarchy), the execution time is presumably shorter. The main concern is: how the LSQ size affects system performance?

From the previous, we come to a conclusion that even relatively small queue (32 entries in our case) is large enough to perceive performance gain. Further increase in queue size does not lead to additional speedup and, in some cases, it negatively affects system performance.

References

[1] Lam M.S. and Wilson R. P., “Limits of Control Flow on Parallelism”, Proceedings of the 19th Annual Symposium on Computer Architecture; May 1992.

[2] Lipasti M. H and Shen J.P., “Exceeding the Dataflow limit via Value Speculation”, Proceedings of the 29th Annual Syposium on Microarchitecture; December 1996.

[3] Lipasti M. H, Wilkerson C. B. and Shen J. P.,”Value Locality and Load Value Prediction,” Proceedings of 7th International Conference on Architectural Support for Programming Languages and Operating System; October 1996

[4] Moshovos A., Breach S.E.,Vijaykumar T. N. and Sohi G. S., ”Dynamic Speculation Synchronization of Data Dependence,” Proceedings of the 24th Annual Symposium on Computer Architecture; June 1997

[5] Gusev M., Mitrevski P., “Modeling and Performance Evaluation of Branch and Value Prediction in ILP Processors”, International Journal of Computer Mathematics, Vol. 80, No. 1, pp. 19-46, 2003

[6] Mitrevski P., Gusev M., “On the Performance Potential of Speculative Execution Based on Branch and Value Prediction”, International Scientific Journal Facta Universitatis, Series: Electronics and Energetics, Vol. 16, No. 1, pp. 83-91, 2003

[7] Patt Y. N., Melvin S.W., Hwu W.W. and Shebanow, “Critical Issue Regarding HPS, A High Performance Microarchitecture,” Proceedings of the 18th Annual ACM/IEEE Workshop on Microprogramming; December 1985

[8] Johnson M., Superscalar Microprocessor Design. Englewood Cliffs N. J.; Printice Hall, 1991

[9] Patterson D.A. and Hennessy J.L. Computer Organization and Design. Morgan Kauffman Publishers, 2005

[10] Sohi G.S., Instruction Issue Logic for High Performance, Interruptible, Multiple Functional Unit, Pipelined Computers, IEEE Transactions on Computers, Volume 39, Issue 3; March 1990

[11] Austin T.M. and Burger D., The SimpleScalar Tool Set, Version 2.0. Technical Raport 1342, Computer Sciences Department, University of Wisconsin, Medison, June 1997

Improving Mobile Complex Service Composition with Graph Technology

Ljupco Antovski[1], Marjan Gusev[1] and Pece Mitrevski[2]

Abstract. **The advances in mobile and wireless technologies enable cost effective usage of e-government services. The case of e-government and m-government in Macedonia is outlined and discussed. The latest findings from the developed platform in the M-GOV projects are presented. The M-GOV platform is based on simple ideas that together provide architecture with high level of flexibility and low levels of technological requirements. The architecture is designed to scale, both from a technical and a financial perspective and to be applicable to integration scenarios from small agency scenarios through to large administration scenarios. The M-GOV architecture is formed of three major components: Citizens access devices, Service Discovery Directory and the Collection of public electronic services from various sources. The built-in M-GOV ontology is discussed. The selection and composition of services based on predefined graphs of belonging is presented. Some results concerning service discovery, composition and consumption of services are outlined.**

Keywords. **Mobility, ontology, service composition, fuzzy logic, graphs, pervasive, electronic public services, m-government**

I. INTRODUCTION

E-government efforts aim to benefit from the use of most innovative forms of information technologies, particularly web-based Internet applications, in improving governments' fundamental functions. These functions are now spreading the use of mobile and wireless technologies and creating a new direction: mobile government (m-government). M-government is defined as the strategy and its implementation involving the utilization of all kinds of wireless and mobile technology, services, applications and devices for improving benefits to the parties involved in e-government including citizens, businesses and all government units [Kushchu, 2003]

.The main point in favour of m-government is the fact that mobile phone's penetration is reaching an 83% rate surpassing fixed telephony subscriptions in Europe (Netsize 2005, IDABC 2006).

[1]Ljupco Antovski and Marjan Gusev are with the Institute of Informatics, Faculty of Science, "Ss. Cyril and Methodius" University, Arhimedova 5, P.O Box 162, 1000 Skopje, Macedonia, Email: anto@ii.edu.mk, marjan@on.net.mk

[2]Pece Mitrevski is with the Faculty of Technical Sciences, University "Sv. Kliment Ohridski", I. L. Ribar bb, 7000 Bitola, Macedonia, E-mail: pece.mitrevski@uklo.edu.mk

Innovative and radical services applied through mobile technology may significantly improve the operations and communication efficiency of governments. Such innovative service is to extend public administrative transactions in ways that actively involve citizens allowing them to communicate and collaborate with public administration systems through their mobile devices in a transparent and trusted environment.

Currently the main focus in the service composition is discovery of appropriate services according to the user's needs. If no service is available, then a complex service should be composed, consisting of several services fulfilling the overall goal of the user. In the M-GOV project we propose fuzzy logic ontology that enables soft selection of services. It enables better flexibility in the selection of partial services.

The main hypothesis of the research is the idea to cache the ontology properties of every service during the design and registration phase and to use these results in selection of services during the composition phase. This approach will enable narrowing the search time in great manner.

II. M-GOV PROJECT

The M-GOV (Mobile Services for Government) project is a research and innovation project at the Institute of Informatics, designed to encourage the access to new mobile and wireless public electronic services. The project is based on close cooperation between Public Authorities, SMEs and Universities. The main goal of M-GOV is to contribute to a development of a new cost-effective open public service platform for mobile citizens. The new platform will support usability, openness, interoperability and scalability. It will introduce the business model of close cooperation among the service providers, public authorities and citizens. The main M-GOV innovation is the Service Discovery Directory. It will enable discovery and instant consumption of new and available M-Government services in the current location of the mobile citizen wherever in Europe.

The M-GOV project aims to meet the following objectives: New models specifying how mobile services for multiple ambiances will be coupled, integrated, assembled and offered by various service providers; New communication platform that can support highly personalized mobile services, while adapting to different networks and protocols; Plug and Play environment for new mobile services; Implementation of optimized interfaces among the service providers, public authorities and citizens; and Innovative Service Discovery Directory.

The Service-oriented architecture is believed to become the future e-government technology solution that promises the

agility and flexibility the business users have been looking for by leveraging the integration process through composition of the services spanning multiple agencies (Channabasavaiah 2004). SOA is an approach to loosely coupled, protocol independent, standards-based distributed computing where software resources available on the network are considered as Services (Arsanjani 2005).

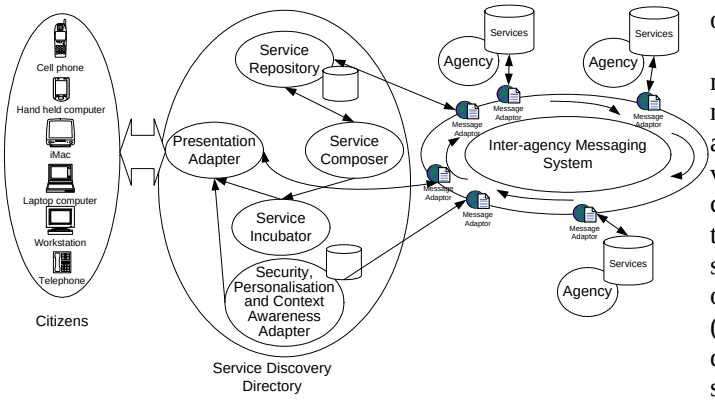


Fig. 1 M-GOV Generic Architecture

The M-GOV architecture as shown on Figure 2 is based on the above architecture and is mainly formed of three major components: Citizens access devices, Service Discovery Directory and the Collection of public electronic services from various sources.

The citizens can access the electronic public services from anywhere, anytime from any Internet-connected device. The Service discovery directory consists of presentation adapter responsible for transferring the data in citizen device's acceptable form, Security Adapter with PKI infrastructure responsible for the security and privacy mechanisms in the framework, Service Repository that minds all the required information for available service consumption, Service Composer responsible for composition of complex services, and Service Incubator, a component responsible for state preservation, personalization, quality of service and adaptation of the current service in use.

The M-GOV platform communicates with the available public electronic services using an innovative circular messaging system. The agencies and the Service Discovery Directory use message adaptors to send and receive messages in predefined format. The goal of the message adaptors is to enable communication among systems working on different platforms.

III. Inter-agency Messaging System

The basic idea for the Inter-agency Messaging System (IAMS) is to use a flexible XML format called Mobile Government Extensible Language (MGML) in binary form (optimized for mobile environment). The flexible format will enable implementation of new services with different content. The format of the new service will be reported in the salutation phase with the Service Discovery Directory and stored in the Service Repository.

The IAMS utilizes the concept of message adapters to wrap services from the technology used by other services. The IAMS provides a variety of message adapters that give services the illusion that all other services use the same messaging transport technology they do. This approach provides a high-level of technical interoperability between agencies that have heterogeneous technologies. Of course, for true interoperability agreement amongst the business owners over the structure and semantics of data is also required.

On the M-GOV IAMS, the synchronous abstraction is also made available to the services without compromising on robustness or reliability by means of a message adapter. This adapter gives the consuming service a synchronous invocation view of any M-GOV exposed service. This message queue is created, maintained and managed by the IAMS on behalf of the second service. It is up to the business owners of second service to pull their messages from the IAMS managed outbox. This has a number of loose coupling advantages (Channabasavaiah 2004). The core of the design is a simple data message routing, auditing and security. In this manner, services can be started, stopped, upgraded without bringing the M-GOV system. This is obviously important in M-Government architectures where very high availability is a required.

IV. Ontology Graph

It is expected that the SOA systems in near future will have to handle the very large number of available Web services which expectably will be several millions.

Most of the implementations use logical reasoning and keep all relevant knowledge in the working memory, which limits the number of resources tremendously.

The most interesting approach for overcoming this is to reduce size of the search space and the number of Web services that need to be taken into consideration for semantic matchmaking between a requested and provided functionality. We find that the most suitable realization is to group available Web services in a way that exhibits the properties of an efficient search graph. One can organize Web services with respect to the provided functionalities that allow achieving a higher scalability of service detection mechanisms. Most adequately, this can be achieved by organizing available Web services in an efficient search graph.

(Stollberg 2007) defines semantic matching as four degrees exact, plug-in, subsume, intersect as different situations wherein the basic matching condition is satisfied, and disjoint denotes that this is not given. We feel that using a more humanoid approach in expressing degree of maching is better. In this manner we propose to use soft fuzzy values for maching condition values. One service can mach with an ontology not only with crisp values (0 or 1) but in the whole spectrum from 0 to 1.

Without any optimization, the computational costs for service detection are in linear time O(n) with n denoting the search space as number of available Web services. The efficiency of the system can be radically improved if the information used in the service design (registration in SDD) phase is structured and latter used in the discovery phase. In this manner we narrow the search space only for the services

that in some degree match a specified goal ($O(n_u)$, where n_u is the number of services that meet the required goal) The design phase includes the registration of the services in the repository with complete ontology specification and fuzzy values for matching degrees. The information is structured in graph consisting of inner nodes (ontology values) and leaf nodes (web services). M-GOV in the discovery phase requests the necessary information from the graph composed in the design phase of the system. Then M-GOV uses a heuristic strategy that allows it to constrain the queries sent to the repository. This heuristic relies on the fuzzy evaluation of the specified preferences.

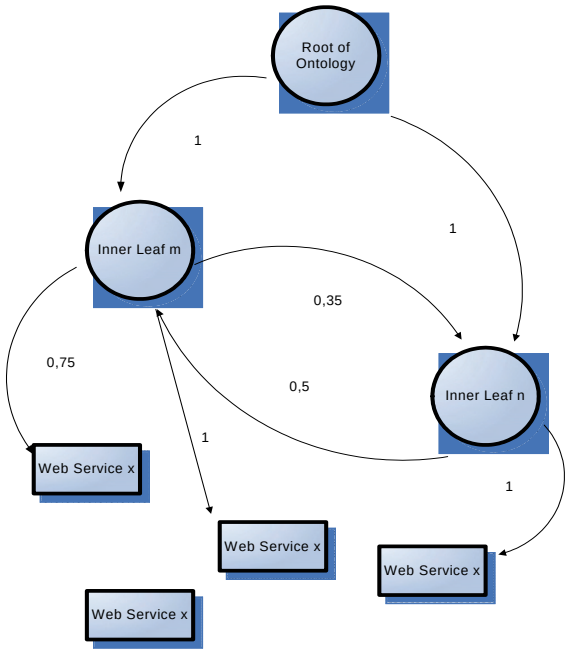


Fig. 2 Ontology Graph

The graph is constructed (Figure 2) with every inner node being a goal template that is described by defined ontology value. Every leaf node is a Web service. The links are mediators that connect nodes and define the matchmaking degree between them, such that the inner nodes are connected by mediators with similarity expressed with fuzzy value from 0-1 and each leaf node or Web service is connected by mediator to the inner node to which it belongs with the greater mediator value also expressed with fuzzy values from 0-1. When a service is registered in the repository, then it is added in the graph. Then all the links-mediators to the specified ontology values are added. If some matching information in the service specification is available for different ontology values, than that information is added also to the graph as inner links..

One service may have matching degrees with several ontology values (complex goal). That web service has links – mediators to all the matching ontology values (goal leafs). This approach improves the search efficiency in the discovery phase, as one should only check if links from all the ontology values in the complex goal lead to a specific service. If no service can be found than a complex services is composed solely on selection of services that match the single specified ontology values (single goals) in the complex goal. In all the cases, the very first web service that has the greatest mediation degree is selected.

For instance, there is ontology for museums in the future European capital. There is an inner leaf 1 in the graph Austrian, with web service 1 for information about a museum only for the 19th century. The mediator is 0,6 There is an inner node 2 Hungarian with no services in the end nodes. The correlation mediator between inner node 1 and 2 is 0,4 bidirectional. If a mobile user searches for a museum for Hungary, there is no mach, but considering he will be offered the web service 1 with belonging value to the ontology of 0,6*0,4=0,24. In the language of fuzzy logic if we have uniform triangular distribution with values as Non Appropriate, Little Appropriate, Appropriate, Very Appropriate and Complete Match, it will be that the service that is found is Very Little Appropriate.

VI. Discovery, Composition and Consumption of Services

The Service Discovery Directory (SDD) contains components for Service Discovery, Service Composition, Service Selection and Service Execution, all of which are involved in the creation of the complex service offerings. Service discovery and composition in M-GOV is not a standalone function. It relies heavily for its innovative qualities within a Security, Personalization and Context Awareness component. Service discovery uses personalization within service selection, to select the service offerings for a particular service type that are most suited for a specific user profile. Personalization uses context parameters.

Service discovery is the mechanism by which services conforming to a certain set of criteria are found in the Service Repository. It will return all services that support any of the discovery mechanisms as used by the M-GOV that conform to the criteria given. If the appropriate service is not found than the Service repository starts the internal service called Search Adapter. It uses a filter engine that evaluates a query against the discovered information from the comprehensive search of services in the M-GOV Platform. The service discovery component is only one step in the overall service composition function. It is usually called several times during the construction of a composite service. This component is responsible for the discovery, registration and deregistration of all services within the M-GOV Platform. When the list of services for one set of criteria has been found, they are then transferred to the Service Composer. This module then interacts with Security, Personalization and Context Awareness Adapter.

In a typical scenario, the user with mobile device (MU) sends a composition request message to the SDD. This is done using the standard request protocol. This message must contain two components, the initial state and the goal to achieve. After receiving the request, the SDD cannot immediately perform the composition and before that, it must contact the repository to retrieve the descriptions of relevant services.

When the repository, its message must include a sequence of OWL complete service descriptions. After receiving the set of services that will be used for composition, the SDD

performs the second step, which refers to internal processing and the planning process itself. Once the planning completed successfully, the SDD converts the plan to a composite service description and informs the MU about the service.

Selecting the set of possible services to form the plan is not an easy task. Several heuristics can be applied. In this section we propose a more informed method for filtering services that makes use of classification of services. We developed a framework for service-based filtering, and then instantiated it to different filters on the basis of context information obtained from the role ontology and the service category derived from the directory structure.

The service composition planning problem can be conceived as follows: Let P = {p1, p2… pn} be the set of all possible composite services for a given service request and S = {s1, s2… sm} the set of input services for proper service composition. The goal of a filter F is to select a given number X of services from S, such that the search is reduced, but the best plan of P is still found. However, it is obvious that this ideal case is not realistic since the problem would be already solved. In principle, one can build up matrices for every possible query.

Once the service has been composed, it is the Service Incubator that looks after the composed service instantiating it as a complete service and monitoring it during its lifetime. The Service Incubator is also responsible for service adaptation. Service adaptation must not be confused with content adaptation. Service Adaptation encompasses content adaptation and more. Service adaptation can be carried out both during service composition and while the service is running. The adaptation of the service can be triggered by changes in context of the user. The adaptation occurs as the system constantly attempts to match the users' needs and the available resources and capabilities of the service, by constantly monitoring changes.

The Service Incubator closely cooperates with the Presentation Adapter. The Presentation Adapter actively responds to the information received from the Service Incubator and adopts the presentation of data to the appropriate quality, quantity and demand of the presentation device used at the given moment. For instance, a citizen reads the flash news from a governmental agency service on his/her PDA while coming to work. As he/her enters her office and logs in to his/her computer, the Presentation Adapter actively responds with enabling a full screen enriched data presentation. But not only the presentation is changed, but the quantity also, because the bigger screen enables presentation of more detailed information. You are requested to strictly follow the author's guidelines.

VII. Final Remarks

The M-GOV platform is based on simple ideas that together provide architecture with high level of flexibility and low levels of technological requirements. The architecture is designed to scale, both from a technical and a financial perspective and to be applicable to integration scenarios from small agency scenarios through to large administration scenarios. The core technology concepts employed all proven technologies that are well suited to use in building a long lived M-Government infrastructure.

The system is implemented in laboratory scale. Future work includes tests on the functionality implemented in comprehensive real environment, and validation of the platform against similar systems, from the perspective of personalised service composition and service adaptation. It is also hoped to extend the adaptation process within the platform to optimize service performance in response to changing context conditions.

We feel that the impact of mobile technologies on government administration is huge as this will have implications for the success of widely adopted M-Government applications. In order to consume the benefits a sound standardized solution as the M-GOV attempted is to be defined.

The problem of composition of complex services has many proposed solutions. All of them are mainly based on QoS parameters selection or on transforming the ontology description to First Order Logic. M-GOV eases the constraints in the design of the services and enables human description of ontology belonging using Fuzzy Logic and Graph technology. We feel that this approach will enable better and easier composition of appropriate services in the m-government area.

References

[1] Ali Arsanjani, How to identify, specify, and realize services for your SOA, IBM Corporation, February 2005

[2] Antovski, Lj. And Gusev, M. (2005), M-Government Framework, Proceedings of the First European Conference on Mobile Government, (editors Ibrahim Kushchu and M. Halid Kuscu), 10-12 July, University Sussex, Brighton UK"

[3] Heeks, R. and Lallana, E. (2004), M-government benefits and challenges,http://www.e-evexchange.org/eGov/mgovprocon.htm

[4] IDABC (2006), eGovernment in the EU member states, eGovernment Observatory, IDABC Report, April 2006, 4th edition

[5] Jason Bloomberg, The SOA Implementation Framework, http://www.zapthink.com, April 2004

[6] Jeff Hanson, Coarse-grained Interfaces Enable Service Composition in SOA, JavaOne, August 2003

[7] Kishore Channabasavaiah and Kerrie Holly, IBM White Paper, Migrating to a Service-Oriented Architecture, April 2004

[8] KIT(2005), National Strategy for Information Society Development, http://www.kit.gov.mk , April 2005

[9] Kushchu, I. and Kuscu, H (2003), "From E-Government to M-Government: Facing the Inevitable", in the proceedings of European Conference on E-Government (ECEG 2003), Trinity College, Dublin

[10] Meta Group White Paper, Practical Approaches to Service-Oriented Architecture, November 2003

[11] Michel Stollberg. Semantic Goal Caching. Technical Report, DERI, 2007.

[12] Ministry of Justice of Macedonia (2006), E-Election List, http://www.izbirackispisok.gov.mk, May 2006

[13] Netsize (2005) 'The Netsize Guide', web: http://www.netsize.com, April 2005

SESSION TST II

Telecommunications Systems and Technology II

Routing Simulation for Wireless Sensor Networks

Zdravko Monov[1] and Ludmila Karakehayova[2]

***Abstract* – This paper describes ARCS – ad-hoc routing and connectivity simulator, a tool that is able to simulate the routing process in distributed sensor networks. After a configuration phase the nodes in the network are ready to route different types of messages. Two versions of the routing algorithm are currently supported. The simulation results produced by ARCS can be used to compare both protocols for a particular deployment. Using the simulator and its graphical module we can assess the capability of the network to route traffic.**

***Keywords* – distributed sensor networks, ad-hoc networks, geographic routing.**

I. INTRODUCTION

Distributed sensor networks (DSN) are made up of a large number of small sensing nodes which cooperatively perform complex tasks. Environmental monitoring, healthcare, building automation, surveillance and rescue missions are applications where wireless ad-hoc networks provide benefits that we would not otherwise be able to obtain.

The interaction between the nodes is based on wireless communication. Sensor-actuator networks employ sensors to gather information and actuators to perform appropriate actions in a area of interest. Since the energy is a scarce and usually non-renewable resource, the functionality of distributed sensor-actuator networks must be viewed from low-power perspective. Normally, the nodes have a limited radio footprint and packets are forwarded in a multihop manner. When a node receives a packet it applies a routing algorithm to select a neighbor for forwarding. Different criteria can guide the local decision. One approach is to choose the closest to the destination neighbor.

Different protocols have been developed for routing in wireless ad-hoc networks [1]. Dynamic topologies, environmental radio propagation conditions, depletion of batteries and congestion make the experiments difficult and expensive in real-world systems [2]. To better understand the functionality of distributed sensor networks simulations are performed. Various simulators are developed to meet different criteria like energy efficiency, minimizing of bottle-necks or optimal area deployment [6]. They allow complex and precise simulations, while keeping the results reliable for any real world system. Designing a new simulator for a general DSN is a complex task that requires the consideration of many parameters.

[1]Zdravko Monov is with the Faculty of German Engineering Education and Industrial Management, Technical University of Sofia, Bulgaria, E-mail: ztm@tu-sofia.bg

[2]Ludmila Karakehayova is with the Technical University of Sofia, Bulgaria, E-mail: lvk@tu-sofia.bg

II. HARDWARE MODEL

A typical node is built around a low-power microcontroller. Wireless transceivers create physical links between nodes. Hardware provides the following low-power mechanisms:

- Each node is capable of determining its coordinates,
- The receiver and transmitter can be individually enabled and disabled and
- The transmit power can be adjusted gradually.

III. ARCS OVERVIEW

Our simulator, ARCS (ad-hoc routing and connectivity simulator), has five major components:

- DSN generator,
- network organization unit,
- route generator,
- routing protocol, and
- computation module.

ARCS' modularity allows easier switching between different protocols and simulation scenarios, while preserving the reliability of the results. The simulator can display the traveled distance for packets, providing valuable information for parameters like energy drain.

The network nodes are randomly distributed in a sensor field. Nodes have a specified wireless cell size. Also, the design entry includes the communication rate, packet data structure, timing and energy parameters. The computation module software is written in the C++ programming language.

Ad-hoc networks are usually homogeneous systems [10]. The internode communication is the largest energy consumer. There are two ways to achieve power-efficient interaction between nodes. First, nodes calculate the distance and tune their transmit power accordingly. Second, nodes stay in a sleeping mode as long as possible. Periodically, nodes wake up and receive the packets buffered for them. Since the internode communications must be synchronized, the nodes operation becomes time constrained [7]. The real-time functionality requires more than just having enough computing power on average - it requires enough power for the worst-case scenario. The network density emerges as the most demanding factor when both variable power levels and synchronization underlie the communication [11].

IV. NETWORK ORGANIZATION AND STRUCTURE

Each node is characterized by its ID and location. The current version of ARCS assumes a static topology. At the beginning of each simulation cycle the nodes are positioned using random coordinates. Deployments could be arranged to match both sparse and dense networks.

The network organization begins with broadcast messages which the nodes transmit after random delays. The messages include the node own ID and location. In addition, each node copies to the message its own database. Fig. 1 shows the network connectivity after a configuration phase. ARCS is able to keep track on the simulation time. By defining the input timing parameters a specific mathematical algorithm is inovked in order to determine the correct behavior of each node.

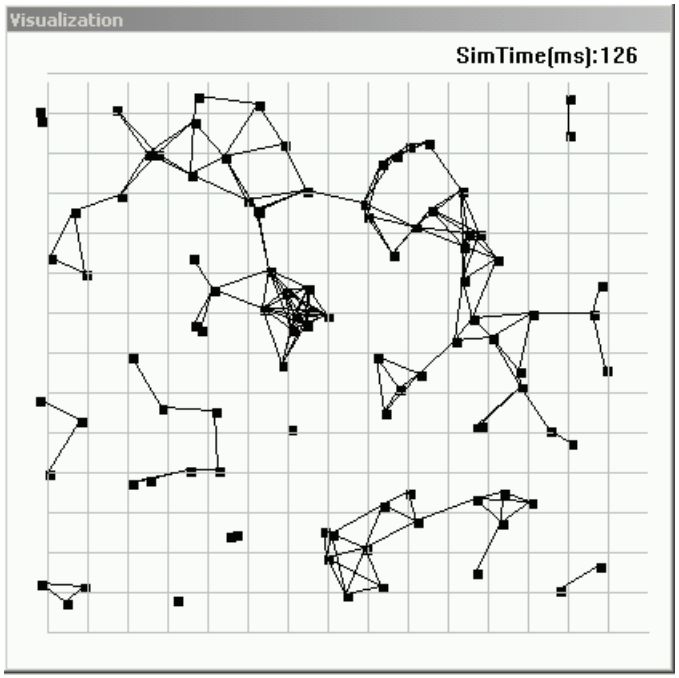


Fig. 1. Network connectivity after a configuration phase

V. ENERGY MODEL

The energy used to send a bit over a distance d may be written as Eq. 1:

$$E = ad^n \tag{1}$$

where a is a proportionality constant [8], [9]. The radio parameter n is a path loss exponent that describes the rate at which the transmitted power decays with increasing distance. Typically, n is between 2 and 4 [8].

$$E = ad^n + b \tag{2}$$

Eq. 2 emerges as a more realistic model. The b constant is associated with specific receivers, CPUs and computational algorithms. The power consumption of a turned on receiver is yet another constant, P_R.

VI. ARCS ROUTING

The route generator randomly selects two nodes as source and destination. The procedure is repeated until the specified number of messages is met. Reaching a dead-end (or void) interrupts the current routing as Fig. 2 presents.

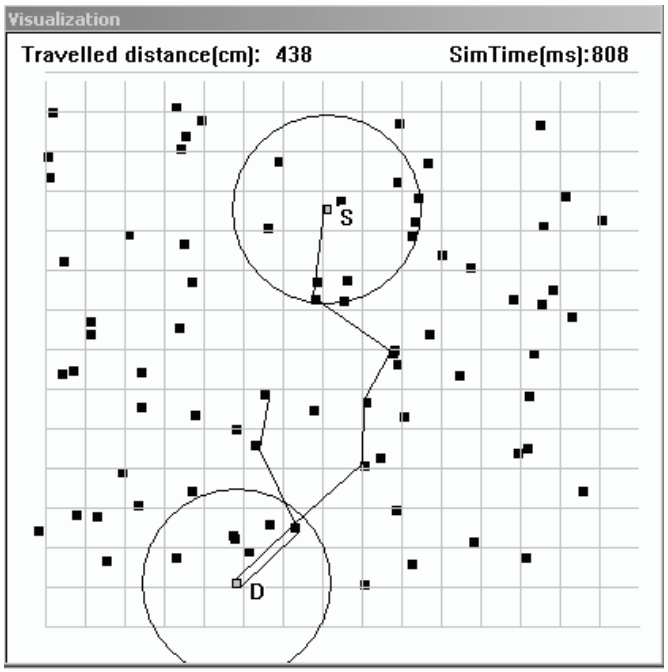


Fig. 2. Unsuccessful routing for the acknowledgement

The simulator is based on two versions of Most Forward within the transmission Range (MFR) routing protocol. MFR assumes that it will be most beneficial to transmit the packet to the closest to the destination node [3]. Fig. 3 indicates the routes when a modification of the MFR protocol is applied. While in the case of geographic routing we may need additional information, the physical position of the participating nodes, on the positive side is the opportunity to forward packets without pre-established routes [4, 5, 12].

The simulation requires only the destination to send an acknowledgement back. Usually, the acknowledgement travels over the same intermediate nodes. ARCS introduces another option – the reverse route is generated by applying MFR again. The difference in routes could be quite significant if the network deployment is reasonably dense. Also the simulation results state a negligible drop in the successful routings, while preserving the nodes from multiple transmissions. Using different routes may be benefical for both power efficiency and security.

The modification of the MFR protocol consists in the different distances that are compared. The next node will be that one, that is farther away from the sender and its position is on the way to the destination.

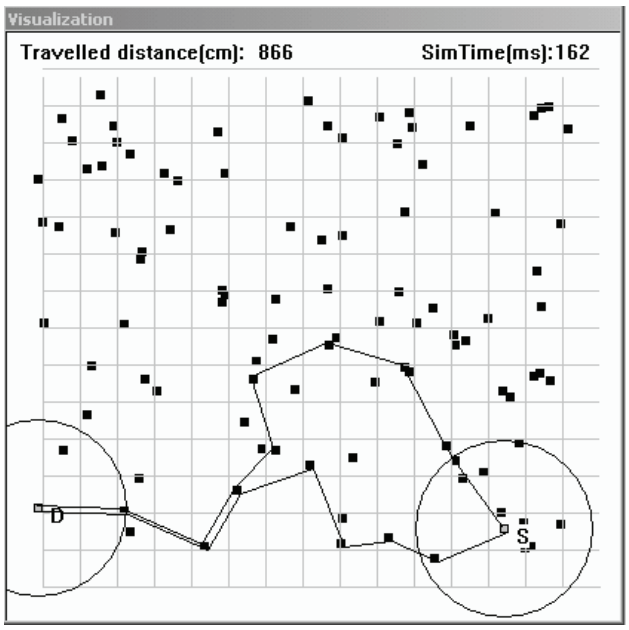


Fig. 3. MFR modification algorithm

For the sake of efficiency, the records in the address table are sorted using a specific criterion – for example, by the distance from the node. The sending node can now easy state which is the farthest node. Then a direction check must be done. If the selected hop doesn't match the corresponding direction, the next node is chosen, and so on. Once a suitable hop is found, the sender transmits the packet.

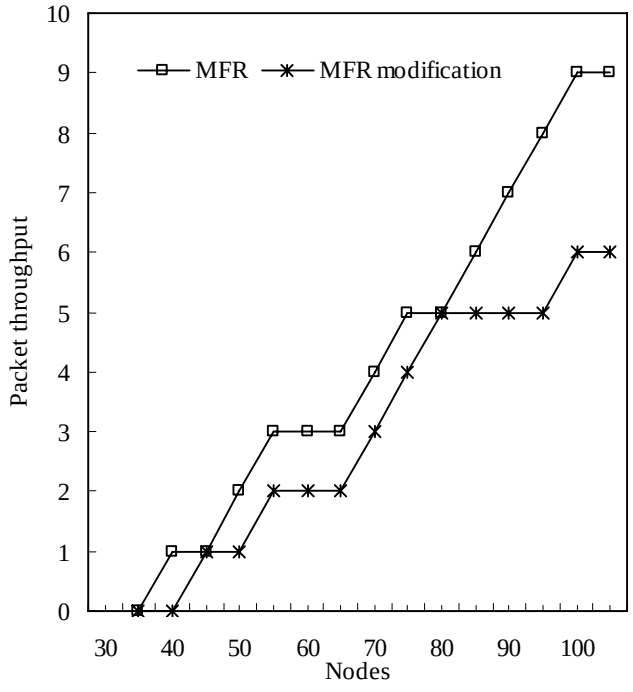


Fig. 4. The fraction of packets delivered for different density of the network

Using ARCS and its graphical presentation of the network connections, we realized that the modification of MFR protocol does not provide the needed efficiency, as Fig. 4 shows [6]. Here, a transmission range of 150 meters is assumed. The area is a square with size 10 km^2, and the baud rate is 1Mb/s. The routes are significantly longer. A longer route means that a larger number of nodes is involved in the routing process. Further more, this leads to increases in the timing parameters – respectively, the packet may not reach the destination in a specific period of time. The general form of the routes generated by the modificated MFR is usually too far from the straight line that connects the initial sender and the final destination.

The transmission range of the nodes also has a significant impact on the network connectivity. While a smaller range may reduce the required energy at overall, it lowers the packet throughput. Fig. 5 shows the results we obtained for different range sizes. We tested the MFR algorithm in terms of routing efficiency. The simulation conditions remain similar to the previous example – area of 10 km^2, transmission speed of 1Mb/s. The nodes were randomly distributed over the whole area. After passing the recognition phase, 10 messages were initiated in a random manner. By increasing the transmission range, more nodes become direct neighbors and therefore more packets are delivered successfully. The downside of a larger radius is that more nodes must wake up to participate during the data transmission. If the signal strength could be adjusted precisely enough, only a small subset of the neighbors would detect the transmission.

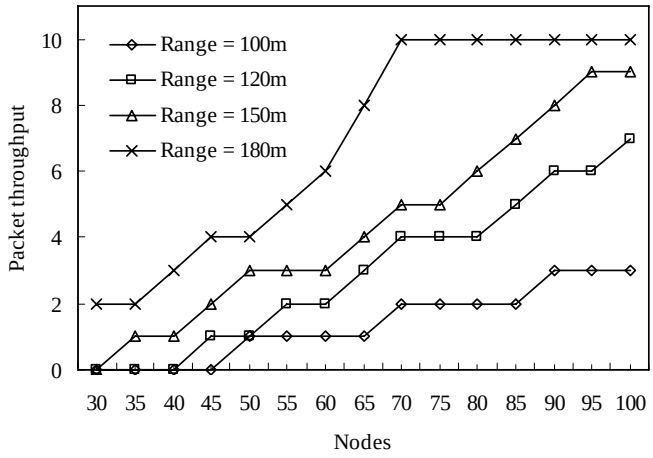


Fig. 5. Routing success rate as a function of the transmission range

VII. CONCLUSION

In this paper we described ARCS, an ad-hoc routing simulator suitable for wireless sensor networks. We provided simulation results that indicate how the packet throughput scales with the network density and communication range. The results showed that the original MFR protocol provides better connectivity for both sparse and dense networks compared to the modified MFR.

REFERENCES

[1] Hu Y. C. and Perrig, A., A survey of secure wireless ad hoc routing, IEEE Security & Privacy, May/June, 2004, pp. 28-39.

[2] Karakehayov Z., Design of distributed sensor networks for security and defense, In Proceedings of the NATO Advanced Research Workshop on Cyberspace Security and Defense: Research Issues, (Gdansk, Poland, September 6-9, 2004). Springer, 2005, pp. 177-192.

[3] H. Takagi and L. Kleinrock, “Optimal transmission ranges for randomly distributed packet radio terminals”, IEEE Trans. Commun., vol. 32, no.3, March, 1984, pp. 246-257.

[4] Karakehayov Z., "Low-power communication for wireless ad-hoc networks", Proceedings ELECTRONICS'2003 International Conference, Sozopol, 2003, pp. 77-82.

[5] D. Puccinelli and M. Haenggi, Wireless sensor networks: applications and challenges of ubiquitous sensing, IEEE Circuits and Systems Magazine, third quarter 2005, pp. 19-29.

[6] Karakehayov, Z., Monov, Z., Karakehayova, L., ARCS: A simulator for distributed sensor networks, Proceedings of the ELECTRONICS'2005 International Conference, Sozopol, September, 2005, pp. 91-96.

[7] Zdravko Karakehayov and Zdravko Monov, Target-aware timing modelling for wireless ad-hoc networks, Proceedings International Scientific Conference Computer Science’2006, October, 2006, Istanbul, pp. 54-59.

[8] J. L. Gao, Energy efficient routing for wireless sensor networks, Ph.D. dissertation, University of California, Los Angeles, 2000.

[9] J. M. Rabaey, M. J. Ammer, J. L. Silva, D. Patel and S. Roundy, "PicoRadio Supports Ad Hoc Ultra-low Power Wireless Networking", IEEE Computer, vol. 33, pp. 42-48, July 2000.

[10] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring and D. Estrin. Habitat monitoring with sensor networks. Communications of the ACM, Special Issue on Sensor Networks, June 2004, Vol. 47, pp. 34-40.

[11] Z. Karakehayov, Design of distributed sensor networks for security and defense, In Proceedings of the NATO Advanced Research Workshop on Cyberspace Security and Defense: Research Issues, (Gdansk, Poland, September 6-9, 2004), edited by J. S. Kowalik, J. Gorski and A. Sachenko, Springer, NATO Science Series II: Mathematics, Physics and Chemistry, Vol. 196, 2005, pp. 177-192.

[12] M. Witt and V. Turau, BGR: Blind geographic routing for sensor networks, In Proceedings of the Third International Workshop on Intelligent Solutions in Embedded Systems, Hamburg, Germany, May, 2005.

Overloaded IP Node With Implemented Priority Queuing QoS Mechanism

Ilija Efnusev[1] and Toni Janevski[2]

***Abstract* – Although QoS in IP is reality now days, its implementation with all of its power is still negligible. More research before putting it into force is needed. Although the research in this paper is just a tiny part of the whole QoS mechanism, still it is an integral part of the whole idea and brings valuable conclusions.**

This paper analyses the priority queuing simulated on a single IP operating node overloaded of incoming traffic. The incoming traffic is divided in different flows with different priorities, which simulates the priority queuing mechanism. The simulation results show the different output traffic parameters for each flow.

***Keywords* – Quality of service (QoS), IP, Simulation, Priority Queuing, Overloaded**

I. INTRODUCTION

With the computerization of the modern world, new applications are emerging every day. Most of this applications demand remote access, shared resources and globalization of information. This indicates that remote transfer of information is becoming more demanding. The transfer of information needs to be faster and faster.

In the moment we can say that IP won the battle for global transfer technology. As this technology was created to be based on "best effort" transfer of information, the fast developing information world brought new challenges in the IP technology [1]. With the new applications, new services emerge. In these services, speed is not the only limit. Different front-end services demand variety of requirements to be met: maximum Packet Delay, maximum Jitter, minimal Bit Rate, maximum Packet Loss etc. This resulted in IP protocol improvement with the new QoS mechanisms [2].

II. QOS ARCHITECTURE

The QoS architecture needs to satisfy different parameters, on different network levels and in different network parts. This brings us to a complex QoS architecture that has to control the specified parameters of a service requested by the end-user application over different network elements. This means that signaling and traffic admission and control has to be established between network elements.

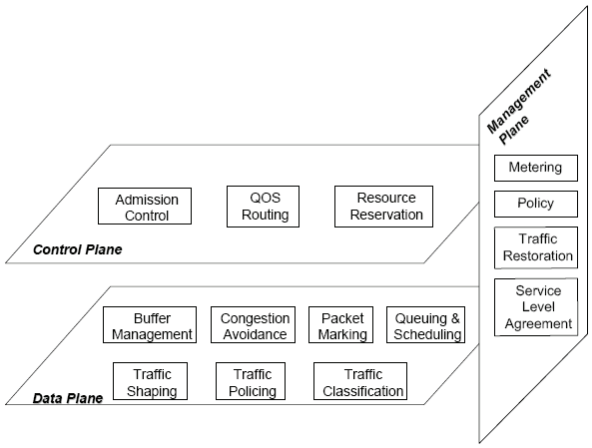


Fig. 1. QoS architectural framework

The complex QoS architecture has been a target of many researchers during the last couple of years. Finally in May 2004 ITU-T came out with the Y.1291 recommendation for the QoS architecture [2]. According to this recommendation the QoS building blocks are organized into three planes, as showed on Fig 1.

The QoS mechanism may refer only to a specific network node (like the Packet marking) or to a network segment (like the QoS routing). This type of architecture requires signaling between network nodes no matter of which part of the network those nodes belong to. The signaling between the QoS segments belonging to the Data plane block is closed in the node (device) and is chosen by the manufacturer, while the signaling in the Control or Management plane is between two or more nodes, so it requires a signaling protocol. The QoS architecture is just a logical framework, so it puts no restrictions of that how the different QoS mechanisms and signaling between them will be realized. Simulating the whole QoS architectural framework is a complicated and demanding process, so only the queuing and scheduling mechanisms will be analyzed in this paper.

III. QUEUING AND SCHEDULING QOS MECHANISM

This mechanism controls which of the incoming packets to transmit to an outgoing link. The incoming traffic could be considered as part of a queuing system, consisted of multiple queues and a scheduler. So by controlling the queuing and scheduling of packets the QoS requirements could be met. This mechanism can still be divided into several approaches:

[1]Ilija Efnusev is Quality of Service Specialist in T-Mobile Macedonia AD. Skopje, Total Quality Management Department, Orce Nikolov bb. Skopje Macedonia, ilija.efnushev@t-mobile.com.mk

[2] Toni Janevski is Prof. Dr. at the "St. Cyril and Methodius" University in Skopje, Faculty of Electrical Engineering and Information Technologies, Dept. of Telecommunications, Karpos 2 bb. Skopje, Macedonia, tonij@etf.ukim.edu.mk

- **First-In, First-Out (FIFO) queuing:** This is when packets are placed into a single queue and they are served in the same order as they arrive in the queue.
- **Fair queuing:** This is when packets are classified into flows and then assigned to queues dedicated to the respective flows. Queues are serviced in round robin method, where empty queues are skipped. This is also referred to as a per-flow or flow-based queuing, where statistically every flow gets the same attention.
- **Priority queuing:** This is when packets are first classified and then placed into different priority queues. Packets are scheduled from the head of a given queue only if all queues of higher priority are empty. Within each of the priority queues, packets are scheduled in first-in, first-out order.
- **Weighted fair queuing:** This is when packets are classified into flows and assigned to queues dedicated to respective flows. A queue is assigned a percentage of output bandwidth according to the bandwidth need of the corresponding flow. By distinguishing variable length packets, this approach also prevents flows with larger packets from being allocated more bandwidth than those with smaller packets.
- **Class-based queuing:** This is when packets are classified into various service classes and then assigned to queues assigned to the service classes, respectively. Each queue can be assigned a different percentage of the output bandwidth and is serviced in round robin. Empty queues are skipped

IV. SIMULATION MODEL

The simulations were made by custom programmed simulator in C. This simulator was made with the purpose of studying the queuing and scheduling QoS mechanisms. The basic model of the simulated node is given on Fig. 2

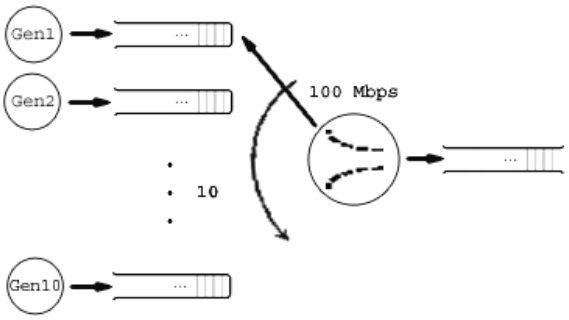


Fig. 2 Simulation Model

A. Simulator characteristics

The custom made simulator has the following characteristics:

- 10 buffers with 10 kB of buffer space
- Packets in each buffer are served by the FIFO queuing.
- If the buffer is full incoming packets are dropped.
- Each buffer together with the generator simulates one flow defined in the ITU QoS architecture.
- The buffers have different priority. The first buffer has the highest, and the 10-th has the lowest priority.
- A buffer with lower priority is only served if all the buffers with higher priority are empty.
- The entering traffic is equally distributed in flows, depending on the number of flows used (ex: If the incoming traffic is 100 Mbps and we have 4 flows, each flow would have speed of 25 Mbps)
- The Node works with speed of 100 Mbps, and serves packet by packet.
- The output buffer is unlimited
- The generators are defined with by their speed. They generate packets with Packet length between 50 and 1500 Bytes (the size of the packet exponentially distributed), and inter-arrival times also exponentially distributed. This simulation model is according the propositions in the reference paper [3].

B. Measured parameters

The simulator measures the following parameters:

- **Transmitted Flow Bit Rate**, measured in %. This is a parameter that shows the % of the Mean Bit Rate of the Flow at the output of the node, compared to the Total Bit rate at the nodes output, according to equation (1).

$$TransmittedFlowBitRate = \frac{OutputFlowMeanBitRate}{TotalNodeOutpuBitRate}100\% \quad (1)$$

- **Lost Packets**, measured in %. This shows the number of packets lost compared to the total incoming packets, per flow. See equation (2).

$$PercentOfLostPackets = \frac{FlowsLostPackets}{Total\ \text{Flows}\ Generated\ \text{Packets}}100\% \quad (2)$$

- **Mean Delay**. This is the average delay of the packets in the flow, calculated as the time difference at the input and output time of the packet in the Node, formula (3). This includes the waiting time in the buffer, plus the serving time by the node.

$$MeanDelay = \frac{\sum_{n=0}^{n=N} (t_{(output)n} - t_{(input)n})}{N} \quad (3)$$

- **Jitter**. This shows the time deviation of the packets at the output from the mean expected output time. This parameter is calculated according to the formula (4).

$$Jitter = \frac{\sum_{n=0}^{n=N} | MeanDelay - (t_{(output)n} - t_{(input)n}) |}{N} \quad (4)$$

V. SIMULATION RESULTS

These simulations were made with the constraint that incoming traffic speed is double than the serving nodes speed. So depending on the number of flows, each flows bit rate is 200Mbps/N, where N is the number of flows. Each flow can be considered as a service which needs different QoS parameters to be satisfied. This simulates what will happen if there is no flow control between the nodes in the network and in one moment the incoming traffic doubles.

Each of the measured parameters in the simulation will be discussed separately.

A. Transmitted Flow Bit Rate

This parameter actually shows which % of the output mean bit rate (the serving rate of the node) has each Flow at the output buffer of the node. The measured results are shown in Fig. 3. From the figure it is clear that the total output Bit Rate (utilization of the output link) is bigger when traffic is segmented in more flows. This is due the fact that from time to time if there is only one buffer, there is a probability that it might be empty although the packets on the input are coming with bigger speed than they are served. While when there are more buffers the node starts processing immediately the packets in the next buffers. Because the packets from lower priority queues are served only when all of the queues with higher priority are empty it can be seen that lower priority flows are rarely served. So they have low transmission rate, as we will se later on Fig.4 they have great packet loss. This leads to a conclusion that, segmenting the flows in more than four flows makes the node unstable, and with terrible performance when overloaded.

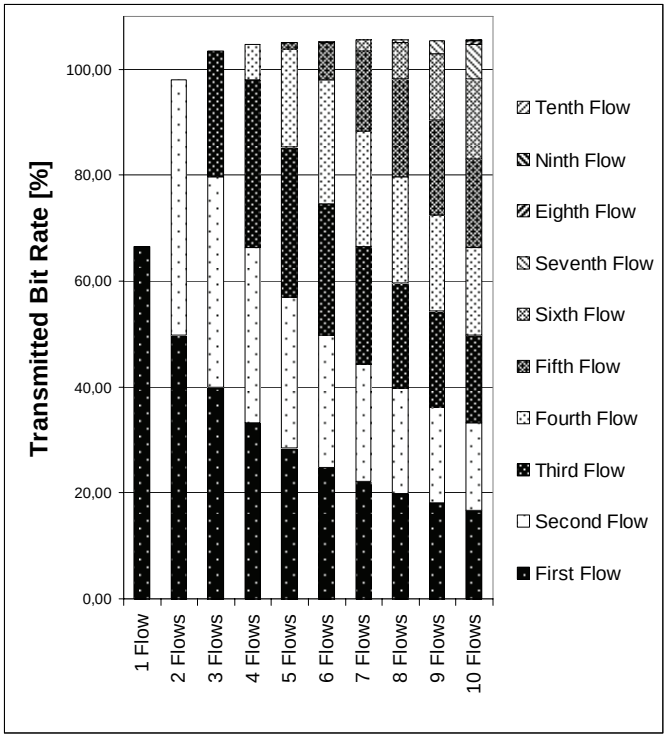


Fig. 3 Transmitted Bit Rate [%]

B. Packet Loss

This parameter shows the percent of the packets that are lost from each flow. Packets get lost if the buffer of the flow is full and new packet is generated by that flow. The results are shown on Fig. 4. As it can be seen in the figure when traffic is separated in more than 4 flows there is packet loss of over 90% in the flows with lower priorities. This means that these flows are blocked and the node doesn't transmit their traffic further. This brings us to the conclusion that flows with lower priorities would have to implement a mechanism of packet retransmission. If the overloading of the node is just for a short time this packets would be retransmitted and the services defined in this flows will continue functioning with some delay. But if the overloading of the node is continuous for these services it would seem like the node is down.

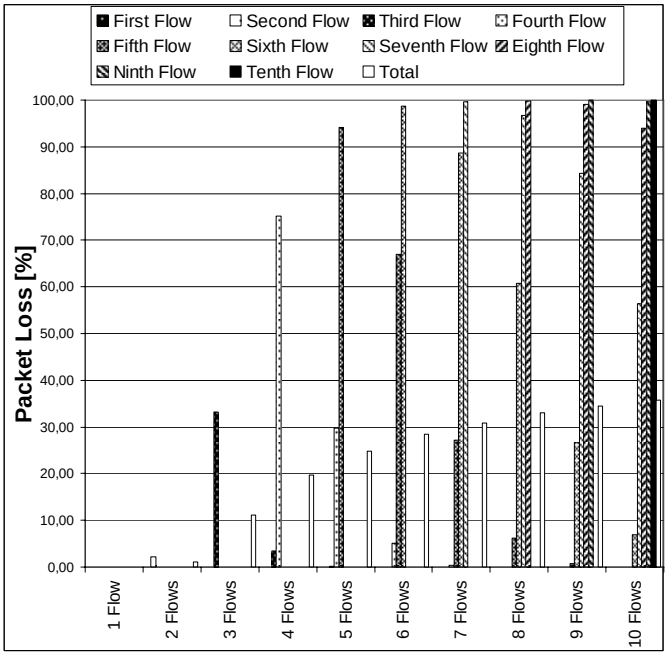


Fig. 4 Packet Loss [%]

B. Mean Delay

This parameter shows the mean delay that the packets experience when served in the node. The delay is as a result of the waiting time in the buffer and the time needed for the packet to be served by the node. The mean delay of each flow is shown on Fig. 5 given in [ms] on a logarithmic scale. Flows with lower priority have enormous delay when the node is overloaded due to their waiting time in the buffer. The highest priority flows tend to the average serving time of the node according to equation (5).

$$ServingTime = \frac{AveragePacketSize[bit]}{NodeOutputSpeed[bit/s]} \qquad (5)$$

From the figure it is clear that the services defined in flows with lower priorities have to be tolerant to packet delay because if we consider that this is just the delay at one node

the total delay in a network would be n times bigger, where n is the number of nodes. When there are more than 4 flows, delays are unbearable for any existing IP service.

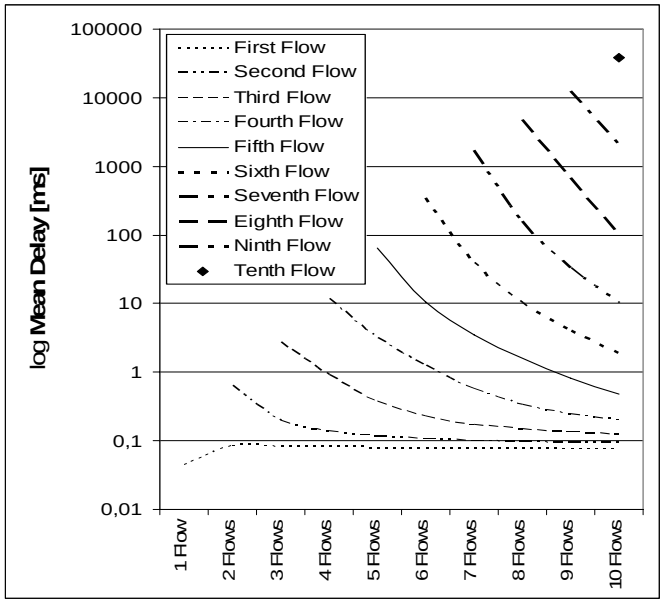


Fig. 5 Mean delay [ms]

B. Jitter

This parameter actually is the mean deviation from the delay of the packets served by the node. The results of the simulation are shown on Fig. 6 in [ms] on a logarithmic scale. The jitter is extreme for flows with lower priority. Services like internet browsing or file download are resistant to jitter and delay but when there are more than 4 flows the jitter is unbearable even for these services. Other real-time services like video and audio streaming are intolerant to jitter so this kind of services must be defined in flows of highest priority.

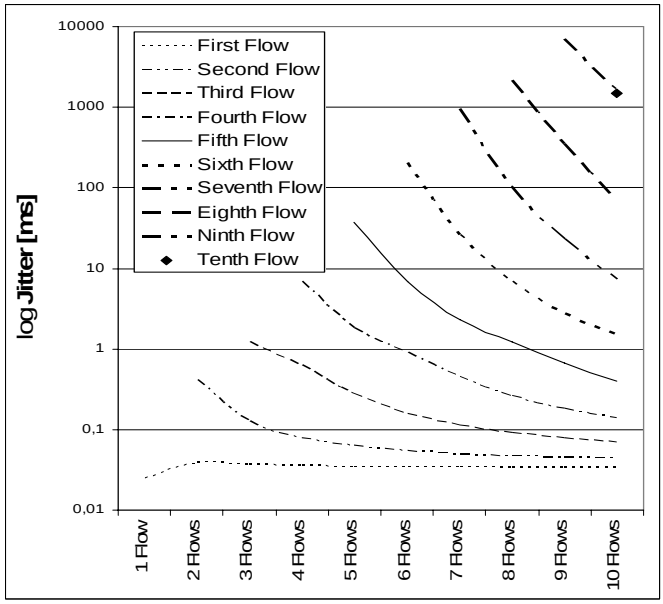


Fig. 6 Jitter [ms]

VI. CONCLUSION

As it can be seen from the simulation results, this kind of implementation of QoS on node level shows deteriorated performance in lower priority flows, when the node is overloaded. This is especially visible when the number of flows is grater (more different service levels are defined). So by the simple implementation of QoS and dividing the traffic in different flows, the low priority flows won't show any performance. As it can be seen in the simulations results, there is a packet loss of over 90% at almost half of the flows. But this is not the only problem. All of the other parameters are also extremely degraded as well. Looking at all the parameters together it can be said that four is the optimal number of flows with different QoS parameters that should be defined. Besides this, all parameters also suggest that a simple implementation of the QoS mechanisms on node level needs to be combined with other QoS mechanisms: flow control, QoS routing and other, to achieve performance and sustain service quality. With this paper it is clearly shown that implementing QoS in just part of the network is not a solution. There has to be a continuity of QoS mechanisms across the network and between networks, so that QoS would really work. Service providers have to become conscious of the power that QoS brings and upgrade their systems to support the QoS mechanisms.

ACKNOWLEDGEMENT

I would like to express my gratitude to the professor Toni Janevski and the Telecommunication Institute at the Faculty of Electrical Engineering and Information Technologies at "St. Cyril and Methodius" University in Skopje, for the shown cooperation and help in my research.

REFERENCES

[1] Toni Janevski, Traffic Analysis and Design of Wireless IP Networks, Artech House Inc. 2003

[2] ITU-T, An architectural framework for support of Quality of Service (QoS) in packet networks (Ref. No.: Y.1291)

[3] Zafer Sahinoglu and Sirin Tekinay, Multimedia Networks: Self-Similar Traffic and Network Performance, New Jersey Institute of Technology.

Planning and Pre-Operational Performance Optimization of WCDMA for Multimedia Wireless Networks

Toni Janevski[1] and Vladimir Nikolic[2]

***Abstract* – In this paper we present important role of detailed planning and optimization of radio network in the design of next generation wireless networks that offers multimedia traffic. Wireless operators that are adapting 3G mobile technologies are facing more and more challenges connected to the design and planning of the factual multi service radio networks as WCDMA itself. These challenges are not independent and must be considered together with the user's demand and performance analyses as key factors of designing process, dimensioning and pre-operational optimization of network infrastructure.**

Keywords – WCDMA, Detailed Planning, Pre-Operational Optimization, Wireless Multimedia Networks, Multimedia Traffic.

I. INTRODUCTION

WCDMA radio network planning process, as defined by 3GPP, basically can be consisted of: modeling of radio network planning tools and preoperational control of network (optimization) [1]. At the beginning we gave overview of the existing simulation approaches that deal with the estimation of the number of simultaneously active users in the next generation multimedia mobile systems on statistical representative way, so the simulation results reflect the system reaction on certain traffic demand. We come up with MATLAB® based simulator performing the distribution of these active users in the network topology at given moment. Than these generated snapshots we used as input for the static WCDMA simulator in order to present the importance of the proper and efficient dimensioning of radio network.

From these reasons we conducted the study for the simulation of the UMTS system for the purpose of validation of the WCDMA planning. As planning tool the MATLAB based NPSW (Nokia®) static simulator is used. In this direction, a couple of user distribution scenarios for a given snapshot are analyzed, compared and evaluated for multiple purposes as identification of the main metrics of the system capacity and verification of the impact of each scenarios on the capacity and coverage improvement.

In the last phase we performed on more study related to the optimization of previously obtained results for the scenarios analyzed during the WCDMA planning. We demonstrated the impact of the antenna selection, antenna tilt, antenna beamwidth and application of the sectorisation on the interference, with the purpose of finding optimal solutions that will improve the capacity and coverage of the mobile systems from next generations, as 3G.

II. GEOGRAPHICAL DISTRIBUTION OF THE SNAPSHOTS CONSISTING ACTIVE USERS

The number of active users for any user profile may be presented as random variable that follows its multiplexing distribution [2]. Once the number of active users is known, they have to be randomly distributed on the spatial plane following the proportions of the geographical user distribution. For the realization of these steps, we create a MATLAB code which on its input uses the number of active mobile stations in a given moment, as well as the dimensions of the area in which the users are located. In the absence of real demographic data for the area with known geographical user distribution (census), in this paper the analyses are performed over the predicted 16.9 km² test area (3G system with 19 three-sectored base stations), whit randomly distributed users. The active users will follow the same random distribution, too. On its output, the code generates *x* and *y* coordinates of the each active mobile station within the test area and are used as input in the NPSW simulator (Network Planning Strategies for Wideband CDMA created by Nokia®) 5.0.0 version [4-6]. For the analyses we assumed a snapshot consisting 11 randomly distributed active users. Based on simulation result all 11 active users are served by the system.

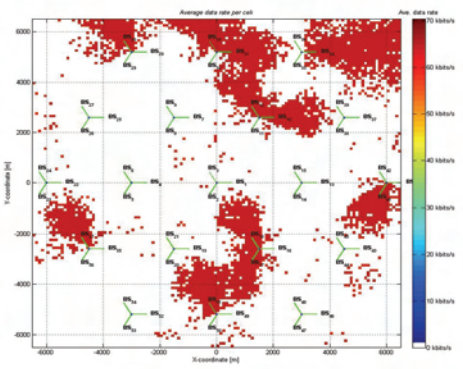


Fig. 1. Average data rate per cell in uplink.

On the Fig. 1 we depicted NPSW graphical simulation results how is distributed the average data rate per cell in uplink on the locations of the 11 served active users. We may

[1]Toni Janevski is with the Faculty of Electrical Engineering and Information Technology, Karpos bb, 1000 Skopje, Macedonia, E-mail: tonij@etf.ukim.edu.mk

[2]Vladimir Nikolic is with the AD Makedonski Telekomunikacii, Orce Nikolov bb, 1000 Skopje, Macedonia, E-mail: vladimir.nikolic@mt.com.mk

conclude that accurate snapshot generation that can serve as proper input to the simulator for WCDMA planning and optimization discussed next with the aim to be estimated the respond of 3G system on offered traffic.

III. Detailed WCDMA Planning with Static Simulations

In this Section in details will be described pre-operational phase of the WCDMA planning process as described by 3GPP [1], [3]. Initial planning (the system dimensioning) provides the very first estimation of the network size and the capacity of the RAN. In the detailed planning phase multiple analyses are performed in order to be achieved the set of requirements. Then optimizing procedures are conducted for interference control by proper configuration of the sites and antennas.

Here the priority will be kept on showing how can be verified detailed planning aspects of WCDMA with the help of static simulations. We will perform UMTS case study by performing a couple of scenarios (different user profiles and distributions), which results we used to discuss the validity of the methodology and its sensitivity to certain demands. The principles of detailed WCDMA radio network planning in this paper are presented with the help of static radio network simulator – NPSW [4-6]. This simulator, is used in this Section and in the Section 4 where are shown the optimization results for WCDMA.

A. Simulation Scenarios Configurations

We chose two basic simulation scenarios that will be evaluated in multiple purposes: Scenario A – One user profile, and Scenario B – Multiple user profiles, both with heterogeneous traffic distribution. The main purpose is to identify the main metrics of the system capacity (throughput per sector, transmit power per sector and uplink load) in case of different users traffic demand. For the simplification and comparison purposes all analyses are based on one snapshot that in given moment of time presents in total 1140 active user in the system. The 13.5 km² test radio network has 10 three-sectored identical base stations (antenna tilt is 7 degrees in all sectors, and azimuths 90, 210 and 330 degrees) [2].

Fig. 2. Graphical representation of Scenario A snapshot.

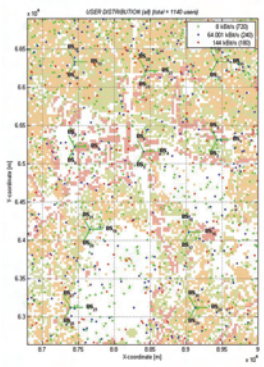


Fig. 3. Graphical representation of Scenario B snapshot.

B. Scenario A – One User Profile, Heterogeneous Traffic Distribution

Regarding the traffic related input parameters for Scenario A, we have: 1140 active users with 8 kbit/s data rate service, heterogeneously distributed in the network layout for which is needed 9120 kbit/s throughput. Simulations were terminated after 8 iterations and graphical results were saved. The graphical result for the snapshot of Scenario A is depicted on the Fig. 2. Other relevant statistics are summarized in Table I.

C. Scenario B: Multiple User Profiles, Heterogeneous Traffic Distribution

In comparison with the previous scenario this is more complex one, since in the same time three different user profiles are active. All the rest configurable parameters are kept the same. Table II consist the simulation inputs. At Fig. 3 we depicted the graphical representation of snapshot for Scenario B as output from NPSW. After analyzes of all graphical results the next discussion is valid:

- Served users: 600 (A: 1116);
- Refused users: 540, based on additional reasons 33% (limitations on the maximal power of the MSs and transmit power of BSs);
- Aggregated throughput = 389*8kbps + 124*64kbps + 87*144kbps = 24.78 Mbit/s (A: 8.72 Mbit/s);
- Uplink load factor: 0.63 (A: 0.42);
- Transmit terminal power is increased because of the increased speed of the users (8.26 dBm).
- The average shot handover overhead factor is 40.7% and it is slightly lower than in Scenario A;
- Soft handover: 20.2% from the total number of served users (121 connections).

IV. Pre-operational Performance Optimization of WCDMA

In this Section we present study for optimization of radio network plan with the aim to perform the optimization early in the planning phase. Since WCDMA based 3G systems are very sensitive regarding the interference, it is from highest interest not to cause or receive a significant part of it. The most efficient ways to improve capacity and coverage performance, e.g. to control the interference in the RNP phase [1], is through setting the site locations and configurations (sectroization) and setting the height, beamwidth and tilt of antennas.

For realization of the current study we use the network topology and snapshot used in the scenario B. Multi path "ITU vehicular A" channel profile is assumed [8]. In addition, in the NPSW simulator we activate 4- and 6- sector configurations. Besides already used antenna with 65º beamwidth (Scenarios A, B) with 3 dB and tilt 7º, we used here four additional and different 3 dB antenna types at 120º, 90º and 33º and additional tilts from 0º, 4º 10º and 14º. First we simulate different antenna tilts in case of different sectorization, for determining the optimal tilt, and after, we describe an environment for determining the capacity as a function of sectorisation and antenna choice.

TABLE I
OVERVIEW OF THE RESULTS OF SCENARIO A

	Avg. Sector	Sum
Requested Connections	38	1140
Served Connections	37.2	1116
SHO connections (22.1 % from MS are connected to two BS)	16.5	493.3
Total connections	53.7	1611
Throughput (kbit/s)	297.6	8928
Total Rate (kbit/s)	429.6	12888
SHO overhead	46.54%	46.54%
Not served - Downlink power per link limit	0.7	21
Not served – too bed CPICH	0.1	3
Not served - Uplink power limit	0	0
Total power of traffic channels (TCH) in downlink (dBm/W)	35.26/3.4	50/102
Uplink load factor	0.42	

TABLE II
INPUT TRAFFIC PARAMETERS FOR SIMULATING SCENARIO B.

Number of 8 Kbit/S users	720
Number of 64 kbit/s users	240
Number of 144 kbit/s users	180
Throughput requested kbit/s	47040
Total simulation iterations	22

A. Optimization of Radio Network by changing the down tilt

Simulations shown that optimal tilt angle is between 7º and 10º. This relatively high optimal angle can be explained by the very high installation height of the antenna (50m). Other-cell to same-cell interference ratio, *i*, goes down as tilt goes up. This is the case since the antenna beamwidth doesn't radiates too much power to other base stations and therefore a bigger amount from the radiated power goes to the area for which it is intention to be served by the analyzed base station. From the other side, the results show that network can serve more users in cases without or with small antenna tilting. Always there is some optimal value of the tilt depending from environment and base stations sites [9-11]. If the angle is too high, the service will decrease and the base station will not be in position to serve a big area as in the case with sufficient tilt. Due to the existence of lobs and wing at the sides of the antenna radiation pattern results show that coverage probability as a function of tilt angle may have certain variations of the *i* value[12]. All of this may be considered on Figs. 4 - 7.

B. Optimization of Radio Network by changing the antenna beamwidth

In the second case we illustrate the capacity improvement as a function of sectorisation when each base station is simulated as site with 3, 4 or 6 sectors. We simulate a couple of scenarios with different antennas beamwidth. We emphasized the validity of the correct antenna selection for the sectored configuration. Some of the graphical results are summarized on the Figs. 8 - 12.

After performed analyses it can be concluded that the network is overloaded in uplink. In all three sectorisation cases, the reason for uplink outages is because some mobile stations suffer from having not enough power. Nevertheless, the downlink is more limiting and more mobile stations go to outage. With the higher sectorisation may be served bigger number of users (bigger capacity). Other observation shows that for each sectorisation case the antenna bemawidth selection plays important role. Better choice is antenna with the bemawidth that is as small as possible, e.g. antenna with 65º is optimal for 3-sectored case and antenna with 33º is best for the 4 and 6 sectored scenarios. In order to achieve higher possible number of served users it is very important to control the interface on the effective way. If the sector overlapping is too big, the interference may pass in other sector directly reducing its capacity.

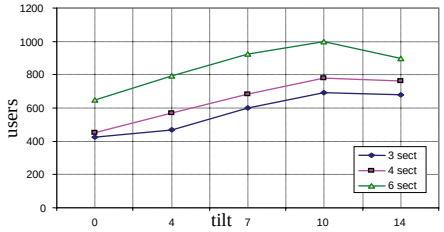


Fig. 4. Impact of the antenna tilt on the capacity.

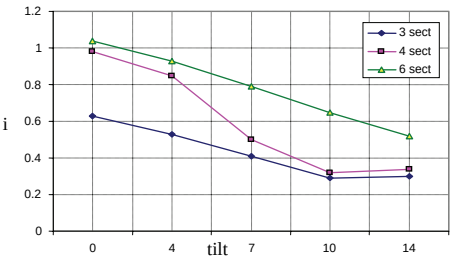


Fig. 5. Impact of the antenna tilt on *i* value.

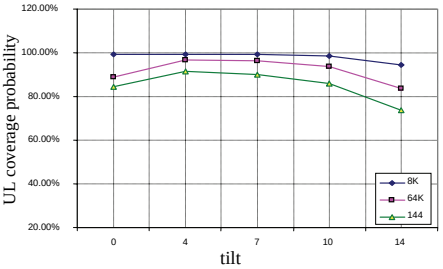


Fig. 6. Impact of the antenna tilt on the UL coverage probability, for 6 sectored configurations and for all services.

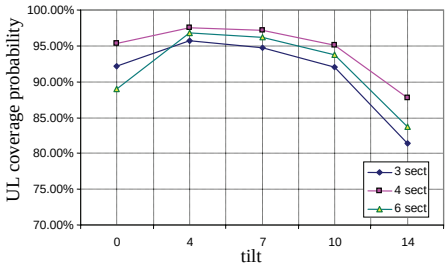


Fig. 7. Impact of the antenna tilt on the UL coverage probability, for 64 kbit/s service.

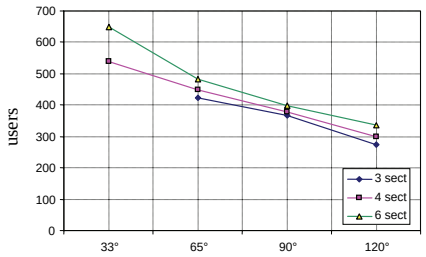


Fig. 8. Impact of the antenna beamwidth on the capacity.

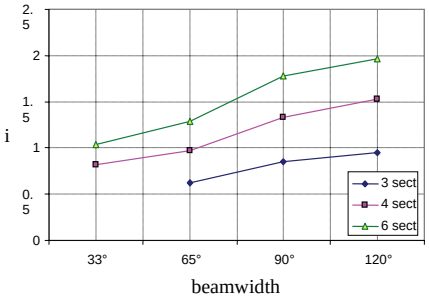


Fig. 9. Impact of the antenna beamwidth on *i* value.

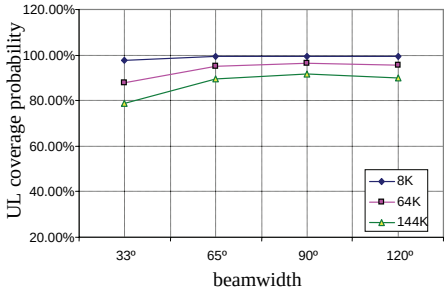


Fig. 10. Impact of the antenna beamwidth on the UL coverage probability, for 4 sectored configurations, all services.

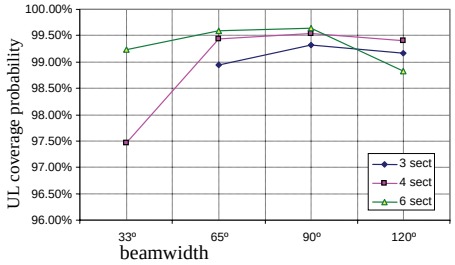


Fig. 12. Impact of the antenna beamwidth on the UL coverage probability, for 8 kbit/s service.

V. Conclusion

Although we can still relay on the fact that data traffic in today mobile networks is in the beginning emerging phase, also it is true that enormous investments are made in order to bring closer those networks to the real multimedia systems. The only mean for turning back the investments includes proper dimensioning of the systems that will handle the data traffic with required QoS, without sacrificing the dominant voice users. All of this combined with the analyzed framework for detailed WCDMA radio network planning and its optimization, leads to conclusion that whenever it will be changed the business strategy of the mobile operator (introduced new services, price policy changes, etc.), it is more than needed support from the planning tool in sense of base station location selection, quality analyses, efficient resource allocation and optimization of the mobile network when with relatively simple means (antenna tilting and correct antenna selection for any scenario) the interference may be controlled and the capacity and coverage may be balanced and improved.

References

[1] Jaana laiho, Achim Wacker, Tomas Novosad (ed.), “Radio Network Planning and Optimization for UMTS”, John Wiley and Sons, LTD 2002.

[2] Leila Zurba Ribeiro, “Traffic Dimensioning for Multimedia Wireless Networks”, Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy In Electrical Engineering, April 17, 2003.

[3] Jaana Laiho, “Radio Network Planning and Optimization for WCDMA”, Thesis for the degree of Doctor of Science in Technology, Helsinki University of Technology Radio Laboratory Publications, Espoo, 2002.

[4] Achim Wacker, Jaana Laiho, Kari Sipila, Kari Heiska, Kai Heikkinen, “NPSW Matlab implementation of static radio network planning tool for WCDMA”, Documentation;Version 5.0.0, 2001.

[5] Achim Wacker, Jaana Laiho, Kari Sipila, Kari Heiska, Kai Heikkinen, “NPSW Matlab implementation of static radio network planning tool for WCDMA”, Scenarios; Version 5.0.0, 2001

[6] Achim Wacker, Jaana Laiho, Kari Sipila, Kari Heiska, Kai Heikkinen, “NPSW Matlab implementation of static radio network planning tool for WCDMA”, Code; Version 5.0.0, 2001

[7] Trung Van Nguyen, “Capacity Improvement Using Adaptive Sectorisation in WCDMA Cellular Systems with Non-Uniform and Packet Mode Traffic”, Submitted in fulfillment for the requirements of the degree of Doctor of Philosophy in Victoria University of Technology, March 2005.

[8] Guidelines for evaluation of Radio Transmission Technologies for IMT-2000, Recommendation ITU-R M.1225, 1997.

[9] Jaana Laiho-Steffens, Achim Wacker, Pauli Aikio, “The impact of Radio Network Planning and Site Configuration on the WCDMA Network Capacity and Quality of Service”, Nokia Networks, 2000 IEEE.

[10] Toskala A., Holma H. and Muszynski P., ETSI WCDMA for UTMS, Proc. ISSSTA 1998, South Africa, September 1998.

[11] Jaana Laiho-Steffens, Achim Wacker, “Radio Network Planning Process and Methods for WCDMA”,

[12] Jarno Niemala, “Impact of Base Station Site and Antenna Configuration on Capacity in WCDMA Cellular Networks”, Masters of Science Thesis, Tampere University of Technology, Tampere, August 7th, 2004.

Signaling on the Most Critical Interfaces in GSM

Zorica Bungurova-Kaziovska [1] and Toni Janevski [2]

Abstract - **In this paper, the analytic model for prediction of signaling on the interfaces, referred as most critical and very often as bottleneck in the core mobile network, and design of signaling network, are given. Gb specific interfaces is out of the scope of this paper.**

Keyword **– Signaling, A-interface, C/D interface**,

I. INTRODUCTION

GSM (Global System for Mobile Communication) is most popular standard for public mobile network, which is designed to provide mobility to the end users of cellular phones and Internet. Common channel signaling system Number 7 is a key element in supporting a large number of applications in telecommunication networks.

Addressing and control information from an origin to destination point in a communication network is referred to as signaling, and is used in selecting a path, providing information on the status of the users and resources, or performing the necessary network supervision and charging function.

II. SIGNALING

The A-interface lies between the MSC (Mobile Switching Center) and the BSS (Base Station Subsystem). The A-interface consists of both user plane (PCM links) and control plane (signaling).

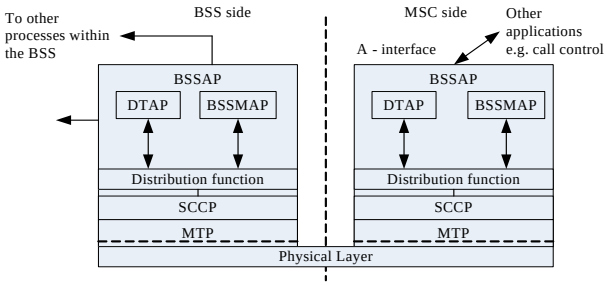


Fig. 1 SS7 stack for BSS Application Part

The SS7 stack is composed of an SCCP (Signaling Connection Control Part) layer on top of the three MTP (Message Transfer Part) layers, MTP3, MTP2 and MTP1.

The signaling part of the A-interface is used to carry information concerning: BSS management, call handling and mobility management. BSSAP [1] (Base Station System Application Part) is application (upper), which sits on an SS7 stack, as it shown in Fig. 1

TCAP is not used for BSSAP. BSSAP is an SCCP user (usually SSN=254 for ETSI) and it often uses the logical connection function in SCCP (Signaling Connection Control Part) in order to identify the MS for signaling messages that follow the initial message. The BSSAP requires both SCCP service: CO and the CL. The CO transfers many or long signaling messages between two nodes. The CL transfers short messages, including routing information, to their destination. The BSSAP application layer is divided into two parts: DTAP (Direct Transfer messages) and BSSMAP (Base Station System Management Application Part). Generally, the CM – Connection Management and MM – Mobility Management layers consist of DTAP messages while the RR – Radio Resource Management and FM – Facilities Management layers consist of BSSMAP messages. Discrimination parameter is used for distribution function.

DTAP (Location Updating, Identity, Authentication, CM service) messages are call control or mobility management between the MSC and the MS, pass through the BSS transparently. Initial MS (DTAP) messages are associated with call control and mobility management and they pass through BSS unchanged, but part of them is analyzed by BSS (Paging, CM Service request). They are encapsulated within the Information Element "Layer 3 Information" in the BSSMAP Complete Layer 3 Information. BSSMAP messages are used for resource management (Assignment, Clear), Ciphering, Handover control as CL and Paging, Circuit management as CO.

C-interface between HLR and GMSC is used for interrogation procedure during Call Setup and a Forward Short Message. D-interface between HLR and VLR is used for management of Subscriber and during Location procedure. The MSC/VLR, HLR, and GMSC communicate via the MAP, using only the CL mode for Location Management, Handling of call and subscriber services, Handover and Transfer of security/authentication data.

To support GSM requirements is especially designed Mobile Application Part – MAP [2]. MAP is a TCAP [3] (Transaction Capabilities Application Part) user and also uses the SCCP [4] and MTP [5] for transmission of information.

Only the MTP network has access to the physical link, for example, cable. It provides the common platform between the different user parts and functional elements. The combination

[1] Zorica Bungurova-Kaziovska, is with Faculty of Electrical Engineering, Dept of Telecommunication, Karpos 2 bb 1 000 Skopje, Macedonia
Zorica.BungurovaKaziovska@t-mobile.com.mk
[2] Toni Janevski, is with Faculty of Electrical Engineering, Dept of Telecommunication, Karpos 2 bb 1 000 Skopje, Macedonia
tonij@etf.ukim.edu.mk

of the MTP and the SCCP is called the Network Service Part (NSP).

In a node, a signaling message can be originated, terminated, or transferred. The signaling method uses an address called a Signaling Point (SP), which is identified by the Network Indicator (NI) and the Signaling Point Code (SPC). Each node in a network must know all its potential receivers (cooperating SP). The NI distinguishes between the different networks, national and international. To generate an originating labeled message, the node uses the SPC of the OWNSP as the Originating Point Code (OPC), and the SPC of the cooperating SP as the Destination Point Code (DPC). Whether terminating or transferring a message, a node always compares the DPC of the incoming message to its OWNSP. If they are not equal, the node must transfer the message. This requires a routing function called MTP routing. The MTP routing ties a DEST to a Link Set (LS). Group of Signaling Links (SL) that directly interconnects two SP is called LS. Each SL in LS receives an individual number called a Signaling Link Code (SLC). SL is one time slot (64kbit/s) within PCM (2Mbit/s). The Signaling Terminal (ST) is connected to the LS via the SLC. Routing is always based on addresses: Calling address (identifies call origination) and Called address (identifies call destination). SCCP addressing is very flexible and makes use of three separate elements: Destination Point Code (DPC), Global Title (GT) and Sub-System Numbers (SSN). SSN and DPC are used for routing if the nodes are directly connected. Usually, SSN and GT are used for routing across mediate nodes. Each node must have its individual calling address, used in national and international transmission. It has the same structure (defined by E.164) as an MSISDN. The structure is CC+NDC+SN.

In AXE (A Switching System for mobile and fixed telephone networks), a specific MAP operation is typically handled by one function block, which is referred to as Application System Elements (ASE). An ASE can only communicate with a compatible peer ASE. A set of Application Service Elements (operations) is called an Application Context. MAP and TCAP belong to the layer 7 in the OSI model. Application Context version 1, 2 and 3 belongs to the Presentation layer (layer 6). Please see Fig. 2.

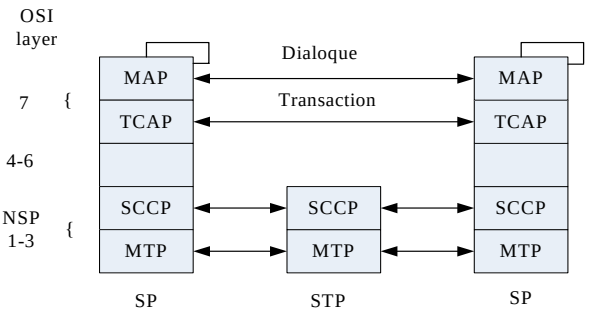


Fig. 2 Functional view of SS7 system

The main purpose of TCAP is support for interactive applications (TC-user) in a distributed environment (across many nodes). TCAP allows TC-users to have several conversations running concurrently. In order to set up and support the dialogue between two TC-users, TCAP provides an association between the local and remote ends called a Transaction. A TCAP message is composed of three portions: transaction, component and dialog. The Transaction portion consists of an originating and/or destination transaction identifier that is used by both TCAP ends to identify the connection. The component portion encapsulates the information sent to and received from TC-users in the form of components. For example, MAP messages to and from TC-users in GSM are encapsulated as components. One or multiple components (MAP messages) can be sent at the same time. The dialogue portion of a TCAP message allows two TC-users to negotiate the terms of a transaction. For example, the dialogue portion is used by MAP TC-users in order to negotiate which MAP version is to be used for a particular transaction. The dialogue portion is optional within TCAP and if used is only present in the first two TCAP messages within a transaction. Components which are used to perform Operation are exchanged within the Structured dialogue (begin, continue and end message). All components within an operation have the same Invoke ID. Dialogue must be terminated or aborted. The MAP protocol only makes use of structured dialogues. TCAP receives data from the TC-users, encapsulates it and hands it over to NSP.

III. Results

The transmission rate for signaling link is 64kbit/s or 8000 octet/s in both directions. The recommended traffic load per SL under normal condition is 0.3 E or 2400 octet/s. Table 1 shows the capacity of the link set. If number of SL in the LS is divisible by 2, the load of the links within the LS is balanced. For redundancy, minimum two links are required towards one destination. Maximum 16 SL can be established towards one destination.

Table 1 Signaling capacity

Number of signaling links	Signaling capacity
1	19,2 kbit/s
2	38,4 kbit/s
3	51,2 kbit/s
4, 5	76,8 kbit/s
6, 7	102,4 kbit/s
8, 9, 10, 11, 12, 13, 14, 15	153,6 kbit/s
16	307,2 kbit/s

Table 2 Measuring report for traffic event

Normal LU	MO	MT	SMS-MO	SMS-MT	HO
0.16	0.15	0.11	0.01	0.03	0.01

Table 2 shows the average value of event/s per subscriber for observed traffic cases, because observed BTS have different busy hour per different service. For sampling was used 15 minutes resolution. Measurements are done within period of 1 year and for all cells within the BSC. SS and HO

produce insignificant signaling load. From the measuring report for traffic event, we can see that bigger part of the signaling load is from LU. All results lead to the same conclusion: 50.7% of load is originated from calls; 1.8% - from handover; 1.65% - from SS; 37% - from LU; 5.3% - from SMS and 2.8% - from internal traffic.

IV. ANALYTIC MODEL

Procedure for calculation of signaling traffic on the common C and D interface

Incoming and outgoing octets per sec., in both directions (query and response) for each traffic event are calculated. Fig. 3 shows the results.

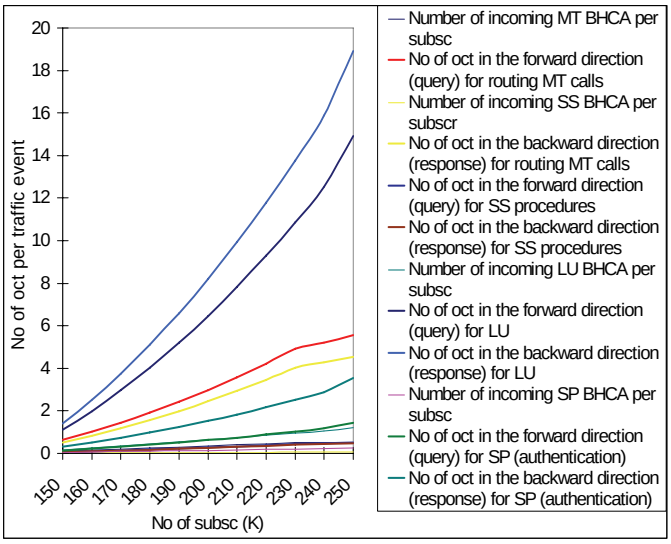


Fig. 3 Estimated no of octet per traffic event on the C/D interface

Table 3 Length of message units on the C/D interface

Direction	Type of event	Length of MSU
forward	MT call	128
forward	LU	179
forward	SS	130
forward	SP	67
backward	MT call	105
backward	LU	227
backward	SS	120
backward	SP	164

Table 3 shows the used length of message signaling unit for observed events on the C/D interface in both directions.

Equations (1) and (2) are used to calculate max number of octets transmitted per second in each direction, and equation (3) - to estimate required number of SL between MSC/VLR and HLR.

$$O_T^I = \left(O_{MT}^I + O_{LU}^I + O_{SS}^I + O_{SP}^I\right) * k \quad (1)$$

$$O_T^O = \left(O_{MT}^O + O_{LU}^O + O_{SS}^O + O_{SP}^O\right) * k \quad (2)$$

Factor k = 1.1 is for SMS and USSD notification.

$$O_T = \max\left(O_T^I, O_T^O\right) = 0.085 * N \quad (3)$$

Fig. 4 shows estimated and measured signaling load on the C/D interface in the six observed measure periods.

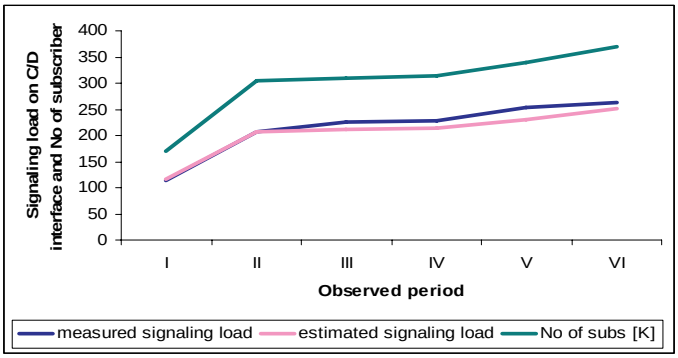


Fig. 4 Estimated signaling load as compared with measured signaling load on the C/D interface

Procedure for calculation of signaling traffic on the A interface

Incoming and outgoing octets per sec., in both directions (query and response) for each traffic event are calculated. Fig. 5 shows the results.

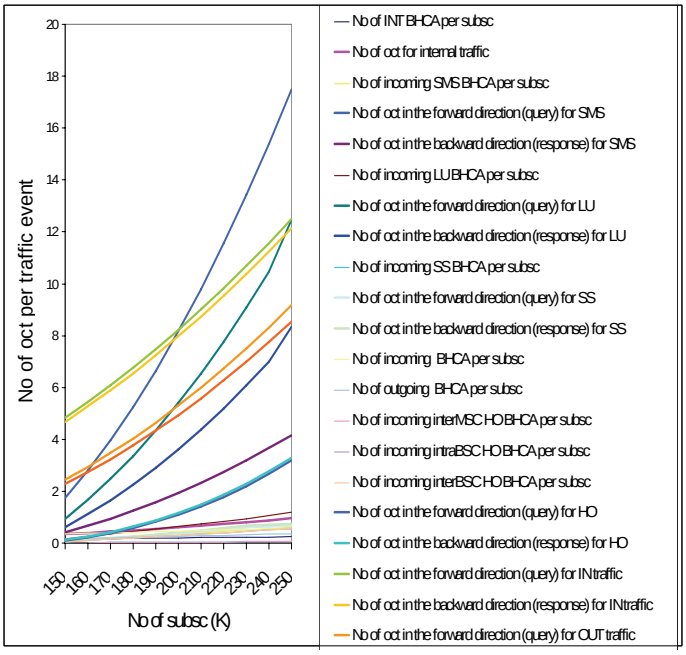


Fig 5 Estimated no of octet per traffic event on the A interface

Table 4 shows the used length of message signaling unit for observed events on the A interface in both directions.

Table 4 Length of message units on the A interface

Direction	Type of event	Length of MSU
forward	call	268
forward	SMS	420
forward	LU	130
forward	SS	165
forward	call/LU/SS	306
forward	inter-MSC HO	82
forward	inter-BSC HO	82
forward	Internal traffic	55
backward	call	260
backward	SMS	100
backward	LU	87
backward	SS	182
backward	call/LU/SS	285
backward	inter-MSC HO	70
backward	inter-BSC HO	90
backward	intra-BSC HO	31

Equations (4) and (5) are used to calculate max number of octets transmitted per second in each direction, and equation (6) - to estimate required number of SL between MSC/VLR and BSC.

$$O^I = O^O_{IN} + O^O_{OUT} + O^O_{HO} + O^O_{SS} + O^O_{LU} + O^O_{SMS} \quad (4)$$

$$O^O = O^I_{IN} + O^I_{OUT} + O^I_{HO} + O^I_{SS} + O^I_{LU} + O^I_{SMS} + O_{INT} \quad (5)$$

$$O_T = \max\left(O^I, O^O\right) = 0.13306 * N \quad (6)$$

Fig. 6 shows estimated and measured signaling load on the A interface in the six observed measure periods.

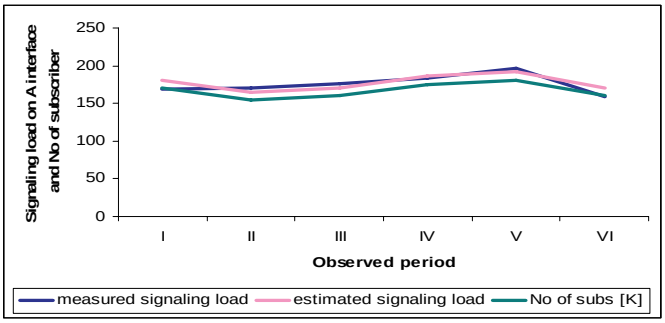


Fig. 6 Estimated signaling load as compared with measured signaling load on the A interface

VI. Discussion

The service area of the observed BSC covers mountain region and valley, and must be define a lot of neighbour relation between cells. Subscribers are very mobile and it was registered high fluctuation of subscribers to the neighbour BSC that cover the major city in the country. Also, in this MSC service area is registered high fluctuation during summer and religious vacations. For this reason is used very high number of attempt per LU (e.g. for normal type of LU - 0.18 event per subscriber in busy hour, instead of measured 0.16. Also, for call was used four time bigger value because of a high number of registered calls without B-answer (snakes). To adjust the real conditions in the model for C/D interface was added factor of 10% for SMS and USSD notification. For Call dependent SS was added a factor of 10% on the C/D interface and 15% on the A-interface, where they have bigger impact on the signaling volume. If Authentication is not performed in the network (during the call, SMS, SS or LAI changing), signaling load will be decreased for 0.5kbit/s on the A and C/D interfaces. LU and call related procedures take the most part of the produced signaling load. The A and C/D interfaces are most critical, because on the other interfaces under the same environment was registered insignificant signaling load.

V. Conclusion

The signaling load is highly dependent on the configuration of terrain and radio design. Bad relation between cells, unbalanced power of the cell, interference, SDCCH configuration in the cell, LAPD concentration factor in the cell and other radio factors can produce a bigger load on the A and also on the C/D interface (e.g. config=4 is set for cell which serves high traffic). A use of different services increases the signaling load (e.g. SMS, USSD notification). Also the signaling load is dependent on the subscriber behavior and the tariff model. This analytic model can be used just for prediction of the signaling load in the normal terms, because signaling in the network has a burst and unpredicted grow under abnormal terms (e.g. accident). Those two observed interfaces can be bottleneck in the network. Signaling load will be rapidly increased with increasing of traffic intensity, number of subscribers, and number of the offered services and applications to the customers. High speed signaling links (bigger capacity of medium and bigger speed) must be used to support needed signaling volume on these interfaces, beside careful and optimal design of the serving area. Special attention is needed for area that covers the major city.

References

[1] Digital cellular telecommunications system (Phase 2+); Signalling transport mechanism specification for the Base Station System - Mobile-services Switching Centre (BSS - MSC) interface (GSM 08.06)

[1] Digital cellular telecommunications system (Phase 2+); Application of the Base Station System Application Part (BSSAP) on the E-interface (GSM 09.08)

[2] Digital cellular telecommunications system (Phase 2+); Mobile Application Part (MAP) specification, GSM 09.02 series

[3]Digital cellular telecommunications system; Guidelines for the modification of the Mobile Application Part (MAP) in phase 2+ (GSM 10.02)

[4] ITU-T recommendation Q.711, Q.712, Q.713, Q.714, Q.716

[5] ITU-T recommendation Q.701, Q.702, Q.703, Q.704, Q.705, Q.706, Q.707

An Implementation of Adaptable "Bottom-up" Model for Calculating Interconnection Costs

A. Kostić-Ljubisavljević [1], V. Radojičić[2] and V. Aćimović-Raspopović[3]

***Abstract* – In this paper we present the results of implementation of adaptable "bottom-up" model for calculating interconnection costs build on the long-term incremental costs methodology. This cost model is suggested to be very convenient for implementation in transition countries. The model needs numerous input parameters for calculation interconnection costs but it can be tested without infrastructure parameters.**

***Keywords* – adaptable bottom-up model, interconnection costs**

I. INTRODUCTION

Various cost allocation methodologies have been proposed for interconnection. Some are based on the operator's accounting and involve allocating historical costs to different services according to criteria prescribed by the regulators. Others estimate costs by reconstituting the networks on the basis of currently available technologies. The latter is generally regarded as the most appropriate one for estimating services interconnection costs. Long Run Incremental Cost (LRIC) methodology estimates the costs incurred, while offering a sub-set of services. The costs considered are those that would be avoided if these services were not offered. For estimating the cost of the interconnection services, the selected increment comprises network elements belonging to the core network, i.e., those shared among all the network's users, excluding network elements contributed to end users.

In this paper we suggest the adaptable *bottom-up* model for calculating interconnection costs as very convenient for implementation in transition countries. Also we proposed mathematical models for calculating demand for PSTN and leased lines services, switching and transmision elements.

The paper is organized in the following way. In Section 2 we briefly discuss the cost determination methodology. In Section 3 LRIC model is presented. Modeling principles are described in Section 4 and model structure is given in Section 5. An example of test results is given in Section 6. Conclusion remarks are given in Section 7.

II. COST DETERMINATION METHODOLOGIES

For telecommunication operators, determining the cost attained to produce a specific service is a delicate and complex task. Therefore, this section defines some of the cost concepts used in telecommunication industry [1].

Direct costs or directly attributable costs are expenses that are incurred when producing a specific service or a series of services or products. They can be fixed or variable. Fixed costs represent the proportion of the organization's expenses that does not depend on, or vary with the activity of the organization, and variable costs are closely related to the level and the development of the organization's production and marketing operations. The difference between fixed and variable costs is time related. Generally, fixed costs are long term oriented. Conversely, variable costs are directly related to the day to day operations of the organization, hence, short-term oriented. However, in the long run, even fixed costs are variables. It is therefore possible to reconcile cost accounting analysis to the economic analysis. *Joint costs* are generated by a family of services or products. From economic point of view, joint costs are costs incurred in fixed proportions every time a service or a product belonging to the same family is produced by the organization. *Common costs* are shared by all the services or products of the company. The sum of the fixed costs, the variable costs, the joint costs and the common costs gives the total production cost or global cost. The global cost or total production cost is directly related to the production volume. Two fundamental cost concepts yield from this definition: the average cost and the marginal cost. *Average cost* is the unit cost obtained by dividing the total cost by the number of units produced. *Marginal cost* is the total cost variation resulting from a variation of the organization's production. Marginal cost is defined as the *incremental cost* resulting from the production of an additional unit. *Historical cost* is the cost value entered in the organization's books. Contrary to historical cost, *forward looking long-term* cost relies on the best available technologies, assuming that the organization's production cycle is globally at optimum.

There are several methodologies in regard to allocation of joint and common costs:

1. The Fully Distributed Costs method (FDC);
2. The Efficient Component Pricing Rule (ECPR) method;
3. The Ramsey-Boiteux and Laffont-Tirole methodology, which rely on demand-price elasticity;
4. Long-term incremental costs methodology.

The cost model analyzed here is relied on the long-term incremental costs methodology.

III. THE LRIC CONCEPT

There are two families of LRIC models: *top down* LRIC and *bottom up* LRIC. *Top down* LRIC is essentially an activity based costing methodology. *Top down* LRIC models derive incremental costs by summing up the costs that can be directly attributed to the service, and adding a mark up that covers a proportion of joint and common costs. The mark up is determined with the help of arbitrary distribution keys. Therefore, the resulting cost is more backward oriented than forward looking. *Bottom-up* LRIC is a methodology for determining forward looking service cost. The methodology involves simulating the cost incurred by an efficiently

[1] Aleksandra Kostić-Ljubisavljević (e-mail: a.kostic@sf.bg.ac.yu)
[2] Valentina Radojičić (e-mail: valentin@sf.bg.ac.yu)
[3]Vladanka Aćimović-Raspopović (e-mail: v.acimovic@sf.bg.ac.yu)
All authors are with University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia

operated network, using best available technology, to provide the service. *Bottom-up* LRIC models are highly recommended for regulatory decisions.

The LRIC methodology takes in account fixed costs caused by the provision of interconnection services, but does not take in account the common costs, which do not vary proportionately with the provision of interconnection services. In practice, the implementation of LRIC models requires a systematic assessment of demand and cost evolutions, as well as the interdependencies that may result. A systematic analysis of network elements load factors is also critical to accurately size the resources needed to convey the incremental traffic. In other words, the LRIC methodology enables the cost determination to be based on forward looking costs, and not on historical costs [2, 3].

The incremental costs of a service somehow represent the cost savings, which result from not implementing that service. The long-run concept involves taking the costs incurred in a long-term perspective. In the long run, production fixed costs can be considered as being variable costs. The long-run incremental costs of a service therefore represent all of the costs that could be avoided if that service were not implemented. Hence, the incremental costs include all the costs directly attributable to service, whether they are variable or fixed.

A telecommunications operator generally offers a wide range of services, whether to its subscribers or to other operators. These services have been traditionally classified according to their commercial nature. However, another service classification approach is based on the use of network elements. In general, there are three major categories of network elements:

- Elements, which are dedicated to end users. These elements are generally found in a local loop network or access network, and include subscriber's connection line to the switch;
- Elements, which are shared among users. These network elements are allocated temporarily to a subscriber and are generally found in the core telephony network.
- Elements shared by users, but used for the provision of complementary or supplementary services.

The services discussed above can be labeled as: access services, for services dedicated to each user, transport services, for services shared among network's users and offered by the core network functions and value added services, for shared services offered by intelligent functions of the network.

The incremental cost methodology distinguishes directly attributable from no attributable costs. It involves ascertaining, whether or not the production cost of "horizontal" service is increased, whenever a "vertical" service is added to the basket of other services produced [7].

The provision of interconnection services does not modify the retail sales service and it does not modify the access service, because the capacity implemented and its maintenance does not have to be adjusted to bear the flow of this additional traffic. In contrast, transport services are affected by the traffic resulting from interconnection services. The increment costs, to be considered for the calculation of interconnection services costs, comprises the core network and excludes the access network's elements which are dedicated to users [8].

IV. Modelling Principles

The LRIC method adopts a long run approach. The reconstruction of all networks elements is assumed during one year, and includes the increment costs. Selecting a forward-looking approach means considering both the best technologies available and their current costs. The model simulates a network that, at a given production level, minimizes the total cost by using best available technologies.

The model has to determine unit interconnection costs. It is, therefore, necessary to select metrics. Traditionally, the traffic flow through the core network is measured in minutes. The cumulated duration of calls, measured in minutes, is considered to be the important parameter in determining costs [4].

However, some interconnection prices and some cost models also take in account the number of calls. In practice, network elements also handle the following non-invoiced periods: the call set-up time, the closing time, and the time spent in handling unsuccessful calls. The sizing of some network elements therefore depends on the number of calls conveyed. This is particularly relevant for switching elements that have to handle call attempts. The number of calls essentially continues to determine the cost incurred by the temporal occupation of non invoiced network elements (waiting time and unanswered calls). The cost per minute is by convention, the representative unit cost.

To sum up, the model re-constructs the costs of the network at two levels. At the level of nodes (switching elements), investment costs are structured in fixed and variable components according to BHE (business hour Erlangs transformed into 2 Mbps) and to the number of subscribers, respectively. At the level of links, calculations are done at two sub-levels:

- Transmission level - at which electronic transmission equipment (mainly on SDH rings) are designed and sized;
- Infrastructure level - at which the link substratum is designed.

The links are sized by the traffic at peak times expressed in capacity (Mbps) and are sized to handle the switched telephone network traffic, as well as leased lines bandwidth. Assumptions regarding the sharing of certain elements of basic infrastructure with other networks subsequently allow for the sharing of costs, which are not fully borne by the core network.

The model is easy to use, and request information that regulators or operators can easily find. However, there remain some limitations. First, the model is designed for "small" networks which do not yet implement complex transit functions or routing algorithms. Second, the model does not seek full optimization when re-building the transmission network. More specifically, the model does not optimize the nodes, as the current network topology is kept unchanged. Third, it does not provide a detailed modeling of the cables and duct networks, although it discriminates among different types of geography (urban, suburban, and rural) or nature of the trenches (wrapped, ducted, and buried). Such a module

could be developed by each regulator, although nodes location optimization may not be critical at current stage of telecommunications network development in developing countries. The proposed cost model provides accurate cost proxy estimates when applied to networks with less than 1.5 million main lines, and when the incumbent's network topology (number and location of nodes) is optimized or close. The lower bound will correspond to a configuration of a fully optimized network (nodes and links), while the upper bound will correspond to a partially optimized network (only links are optimized).

V. Model Structure

In general, the model requires the entry of substantial amount of information characterizing the networks which are interconnecting (topology, architecture principles and rules, traffic matrix and patterns, costs elements, etc.). To palliate the frequent gaps in the information systems, the model proposes default values that the user could consider if needed. This applies mainly to the routing factors that are rarely documented in most countries. The logic of the model is: [5]

1. The main input elements of network are nodes and links;
2. Each service uses these elements in different proportions. The routing factors represent the average number of times a given element is used by the service considered. The model then calculates the total load supported for each network element;
3. The model calculates the size of network elements (transmission elements) within the framework of the selected topology;
4. The model adds up all corresponding network element costs and calculates the per minute cost for each network element;
5. The model calculates the interconnection costs on the basis of the routing factors.

The network conveys interconnection service according to the extent and complexity of the service. The model calculates the investment cost, and also the cost of operation and maintenance incurred by network elements. These costs are distributed over four components: investment, operating, attributable and common costs.

Common costs are expressed as a percentage of the attributable costs. The attributable costs are the costs directly caused by the interconnection service and would have been avoided by the provider. They consist of operating and investment costs. The operating costs are composed of two terms: maintenance and operating cost related to the network element (spare parts, equipment section of preventive and corrective maintenance, energy consumed, etc.) and staff costs related to operation and maintenance activities. The staff cost is distributed among different network elements according to parameters specified by the user. The investment costs are calculated considering the estimated volume of traffic to be handled by each network element at peak hour. The size of each network element is accordingly derived using engineering procedures. For each network element, the model calculates the investment cost incurred and generates an investment annuity. Operating and non attributable costs are then allocated to it to obtain a global cost per minute for each element. For each service, relevant costs per minute for network element are added up to obtain the interconnection cost. The latter value is adjusted using the gradient of retail prices to derive the interconnection rate.

The model calculates the interconnection costs over:

- Switching elements are considered to be the nodes. Their investment costs include a fixed component and a variable one which depends on the switching system BHE and the number of subscribers connected.
- The links connecting the different network's nodes. The model differentiates the infrastructure and transmission layers:
  - The transmission layer enables the user to design the electronic transmission equipment by choosing and sizing the capacity of the most appropriate technology to be rolled out (SDH rings, PDH technology, etc);
  - The infrastructure layer enables the determination of the links' substratum or the physical elements that will support the transmission link.

The links are sized by the traffic carried at peak hour expressed in Mbit/s. The transmission links size is determined, to enable normal flow of traffic expected from switching centers and leased lines connected to them. Certain infrastructure elements are shared by different categories of traffic; this is particularly applicable to SDH rings in the access network. The model enables such costs to be shared. For example, SDH rings in the access network will be used to carry core network traffic flow and the access network traffic.

VI. Example

We tested the previously described model on hypothetical local network. Input parameters of the model are:

1. **Node and subscriber line informations** (3 tandem switches, 15 local switches, 30 remote concentrator units);

2. **Traffic informations** (number of minutes and number of successful calls in local, long distance, international, calls to/from mobiles, interconnection, number of leased lines and others). We used recommended values for traffic statistic of successful and unsuccessful calls.

Demand generated by the current number of billed calls can be expressed as:

$$D_c = \sum_n B_c \tag{1}$$

where: *Dc* - current demand measured in number of calls, *n* - number of PSTN services, *Bc* - number of billed calls.

Demand generated by current number of billed calls in minutes can therefore be expressed as:

$$D_m = \sum_n B_m + (B_c \cdot T_h) \tag{2}$$

where: D_m - current demand measured in minutes, *n* - number of PSTN services, B_m - number of billed minutes, B_c - number of billed calls, T_h - holding time in minutes.

Demand for leased lines is expressed in 64bit/s equivalents and is converted into Mbit for dimensioning transmission electronics. This can be expressed as:

$$D_L = (n/\psi) \cdot 2 \tag{3}$$

where: D_L - existing demand for leased lines in Mbit, *n* - number of 64kbit equivalent leased lines, ψ - number of 64 kbit channels in a 2 Mbit.

3. Percentage of subscriber lines connected to network elements;

4. **Establishing demand for switching elements** (we used recommended values for utilization level of switching nodes);

The total demand for the switching element measured in number of calls is expressed as:

$$T_{dcE} = \sum_{i}^{n} T_{dci} \cdot R_i \quad (4)$$

where: T_{dcE} - total demand for the switching element for all narrow band calls, n - all narrow-band services, R - routeing factor for switching element, i - narrow band call service,-T_{dc} - total demand (in calls) for narrow-band service.

The total demand for the switching element measured in minutes can be expressed as:

$$T_{dmE} = \sum_{i}^{n} T_{dmi} \cdot R_i \quad (5)$$

where: T_{dmE} - total demand for the switching element for all narrow band minutes, n - all narrow-band services, R - routeing factor for switching element, i - narrow band call service, T_{dm} - total demand (in minutes) for narrow-band service

5. **Establishing Demand for Transmission Elements-** Traffic (in minutes) making use of the transmission network is converted to BHE that are used to dimension the transmission network. BHEs have to be converted to Mbit as follows:

$$D_{SW} = BHE /(g * \psi / 2) \quad (6)$$

where: D_{SW} - existing demand for switched traffic in Mbit *BHE* - total switched traffic in BHE, g - circuit efficiency (or erlangs per circuit), ψ - number of 64 kbit/s channels in a 2Mbit;

The adjusted total demand for the transmission element expressed in Mbit is:

$$ATD_{TE} = TD_{TE} / f \quad (7)$$

Where ATD_{TE} - total (adjusted) Mbit at a transmission element; TD_{TE} - total demand for the transmission element in Mbit, f - percentage of capacity utilised;

Adaptive interconnection model is tested without infrastructure parameters and results are shown in Tables 1, 2.

Table 1. LRIC including common costs

	RCU	LS	TS
Equipment Costs (euros)			
network	966,019	2,279,909	986,340
non-network	47,010	110,950	53,977
total equipment costs	1,013,029	2,390,859	1,040,317
Operating Costs (euros)			
network	540,256	1,016,933	440,118
non-network	159,889	300,558	214,905
total operating costs	700,146	1,317,491	655,023
Total costs (inc working capital)	1,884,493	4,079,185	1,864,874
	RCU	LS	TS
Unit Cost	2.122774	8.436784	4.008328

Table 2. LRIC excluding common costs

	RCU	LS	TS
Equipment Costs (euros)			
network	666,481	1,807,705	986.340
non-network	47,010	110,950	53,977
total equipment costs	713,491	1,918,655	1,040,317
Operating Costs (euros)			
network	346,216	713,746	440,118
non-network	159,889	300,558	214,905
total operating costs	506,106	1,014,304	655,023
Total costs (inc working capital)	1,341,556	3,226,255	1,864,874
	RCU	LS	TS
Unit Cost	1,341,556	3,226,255	4.008328

It could be possible testing different network configuration (maximum number of nodes on a SDH ring is 16), with great range of input parameters.

VII. CONCLUSION

The analyzed cost model is a tool provided to regulators and operators to help them to determine economic oriented interconnection rates for terminating and departing traffic from, or to, fixed and mobile networks. It also ratifies crucial assumptions pertaining to traffic demand, network optimization, and cost allocation between final and intermediary services. LRIC model is generally accepted as best practice. This approach is most efficient in calculating interconnection costs based on current technology rather than exiting book assets. It usually leads to lower interconnections rates and stimulates competition but provides lower revenues to incumbent operator.

In this paper we demonstrate the results of use of software tool based on adaptable interconnection model. We choose hypothetical network example on local national level and calculate LRIC including common costs, LRIC excluding common costs and common costs. In practice, this cost modeling enables regulators and operators to improve their knowledge on the industry cost structure, and its efficiency frontier.

ACKNOWLEDGMENTS

This paper resulted from researching project *Planning, Forecasting and Pricing in Telecommunication Networks* supported by Serbian Ministry of Science and Environmental Protection (TP-6106A).

REFERENCES

[1] M. Armstrong, "Network Interconnection in Telecommunications," *Economic Journal* 108 (1998), pp. 545-564

[2] M. Armstrong, C. Doyle and J. Vickers, "The Access Pricing Problem: A Synthesis," *Journal of Industrial Economics* 44 (1996), pp. 131-150

[3] J. J. Laffont, "Translating Principles Into Practice", EDI Regulatory Reform Discussion Paper, Washington, (1998)

[4] http://www.worldbank.org

[5] http://www.europa.eu.int/ISPO/infosoc/telecompolicy/en/lricmain.pdf

[6] J.J. Laffont and J. Tirole, *Competition in Telecommunications*, Munich Lectures, The MIT Press, 2nd edition, 2001

[7] J.J. Laffont, P. Rey and J. Tirole, "Network Competition: I. Overview and Non discriminatory Pricing; II. Discriminatory Pricing," *RAND Journal of Economics* 29 (1998), pp. 1-56

[8] InfoDev, *Telecommunications Regulation Handbook*, Module 3: "Interconnection ",Table 3-2, The World Bank, Washington, DC ,2000.

TCP Congestion Control and Asymmetric Networks

Georgi V. Hristov[1] and Mihail P. Iliev[2]

***Abstract* – As the number of Internet users grows, the demand for faster networks becomes larger. To meet these demands, some new network technologies are emerging. Some of these technologies provide asymmetric bandwidth for uplink (from the client to the server) and downlink (from the server to the client). These technologies are considered to be suitable for the Internet access. It is because the user's access to the Internet is essentially asymmetric. The user usually retrieves the information from the Internet through WWW (World Wide Web) service or file transfer service. Both HTTP (Hyper Text Transfer Protocol) for WWW service and FTP (File Transfer Protocol) for file transfer service use TCP - the most popular transport layer protocol in the Internet. The problem is that TCP has not been designed for asymmetric networks.**

Keywords – TCP, asymmetric network, simulation, emulation

I. INTRODUCTION

There are various types of asymmetry, which can be related to [1, 2, 6]:

Bandwidth: The bandwidth is different in the two directions (downstream and upstream).

Latency: The latency is different in the two directions. Normally caused by data being sent through two different transmission media. The latency in this case is viewed as the delay of physical medium, i.e., the combination of propagation and processing delay.

Quality: The error rate is different in the two directions, also normally caused by two different transmission media.

We describe some known asymmetric technologies to exemplify different types of asymmetry. A given asymmetric technology is often affected by a combination of different types of asymmetry.

Cable TV networks have wide downstream bandwidth, but often there is no return channel. Some solutions use ISDN modems as return channel, which introduces a significant bandwidth asymmetry. Others use a limited return channel via cable net, which is shared with everyone else.

The technology clearly exhibits bandwidth asymmetry. The technology also exhibits a degree of latency asymmetry, which is dependent on the return channel. If ISDN is used, the two medias differ and thus the latency will probably also differ. The latency of the return channel via cable net is often also affected, due to guard times between transmissions and contention intervals on the shared media.

As the name implies, the **ADSL technology** is based on asymmetric lines, where bandwidth asymmetry is the primary factor. In ADSL the downstream capacity is larger than the upstream capacity.

There is a vast number of **satellite links** with different kinds of feedback channels. Due to the many different types of feedback channels, satellite links can exhibit all types of asymmetry. The two major asymmetries for satellite links are bandwidth and latency asymmetry. The physical distance to the satellite often introduces large propagation latency (roughly 300 ms one-way).

Some **wireless/radio** technologies suffer from media access asymmetry. Especially technologies using a central base station and several mobile clients as this requires a RTC/CTS (Ready To Send/Clear To Send) protocol, with a large turn-around time caused by the radio. The drop rate is also significantly higher in wireless networks, thus some kind of quality asymmetry is also likely to occur.

II. THEORY OF ASYMMETRY AND EFFECT ON TCP

The network model which we use in the analysis and simulation is depicted in Figure 1. The model consists of a server, a client, and two links; downlink from the server to the client and uplink from the client to the server. The uplink and the downlink have asymmetric bandwidth, denoted as μ_f [data segments/s] and μ_r [ACK segments/s], respectively. Note that μ_f and μ_r are represented by units of data/ACK segment, respectively. The buffer sizes are denoted as B_f [data segments] and B_r [ACK segments], and the propagation delays between the server and client are τ_f [s] and τ_r [s]. Whether the upstream capacity is sufficient, depends on the ratio between the downstream and upstream bandwidth divided by the ratio between the size of data packets and ACK packets. For one-way transfers, this is defined as the normalized bandwidth ratio, k:

$$k = \frac{Downstream/Upstream}{Data_{size}/ACK_{size}} = \frac{\mu_f}{\mu_r} \qquad (1)$$

[1]Georgi V. Hristov is with the Department of Communication Technique and Technologies, 8 Studentska Str., 7017 Rousse, Bulgaria, E-mail: ghristov@mbox.contact.bg

[2] Mihail P. Iliev is with the Department of Communication Technique and Technologies, 8 Studentska Str., 7017 Rousse, Bulgaria, E-mail: miliev@ecs.ru.acad.bg

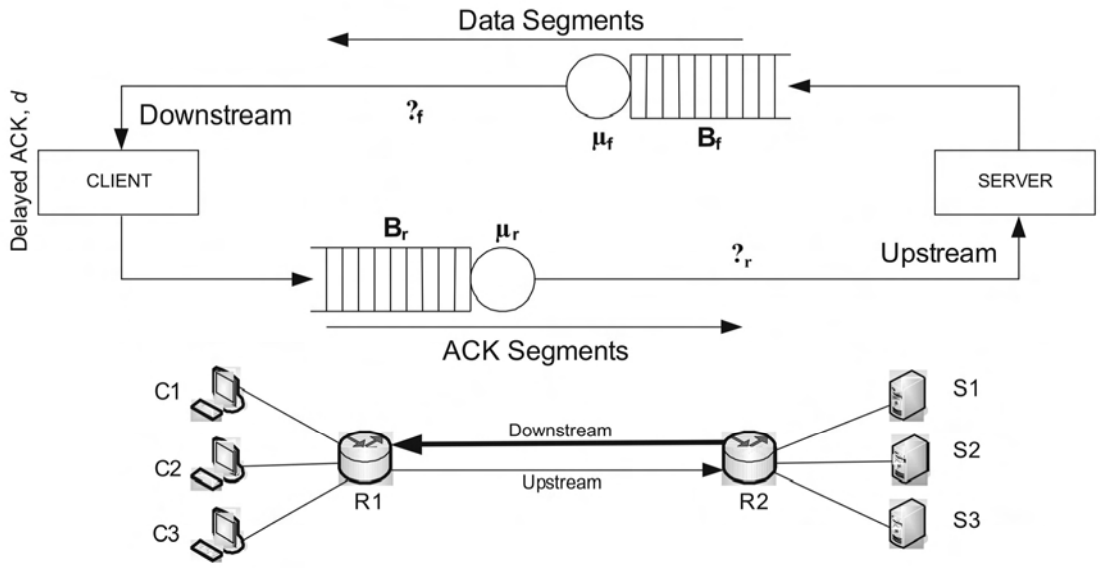


Fig. 1. Network Model

The value ***k*** indicates, how many data packets need to be acknowledged per ACK packet, to avoid saturation of the upstream by ACK packets. In other words, if there is more than one ACK packet for every k data packets, the upstream will get saturated (with ACK packets) before the downstream.

The TCP connection's delayed ACK factor, ***d*** defines the number of TCP data segments acknowledged by an ACK packet. Therefore, the ***k*** value must be evaluated in relation to the ***d*** factor. We can conclude if $k > d$ the upstream will be saturated by ACK packets. The delayed ACK algorithm specifies that TCP should send an ACK packet (at least) every second data packet. This means that 2 is an upper bound on ***d***. Accordingly , it is a minimum requirement that $k < d < 2$ [3].

The above scenario only holds for one-way transfers, but most real-life traffic is bidirectional. The upstream link might be used for data packets, while at the same time delivering ACK packets. The competing reverse traffic consumes a part of the upstream capacity, effectively increasing the degree of bandwidth asymmetry.

Variable latency affects the smoothness of a TCP data flow. TCP measures the Round- Trip Time (RTT) on the path to estimate the path Retransmission Time Out (RTO) (calculated from a smoothed RTT estimate and a linear deviation). In the event of (multiple) packet loss, the RTO needs to be accurate in order to respond in time to avoid unnecessary idle periods.

The line quality and packet losses generally affect TCP and is not only an asymmetric problem. Loss of ACK packets is less significant than loss of data packets, as ACK packets are cumulative (ACK'ing all outstanding data), which generally results in stretched ACKs.

Media access asymmetry can be manifested in several ways (which we will not describe in detail). It is generally defined in literature an uneven access to a shared medium by a distributed set of nodes.

III. Improving TCP performance over asymmetric networks

Several researchers have studied TCP performance on asymmetric networks [1, 2, 4, 6]. To solve the problems caused by bandwidth asymmetry, authors proposed such strategies as ACK Congestion Control (ACC), ACK Filtering (AF), Sender Adaptation (SA), and ACK Reconstruction (AR) to reduce the congestion in the upstream direction and improve TCP performance.

ACK filtering (AF) is a TCP-aware link-layer technique that reduces the number of TCP ACKs sent on the reverse channel. The challenge is to ensure that the sender does not stall waiting for ACKs, which can happen if ACKs are removed indiscriminately on the reverse path. AF removes only certain ACKs without starving the sender by taking advantage of the fact that TCP ACKs are cumulative. As far as the sender's error control mechanism is concerned, the information contained in an ACK with a later sequence number subsumes the information contained in any earlier ACK.

ACK congestion control (ACC) is an alternative to ACK filtering that operates end-to-end rather than at the upstream bottleneck router. The key idea in ACC is to extend congestion control to TCP ACKs, since they do make non-negligible demands on resources at the bandwidth-constrained upstream link. ACKs occupy slots in the reverse channel buffer, which capacity is often limited to a certain number of packets (rather than bytes).

ACC and AF alleviate the problem of congestion on the reverse bottleneck link by decreasing the frequency of ACKs, with each ACK potentially acknowledging several data packets. This can cause problems such as sender burstiness and a slowdown in congestion window growth. Sender

adaptation is an end-to-end technique for alleviating this problem. A bound is placed on the maximum number of packets the sender can transmit back-to-back, even if the window allows the transmission of more data. If necessary, more bursts of data are scheduled for later points in time computed based on the connection's data rate. The data rate is estimated as the ratio cwnd/srtt, where cwnd is the TCP congestion window size and srtt is the smoothed RTT estimate. Thus, large bursts of data get broken up into smaller bursts spread out over time.

ACK reconstruction is a technique to reconstruct the ACK stream after it has traversed the reverse direction bottleneck link. AR is a local technique designed to prevent the reduced ACK frequency from adversely affecting the performance of standard TCP sender implementations. This enables us to use schemes such as ACK filtering or ACK congestion control without requiring TCP senders to be modified to perform sender adaptation. This solution can be easily deployed by Internet Service Providers (ISPs) of asymmetric access technologies in conjunction with AF to achieve good performance.

IV. Analysis and Simulation Results

In this section, we use simulation and emulation to demonstrate the performance of TCP in asymmetric networks.

We expect bidirectional traffic to have a negative impact on the throughput achieved by the downstream transfer as the reverse ACK traffic is competing with the upstream transfer, which effectively increases the k value.

To illustrate and verify that the downstream TCP traffic is affected by upstream traffic a simple test is performed with a single continuous downstream transfer and short upstream transfers. The downstream transfer is allowed to run alone after which an upstream transfer is started. The results is shown in Figure 2.

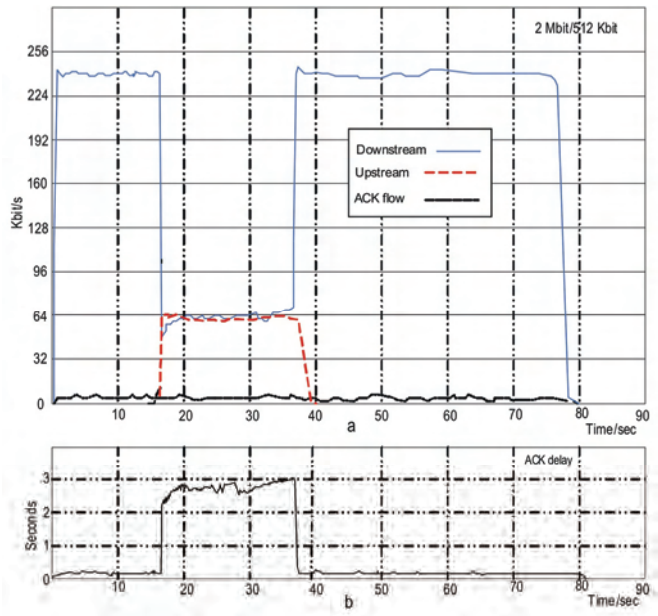


Fig. 2. Illustrating the downstream throughput problem on asymmetric link (a) and ACK delay (b)

In this item, we consider TCP level performance. Figure 3 shows the throughput of TCP as a function of the asymmetry factor of the network, $k = \mu_f / \mu_r$. We change μ_f while fixing μ_r.to determine k . In the figure, we also plot emulation results to confirm accuracy of our simulation.

We can see from this figure that when the asymmetry factor k is small, the throughput increases in proportion to k. Then, it is suddenly decreased and kept constant when the nominal value of k is too big. That is, if k is smaller than the downlink buffer size B_f , segment loss occurs only when the window size exceeds the sum of bandwidth-delay product of the connection and the uplink and downlink buffer sizes. In other words, the throughput increases in proportion to the downlink bandwidth, μ_f. If k is larger than B_f , on the other hand, segment loss takes place when the size of the burst generated by the server exceeds the downlink buffer size. It is independent of μ_f, and therefore the throughput is kept constant.

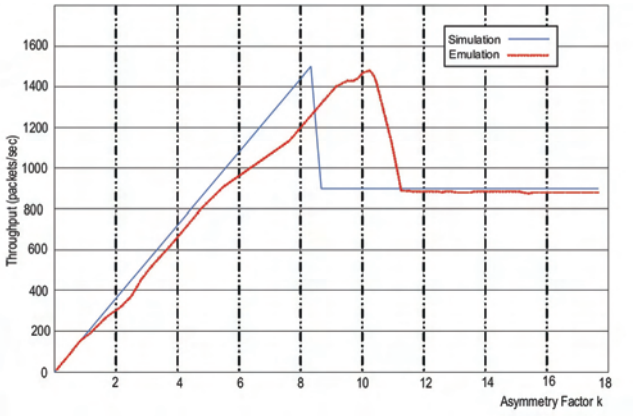


Fig 3. Throughput vs. asymmetry factor k

We now discuss the results of the second set of experiments, designed to study the effects of performance, and to investigate how Reno, AF and ACC alter the inter-ACK spacing and how SA and AR prevent burst transmissions.

AF/AR and AF/SA perform the best, achieving throughputs between 15% and 21% better than Reno TCP. The degree of burstiness reduces significantly and the reverse router queue is no longer perpetually full. This is shown in Figure 4, which depicts the time-evolution of congestion windows for the different schemes. It is clear from these results that reducing the frequency of ACKs alone is not sufficient—AF alone performs even worse than Reno and techniques like SA or AR are essential to achieve good performance.

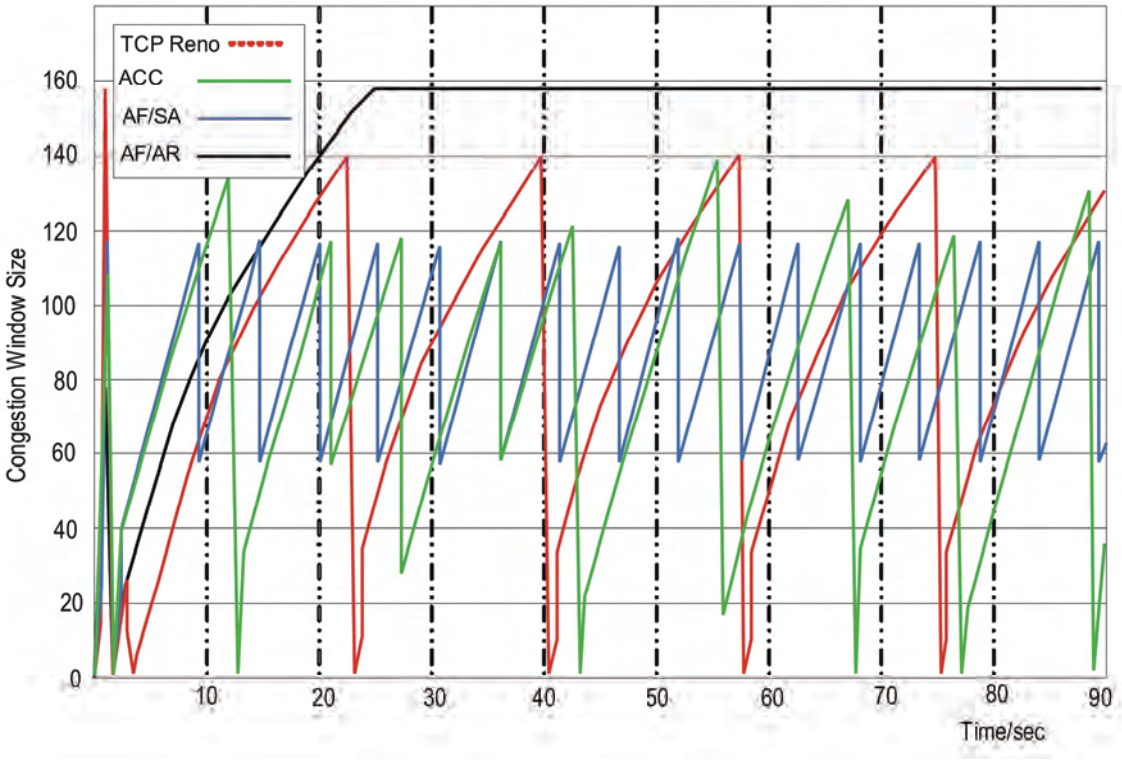


Fig. 4. Congestion Window evolution for the different scheme

V. Conclusion

In asymmetric networks, the slower upstream channel is the primary bottleneck for the TCP throughput over the faster downstream channel. The amount of throughput degradation increases with the normalized asymmetry factor ***k***. For unidirectional data transfer, the *delayed TCP ACK* mechanism helps in reducing the congestion over the slower upstream channel resulting in better TCP throughput over the forward channel. Bi-directional or two-way data transfer further increase the asymmetry problem. AF/AR and AF/SA perform the best, achieving throughputs between 15% and 21% better than standard TCP version. It is clear from these results that reducing the frequency of ACKs alone is not sufficient—AF alone performs even worse than standard TCP version and techniques like SA or AR are essential to achieve good performance.

Acknowledgement

This work is a part of the research project BY-TH-105/2005 of Bulgarian Science Fund at Ministry of Education and Science.

References

[1] Hossam Afifi, Omar Elloumi, Gerardo Rubino, “A Dynamic Delayed Acknowledgement to Improve TCP Performance for Asymmetric Links” Computers and Communications, *ISCC '98*. Proceedings. Third IEEE Symposium on , 1998

[2] H. Balakrishnan, V.N. Padmanabhan and R.H. Katz, “ The effects of TCP/IP performance,” Proceedings of *INFOCOM'99*

[3] T.V. Lakshman “Window-based error recovery and flow control with a slow acknowledgement channel: a study of TCP/IP performance.” *INFOCOM '97*.

[4] Henderson, T.R.; Sahouria, E.; McCarme, S.; Katz, “On Improving the Fairness of TCP Congestion Avoidance” *GLOBECOM'98*. Volume: 1 , 1998 , Page(s): 539 -544

[5] Go Hasegawa, Masayuki Murata, and Hideo Miyahar, “ Fairness and Stability of Congestion Control Mechanism of TCP” *INFOCOM '99*. Volume: 3 , 1999 , Page(s): 1329 -1336 vol.3

[6] T.C. Kwok, “Residential Broadband Architecture over ADSL andG.Lite (G.992.2): PPP over ATM,” *IEEE Communications Magazine*, vol. 37, no. 5, pp.84-89, May 1999.

Probability Density Function of M-ary FSK Signal in the Presence of Impulse Noise and Nakagami Fading

Mihajlo Stefanović[1], Dragana Krstić[2], Nataša Kapacinović[3], Edis Mekić[4] and Fatih Dizdarević[5]

Abstract **- In this paper the receiver for the demodulation of M-FSK signals in the presence of Gaussian noise, impulse noise and Nakagami fading is considered. Probability Density Function of M-ary FSK Signal in the Presence of Impulse Noise and Nakagami Fading is derived. These interferences can seriously degrade the communication systems performance.**

Keywords **-** ***M*****-ary Frequency Shift Keying, Impulse noise, Nakagami Fading, Probability Density function.**

I. INTRODUCTION

In this paper we consider a system for coherent demodulation of *M*-ary FSK signals in the presence of Gaussian noise, impulse noise and variable signal amplitude. These annoyances can seriously degrade the performance of communication systems [1]-[3]. In the paper [4], the performance evaluation of several types of FSK and CPFSK receivers was investigated in detail using the modified moment's method. Also, the error probability of the cross-correlator receiver for binary digital frequency modulation detection is studied using theoretical analysis and computer simulations [5]. In [6] average bit-error probability performance for optimum diversity combining of noncoherent FSK over Rayleigh channels is determined. Performance analysis of wide-band *M*-ary FSK systems in Rayleigh fading channels is given in [7]. The influence of impulse noise in noncoherent *M*-ary digital systems is observed in the paper [8].

In order to view an influence of the Gaussian noise, impulse noise and variable signal amplitude on the performances of an *M*-ary FSK system, we are derived the probability density function of *M*-ary FSK receiver output signal, the joint probability density function of output signal and its derivative, and the joint probability density function of output signal at two time instants. The bit error probability, the signal error probability and outage probability can be determined by the probability density function of output signal. Also, the moment generating function, the cumulative distribution of output signals and the moment and variance of output signals can be derived by probability density function of output signals. An average level crossing rate and an average fade derivation of an output signal process can be calculated by the joint probability density of an output signals and its derivative. The expression for calculation autocorrelation function can be derived by joint probability density function of an output signals at two time instants and by using Winner-Hinchine theorem the spectral power density function of an output *M*-ary FSK signal can be obtain. Based on this, the results obtained in this paper have a great significance.

This paper is organized as follows: first section is introduction. In second section the model of the *M*-ary FSK system is defined. The expressions for the probability density function of the output signal is obtained in the third section. In the next, fourth section, the numerical results in the case $M=2$ are given. The last section is the conclusion.

II. MODEL OF THE M-ARY FSK SYSTEM

The model of an M-ary FSK system, which we consider in this paper, is shown at Fig. 1. This system has *M* branches. Each branch consists of the bandpass filter and correlator. The correlator is consisting of multiplier and lowpass filter.

The signal at the input of the receiver is digital frequently modulated signal corrupted by additive Gaussian noise, impulse noise and Nakagami fading.

Transmitted signal for the hypothesis H_i is:

$$s(t) = A\cos\omega_i t \tag{1}$$

where *A* denotes the amplitude of the modulated signal and has Nakagami distribution.

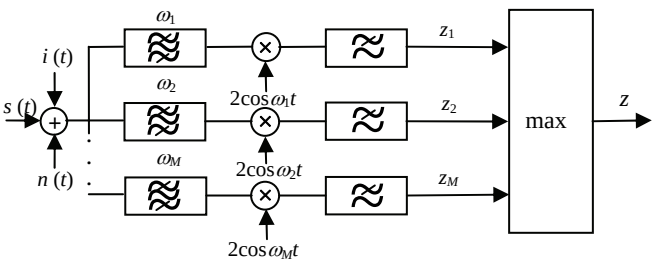


Fig. 1. Block diagram of the system for coherent demodulation of *M*-ary FSK signal

Gaussian noise at the input of the receiver is given with:

[1]Mihajlo Č. Stefanović is with Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: misa@elfak.ni.ac.yu

[2]Dragana Krstić is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail:dragana@elfak.ni.ac.yu

[3]Nataša D. Kapacinović is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia, e-mail: nale@elfak.ni.ac.yu

[4]Edis Mekić is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia.

[5]Fatih Dizdarević is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Nis, Serbia

$$n(t)=\sum_{i=1}^{M} x_i \cos\omega_i t + y_i \sin\omega_i t\,, \quad i=1, 2,...M \tag{2}$$

where x_i and y_i are components of the Gaussian noise, with zero means and variances σ^2.

The pulse interference *i(t)* can be written as:

$$i(t)=\sum_{i=1}^{M} (A_i+cn)\cos(\omega_i t+\theta_i) \tag{3}$$

c is constant and *n* has Poisson's distribution:

$$p(n)=\frac{\lambda^n}{n!}e^{-\lambda} \tag{4}$$

λ is intensity of impulse process. Phases θ_i, i=1,2,...,M have uniform probability density function.

These signals pass first through bandpass filter whose central frequencies ω_1, ω_2,..., ω_M correspond to hypotheses H_1,H_2,...,H_M. Then, after multiplying with signal from the local oscillator, they pass through lowpass filter which cut all spectral components which frequencies are greater than the border frequency of the filter.

If z_1,z_2, …z_M are the output signals of the branch of receiver, then *M*-FSK receiver output signal is:

$$z=\max\{z_1, z_2...z_M\} \tag{5}$$

The probability density of output signal is

$$p_z(z)=\sum_{i=1}^{M} p_{z_i}(z)\cdot\prod_{\substack{j=1\\ j\neq i}}^{M} F_{z_j}(z) \tag{6}$$

III. PROBABILITY DENSITY FUNCTION

In the case of the hypothesis H_1, transmitted signal is:

$$s(t)=A\cos\omega_1 t \tag{7}$$

while the output branch signals of the receiver are:

$$z_1 = A + x_1 + A_1\cos\theta_1 \tag{8}$$

$$z_k = x_k + A_k\cos\theta_k\,, \qquad k=2,3,\ldots,M \tag{9}$$

It is necessary to define the probability density functions on the output of branches and the cumulative density of these signals to obtain output probability density function of *M*-ary FSK receiver.

The conditional probability density functions for the signals z_1, z_2,..., z_M are:

$$p_{z_1/A,\theta_1}(z)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(z-A-(A_1+cn)\cos\theta_1)^2}{2\sigma^2}} \tag{10}$$

σ denotes standard deviation.

$$p_{z_k/A,\theta_k}(z)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(z-(A_k+cn)\cos\theta_k)^2}{2\sigma^2}}\,, \quad k=2,3,\ldots,M \tag{11}$$

By averaging (11) and (12) we obtain the probability density functions of the signals on the output of the branches:

$$p_{z_1}(z)=\int_0^{\infty}\int_{-\pi}^{\pi}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(z-A-(A_1+cn)\cos\theta_1)^2}{2\sigma^2}}\cdot$$

$$\cdot\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\frac{2}{\Gamma(m_1)}\left(\frac{m_1}{\Omega_1}\right)A^{2m_1-1}e^{-\frac{m_1}{\Omega_1}A^2}dA\frac{1}{2\pi}d\theta_1 \tag{12}$$

$$p_{z_k}(z)=\int_0^{\infty}\int_{-\pi}^{\pi}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(z-(A_k+cn)\cos\theta_k)^2}{2\sigma^2}}\cdot$$

$$\cdot\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\frac{2}{\Gamma(m_k)}\left(\frac{m_k}{\Omega_k}\right)A^{2m_k-1}e^{-\frac{m_k}{\Omega_k}A^2}dA\frac{1}{2\pi}d\theta_k \tag{13}$$

The cumulative distributions of the signals z_1, z_2,..., z_M are:

$$F_{z_1}(z)=\int_{-\infty}^{z}\int_0^{\infty}\int_{-\pi}^{\pi}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(r-A-(A_1+cn)\cos\theta_1)^2}{2\sigma^2}}\cdot$$

$$\cdot\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\frac{2}{\Gamma(m_1)}\left(\frac{m_1}{\Omega_1}\right)A^{2m_1-1}e^{-\frac{m_1}{\Omega_1}A^2}dA\frac{1}{2\pi}d\theta_1 dr \tag{14}$$

$$F_{z_k}(z)=\int_{-\infty}^{z}\int_0^{\infty}\int_{-\pi}^{\pi}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(r-(A_k+cn)\cos\theta_k)^2}{2\sigma^2}}\cdot$$

$$\cdot\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\frac{2}{\Gamma(m_k)}\left(\frac{m_k}{\Omega_k}\right)A^{2m_k-1}e^{-\frac{m_k}{\Omega_k}A^2}dA\frac{1}{2\pi}d\theta_k dr \tag{15}$$

The probability density function of the *M*-ary FSK receiver output signal in the case of the hypothesis H_1 can be obtained from:

$$p_z(z)=\sum_{i=1}^{M} p_i(z)\cdot\prod_{\substack{j=1\\ j\neq i}}^{M} F_j(z) \tag{16}$$

IV. NUMERICAL RESULTATS

We now consider the dual branch FSK receiver because of its easy implementation and very good performances. It is employed in many practical telecommunication systems.

The probability density function, in case of dual branch, has form:

$$p_z(z)=p_{z_1}(z)\cdot F_{z_2}(z)+p_{z_2}(z)\cdot F_{z_1}(z) \tag{17}$$

The probability density functions $p(z)$ versus output signal *z*, with and without interference and impulse noise, for various

values of fading severity parameter m and average power Ω are shown at Figs. 2 and 3, respectively.

At Figs. 4. to 6. probability density functions $p(z)$ versus output signal z, are given depends on parameter λ.

The probability density functions $p(z)$ versus output signal z, when dependence is on standard deviation σ.

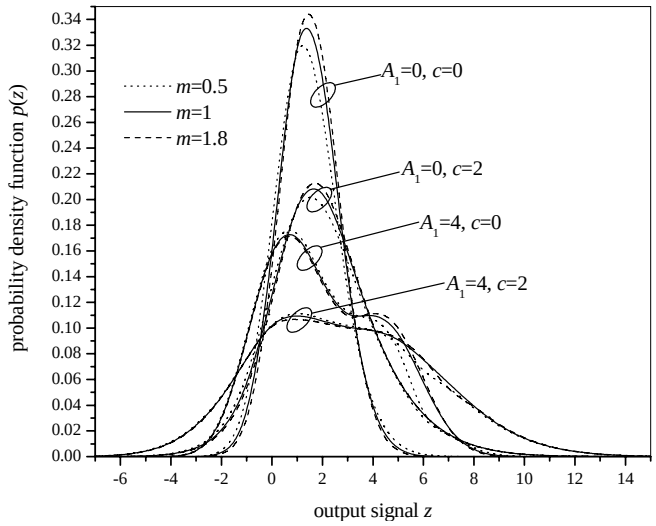


Fig. 2. The probability density functions $p(z)$ for the parameters σ=1, λ=1 and Ω=2

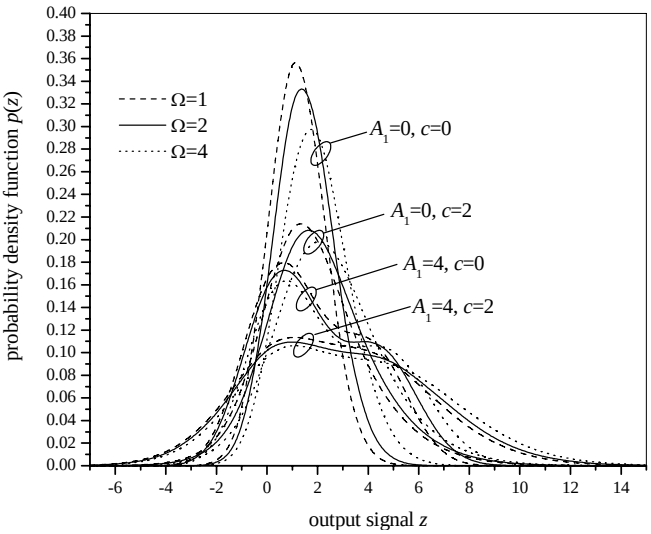


Fig. 3. The probability density functions $p(z)$ for the parameters σ=1, λ=1 and m=1

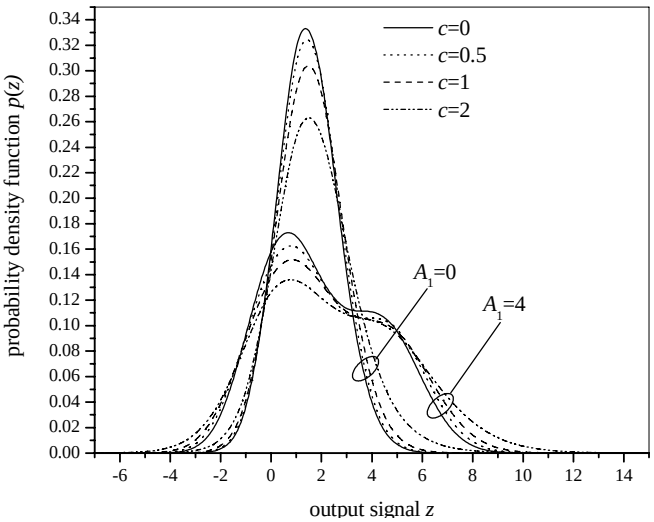


Fig. 4. The probability density functions p(z) for the parameters σ=1, λ=0.5, Ω=2 and m=1

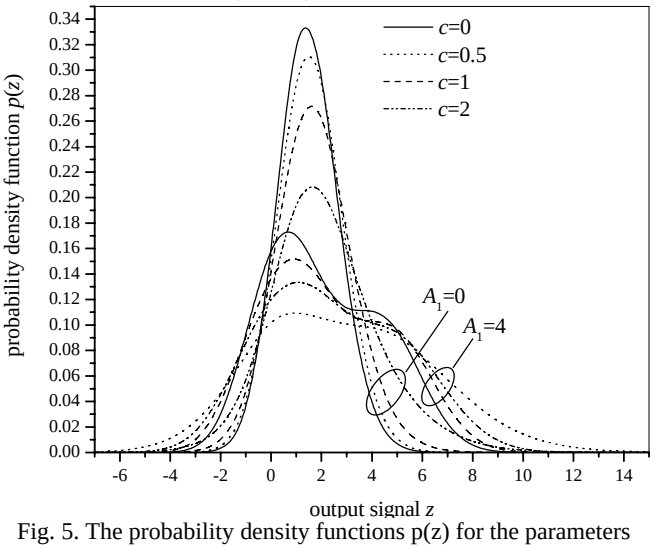


Fig. 5. The probability density functions p(z) for the parameters σ=1, λ=1, Ω=2 and m=1

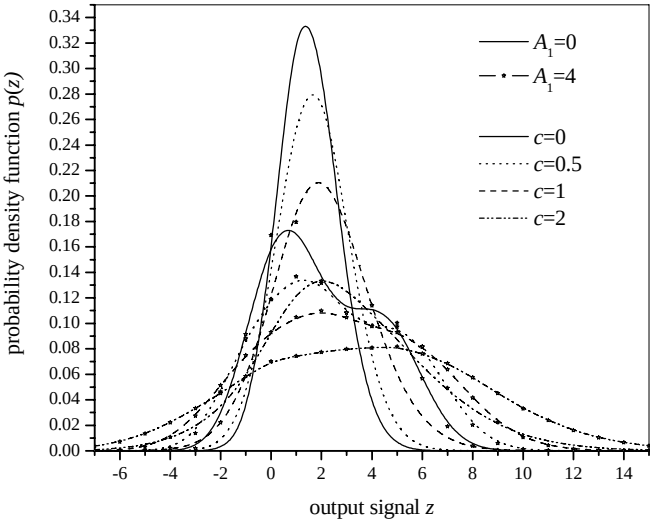


Fig. 6. The probability density functions p(z) for the parameters σ=1, λ=2, Ω=2 and m=1

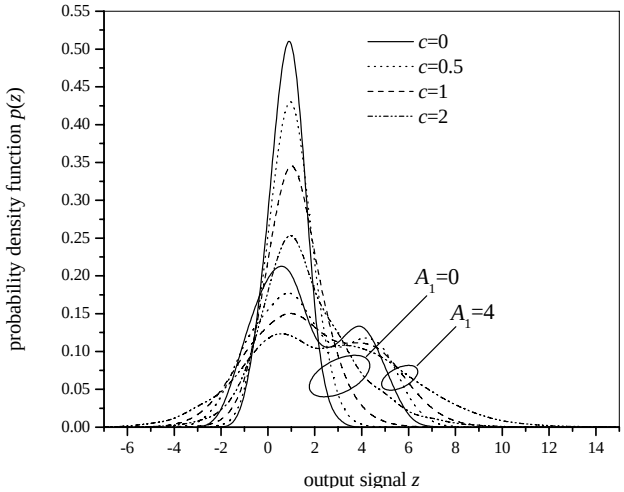


Fig. 7. The probability density functions p(z) for the parameters σ=0.5, λ=1,Ω=2 and m=1

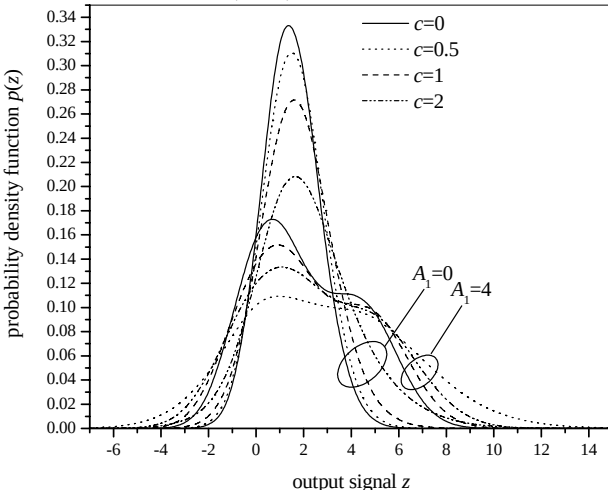


Fig. 8. The probability density functions p(z) for the parameters σ=1, λ=1, Ω=2 and m=1

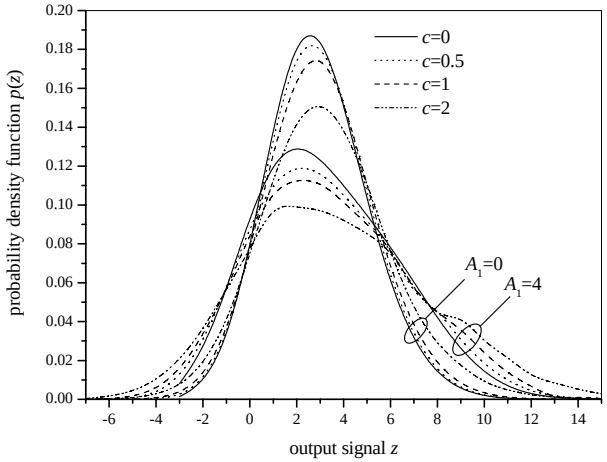


Fig. 9. The probability density functions p(z) for the parameters σ=2, λ=1, Ω=2 and m=1

V. CONCLUSION

The statistical characteristics of signal at the output of receiver for coherent FSK demodulation are derived in this paper. The input signal of the receiver is digital frequently modulated signal corrupted by additive Gaussian noise, impulse noise and Nakagami fading. The interferences appear in each receiver branch. In this paper the probability density function of *M*-ary FSK receiver output signal is derived. The bit error probability and the outage probability can be determined by the probability density function of output signal.

REFERENCES

[1] T. Okoshi and K. Kikuchi, *Coherent Optical Fiber Communications*, Tokyo, Japan: KTW Scientific, 1988.

[2] J. M. Senior, *Optical Fiber Communications*, New York: Prentice Hall, 1992.

[3] J. G. Proakis, *Digital Communications 2nd ed*. New York, McGraw-Hill, 1989.

[4] H. Miin-Jong and B. Stephen Wicker, "Performance evaluation of FSK and CPFSK optical communication systems: a stable and accurate method", J. Lightwave Technol., vol. 13, no. 8, Aug. 1995.

[5] K. A. Farrell and P. J. McLane, "Performance of the cross-correlator receiver for binary digital frequency modulation", IEEE Trans. Commun., vol. 45, no 5, May 1997.

[6] M..K Simon, and M. Alouini, "Average Bitt-Error Probability Performance for Optimum Diversity Combining of Noncoherent FSK Over Rayleigh channels", ", IEEE Trans. on Commun., vol. 51, no 4, pp.566-569, April 2003.

[7] Y. K. Jeong, K. B. Lee, "Performance analysis of wide-band M-ary FSK systems in Rayleigh fading channels", IEEE Trans. on Commun., vol.48, Issue 12, pp. 1983–1986, Dec. 2000.

[8] H. Huynh, M. Lecours, "Impulsive Noise in Noncoherent M-ary Digital Systems", IEEE Trans. on Commun., vol. 23, Issue 2, pp. 246 – 25, Feb 1975.

Performances of Receiver with MRC, SC and SSC Combiner in the Presence of Fading and Gaussian Noise

Mihajlo Stefanović, Jelena Anastasov, Bojana Palović, Ilija Temelkovski and Vladimir Veličković

Abstract **- In this paper work we consider statistical characteristics of signal at the output of combiner with two diversity branches in channel with Nakagami-m fading. We analyze**
MRC, SC i SSC combining. Fading is identical and independent.

I. INTRODUCTION

In this work we consider diversity system with two branches and predetection combining. Fading appears due to conveying signals over multiple paths. Amplitudes of signals over the branches are independent. Probability densities of the amplitudes are Nakagami-m with same parametres. In this paper is analyzed diversity system in presence of Nakagami-m fading which is not correlated between branches. At first we consider signal at output of MRC combiner. Input signals are in phase. Their squires of signal/noise ratios are collected and this sum is signal at output of MRC combiner. Probability density function of signal at output is combination of probability density functions of signals at its inputs.

Characteristic of signal function at output is product of characteristics of signal functions at inputs of MRC combiner. Based on probability density of signals at output, average value and variance of signals at output of combiner can be calculated. Average value and variance of output signal can be obtained by characteristic functions of signals at output of combiner. The output signal of SC combiner is equal to the value of its input which is higher. There is probability that any of input signals is higher than other. If the signal from the first input is higher than signal from the second, then the probability density function of output signal is equal to the probability density function of signal from first input multiplicated by the probability that the signal at second input is lower than the signal at first input. At this way the probability density function at the output of SC combiner is determinated. Characteristic function of output signal and its momentums can be also determinated.

If we analyze SSC combinig we consider that combiner first examines signal from one input. If this signal is higher than some beforehand determinated threshold then it is forwarded to circuit for demodulation. If this signal is lower than threshold than the signal from other branch is forwarded to circuit for demodulation. For example, in this paper, is determinated probability density function of signal at output of combiner. For calculating these probabilities we used Gilbert model.

Faculty of Electronic Engineering, Niš, Serbia
Email: anas-j@bankerinter.net; jobo@medianis.ne

II. MRC COMBINER

Model MRC is presented on the figure 1. Input signals are r_1 i r_2, and the output signal is r. The squares of signals r_1 i r_2 are $\gamma_1 = r_1^2$ i $\gamma_2 = r_2^2$.
Signal power at the output of combiner is γ:
$\gamma = \gamma_1 + \gamma_2$ ili $\gamma_1 = \gamma - \gamma_2$.

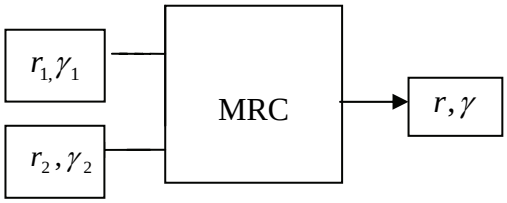


Fig.1. Model of MRC combiner

Conditional probability function of the signal γ at the output of the combiner is:

$$p(\gamma/\gamma_1) = \left|\frac{d\gamma_1}{d\gamma}\right| p_{\gamma_1}(\gamma - \gamma_2) \qquad (1)$$

The probability density function (PDF) of the signal can be acquired by averaging the previous function:

$$p_\gamma(\gamma) = \int_0^\gamma p_{\gamma_1}(\gamma - \gamma_2) p_{\gamma_2}(\gamma_2) d\gamma_2 \qquad (2)$$

Because $\frac{dr_1}{dr} = 1$, when the Nakagami-m fading is present, the probability density function of the amplitude of signals at the input of combiner are equal.

$$p_{r_1}(r_1) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m r_1^{2m-1} e^{-\frac{m}{\Omega} r_1^2} \quad (3)$$

$$p_{r_2}(r_2) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m r_2^{2m-1} e^{-\frac{m}{\Omega} r_2^2} \quad (4)$$

We discuss the case when the fading between the branches is independent and identical. Parameter m can take the following values (0.5, ... ∞). For $m=1$ Nakagami-m fading becomes Rayleigh fading, and for $m=\frac{1}{2}$ Nakagami-m feding becomes Gaussian fading. For $m\to\infty$ there is no fading. According to: $\gamma_1 = r_1^{\,2}$, $r_1=\sqrt{\gamma_1}$, $\frac{dr_1}{d\gamma_1}=\frac{1}{2\sqrt{\gamma_1}}$, $r_1\ge 0$, $\gamma_1\ge 0$, we acquire:

$$p_{\gamma_1}(\gamma_1)=\frac{1}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m {\gamma_1}^{m-1} e^{-\frac{m}{\Omega}\gamma_1} \quad (5)$$

At the same way we acquire:

$$p_{\gamma_2}(\gamma_2)=\frac{1}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m {\gamma_2}^{m-1} e^{-\frac{m}{\Omega}\gamma_2} \quad (6)$$

Probability density function of the signal γ is:

$$p_\gamma(\gamma)=\int_0^\gamma \frac{1}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m (\gamma-\gamma_2)^{m-1} e^{-\frac{m}{\Omega}(\gamma-\gamma_2)} \frac{1}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m {\gamma_2}^{m-1} e^{-\frac{m}{\Omega}\gamma_2} d\gamma_2 = \frac{1}{\Gamma^2(m)}\left(\frac{m}{\Omega}\right)^{2m} e^{-\frac{m}{\Omega}\gamma}\gamma^{2m-1}\sum_{i=0}^{m-1}\binom{m-1}{i}\frac{1}{i+m} \quad (7)$$

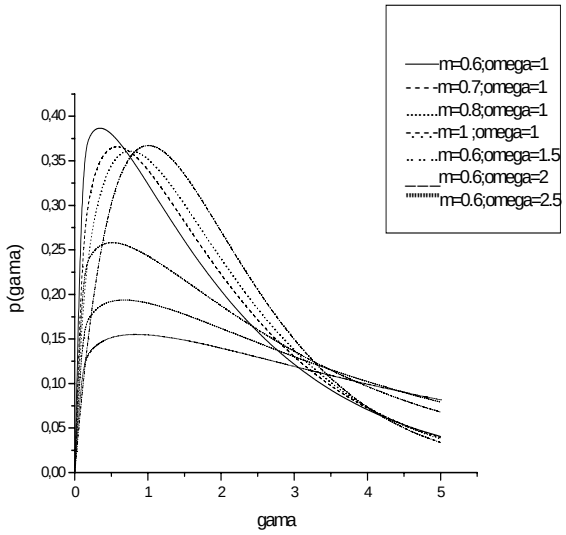


Fig.2.Probability density function of the signal γ

Characteristic function of the signal at the output of the combiner is :

$$M_\gamma(s)=\int_0^\infty e^{\gamma s} p_\gamma(\gamma)d\gamma = \int_0^\gamma e^{\gamma s}\frac{1}{\Gamma^2(m)}\left(\frac{m}{\Omega}\right)^{2m} e^{-\frac{m}{\Omega}\gamma}\gamma^{2m-1} \sum_{i=0}^{m-1}\binom{m-1}{i}\frac{1}{i+m} = \frac{1}{\Gamma^2(m)}\left(\frac{m}{\Omega}\right)^{3m} \sum_{i=0}^{m-1}\binom{m-1}{i}\frac{1}{i+m}\frac{\Omega^{2m}}{(m-\Omega s)^{2m}}\Gamma(2m) \quad (8)$$

Momentum of n-order of the signal at the output of combiner is :

$$m_n=\gamma^n=\int_0^\infty \gamma^n p_\gamma(\gamma)d\gamma = \int_0^\infty \gamma^n \frac{1}{\Gamma^2(m)}\left(\frac{m}{\Omega}\right)^{2m} \gamma^{2m-1} e^{-\frac{m}{\Omega}\gamma}\sum_{i=0}^{m-1}\binom{m-1}{i}\frac{1}{i+m}d\gamma = \frac{1}{\Gamma^2(m)}\left(\frac{m}{\Omega}\right)^{2m}\sum_{i=0}^{m-1}\binom{m-1}{i}\frac{1}{i+m} \left(\frac{m}{\Omega}\right)^{n+2m}\Gamma(n+2m) \quad (9)$$

For n=1 the average value of the signal at the output of MRC combiner is determinated like:

$$m_1=\frac{1}{\Gamma^2(m)}\sum_{i=0}^{m-1}\frac{1}{i+m}\left(\frac{\Omega}{m}\right)^{2m}\Gamma(2m+1)\left(\frac{\Omega}{m}\right) \quad (10)$$

For n=2 mean squared value is acquired:

$$m_2=\frac{1}{\Gamma^2(m)}\sum_{i=0}^{m-1}\binom{m-1}{i}\frac{1}{i+m}\Gamma(2m+1)\left(\frac{\Omega}{m}\right)^2 \quad (11)$$

Cumulative probability function of the output signal is:

$$F_\gamma(\gamma)=\int_0^\gamma p_\gamma(x)dx = \int_0^\gamma \frac{1}{\Gamma^2(m)}\left(\frac{m}{\Omega}\right)^{2m} x^{2m-1} e^{-\frac{m}{\Omega}x}\sum_{i-0}^{m-1}\binom{m-1}{i}\frac{1}{i+m}dx = \left(\frac{m}{\Omega}\right)^{2m}\Gamma\left(2m,\frac{m}{\Omega}\gamma\right)\frac{1}{\Gamma^2(m)} \sum_{i=0}^{m-1}\binom{m-1}{i}\left(\frac{1}{i+m}\right), \quad (12)$$

where the function $\Gamma\left(2m, \frac{m}{\Omega}\gamma\right)$ is incomplete gamma function. The probability density function of the signal at the output of the combiner can be used for evaluating the probability of the signal by the formula :

$$P_0 = \int_0^{\gamma_t} p = F_\gamma(\gamma_t) = \frac{1}{\Gamma^2(m)} \sum_{i=0}^{m-1} \frac{1}{i+m} \Gamma\left(2m, \frac{m}{\Omega}\gamma_t\right) \quad (13)$$

III. SC COMBINER

The output signal of SC combiner is determinated like maximum of its input signals:

$$\gamma = \max(\gamma_1, \gamma_2)$$

When the signal γ_1 is higher than the signal γ_2 than the probability density function of the signal γ is determinated like production of its probability density functions. When the signal γ_2 is higher than the signal γ_1 then the probability density function of the signal γ is counted like production of probability density function (PDF) of the signal γ_2 and the probability that the signal γ_1 is lower than the signal γ_2. According to this:

$$\begin{aligned} p_\gamma(\gamma) &= p_{\gamma_1}(\gamma) P(\gamma_2 < \gamma_1) + \\ & p_{\gamma_2}(\gamma) P(\gamma_1 < \gamma_2) = \\ &= p_{\gamma_2}(\gamma) F_{\gamma_2}(\gamma) + p_{\gamma_1}(\gamma) F_{\gamma_1}(\gamma) \end{aligned} \quad (14)$$

Cumulative probability function of the signal γ_1 is:

$$F_{\gamma_1}(\gamma) = \int_0^{\gamma_1} p_{\gamma_1}(x)dx = \int_0^{\gamma_1} \frac{1}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m x^{m-1} e^{-\frac{m}{\Omega}x} dx = \frac{1}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m \left(\frac{\Omega}{m}\right)^m \Gamma\left(m, \frac{m}{\Omega}\gamma_1\right) \quad (15)$$

At the same way, the cumulative probability function of the signal γ_2 is acquired:

$$F_{\gamma_2}(\gamma_2) = \int_0^{\gamma_2} p_{\gamma_2}(x)dx = \frac{1}{\Gamma(m)} \Gamma\left(m, \frac{m}{\Omega}\gamma_2\right) \quad (16)$$

According to all of this the probability density function of the output signal from SC combiner is:

$$p_\gamma(\gamma) = \frac{2}{\Gamma^2(m)} \left(\frac{m}{\Omega}\right)^m \gamma^{2m-1} e^{-\frac{m}{\Omega}\gamma} \Gamma\left(m, \frac{m}{\Omega}\gamma\right) \quad (17)$$

Joint cumulative probability of the input signals γ_1 and γ_2 is:

$$F_{\gamma_1\gamma_2}(\gamma_1\gamma_2) = F_{\gamma_1}(\gamma_1) F_{\gamma_2}(\gamma_2) \quad (18)$$

Cumulative probability function of the signal γ at the output of the SC combiner is :

$$F_\gamma(\gamma) = F_{\gamma_1\gamma_2}(\gamma_1\gamma_2) = F_{\gamma_1}(\gamma_1) F_{\gamma_2}(\gamma_2) = \frac{1}{\Gamma^2(m)} \Gamma^2\left(m, \frac{m}{\Omega}\gamma\right) \quad (19)$$

Cumulative probability function of the output signal γ can be also accomplished at the following way :

$$p_\gamma(\gamma) = \frac{dF_\gamma(\gamma)}{d\gamma} = \frac{d\left(F_{\gamma_1}(\gamma) F_{\gamma_2}(\gamma)\right)}{d\gamma} = p_{\gamma_1}(\gamma) F_{\gamma_2}(\gamma) + p_{\gamma_2}(\gamma) F_{\gamma_1}(\gamma) \quad (20)$$

IV. SSC COMBINER

At first, the signal at one branch is examined, analyzing SSC combiner. If this signal is higher than some beforehand determinated threshold then it is forwarded to circuit for demodulation of signal. If this signal is lower than threshold then the signal from other branch is forwarded to circuit for demodulation. Probability p_1 is the probability that the signal from first branch is firstly examined and p_2 is probability that the signal from second branch is firstly examined. Let assume $\gamma < \gamma_T$. If the signal from the first branch is firstly examined than the signal from the second branch is forwarded to circuit for demodulation because of the given condition $\gamma < \gamma_T$. Based on this:

$$p_\gamma(\gamma) = p_1 p_{\gamma_2}(\gamma) + p_2 p_{\gamma_1}(\gamma) \quad (21)$$

If the condition $\gamma > \gamma_T$ is filled at least in one of two branches then the signal is higher than γ_T . The signal from the first branch is firstly examined. If the signal is higher than γ_T then it is forwarded to the circuit for demodulation. If the signal is lower than γ_T then the signal from the second branch is forwarded to the circuit for demodulation because of the condition $\gamma > \gamma_T$. The same thing is done when the signal from the second branch is firstly examined. Based on this the probability density function at the output of SC combiner, according to $\gamma > \gamma_T$, is determinated:

$$\begin{aligned} p_\gamma(\gamma) &= p_1 p_{\gamma_1}(\gamma) + p_1 p_{\gamma_2}(\gamma) F_{\gamma_1}(\gamma_T) + \\ & p_2 p_{\gamma_2}(\gamma) + p_2 p_{\gamma_1}(\gamma) F_{\gamma_2}(\gamma_T) \end{aligned} \quad (22)$$

Moreover :

$$p_\gamma(\gamma) = p_1 \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)\gamma^{m-1}e^{-\frac{m}{\Omega}\gamma} + p_2 \frac{2}{\Gamma(m)}\gamma^{m-1}e^{-\frac{m}{\Omega}\gamma}$$
$$\frac{1}{\Gamma(m)}\Gamma\left(m, \frac{m}{\Omega}\gamma_T\right)$$
$$+ p_2 \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)\gamma^{m-1}e^{-\frac{m}{\Omega}\gamma} + p_1 \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)\gamma^{m-1}e^{-\frac{m}{\Omega}\gamma} \quad (23)$$
$$\Gamma\left(m, \frac{m}{\Omega}\gamma_T\right) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)\gamma^{m-1}e^{-\frac{m}{\Omega}\gamma}(p_1 + p_2)$$
$$\left[1 + \Gamma\left(m, \frac{m}{\Omega}\gamma_T\right)\right]$$

Probabilities p_1 i p_2 are determinated by the Gilbert's model which is presented on figure 3:

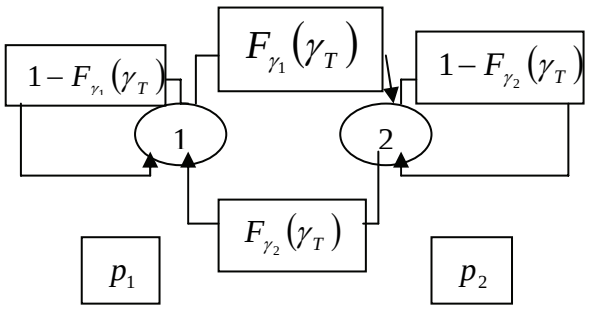


Fig.3. Gilbert's model

Equations of conditions for this model are :

$$p_1 = p_1\left[1 - F_{\gamma_1}(\gamma_T)\right] + p_2 F_{\gamma_2}(\gamma_T)$$
$$p_2 = p_1 F_{\gamma_1}(\gamma_T) + p_1\left[1 - F_{\gamma_2}(\gamma_T)\right] \quad (24)$$
$$p_1 + p_2 = 1$$

Evaluating these equations it can be acquired:

$$p_1 = \frac{F_{\gamma_2}(\gamma_T)}{F_{\gamma_1}(\gamma_T) + F_{\gamma_2}(\gamma_T)}$$
$$p_2 = \frac{F_{\gamma_1}(\gamma_T)}{F_{\gamma_1}(\gamma_T) + F_{\gamma_2}(\gamma_T)} \quad (25)$$

V. CONCLUSION

In this paper we consider diversity system with two branches and predetection combining. There is the Nakagami-m fading at input branches of the combiner. The spatial diversity system is used for reducing the influence of the fading on system performances. There are several ways of combining. The amplitude of signals can be collected, but also signal to noise ratios. In both cases it is needed to bring signals in phase at input. Because of this, these two ways of combining are expensive for realization. The process of combining where the signal to noise ratios are collected is the optimal. There is another process of combining, where the signal on input which is minimal is forwarded to the circuit for determination. The process isn't the optimal one but is simple for realization. The process of combining, where the signal from the input is examined and if it is higher than the threshold then it is forwarded to the circuit for demodulation and if it is lower then the other one is forwarded to the circuit, is the simplest for realization. In this paper are also determinated statistical characteristic of output signal of MRC, SC and SSC combiner. The most important is probability density function which is used to determinate bit error rate and state probability of the signal.

REFERENCES

[1] J.G.Proakis, "On the probability of error for multichannel reception of binary signals", IEEE Trans, Commun.Technol.vol.COM-16.Feb.1968.

[2] W.C.Lindsey, "Error probabilities for Ricean fading multichannel reception of binary and N-ary signals", IEEE Trans.Inform.Tlecry, vol.IT-10.Oct 1964.

[3] J.F.Wend and S.H.Leung, "Analysis of DPSK with equal gain combining in Nakagami fading channel", Electron.Lell.vol.33,Apr.1997.

[4] J.G.Proakis, Digital Communications, New York;Mc Graw-Hill.1989.

Evaluation of a Congestion Control Transport Protocol for Real Time Applications in NGN

Stanimir I. Statev[1] and Seferin T. Mirtchev[2]

***Abstract*: In recent years, real-time applications, such as video streaming, IP telephone, TV conference, and network game, become popular rapidly by increasing speed of the network, or the rising demand for multimedia applications in Next Generation Network (NGN) and IP networks. Either UDP (User Datagram Protocol) or TCP (Transmission Control Protocol) has been used as a transport layer protocol for real-time applications. TCP is not well suited for these applications, since reliable in-order delivery and congestion control can cause arbitrarily long delays. UDP avoids long delays, but UDP applications that implement congestion control must do so on their own. This paper deals with the Datagram Congestion Control Protocol (DCCP): a new transport protocol currently being standardized by IETF that provides a congestion-controlled flow of unreliable datagrams. In this paper the DCCP is described and is its main parameters are examined and compared with TCP and UDP.**

***Keywords:* DCCP, congestion control, transport protocols, NGN, UDP, TCP**

I. INTRODUCTION

In recent years, real-time applications, such as video streaming, IP telephone, TV conference, and network game, become popular rapidly by increasing speed of the network, or the rising demand for multimedia applications in Next Generation Network (NGN) and IP networks [8]. Generally, either UDP (User Datagram Protocol) or TCP (Transmission Control Protocol) has been used as a transport layer protocol for real-time applications. Since the Internet is a best-effort network where multiple users share the network bandwidth, all network applications need to have a mechanism for adapting to congestion status of the network. However, UDP is simply a protocol for datagram transfer, and does not have a mechanism for controlling network congestion. Hence, when a real-time application uses UDP as a transport layer protocol, it is necessary for the application to implement a certain congestion control mechanism at an application layer for preventing congestion collapse of the network [3]. TCP has a mechanism for adjusting the packet transmission rate according to the available bandwidth of the network by performing congestion control between source and destination hosts.

However, TCP is a transport layer protocol originally designed for data transfer applications that can tolerate a certain amount of transmission delay.

[1] Stanimir Statev is with the Faculty of Communication Technology, bul. Kl. Ohridsky 8, 1000 Sofia, Bulgaria, s_statev@tu-sofia.bg

[2] Seferin T. Mirtchev is with the Faculty of Communication Technology, bul. Kl. Ohridsky 8, 1000 Sofia, Bulgaria, stm@tu-sofia.bg

Since the congestion control mechanism of TCP is the AIMD (Additive Increase and Multiplicative Decrease) window flow control, the packet transmission rate from a source host fluctuates at the time scale of approximately round-trip time [9]. Although such fluctuation is not a problem when using TCP with nonreal time applications such as data transfer applications, it becomes a serious problem in real-time applications such as video streaming.

Fast-growing Internet applications like streaming media and telephony prefer timeliness to reliability, making TCP a poor fit. Unfortunately, UDP, the natural alternative, lacks congestion control. High-bandwidth UDP applications must implement congestion control themselves—a difficult task—or risk rendering congested networks unusable. The outcome, the Datagram Congestion Control Protocol or DCCP, adds to a UDP-like foundation the minimum mechanisms necessary to support congestion control. The resulting protocol sheds light on how congestion control interacts with unreliable transport, how modern network constraints impact protocol design, and how TCP's reliable byte stream semantics intertwine with its other mechanisms, including congestion control.

The Datagram Congestion Control Protocol is a new transport protocol currently being standardized by IETF that provides a congestion-controlled flow of unreliable datagrams. Recent years have seen dramatic increases of the use of multimedia applications on the Internet. Delay-sensitive applications, such as streaming media and video telephony, typically prefer timeliness to reliability. These applications have historically used UDP for transport and implemented their own congestion control mechanisms - a difficult task, or no congestion control at all. This can easily cause compatibility problems or congestion collapse in the networks. DCCP makes it easy to deploy these applications without risking congestion collapse. It aims to add to a UDP-like foundation the minimum mechanisms necessary to support congestion control. In short, DCCP is an unreliable transport protocol like UDP, but it has congestion control like TCP. The protocol can be extended by adding new congestion control algorithms, e.g. TCP-like congestion control or TCP-Friendly Rate Control (TFRC) as profiles, in order to customize the congestion control for applications with different characteristics. It is highly promising that DCCP will become the de facto standard for multimedia rich content delivery over IP-based networks. On the other hand, as DCCP provides many features and options to users, and has been criticized for the resulting complexity.

The main task in this paper is to describe the DCCP protocol, to examine its main parameters and characteristics and to compare with TCP and UDP.

II. DCCP Main Functions and Parameters

The Datagram Congestion Control Protocol (DCCP) is a transport protocol that implements bidirectional, unicast connections of congestion-controlled, unreliable datagrams. Connections start and end with three-way handshakes. A choice of modular congestion control mechanisms is available. Two mechanisms are currently specified: TCP-like Congestion Control [7] and TCP-Friendly Rate Control (TFRC) [5]. DCCP is easily extensible to further forms of unicast congestion control. Specifically, DCCP provides the following:

- Unreliable flows of datagrams.
- Reliable handshakes for connection setup and teardown.
- Reliable negotiation of options, including negotiation of a suitable congestion control mechanism.
- Mechanisms allowing servers to avoid holding state for unacknowledged connection attempts and already-finished connections.
- Congestion control incorporating Explicit Congestion Notification (ECN) and the ECN Nonce [4].
- Acknowledgement mechanisms communicating packet loss and ECN information. Acks are transmitted as reliably as the relevant congestion control mechanism requires, possibly completely reliably.
- Optional mechanisms that tell the sending application, with high reliability, which data packets reached the receiver, and whether those packets were ECN marked, corrupted, or dropped in the receive buffer.
- Path Maximum Transmission Unit (PMTU) discovery [4].

DCCP datagram begin with the 16-byte generic header is shown in Figure 1.

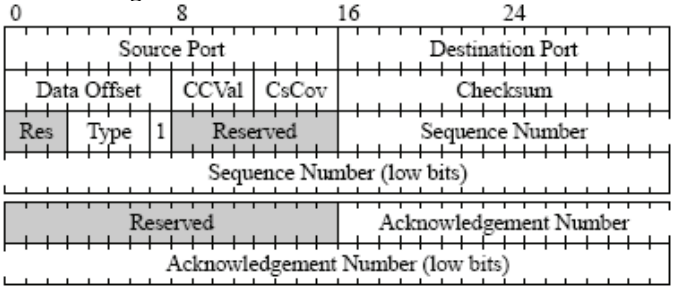


Fig.1. DCCP datagram

The Port fields resemble those in TCP and UDP. Data Offset measures the offset, to the start of packet data. Since this field is 8 bits long, a DCCP header can contain more than 1000 bytes of option. The Type field gives the type of packet, and is somewhat analogous to parts of the TCP flags field. The names in Figure 1 correspond to packet types, of which DCCP specifies ten. Many packet types require additional information after the generic header, but before options begin; this design choice avoids cluttering the universal header with infrequently-used fields. Even the acknowledgement number is optional, potentially reducing header overhead for unidirectional flows of data. CsCov - Checksum Coverage determines the parts of the packet that are covered by the Checksum field. Sequence and acknowledgement numbers are 48 bits long, although some packet types also permit a compact form to be used. CCVal is used for CCID assignment.

Every DCCP packet starts with a fixed-size generic header. Particular packet types include additional fixed-size header data; for example, DCCP-Ack include an Acknowledgement Number, DCCP options and any application data follow the fixed-size header. The packet types are as follows:

DCCP-Request - Sent by the client to initiate a connection (the first part of the three-way initiation handshake).

DCCP-Response - Sent by the server in response to a DCCP-Request (the second part of the three-way initiation handshake).

DCCP-Data - Used to transmit application data.

DCCP-Ack - Used to transmit pure acknowledgements.

DCCP-DataAck - Used to transmit application data with piggybacked acknowledgement information.

DCCP-CloseReq - Sent by the server to request that the client close the connection.

DCCP-Close - Used by the client or the server to close the connection; elicits a DCCP-Reset in response.

DCCP-Reset - Used to terminate the connection, either normally or abnormally.

DCCP-Sync and DCCP-SyncAck - Used to resynchronize sequence numbers after large bursts of loss.

The three-way handshakes during initiation and termination are presented in Figure 2.

Client Server

(1) Initiation

DCCP-Request -->

<-- DCCP-Response

DCCP-Ack -->

(2) Data transfer

DCCP-Data, DCCP-Ack, DCCP-DataAck -->

<-- DCCP-Data, DCCP-Ack, DCCP-DataAck

(3) Termination

<-- DCCP-CloseReq

DCCP-Close -->

<-- DCCP-Reset

Fig.2. DCCP connection example

III. Congestion Control

As a congestion control framework, DCCP gives the application a choice of congestion control mechanisms. Some applications might prefer TFRC congestion control, avoiding TCP's abrupt halving of the sending rate in response to congestion, while others might prefer a more aggressive TCP-like probing for available bandwidth. The choice is made via Congestion Control IDs (CCIDs), which name standardized congestion control mechanisms. A connection's CCIDs are negotiated at connection startup. This section describes the two CCIDs that have currently been developed, congestion control issues exposed by DCCP's target applications that remain to be solved, and more general problems relating to congestion

control, including misbehaving receivers and non-congestion loss.

DCCP's CCID 2 provides a TCP-like congestion control mechanism, including the corresponding abrupt rate changes and ability to take advantage of rapid fluctuations in available bandwidth. CCID 2 acknowledgements use the Ack Vector option, which is essentially a version of TCP's SACK. Its congestion control algorithms like wise follow those of SACK TCP, and maintain similar variables: a congestion window "cwnd", a slow-start threshold, and an estimate of the number of data packets outstanding [7].

TFRC congestion control in DCCP's CCID 3 uses a different approach. Instead of a congestion window, a TFRC sender uses a sending rate. The sender uses this loss event rate to determine its sending rate; if no feedback is received for several round-trip times, the sender halves its rate. TFRC [5] is an equation based rate-controlled congestion control mechanism that is designed to be reasonably fair when competing for bandwidth with TCP and TCP-like flows. The advantage of TFRC is that it has a lower variation in throughput over time compared to TCP, which makes it more suitable for applications where such a smooth sending rate is desired. The basic idea behind the TFRC mechanism is that the rate at which the sender can transmit data packets is restricted by the allowed sending rate, which is calculated from the current loss event rate and the roundtrip time. It is a receiver-based mechanism with the calculation of the congestion control information in the receiver rather than in the sender. Upon receiving an acknowledgement, the sender updates its allowed sending rate based on the information received. The sending endpoint also estimates the round-trip time and calculates the nofeedback timer.

Together with the packet size s, expressed in bytes, and an estimate RTT, in seconds, of the round-trip time, the loss vent rate p can be plugged into the throughput equation, based on the throughput obtained by TCP Reno:

$$X = \frac{s}{RTT\left(\sqrt{\frac{2p}{3}} + 12p\sqrt{\frac{3p}{8}}(1+32p^2)\right)},$$

where X is the allowed sending rate in bytes/second. Afterward, the sending rate is updated in correspondence with the long term round-trip time to provide a better congestion avoidance strategy [5].

IV. Results from Simulations

For our simulations, we nave used the Network Simulator (NS-2) [1], a popular discrete event simulator, targeted at networking research. A DCCP module for NS was implemented by Mattsson [2] and will be used to obtain the results presented in this section.

Figure 3 and Figure 4 presents the behaviour of DCCP with CCID2 and CCID3 and queue length, between two nodes with bottleneck 20Mbps, delay 20ms and drop tail 30 length queue positions. The differences between these two congestion control mechanisms could be clear detected.

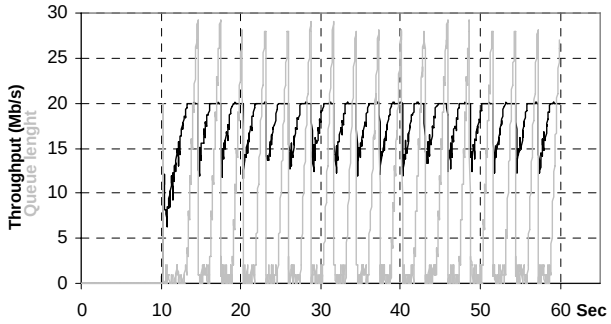


Fig.3. DCCP with CCID2

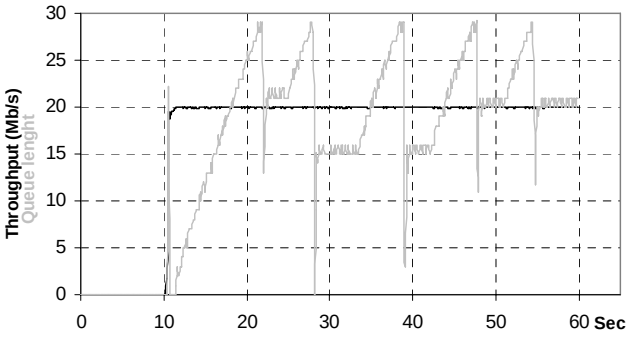


Fig.4. DCCP with CCID3

Both variants of DCCP are compared with TCP and UDP illustrates in Figure 5 and Figure 6.

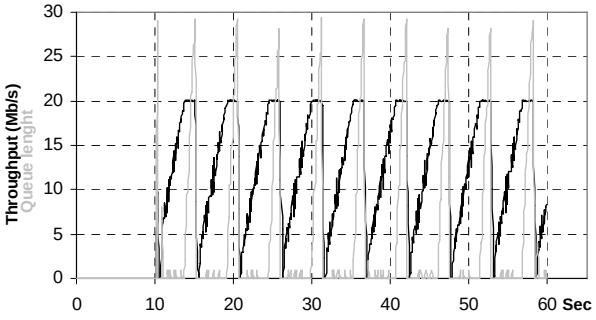


Fig.5. TCP

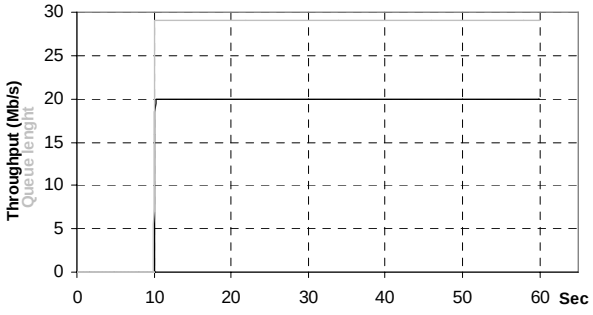


Fig.6. UDP with 25Mbps input flow

The results from the simulations are displayed in Table1. We compare average packet delay and lost packet percent for DCCP CCID2, DCCP CCID3, TCP and UDP in variants – network without losses and network with 1% packet loss. In

the case of network with packet loss, the transmission time for TCP flows is grater than DCCP and UDP case because of retransmission of lost packets.

TABLE I

Protocol	No losses		1% Packet loss	
	Packets delay, sec	Lost packets, %	Packets delay, sec	Lost packets, %
DCCP CCID2	0.02408	0.13673	0.02091	1.14353
DCCP CCID3	0.02526	0.09519	0.02049	1.03517
TCP	0.02147	0.04256	0.18390	0.05357
UDP	0.03191	20.0326	0.032372	23.8585

For the next simulations we use the network setup shown in Figure 7. In this setup three flows – DCCP, TCP and UDP share a bottleneck link. The bandwidth of DCCP and TCP flows depend on available bandwidth while UDP flow does not change. The results from the simulation with different DCCP and TCP flows are presented in Figure 8 and Figure 9.

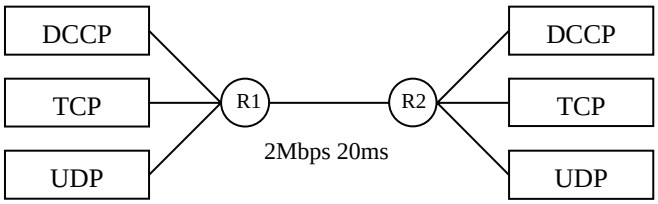


Fig.7. Simulation topology

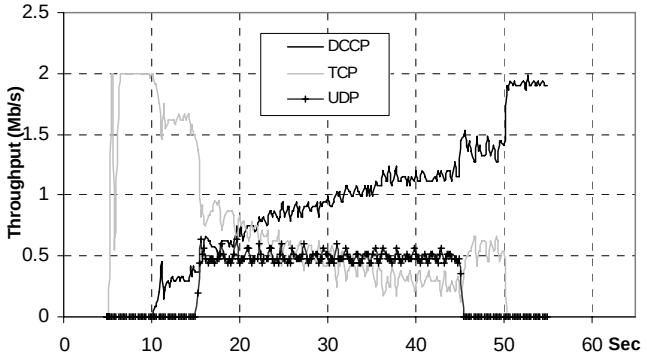


Fig.8. DCCP with CCID3, TCP Reno an UDP

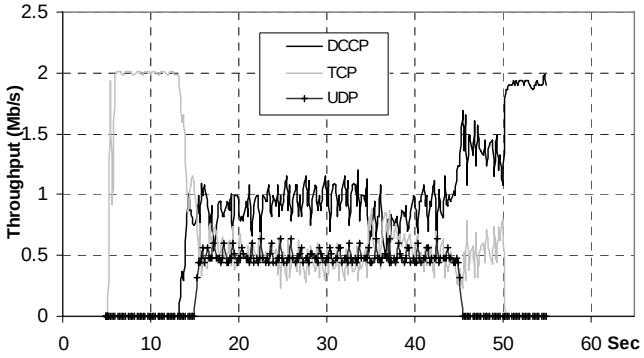


Fig.9. DCCP with CCID2, TCP Sack an UDP

V. CONCLUSION

In this paper the behavior of DCCP is studied. This protocol is compared with the others very popular transport protocols TCP and UDP using the parameters packet loss and average delay.

The results from simulations have confirmed that DCCP could be the proper solution for application that can benefit from control over the tradeoffs between delay and reliable in-order delivery such as streaming media.

As part of future work, we plan to impellent DCCP in real network and obtain results for real-time applications.

ACKNOWLEDGEMENT

This paper is sponsored by the Ministry of Education and Science of the Republic of Bulgaria in the framework of project ВУ-ТН-105/2005 “Multimedia Telecommunications Networks Planning with Quality of Service and Traffic Management”.

REFERENCES

[1] The Network Simulator – NS-2.qq.

[2] N.-E. Mattsson. A DCCP module for NS-2. Sept. 2004. – Master thesis.

[3] E. Kohler, S. Floyd, and M. Handley, “Designing DCCP: Congestion Control without Reliability,” in *Proc. ACM SIGCOMM*, Pisa, Italy, 2006.

[4] E. Kohler, M. Handley, and S. Floyd, “Datagram Congestion Control Protocol (DCCP),” RFC 4340, Mar. 2006.

[5] S. Floyd, E. Kohler, and J. Padhye. Profile for DCCP Congestion Control ID 3: TRFC Congestion Control. Request For Comments 4342, IETF, Mar. 2006.

[6] S. Floyd, M. Handley, and E. Kohler. Problem Statement for the Datagram Congestion Control Protocol (DCCP). RFC 4336 (Informational), Mar. 2006.

[7] S. Floyd and E. Kohler. Profile for DCCP Congestion Control ID 2: TCP-like Congestion Control. Request for Comments 4341, IETF, Mar. 2006.

[8] ITU-T Recommendation Y.2001, *General overview of NGN*, 12/2004.

[9] М.Илиев, Г. Христов, В. Николов, „Анализ и моделиране на телекомуникационни мрежи и протоколи” , Русенски Университет „Ангел Кънчев” 2006.

SESSION RCMTA II

Radio Communications, Microwave Technique and Antennas II

Resonance and Dispersion Characteristics of Microstrip Slow-Wave Open-Loop Resonator

Marin V. Nedelchev[1] and Ilia G. Iliev[2]

Abstract: **This paper represents a theory of admittance loaded slow-wave open loop resonators. The main resonator is extended to periodically loaded line. A closed form formulas for the fundamental and the first spurious resonance frequency are derived. Based on the Floquet's theorem, a dispersion equation of the slow-wave resonator is derived. The resonance characteristics are studied for different impedances of the loading stubs. Dispersion characteristics, defining the filtering function of the periodically loaded line, are also studied for different impedances and electrical length of the loading stubs.**

Keywords**: Dispersion equation, resonance characteristics, slow-wave open loop resonator.**

I. INTRODUCTION

Radio frequency and microwave planar bandpass filters are presently required in wide variety of applications of wireless communication systems, WLAN, software radio etc.
The compact sizes of the filters, combined with wider upper stopband are crucial factors for some wireless applications. However most of the microstrip bandpass filters are large in size and their first spurious frequencies appear at $2f_0$ or $3f_0$, where f_0 is the filter central frequency. Many authors propose using stepped impedance resonators [1,2] to shift the first spurious resonance frequency not on a multiple of the fundamental resonance. In [3] is presented a theory and experiment of slow-wave open loop resonators (Fig.1a) and their extension for capacitively loaded transmission line. This theory is a special case of a periodically admittance loaded transmission line. In [4] is proposed a slow-wave resonator loaded with different loading capacities (Fig.1b). [4] claims for higher unloaded quality factor than the slow-wave resonator presented in [3].

(a) (b)

Fig.1 Topologies of slow-wave open loop resonators

This paper presents a general theory of slow-wave open loop resonator. Based on the circuit theory, closed form formulas for the fundamental resonance frequency and the first spurious resonance frequency are derived. Using the electrical parameters of the transmission lines, a general dispersion equation is derived.

The dispersion effect conducts the filtering function of the slow-wave resonator. It is studied four cases for slow-wave resonators. The obtained results are graphically shown and are applicable for practical design.

II. RESONANCE CHARACTERISTICS OF SLOW-WAVE OPEN LOOP RESONATOR

The slow-wave open-loop resonator consists of a transmission line loaded on both sides with admittances Y_1 and Y_2. The main parameters of the transmission line are the characteristic impedance Z_c, length of the line l, and the propagation constant k. The resonator is shown on Fig.2.

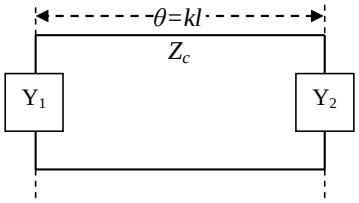


Fig.2 Admittance loaded slow-wave resonator

The electric length of the unloaded line becomes $\theta = \gamma l$, where $\gamma = jk$. The electrical characteristics of the resonator may be described by the ABCD matrix. The overall matrix of the resonator is a product of the three ABCD matrices - admittances Y_1, Y_2 and the unloaded transmission line.

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

where

$$[ABCD]_1 = \begin{bmatrix} 1 & 0 \\ Y_1 & 1 \end{bmatrix}, \; [ABCD]_3 = \begin{bmatrix} \cosh\theta & Z_c \sinh\theta \\ \frac{1}{Z_c}\sinh\theta & \cosh\theta \end{bmatrix},$$

$$[ABCD]_3 = \begin{bmatrix} 1 & 0 \\ Y_2 & 1 \end{bmatrix}.$$

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

$$A = \cosh\theta + Y_2 Z_c \sinh\theta \quad (2a)$$

$$B = Z_c \sinh\theta \quad (2b)$$

[1]Marin Veselinov Nedelchev – Assistant, PhD in Dept. of Radiotechnic in Faculty of Communications and Communication Technologies in TU –Sofia E-mail mnedelchev@tu-sofia.bg

[2]Ilia Georgiev Iliev – Assoc. Professor, PhD in Dept. of Radiotechnic in Faculty of Communications and Communication Technologies in TU –Sofia E-mail: igiliev@tu-sofia.bg

$$C = (Y_1 + Y_2)\cosh\theta + \left(Y_1 Y_2 Z_c + \frac{1}{Z_c}\right)\sinh\theta \qquad (2c)$$

$$D = Z_c Y_1 \sinh\theta + \cosh\theta \qquad (2d)$$

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

If a standing wave is excited in the slow-wave resonator and taking into account the boundary conditions $I_1 = 0$ and $I_2 = 0$, consequently for non-zero voltages V_1 and V_2 it is required that:

$$\frac{C}{A} = \frac{I_1}{V_1}\bigg|_{I_2=0} = \frac{I_2}{V_2}\bigg|_{I_1=0} = 0 . \qquad (3)$$

It is clear that for the fundamental resonance:

$$A = \frac{V_1}{V_2}\bigg|_{I_2=0} = -1 . \qquad (4)$$

Solving the Eqs (3) and (4) it is derived the following quadratic equation:

$$(Y_1 + Y_2)\tan h^2\frac{\gamma l}{2} - \frac{2}{Z_c}\tan h\frac{\gamma l}{2} + Y_1 - Y_2 = 0 \qquad (5)$$

The solutions for the electrical length of unloaded transmission line are derived from the Eq.(5).

$$\tan h\frac{\gamma_0 l}{2} = \frac{1/Z_c \pm \sqrt{1/Z_c^2 - (Y_1^2 - Y_2^2)}}{Y_1 + Y_2} \qquad (6).$$

For the first spurious resonance we have:

$$A = \frac{V_1}{V_2}\bigg|_{I_2=0} = 1 \qquad (7).$$

Then solving the system of Eqs. (3) and (7) leads to the following quadratic equation:

$$(Y_2 - Y_1)\tan h^2\frac{\gamma l}{2} + \frac{2}{Z_c}\tan h\frac{\gamma l}{2} + Y_1 + Y_2 = 0 \qquad (8)$$

The solutions for the electrical length of the unloaded transmission line of the slow-wave resonator ($Y_1 \neq Y_2$) are:

$$\tan h\frac{\gamma_1 l}{2} = \frac{-1/Z_c \pm \sqrt{1/Z_c^2 - (Y_1^2 - Y_2^2)}}{Y_2 - Y_1} \qquad (9).$$

In the most practical cases the loading admittances Y_1 and Y_2 are equal ($Y_1 = Y_2 = Y$) [1,3]. Then the transmission line electrical length for the fundamental resonance is derived from Eq.(6) as:

$$\tan h\frac{\gamma_0 l}{2} = \frac{1}{Z_c Y} \qquad (10)$$

For the spurious resonance the equation is:

$$\tan h\frac{\gamma_1 l}{2} = -Z_c Y \qquad (11)$$

For microstrip slow wave resonators often the loading admittances are open circuited stubs. In the case of open stub the input admittance is $Y_{in} = Y_{opst}\tan h\theta_{opst}$. Assuming a lossless lines ($\alpha = 0$) the propagation constant is a pure imaginary value. Then the fundamental resonance condition becomes:

$$\theta_0 = 2arc\tan\left(Y_{opst} Z_c tg\theta_{opst}\right) \qquad (12)$$

The electrical length of the unloaded transmission line for the first spurious resonance is found as:

$$\theta_1 = 2\pi - 2arc\tan\left(Y_{opst} Z_c tg\theta_{opst}\right) \qquad (13)$$

It is clearly seen that for unloaded resonator, the resonator electrical length for the fundamental resonance is $\theta_0 = \pi$, and for the first spurious resonance- $\theta_1 = 2\pi$.

It is shown on Fig.3 the dependence of the overall electrical length for the fundamental resonance of the slow-wave resonator ($\theta_{t0} = 2\theta_{opst} + \theta_0$) as a function of the open stub length (θ_{opst}) for a constant impedance ratio $K = Z_c / Y_{opst}$.

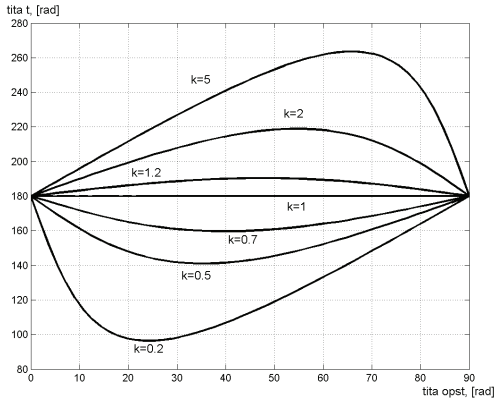


Fig.3 Electrical length of the resonator for the fundamental frequency

If the impedance ratio $K < 1$ the resonator is shorter than the halfwave unloaded resonator. This fact is applicable for small sized resonators used in mobile communication systems. But for $K > 1$, the resonator is longer than a halfwave resonator. It is also seen that for big or small values of the impedance ratio K, it is well pronounced a maximum of the resonator length θ_t.

Fig.3 shows the dependence of the slow-wave resonator electrical length for the first spurious resonance as a function of as a function of the open stub length θ_{opst} for a constant impedance ratio K.

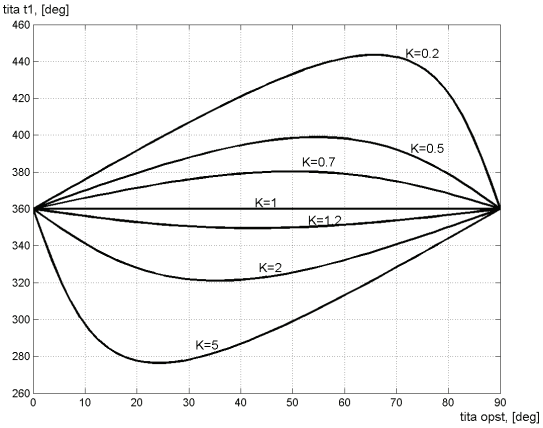


Fig.4 Electrical length of the resonator for the first spurious frequency

It is clearly seen form Fig.4 that for unloaded halflength resonator, the overall electrical length is $2\pi\ rad$ for the first spurious resonance. For impedance ratio $K<1$, the resonator electrical length is larger than halfwave resonator. It is obvious that for larger impedance ratio the resonator length for the first spurious resonance is shorter.

Fig.4 shows the dependence of the ratio of the resonator overall lengths θ_{t1}/θ_{t0} as a function of the electrical length of the open stub for various impedance ratios K. For $K=1$ (halfwave resonator) it is seen that $\theta_{t1}/\theta_{t0}=2$ i.e. the first spurious frequency is twice the fundamental resonance frequency.

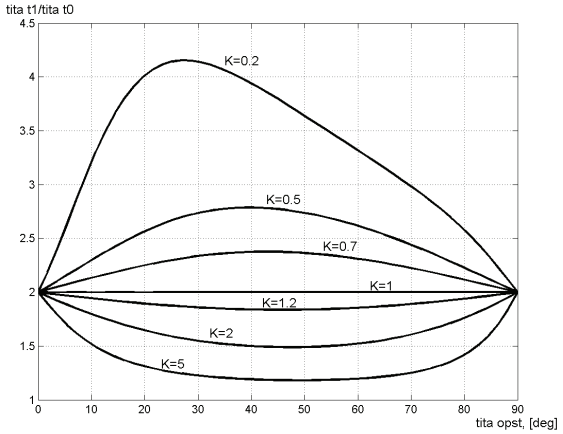


Fig.5. Ratio of the resonator lengths θ_{t1}/θ_{t0} as a function of the electrical length of the open stub.

To control the first spurious frequency, it is necessary to choose a proper impedance ratio. In the most common cases the first spurious resonance is not a multiple of the fundamental resonance and it is greater than twice the fundamental. From Fig.5 it is clear that in these cases the impedance ratio should be less than 1. Then the length of the resonator is less than the halfwave resonator.

III. Dispersion Equation

We may consider the slow-wave resonator from Fig.1 for a unit cell of as periodically loaded transmission line. This assumption allows explaining the physical mechanism of the filtering characteristics of the periodically loaded transmission line. Let's β is the propagation constant in the periodic admittance loaded transmission line. Then applying the Floquet's theorem [5] i.e.:

$$\begin{aligned} V_2 &= e^{-j\beta d}V_1 \\ -I_2 &= e^{-j\beta d}I_1 \end{aligned} \tag{14}$$

to the [ABCD] matrix of an unit cell results in:

$$\begin{bmatrix} A-e^{-j\beta d} & B \\ C & D-e^{-j\beta d} \end{bmatrix}\cdot\begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{15}$$

The non-zero solution for the output voltage V_2 and current $-I_2$ exists only if the determinant is zero. This leads to the following equation:

$$\left(A-e^{-j\beta d}\right)\left(D-e^{-j\beta d}\right)-BC=0 \tag{16}$$

Substituting Eq.(2a-2d) in (16) and taking into account the [ABCD] matrix property $AD-BC=1$ we derive the dispersion equation of the slow wave open loop resonator.

$$\cosh(\beta d)=\cosh(\theta)+Z_c\frac{(Y_1+Y_2)}{2}\sinh(\theta) \tag{17}$$

If the transmission line is loaded with equal open stubs $Y_1=Y_2=Y$, the dispersion equation (17) becomes:

$$\cosh(\beta d)=\cosh(\theta)+Z_cY\sinh(\theta) \tag{18}$$

For the fundamental frequency substituting (12) in (18) it is derived $\cosh(\beta_0 d)=-1$. For the first spurious frequency it is seen that $\cosh(\beta_1 d)=1$. Here the propagation constants are $\beta_0=\omega_0/v_{p0}$ and $\beta_1=\omega_0/v_{p1}$, where v_{p0} and v_{p1} are the phase velocities for the fundamental and the first spurious resonance. If there is no dispersion, the phase velocity is a constant. This is true for unloaded transmission line. The phase velocity is frequency dependant for periodically admittance loaded transmission line. This fact determines the filtering function of the periodically loaded transmission lines.

There are studied two main types of slow wave resonators - with impedance ratios $K<1$ and $K>1$.

The first case is for slow wave open loop resonator with $\theta_{opst}=\pi/6\ rad, Z_{opst}=1/Y_{opst}=20\Omega$ and $Z_{opst}=1/Y_{opst}=40\Omega$. The characteristic impedance of the main transmission line is $Z_c=50\Omega$. The dispersion curves are shown on Fig.6.

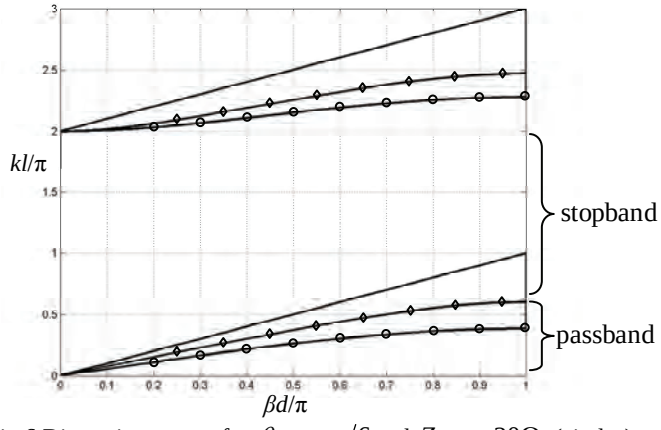


Fig.6 Dispersion curves for $\theta_{opst}=\pi/6\ rad, Z_{opst}=20\Omega$ (circles) and $Z_{opst}=40\Omega$ (diamonds), halfwave resonator-solid line

It is seen from Fig.5 that increasing the characteristic impedance of the loading stub leads to decreasing of the dispersion effect. For $Z_{opst}=40\Omega$, the passband is wider than the case with $Z_{opst}=20\Omega$. For unloaded transmission line resonator (halfwave resonator case) it is seen that the first spurious passband is placed on twice the fundamental resonance frequency.

The next studied resonator is with loading stubs with electrical length $\theta_{opst}=\pi/12\ rad$ and characteristic

impedances $Z_{opst} = 20\Omega$ and $Z_{opst} = 40\Omega$. The results are shown on Fig.7.

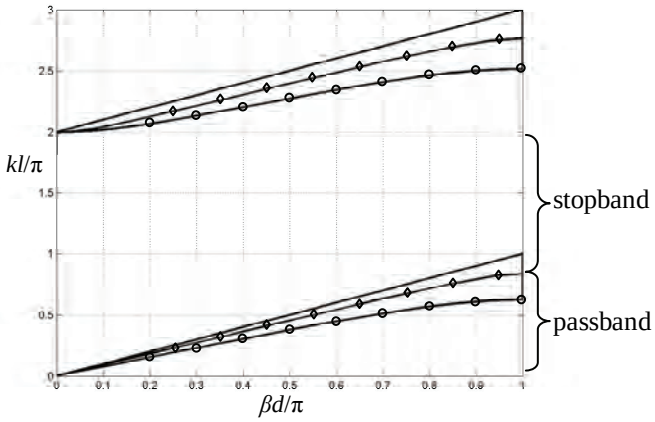


Fig.7 Dispersion curves for $\theta_{opst} = \pi/12\, rad$, $Z_{opst} = 20\Omega$ (circles) and $Z_{opst} = 40\Omega$ (diamonds), halfwave resonator-solid line

Comparing both Fig.6 and Fig.7, it is clearly seen that longer loading stubs lead to well pronounced dispersion effect. This effect results in better filtering functions and a possibility to control the passband and the stopband. Moreover choosing proper characteristic impedance of the loading stubs and their electrical length, it may control the first spurious passband.

The dispersion curves for $K < 1$ and $\theta_{opst} = \pi/6\, rad$ are shown on Fig.8.

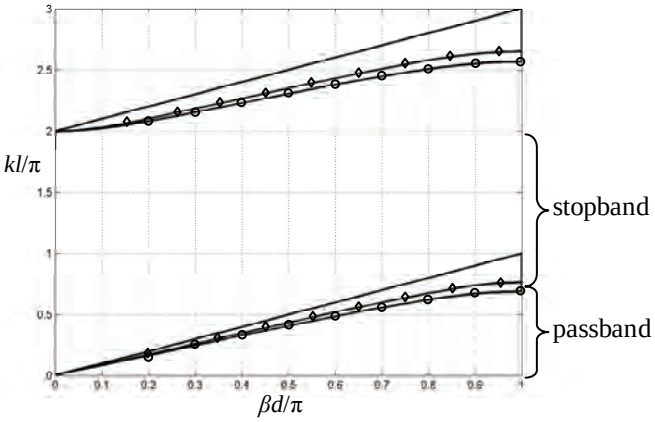


Fig.8 Dispersion curves for $\theta_{opst} = \pi/6\, rad$, $Z_{opst} = 55\Omega$ (circles) and $Z_{opst} = 75\Omega$ (diamonds), halfwave resonator-solid line

It can be found from Fig.8 that the dispersion effect is less pronounced than the case for $K > 1$ (Fig.6 and Fig.7). However the slope of the dispersion curve for the fundamental resonance is steeper than the slope for the first spurious resonance. This fact forms narrower spurious passband than the fundamental passband. If the characteristic impedance is bigger ($Z_{opst} = 75\Omega$), the dispersion effect is closer to the halfwave unloaded resonator.

The last studied case is for slowwave resonator loaded with open stubs, having characteristic impedance greater than the main transmission line and electrical length $\theta_{opst} = \pi/12\, rad$.

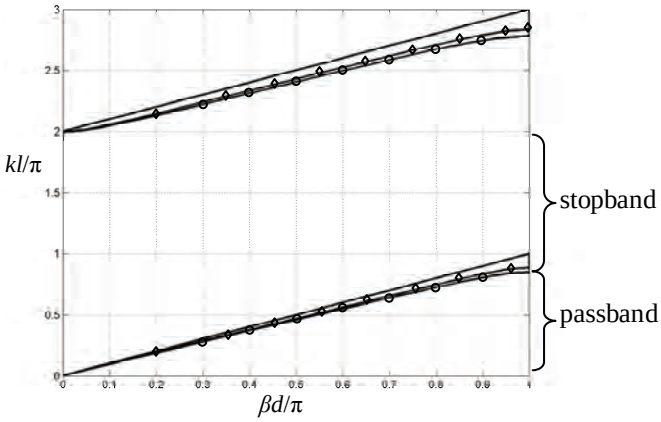


Fig.9 Dispersion curves for $\theta_{opst} = \pi/12\, rad$, $Z_{opst} = 55\Omega$ (circles) and $Z_{opst} = 75\Omega$ (diamonds), halfwave resonator-solid line

It is clearly seen from Fig.9 that the dispersion effect is less pronounced for short high impedance open stubs. The passband is almost equal to the passband of halfwave resonator.

V. Conclusion

This paper presents a general theory of slow-wave open loop resonator. Based on the circuit theory, closed form formulas for the fundamental resonance frequency and the first spurious resonance frequency are derived. Using the electrical parameters of the transmission lines, a general dispersion equation is derived. The theoretical view of the admittance loaded transmission line resonator gives insight for the main properties of this type of resonator. The filters designed with slow-wave resonators will have wider stopband characteristics, due to the dispersion effect.

References

[1] Makimoto M. and S. Yamashita. "Bandpass Filters Using Parallel Coupled Stripline Stepped Impedance Resonators." 1980 Transactions on Microwave Theory and Techniques Dec. 1980 pp.1413-1417.

[2] Sheta A.F., J.P. Coupez, G. Tanne, S. Toutain and J.P. Blot. "Miniature Microstrip Stepped Impedance Resonator Bandpass Filters and Diplexers for Mobile Communications." 1996 MTT-S International Microwave Symposium Digest pp.607-610.

[3] Hong, Jia-Sheng and M.J. Lancaster. "Theory and experiment of novel microstrip slow-wave open-loop resonator filters." 1997 Transactions on Microwave Theory and Techniques 45.12 (Dec. 1997, Part II [T-MTT] (1997 Symposium Issue)): 2358-2365

[4] Akkaraekthalin, Jantree, Microstrip Slow-wave Open-Loop Resonator Filters with Reduced Size and Improved Stopband Characteristics, ETRI Journal, Vol.28, Number 5, Oct.2006, pp.604-617.

[5] Collin, Foundation for Microwave Engineerig, McGraw Hill, NY, 2002

Relative Influence of Some Stochastic Factors on Bit-Error Rate of Ground-to-Ground Free Space Optics

Boncho G. Bonev[1]

***Abstract* – In this paper an equation for full Bit-Error Rate calculation of ground-to-ground free space optics according to partial Bit-Error Rates, caused by signal current fluctuations and quantum noises of signal and background currents has been suggested. Fluctuations of signal current are produced by atmosphere transmittance fluctuations, mechanical vibrations of antennas and atmospheric turbulence. Relative influence of this factors on full Bit-Error Rate is investigated.**

***Keywords* – Bit-Error Rate, Free Space Optics, Atmospheric Transmittance Fluctuation, Atmospheric Turbulence and Quantum Noise.**

I. INTRODUCTION

The interest in Free-Space Laser Communications (Free-Space Optics – FSO) is increased in the last years. The great importance of Internet in modern world and some disadvantage of microwave and cable networks lead to widely usage of FSO in local area networks. This is caused to a great extent by the fact that FSO's technical and economical parameter most fully correspond to the requirements of that type of networks – quick and easy installation, portability, relatively low price and high speed information flows. Through Free-Space Optics the "last mile" problem is resolved. Furthermore FSO are irreplaceable for connections of distant light-guide systems, in cases when for infrastructure or economical reasons is impossible to use another light-guide.

The utilization of FSO is related to some important problems, which decision is subject to a number of researches [1-8]. They are result of simultaneous influence of many factors that affect the spatial structure of the laser beam - the atmospheric transmittance fluctuations, the propagation direction fluctuations caused by mechanical vibrations [6] of antennas and atmospheric turbulence [6,9-11]. Jointly influence of these factors and their stochastic character on one hand and the quantum noises of signal and background currents on the other have formed Bit-Error Rate (BER) of FSO.

The role of each of this factors in total effect of their jointly influence has not been examined in the literature. Therefore in this paper relative influence of mentioned factors on BER has been investigated.

[1]Boncho G. Bonev is with the Faculty of Communications and Communication Technologies, Technical University - Sofia, 8 Kl. Ohridski Blvd., 1756 Sofia, Bulgaria, E-mail: bbonev@tu-sofia.bg

II. THEORETICAL ANALYSIS

FSO shown in Fig.1 consists of laser and pulse-code modulator (PCM) with output optical power Φ_L, transmitting antenna TA with transparency τ_1, radiated Gaussian laser beam with initial radius r_0, receiving antenna RA with radius R_2 and transparency τ_2, interference filter IF and photodetector PhD with radius R_d.

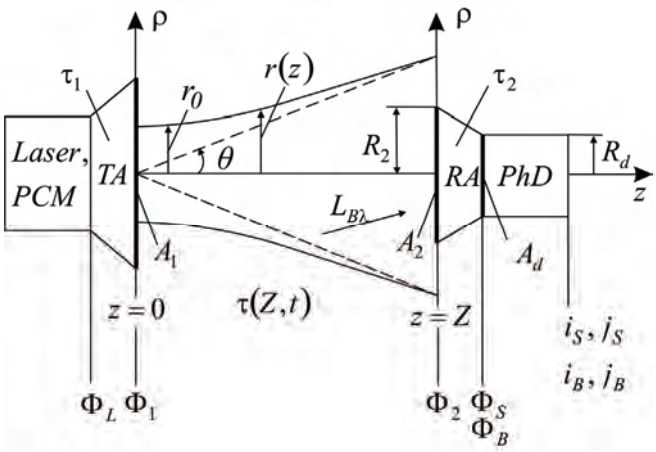


Fig. 1

Let us assume that the values of these parameters of laser communication system – the frequency bandwidth of pulse-code signal Δf, the working wavelength λ, the horizontal link altitude H and the distance between the transmitter and the receiver Z are known.

The average value of signal current over quantum fluctuations is given by [14]

$$\langle i_S \rangle(Z,\rho) = U(Z)\tau(Z)\psi(Z,\rho),$$

$$U(Z) = \frac{2\tau_1\tau_2 S_i \Phi_L R_2^2}{\varepsilon r^2(Z)},$$

$$\varepsilon = 1 - \frac{1}{exp(2)} \approx 0{,}865, \qquad (1)$$

where $\tau(Z)$ is the atmospheric transmittance and the function

$$\psi(Z,\rho)=\exp\left[-2\frac{\rho^2}{r^2(Z)}\right] \quad (2)$$

gives casual linear deviation ρ of laser beam in relation to receiving antenna centre, caused by mechanical vibration of antennas and atmospheric turbulence [14].

Gaussian laser beam radius in receiving antenna plane is presented with the equation

$$r(Z)=r_0\sqrt{1+\left(\frac{K\lambda}{\pi r_0^2}\right)^2 Z^2}\,, \quad (3)$$

where the coefficient K provides a simple and adequate account of both real beam divergence and additional beam expansion resulting from the radiance scatter from relatively small (in relation to $2r(Z)$) aerosol inhomogeneties and turbulence eddies [12].

The absolute photodetector's sensitivity is computed from relation

$$S_i=\frac{e\eta\lambda}{hc}\,, \quad (4)$$

where η is quantum efficiency of the photodetector and $e=1{,}6.10^{-19}$ C, $h=6{,}626.10^{-34}$ J.s, $c=3.10^8$ m/s.

Signal-to-Noise Ratio (SNR) of the laser communication system is represented as

$$SNR=\frac{\langle i_S\rangle^2(Z)}{\sigma_{i_S}^2(Z)+\sigma_{j_S}^2(Z)+\sigma_{j_B}^2(Z)}\,. \quad (5)$$

The averaging of $\langle i_S\rangle$ over fluctuations of atmospheric transmittance τ and over function $\psi(Z,\rho)$ can be given by expression

$$\langle\langle i_S\rangle\rangle(Z)=\langle i_S\rangle(Z,\langle\rho\rangle=0)=$$
$$=\langle i_S\rangle(Z)=U(Z).\langle\tau\rangle(Z).\langle\psi\rangle(Z). \quad (6)$$

Dispersion of signal current determined by the fluctuations of τ and $\psi(Z,\rho)$ is

$$\sigma_{i_S}^2(Z)=$$
$$=U^2(Z)\left[\sigma_\tau^2(Z)\sigma_\psi^2(Z)+\langle\psi\rangle^2(Z)\sigma_\tau^2(Z)+\langle\tau\rangle^2(Z)\sigma_\psi^2(Z)\right]. \quad (7)$$

Schottky's formula gives the variances of the quantum fluctuations of signal and background currents

$$\sigma_{j_S}^2(Z)=2e\langle i_S\rangle(Z)\Delta f \quad (8)$$

and

$$\sigma_{j_B}^2=2e\pi^2\tau_2 S_i L_{B\lambda} R_2^2 R_d^2\frac{(\Delta\lambda)_{IF}\Delta f}{f_{eq}^2}\,, \quad (9)$$

where $L_{B\lambda}$ is the brightness of background radiation, $(\Delta\lambda)_{IF}$ is the bandwidth of the interference filter before photodetector and f_{eq} is equivalent focal length of the receiving antenna.

Statistical moments of atmospheric transmittance τ is given by relations [13]

$$\langle\tau\rangle(Z)=\exp\left(-\langle\alpha\rangle Z\right), \quad (10)$$

$$\sigma_\tau^2(Z)=\langle\alpha\rangle^2\Delta_\alpha^2 HZ\exp\left(-2\langle\alpha\rangle Z\right).$$

For calculation of coefficient of extinction $\langle\alpha\rangle$ the following expression is used [15]

$$\langle\alpha\rangle,\mathrm{km}^{-1}=\frac{3{,}91}{S_M,\mathrm{km}}\left(\frac{\lambda,\mu\mathrm{m}}{0{,}55}\right)^{-0{,}585(S_M,\mathrm{km})^{1/3}}\,, \quad (11)$$

where S_M is meteorological visibility.

With $\Delta_\alpha\in[0{,}05\,;0{,}5]$ is represented the variation coefficient of spatial heterogeneity of α [16] and

$$\Delta_\alpha=\frac{\sigma_\alpha}{\langle\alpha\rangle}\,. \quad (12)$$

Eq. (5) can be transformed like

$$SNR=\frac{1}{\frac{1}{\frac{\langle i_S\rangle^2(Z)}{\sigma_{i_S}^2(Z)}}+\frac{1}{\frac{\langle i_S\rangle^2(Z)}{\sigma_{j_S}^2(Z)}}+\frac{1}{\frac{\langle i_S\rangle^2(Z)}{\sigma_{j_B}^2}}}=$$
$$=\frac{1}{\frac{1}{SNR_{i_S}}+\frac{1}{SNR_{j_S}}+\frac{1}{SNR_{j_B}}}\,, \quad (13)$$

where SNR_{i_S}, SNR_{j_S} and SNR_{j_B} are signal-to-noise ratios, caused by examining factors – fluctuation of signal current due to the atmospheric transmittance and laser beam direction fluctuations, quantum fluctuations of signal current and quantum fluctuations of background current.

With respect to Eqs. (6), (7), (8) and (13) for SNR_{i_S}, SNR_{j_S} and SNR_{j_B} is obtained

$$SNR_{i_S}=\frac{\langle i_S\rangle^2(Z)}{\sigma_{i_S}^2(Z)}=$$
$$=\frac{\langle\tau\rangle^2(Z).\langle\psi\rangle^2(Z)}{\sigma_\tau^2(Z).\sigma_\psi^2(Z)+\langle\psi\rangle^2(Z).\sigma_\tau^2(Z)+\langle\tau\rangle^2(Z).\sigma_\psi^2(Z)}\,, \quad (14)$$

$$SNR_{j_S} = \frac{\langle i_S \rangle^2(Z)}{\sigma_{j_S}^2(Z)} = \frac{\langle i_S \rangle(Z)}{2e\Delta f} =$$

$$= \frac{U(Z).\langle \tau \rangle(Z).\langle \psi \rangle(Z)}{2e\Delta f}, \qquad (15)$$

$$SNR_{j_B} = \frac{\langle i_S \rangle^2(Z)}{\sigma_{j_B}^2} \qquad (16).$$

In Eq. (14) SNR_{i_S} depends only on atmospheric communication channel condition, but SNR_{j_S} and SNR_{j_B} from Eqs. (15) and (16) depend on both atmospheric communication channel condition and power parameters of communication system.

According to [14] partial Bit-Error Rates, produced of mentioned three factors can be calculated by equations

$$BER_{i_S} = \frac{1}{2}\operatorname{erfc}\left(\frac{SNR_{i_S}}{2\sqrt{2}}\right),$$

$$BER_{j_S} = \frac{1}{2}\operatorname{erfc}\left(\frac{SNR_{j_S}}{2\sqrt{2}}\right), \qquad (17)$$

$$BER_{j_B} = \frac{1}{2}\operatorname{erfc}\left(\frac{SNR_{j_B}}{2\sqrt{2}}\right).$$

The Bit-Error Rate of system can be expressed with partial BERs by means of the equation

$$BER = \frac{1}{2}\operatorname{erfc}\left\{\sum_{k=i_S;j_S;j_B}\frac{1}{[\operatorname{erfc}\operatorname{inv}(2.BER_k)]^2}\right\}^{-\frac{1}{2}} = \frac{1}{2}\operatorname{erfc}\left(\frac{SNR}{2\sqrt{2}}\right). \qquad (18)$$

In Eq. (18) $X = \operatorname{erfcinv}(Y)$ is the inverse function of $Y = \operatorname{erfc}(X)$, which is integrated in MATLAB 7.1.

III. Numerical Results

The results of calculation of BER_{i_S}, BER_{j_S} and BER in dependence on meteorological visibility S_M and dispersion of linear deviation of laser beam in transmitter's plane σ_ρ^2 are represented on Fig. 2 and Fig. 3. The following typical for FSO input data have been assumed: $Z = 10\,\text{km}$; $H = 50\,\text{m}$; $\Delta_\alpha = 0{,}35$; $\tau_1 = 0{,}7$; $\tau_2 = 0{,}6$; $\Phi_L = 10\,\text{mW}$; $R_2 = 10\,\text{cm}$; $\Delta f = 1\,\text{GHz}$; $\lambda = 1{,}55\,\mu\text{m}$; $\eta = 0{,}7$; $r_0 = 3\,\text{mm}$; $K = 10$; $L_{B\lambda} = 0{,}01\,\text{W/m}^2\text{sr}\,\text{Å}$; $R_d = 5\,\mu\text{m}$; $f_{eq} = 0{,}5\,\text{m}$; $(\Delta\lambda)_{IF} = 20\,\text{Å}$.

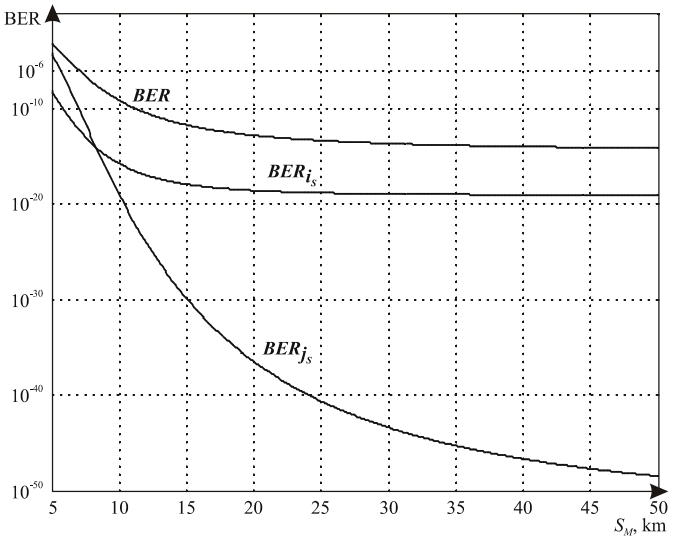


Fig. 2. Dependence of BER_{i_S}, BER_{j_S} and BER on meteorological visibility S_M for $\sigma_\rho^2 = 5{,}76\,\text{m}^2$

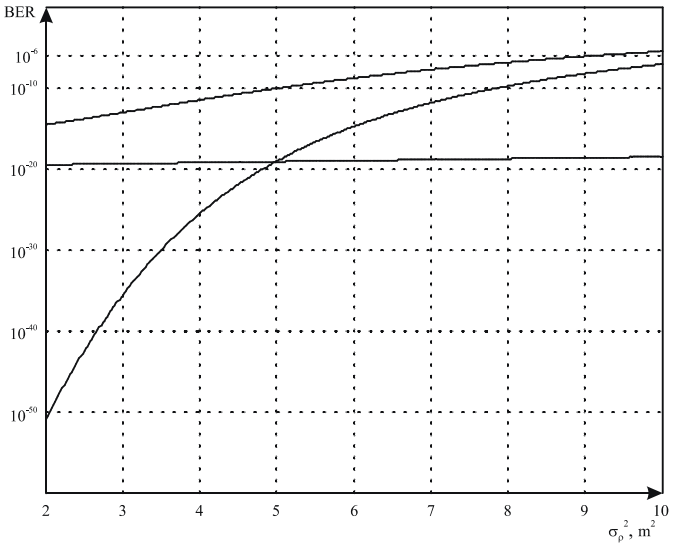


Fig. 3. Dependence of BER_{i_S}, BER_{j_S} and BER on dispersion σ_ρ^2 for $S_M = 10\,\text{km}$

IV. Conclusion

The calculations with MATLAB 7.1. show that BER_{j_B} is too small and therefore is not displayed on Fig. 2 and Fig. 3. This is confirmed by [7]. The exception is the case of direct sunlight hitting the photodetector. Direct sunlight may cause break of communication link. That can be escape with close to North-South orientation of antennas.

Fig. 2 illustrates the natural decrement of Bit-Error Rate with the meteorological visibility S_M increase. That is strongly expressed in BER_{j_S} because of more weaker influence of quantum noise in case of high signal current values. The value of BER follows BER_{i_S} but is five to six orders greater. Despite too small values of BER_{j_S} in case of

high meteorological visibility, it produces sensible influence on full Bit-Error Rate.

In Fig. 3 BER_{i_S} is practically permanent, while BER_{j_S} has changed in wide number of values. When the dispersion of linear deviation of laser beam σ_ρ^2 has great values, BER_{j_S} vastly exceeds BER_{i_S} and has a decisive role in forming of BER. That can be explained with increasing of quantum noises when the values of signal current are smaller in case of greater linear deviation σ_ρ^2.

From investigations can be concluded that signal current fluctuations due to atmospheric transmittance, mechanical vibrations of antennas and atmospheric turbulence and quantum noises of detection has a great part in Bit-Error Rate forming. Therefore they can be taken into consideration in FSO planning. Influence of quantum fluctuations caused by background current can be ignored, when the suitable orientation of communication system antennas is used.

REFERENCES

[1] Kim, I., R. Stieger, J. Schuster, C. Moursund, B. Riley, J. Koontz, H. Hakakha, M. Barclay, P. Adhikari, E. Korevaar, Portable high-data-rate lasercom system, Military Communications Conference, 1998. MILCOM 98. Proceedings., IEEE, Vol.2, p. 638-642, October 1998.

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

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

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

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

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

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

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

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

[10] Uysal, M., J. Li, M. Yu, Error Rate Performance Analysis of Coded Free-Space Optical Links over Gamma-Gamma Atmospheric Turbulence Channels, IEEE Transactions on Wireless Communications, Vol. 5, No. 6, June 2006.

[11] Uysal, M., J. Li, BER Performance of Coded Free-Space Optical Links over Strong Turbulence Channels, IEEE Vehicular Technological Conference (VTC Spring), Milan, Italy, May, 2004.

[12] Ferdinandov, E., B. Pachedjieva. Formation of Bit-Error Rate in the Free–Space Laser Communication Systems – methods of research., E+E, Vol. 9-10, 2006, p. 12-16, ISSN 0861-4717. (in Bulgarian)

[13] Ferdinandov, E., B. Pachedjieva, B. Bonev, K. Dimitrov. Influence of the Atmospheric Transparency Fluctuations on Bit–Error Rate in ground-to-ground Free–Space Laser Communication Systems. E+E, Vol. 11-12, 2006, p. 46-52, ISSN 0861-4717. (in Bulgarian)

[14] Ferdinandov E., B. Pachedjieva, B. Bonev, Sl. Saparev, Jointly Influence of Heterogeneous Stochastic Factors on Bit-Error Rate in ground-to-ground Free–Space Laser Communication Systems, Optics Communications 207, p.121-127, 2007.

[15] H. Willebrand, B. Ghuman, Free-Space Optics: Enabling Optical Connectivity in Today's Networks, SAMS Publ., Indianapolis, 2002.

[16] Zuev V., G. Matvienko, I. Samokhvalov, Izvestiya Akademii Nayk USSR (seriya Physika Atmosphere and Ocean) 12 (1976) 1243-1250. (In Russian)

Feed Type Influence Over the Bandwidth of Microstrip Patch Antennas

Iva N. Stefanova[1]

***Abstract* –An investigation of the influence of the different type of feed structure for microstrip antennas is made in this paper. A comparison is made of excitation by means of a transmission line, by proximity coupling and by aperture coupling. The analyzed case is that of two parallel-fed rectangular microstrip resonators. The feed structure types that are considered here are compared by means of voltage standing wave ratio.**

***Keywords* – microstrip patch resonators, antenna arrays, feed type structure.**

I. INTRODUCTION

Microstrip patch antennas are dynamic developing section of the antenna field mostly because of their number of advantages – low weight, simple manufacturing, easy installation, aerodynamic profile and naturally low cost. Major disadvantage of the microstrip antennas are their not very good electrical characteristics – low efficiency, low power, poor polarization purity, spurious feed radiation and narrow frequency bandwidth [1]. There are different methods which can be used to increase the efficiency and the bandwidth [2].

The microstrip antenna consist of a very thin metallic strip (patch), placed a small fraction of wavelength above a ground plane. They are separated by a dielectric layer called substrate. The microstrip patch is designed so that its pattern maximum is perpendicular to the patch. This is achieved by proper choosing of the field configuration beneath the patch.

There are numerous dielectric materials that can be used for substrates in microstrip antennas and their dielectric constants most often are in the range of $2 \le \varepsilon_r \le 10$. Most desirable for antennas are thick substrates with dielectric constant in the lower part of the range because they provide higher efficiency, wider bandwidth, but also larger element size [3]. Thin substrates with higher dielectric constants are used in microwave circuitry where tight bounding of the fields is desired in order to minimize the undesired radiation and the element size is smaller. In this case there are greater losses, they are less efficient and have relatively smaller bandwidths. Since microstrip antennas are often integrated with other microwave circuits, a compromise has to be made between good antenna and circuit performance.

The microstrip patch can have various shape [1]. The most often used shapes are rectangular and circular, because they are easy to manufacture and analyze, and have good radiation characteristics.

In this paper the influence of the different type of the feed structure of microstrip patch resonators is investigated. A comparison is made of excitation by means of a transmission line, by proximity coupling and by aperture coupling (fig. 1). The cases considered are that of a single rectangular patch and also that of two parallel-fed rectangular microstrip resonators. The comparison is made by means of voltage standing wave ratio (VSWR).

II. INVESTIGATION OF THE INFLUENCE OF THE FEED TYPE STRUCTURE

When a choice is made of a type of feed structure of microstrip patch antennas, one of the most important considerations is the bandwidth that any of the possible methods can provide. Next comes the possibilities of ensuring a matching (how many states of freedom are there, or how many are the parameters that can be varied in order to achieve a better impedance match). On a third place is the price of manufacturing the given type of feed structure.

The most common feed configurations for microstrip antennas are: with microstrip line, with coaxial probe, proximity coupling and aperture coupling. The microstrip line feed is the first type of feed used for microstrip patches and arrays [4]. It is a conducting feed strip placed above the substrate like the patch and connected to the patch in a position inside its borders (fig 1–a). This method is simple to fabricate, easy to match by controlling the inset position, and simple to model, but has an inherent asymmetry that generates higher order modes which lead to cross-polarized radiation. In order to overcome some of these problems the noncontacting feed methods can be used. The noncontacting feed methods use electromagnetic coupling to send energy between the patch and the feed line. The most common noncontacting feed techniques are the proximity coupling and the aperture coupling. In the both there are two substrates which allow certain independent optimization of the antenna and feed mechanisms, but make the fabrication difficult and the antenna price higher.

The proximity coupling is achieved by placing the patch and the feed structure on different levels. The structure consists of two dielectric substrates as the ground plane is at the bottom of the lower substrate, the patch is on top of the upper substrate and the microstrip feed line is between the two substrates (fig. 1–b). The spurious feed radiation can considerably be reduced by use of thin dielectric with high

[1]Iva N. Stefanova is with the Faculty of Communications and Communication Technologies, Department of Radiotechniques, Technical University – Sofia, 8, Kliment Ohridski street, 1000 Sofia, Bulgaria, E-mail: stefanova.iva@abv.bg

value of dielectric constant, while at the same time the upper substrate can be thicker and with lower dielectric constant so that the radiation of the antenna can be facilitated. To a certain extent the transmission line and the patch can be optimized separately in this way. However a structure with two dielectric substrates is more difficult to analyze because the simple models that were developed for a case of one substrate only cannot be used. The radiation efficiency can be increased by

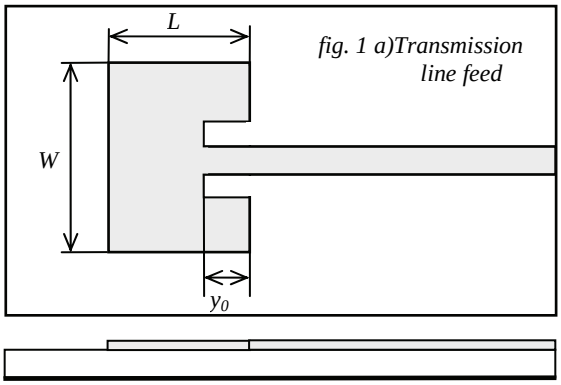


fig. 1 a)Transmission line feed

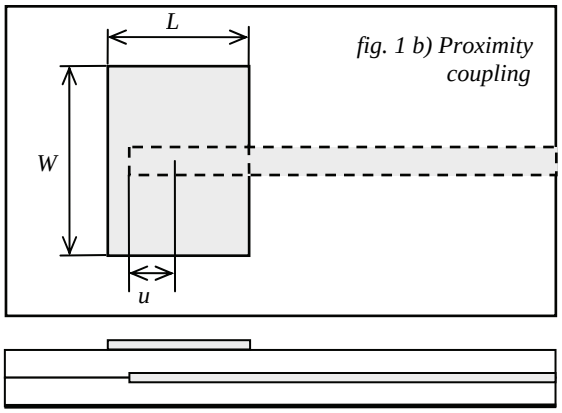


fig. 1 b) Proximity coupling

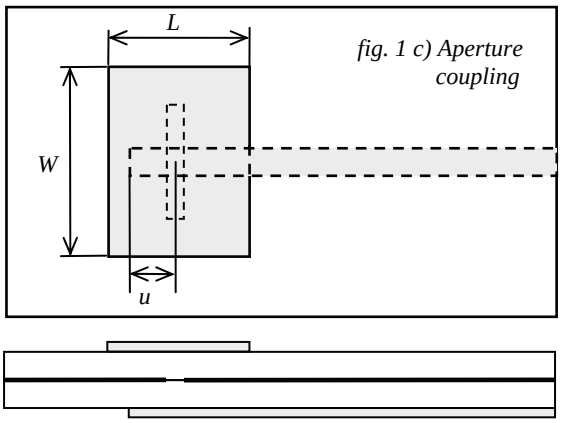


fig. 1 c) Aperture coupling

Fig. 1

increasing the distance between the ground plane and the patch. The bandwidth also depends on this distance, and the overlap of the line and the patch has influence over the resonant frequency [5]. The resulting structure becomes more difficult to manufacture because of the presence of two dielectric substrates instead of one, but there is no need of soldering of different conductors as it is in the case of contact feed methods. A disadvantage is the fact that the transmission feed line is not open and there is not direct access to it, so that elements cannot be added to the feed circuit.

The feed by proximity coupling has somewhat large bandwidth and low spurious radiation. The length of the feeding stub and the width-to-line ratio of the patch can be used to control the match. With this type of feed a bandwidth of up to 13% can be achieved [6].

The aperture coupling is most difficult of all to fabricate and it also has relatively narrow bandwidth. However it is easier to model and has moderate spurious radiation. The aperture coupling consists of two dielectric substrates, separated by a ground plane. On the bottom side of the lower substrate is the transmission feed line and the microstrip patch is on top of the upper substrate (fig 1–c). The energy is coupled between the line and the patch through a slot on the ground plane separating the two substrates [7], [8]. In this way the radiating part and the feed part of the antenna are completely separated. For optimization of the mentioned above parts typically a thick substrate with low dielectric constant is used for the top substrate, and material with high dielectric constant for the bottom substrate. The slot ensures coupling of the patch and the feed line but shouldn't radiate, so that energy won't be dissipated towards the backside of the antenna. The slot dimensions must be chosen so that there is no resonance in the working range, and the slot itself must be situated on enough distance from the edge of the patch.

For feeding by aperture coupling the substrate electrical parameters, feed line width and slot size and position can be used to optimize the design [9]. Usually matching is performed by controlling the feed line and the length of the slot [8]. A change in the substrate dielectric constant can lead to an increase in the electromagnetic coupling and does not change the resonant frequency. Increase in the thickness of one of the dielectrics results in decrease of the electromagnetic coupling, which can be compensated by enlargement of the slot.

In the case of aperture coupling the manufacture is harder because two dielectric substrates are involved, connected by a common ground plane. In the metallization a slot must be cut or etched. However there is no need of breaking through dielectrics or soldering of conductors. The feed line is open microstrip transmission line to which additional components easily can be added.

In this paper a rectangular microstrip resonator, calculated for work on 10 GHz and excited by the mentioned above feed types is investigated. A single patch element is examined as well as two parallel-fed patches, so the mutual influence of the resonators and the feed circuit can be established after comparing the results obtained. A case of two elements is investigated since the antenna array theory shows that the mutual influence in a large number of elements can be accounted in pairs and summed afterwards.

III. RESULTS

There are number of methods for taking account of bandwidth. In this investigation a comparison is made in VSWR on a level of 2, which is a common case often used in

practice. The three types of feed are analyzed by means of electromagnetic simulator that uses for calculation the Finite Elements Method. A multitude of simulations with varying of the physical parameters have been conducted till an optimal match in each of the examined cases has been obtained. On figs. 2 – 7 are shown results that illustrate the obtained conclusions.

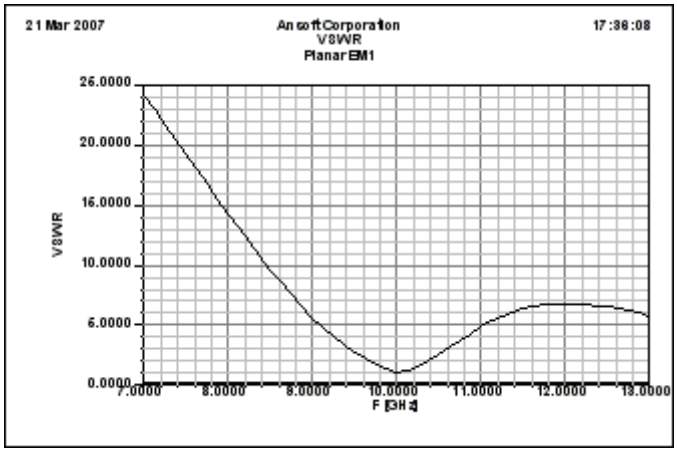


Fig. 2, VSWR, one resonator, transmission line feed.

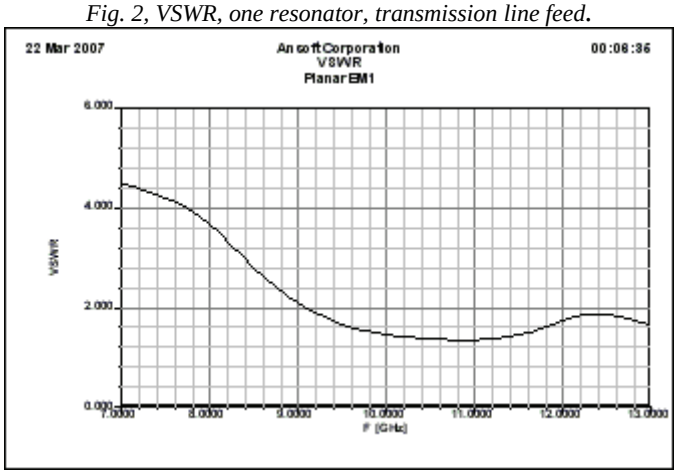


Fig. 3, VSWR, one resonator, proximity coupling.

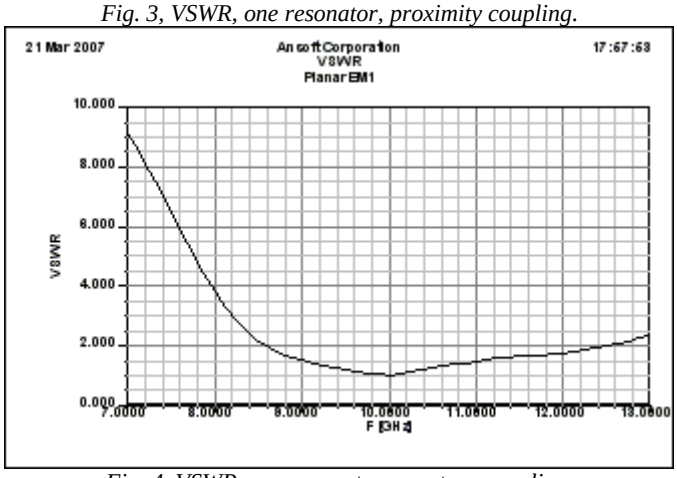


Fig. 4, VSWR, one resonator, aperture coupling.

The voltage standing wave ratio (VSWR) for the three types of feed structure and a single patch are shown on figs 2 – 4. The same types of feed structure, but in the case of two rectangular microstrip parallel-fed patches are shown on figs. 5 – 7.

IV. DISCUSSION

As can be seen from figs. 2 – 4, the results concur with the theoretical investigation. Narrowest bandwidth has the case of transmission line feed, the aperture coupling has considerably larger bandwidth, and largest bandwidth has the case of proximity coupling. The same tendency can be seen in VSWR for two parallel-fed patches (figs. 5 – 7).

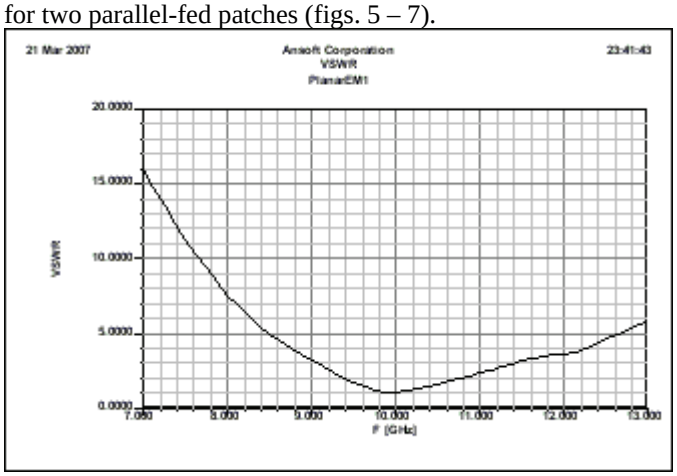


Fig. 5, VSWR, two resonators, transmission line feed.

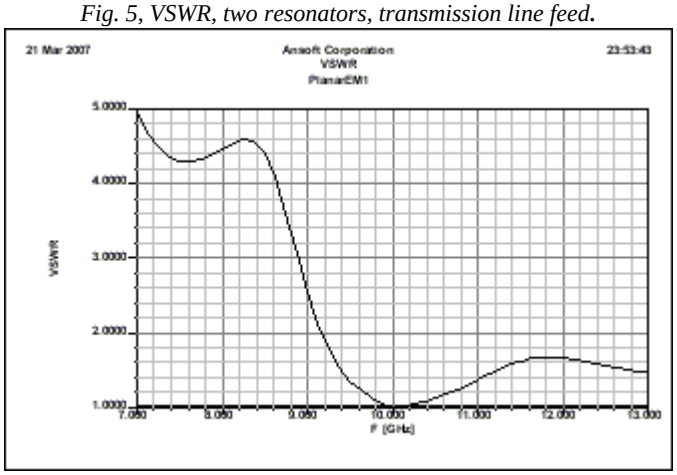


Fig. 6, VSWR, two resonators, proximity coupling.

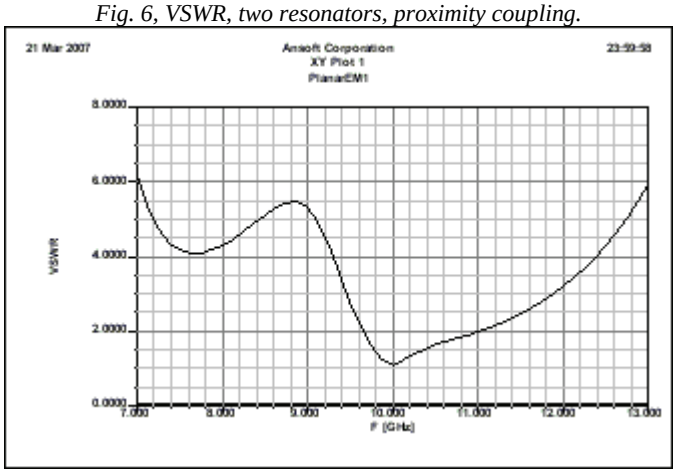


Fig. 7, VSWR ,two resonators, aperture coupling.

When the frequency is changed in the investigated range, the change in VSWR is defined by the change of a number of factors. The first of them is the physical dimensions of the patch (its width and length W and L). The second defining factor is the method used to ensure match – line inset in the

patch y_0 for the case of transmission line feed and line inset u under the patch for the other two cases, as well as the slot dimensions for the case of aperture coupling.

Third defining factor is the mutual influence of the two resonators which can be expressed by means of the mutual resistance R_{12} [1]. It can be accounted that some of the factors mentioned above have contradictory influence – that is, while some of them lead to increase in the resistance, others have the opposite effect. The tendency the bandwidth to become narrower with increase of the number of the elements, that can be expected, is not observed in the case of transmission line feed (figs. 2 and 5).

V. Conclusion

The aim of the present investigation is to evaluate the bandwidth of two parallel-fed rectangular microstrip resonators for three types of the feed structure – transmission line feed, proximity coupling and aperture coupling. The influence of the parallel position of the two patches is taken into account.

From the results obtained in this study it can be seen that the optimal case which combines simplicity of manufacture and maximum bandwidth appears to be the case of proximity coupling. The achieved results can be used during the selection of the feed type structure in microstrip antenna array design.

References

[1] C. A. Balanis, Willey, J., *Antenna Theory (Analysis and Design)*, New York, 1998;

[2] D. M. Pozar, "A Review of Bandwidth Enhancement Techniques for Microstrip Antennas" in Microstrip Antennas, The Analysis and Design of Microstrip Antennas and Arrays, D. M. Pozar and D. H. Schaubert (Eds.), New York, IEEE Press, pp157-166, 1995;

[3] K. R. Carver, and J. W. Mink, "Microstrip Antenna Technology", IEEE Trans. Antenna Propaga., vol. AP-25, no.1, pp. 2-24, Jan. 1981;

[4] R. E. Munson, "Conformal Microstrip Antennas and Microstrip Phased Arrays", IEEE Transactions on Antennas and Propagation, vol. 22, pp. 74-78, January 1974;

[5] H. G. Oltman and D. A. Huebner, "Electromagnetically Coupled Microstrip Dipoles", IEEE Transactions on Antennas and Propagation, vol. 29, no. 1, pp. 151-157, January 1981;

[6] D. M. Pozar and B. Kaufman, "Increasing the Bandwidth of a Microstrip Antenna by Proximity Coupling", Electronic letters, v. 23, no. 8, pp. 368-369, April 1987;

[7] G. Gronau and I. Wolff, "Aperture-Coupling of a Rectangular Microstrip Resonator", Electronics letters, v. 22, pp. 554-556, May 1986;

[8] P. L. Sullivan and D. H. Schaubert, "Analysis of an Aperture-Coupled Microstrip Antenna", IEEE Transactions on Antennas and Propagation, vol. 34, no. 8, pp. 977-984, August 1986;

[9] R. Oostlander, Y. M. M. Antar, A. Ittipiboon and M. Cuhaci, "Aperture Coupled Microstrip Antenna Element Design", Electronics letters, v. 26, no. 4, pp. 224-225, February 1990.

Modeling of Multiconductor Shielded Microstrip Lines

Sarhan M. Musa[1] and Matthew N. O. Sadiku[2]

Abstract – **This paper presents a modeling of multiconductor transmission lines using COMSOL, a finite element package. We determined the capacitance coupling of the three-strip, six-strip, and eight-strip shielded microstrip lines. We compared our results with those obtained by other methods and found them to be close.**

Keywords – **Multiconductor Shielded microstrip line, Capacitance per unit length, Modelling and simulation.**

I. INTRODUCTION

Today, electromagnetic propagation on multiple parallel transmission lines has been a very attractive area in computational electromagnetics. Multiple parallel transmission lines have been successfully applied and used by designers in compact packaging, semiconductor device, high speed interconnecting buses, monolithic integrated circuits, and other applications.

Many authors have already calculated the multiconductor line capacitances using different approaches. Those approaches include analytical approach [1-2], the method of lines [3], finite element method [4-5], quasi-TEM spectral domain analysis [6], integral-equation computer-solution technique [7-9], conformal mapping technique [10], and the moment method [11].

In this paper, we consider systems of three-strip, six-strip, and eight-strip (multiconductor) shielded microstrip lines. Using COMSOL, a finite element package, we performed the simulation of these systems of microstrip lines.

II. CAPACITANCE MODEL

For coupled multiconductor microstrip lines, it is convenient to write as [12-13]:

$$Q_i = \sum_{j=1}^{m} C_{sij} V_j \quad \text{(i = 1, 2, ….., m)} \qquad (1)$$

where Q_i is the charge per unit length, V_j is the voltage of j th conductor with reference to the ground plane, C_{sij} is the short circuit capacitance between i th conductor and j th conductor. The short circuit capacitances can be obtained either from measurement or from numerical computation. From the short circuit capacitances, we obtain

$$C_{ii} = \sum_{j=1}^{m} C_{sij} \qquad (2)$$

where C_{ii} is the capacitance per unit length between the i th conductor and the ground plane. Also,

$$C_{ij} = -C_{sij}, \qquad j \neq i \qquad (3)$$

where C_{ij} is the coupling capacitance per unit length between the i th conductor and j th conductor. The coupling capacitances are illustrated in Fig. 1.

For m -strip line, the per-unit-length capacitance matrix is given by

$$C = \begin{bmatrix} C_{11} & -C_{12} & \cdots & -C_{1m} \\ -C_{21} & C_{22} & \cdots & -C_{2m} \\ \vdots & \vdots & & \vdots \\ -C_{m1} & -C_{m2} & \cdots & C_{mm} \end{bmatrix} \qquad (4)$$

Also, we can determine the characteristic impedance matrix for m -strip line by

$$Z_o = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1m} \\ Z_{21} & C_{22} & \cdots & Z_{2m} \\ \vdots & \vdots & & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mm} \end{bmatrix} \qquad (5)$$

where Z_o is the characteristic impedance per unit length.

III. RESULTS AND DISCUSSION

Using COMSOL for modeling and simulation of the lines involves taking the following steps:

1. Develop the geometry of the line, such as shown in Fig. 2.
2. We take the difference between the conductor and dielectric material
3. We select the relative permittivity as 1 for the difference in step 2.
4. We add a dielectric region under the inner conductors with specified relative permittivity.
5. For the boundary, we select the outer conductor as ground and the inner conductors as ports.
6. We generate the finite element mesh, and then we solve the model.
7. As postprocessing, we select Point Evaluation and choose capacitance elements to find the coupling capacitance per unit length of the line.

Authors are with Prairie View A&M University, Houston, USA
E-mail: smmusa@pvamu.edu; sarhanhamed@gmail.com

These steps were taken for the following three cases.

A.Three-strip line

Figure 2 shows the cross section for three-strip line with the following parameters:

a = width of the outer conductor = 13 mm
b = height of the free space region (air) = 4 mm
h = height of the dielectric region = 2 mm
w = width of each inner strip = 2 mm
t = thickness of each inner strip = 0.01 mm
D = distance between the outer conductor and the first strip = 2.5 mm
s = distance between two consecutive strips = 1 mm

Figure 3 shows the finite element mesh, while Fig. 4 illustrates the potential distribution along line y = h. Table I shows the finite element results for the three-strip line. Unfortunately, we could not find any work in the literature to compare our results.

TABLE I
CAPACITANCE VALUES (IN p F / m) FOR THREE –STRIP SHIELDED MICROSTRIP LINE

Methods	C_{11}	C_{21}	C_{31}
COMSOL[16]	163.956	-27.505	-0.4301

B. Six-strip line

Figure 5 shows the cross section for six-strip line with the following parameters:

a = width of the outer conductor = 15 mm
b = height of the free space region (air) = 2 mm
h = height of the dielectric region = 8 mm
w = width of each inner strip = 1 mm
t = thickness of each inner strip = 0.01 mm
D = distance between the outer conductor and the first strip = 2 mm
s = distance between two consecutive strips = 1 mm

Figure 6 shows the finite element mesh, while Fig. 7 depicts the potential distribution along line y = h. The capacitance values for six-strip shielded microstrip line are compared with other methods as shown in Table II, where "iterative" refers to an iterative method [14] and ABC refers to the asymptotic boundary condition [15]. It is evident from the table that the finite element methods based on [4] and [16] closely agree. The finite element methods seem to be more accurate than the iterative and ABC techniques.

C. Eight-strip line

Figure 8 shows the cross section for eight-strip line with the following parameters:

a = width of the outer conductor = 175 mm
b = height of the free space region (air) = 100 mm
h = height of the dielectric region = 16 mm
w = width of each inner strip = 1 mm
t = thickness of each inner strip = 0.01 mm
D = distance between the outer conductor and the first strip = 80 mm
s = distance between two consecutive strips = 1 mm

Figure 9 shows the finite element mesh, while Fig. 10 depicts the potential distribution along line y = 20mm. The capacitance values for eight-strip shielded microstrip line are compared with other methods as shown in Table III, where other authors used analytic approach [17] and Fourier series expansion [18]. It is evident from the table that the results from finite element method (COMSOL) closely agree with the analytic approach.

IV. CONCLUSION

In this paper, we have presented the modeling of three-strip, six-strip, and eight-strip (multiconductor) shielded microstrip lines. The results obtained using COMSOL for the coupling capacitance per unit length agree well those found in the literature.

REFERENCES

.
[1] D. Homentcovshi et al., "An analytic solution of the coupled stripline-like microstrip line problem," *IEEE Transactions on Microwave Theory and Techniques*, vol. 36, no. 6, June 1988, pp. 1002-1007.
[2] D. Homentcovshi, G. Ghione, and C. Naldi, "Analytic determination of the capacitance matrix of planar or cylindrical multiconductor lines on multilayered substrates," *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, no. 2, Feb. 1995, pp. 363-373.
[3] M. Davidovitz, "Calculation of multiconductor microstrip line capacitances using the semidiscrete finite element method," *IEEE Microwave and Guided Wave Letters*, vol. 1, no. 1, Jan. 1991, pp. 5-7.
[4] A. Khebir, A. B. Kouki, and R. Mittra, "High order asymptotic boundary condition for the finite element modeling of two-dimensional transmission line structures," *IEEE Transactions on Microwave Theory and Techniques*, vol. 38, no. 10, Oct. 1990, pp. 1433-1438.
[5] Z. Pantic and R. Mittra, "Quasi-TEM analysis of microwave transmission lines by the finite-element method," *IEEE Transactions on Microwave Theory and Techniques*, vol. 34, no. 11, Nov. 1986, pp. 1096-1103.
[6] E. Drake, F. Medina, and M. Horno, "Improved quasi-TEM spectral domain analysis of boxed coplanar multiconductor microstrip lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, no. 2, Feb. 1993, pp. 260-267.
[7] A. E. Ruehli and P. A. Brennan, "Efficient capacitances for three-dimensional multiconductor systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 21, no. 2, Feb. 1973, pp. 76-82.

[8] W. T. Weeks, "Calculation of coefficients of capacitance of multiconductor transmission lines in the presence of a dielectric interface," *IEEE Transactions on Microwave Theory and Techniques*, vol. 18, no. 1, Feb. 1970, pp. 35-43.

[9] W. Delbare and D. De Zutter, "Space-domain Green's function approach to the capacitance calculation of multiconductor lines in multilayered dielectrics with improved surface charge modeling, *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 10, Oct. 1989, pp. 1562-1568.

[10] C. Wan, "Analytically and accurately determined quasi-static parameters of coupled microstrip lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, no. 1, Jan. 1996, pp. 75-80.

[11] C. Wei et al, "Multiconductor transmission lines in multilayered dielectric media," *IEEE Transactions on Microwave Theory and Techniques*, vol. 32, no. 4, April 1984, pp. 439-450.

[12] M. S. Lin, "Measured capacitance coefficients of multiconductor microstrip lines with small dimensions," *IEEE Transactions on Microwave Theory and Techniques*, vol. 13, no. 4, Dec. 1990, pp. 1050-1054.

[13] F. Y. Chang, "Transient analysis of lossless coupled transmission lines in a nonhomogeneous dielectric media," *IEEE Trans. Microwave Theory Tech.*, vol 18, no 9, pp. 616-626, Aug. 1970.

[14] P. N. Harms, C. H. Chan, and R. Mittra, "Modeling of planar transmission line structures for digital circuit applications," Arch. Eleck. Ubertragung., vol. 43, 1989, pp. 245-250.

[15] A. Kherbir, A. B. Kouki, and R. Mittra, "Absorbing boundary condition for quasi-TEM analysis of microwave transmission lines via the finite element method," *J. Electromagn. Waves and Appl.*, vol. 4, no. 2, 1990.

[16] www.comsol.com

[17] D. Homentcovschi, G. Ghione, C. Naldi, and R. Oprea," Analytic determination of the capacitance matrix of planar or cylindrical multiconductor lines," *IEEE Trans. Microwave Theory Tech.*, pp. 363-373, Feb. 1995.

[18] M. K. Amirhosseini," Determination of capacitance and conductance matrices of lossy shielded coupled microstrip transmission lines," *Progress In Electromagnetics Research*, PIER 50, pp. 267-278, 2005.

TABLE II
CAPACITANCE VALUES (IN p F / m) FOR SIX –STRIP SHIELDED MICROSTRIP LINE

Methods	C_{11}	C_{21}	C_{31}	C_{41}	C_{51}	C_{61}
Iterative [14]	66.8	-27.9	-5.49	-2.08	-0.999	-0.704
Finite Element [4]	84.8	-26.4	-3.71	-1.17	-0.456	-0.812
ABC [15]	68.6	-31.5	-6.00	-2.25	-0.792	-0.602
COMSOL [16]	80.4	-23.9	-3.61	-1.15	-0.451	-0.180

TABLE III
CAPACITANCE VALUES (IN p F / m) FOR EIGHT –STRIP SHIELDED MICROSTRIP LINE

Method	C_{11}	C_{21}	C_{31}	C_{41}	C_{51}	C_{61}	C_{71}	C_{81}
Analytic [17]	127.776	-58.446	-13.024	-5.721	-3.104	-1.892	-1.282	-1.211
Fourier series [18]	126.149	-57.066	-12.927	-5.684	-3.086	-1.875	-1.264	-1.185
COMSOL [16]	128.204	-58.759	-13.064	-5.739	-3.1206	-1.902	-1.290	-1.226

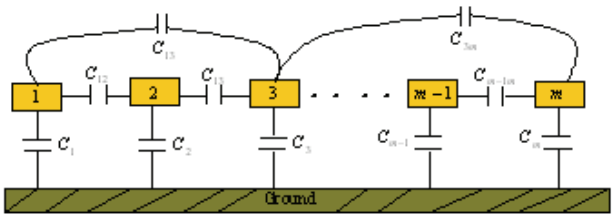


Fig. 1 The per-unit length capacitances of a general m -conductor transmission line.

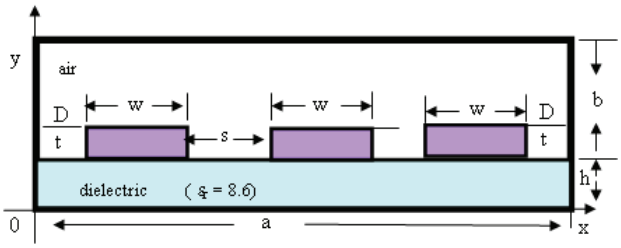


Fig. 2 Cross-section of the three-strip transmission line.

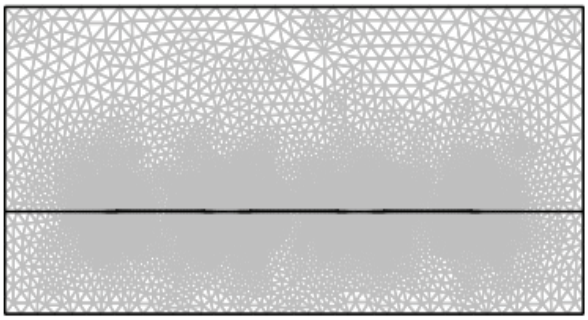


Fig. 3 Mesh for the three-strip transmission line.

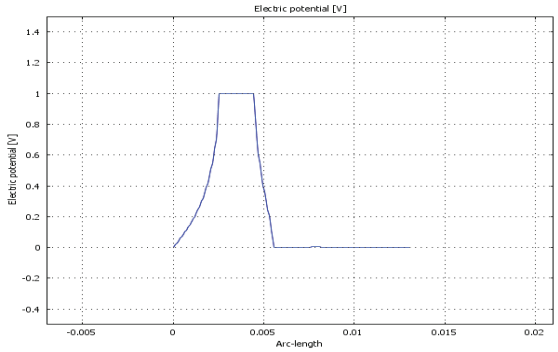


Fig. 4 Potential distribution along the air-dielectric interface (y = h) for the three-strip transmission line.

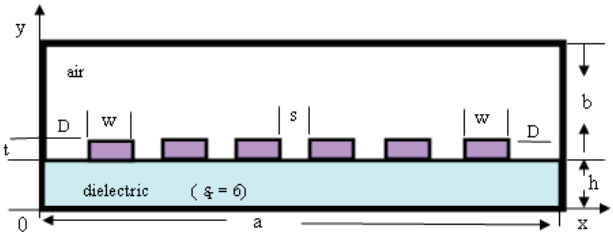


Fig. 5 Cross-section of the six-strip transmission line.

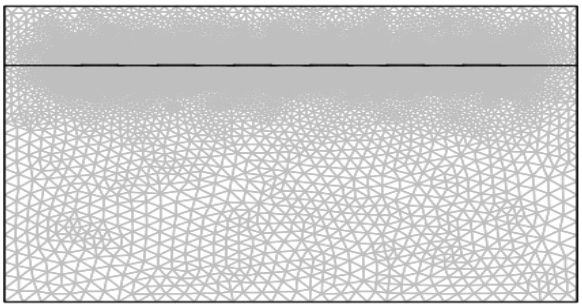


Fig. 6 Mesh for the six-strip transmission line.

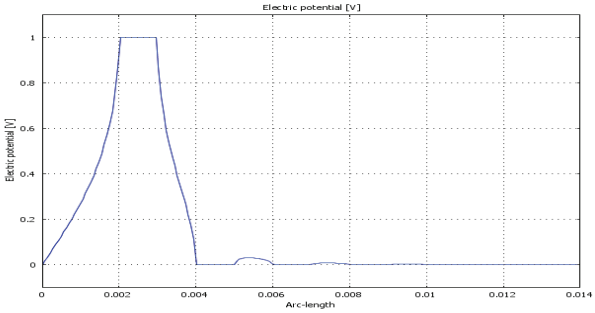


Fig. 7 Potential distribution along the air-dielectric interface (y = h) for the six-strip transmission line.

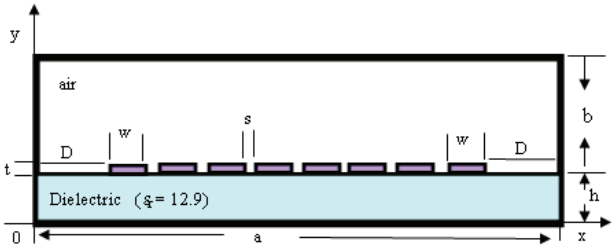


Fig. 8 Cross-section of the eight-strip transmission line.

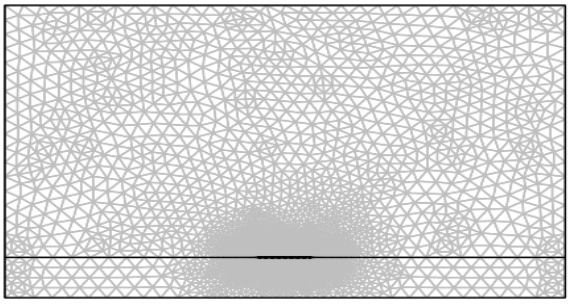


Fig. 9 Mesh for the eight-strip transmission line.

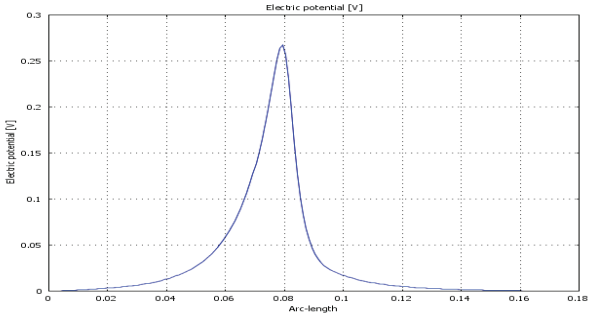


Fig. 10 Potential distribution along the air-dielectric interface (y = 20 mm) for the eight-strip transmission line.

Substrate's Parameters Influence over the Directivity and Bandwidth at Microstrip Planar Arrays

Nikola I. Dodov[1] and Mario O. Krastev[2]

***Abstract* – In this paper are studied dependencies on dielectric substrate's parameters over the directivity and bandwidth at microstrip planar antenna arrays. It has been provided numerical calculations based on theoretical models for microstrip rectangular patch antennas.**

***Keywords* – Microstrip antennas, Dielectric substrate, Bandwidth, Directivity**

I. INTRODUCTION

Microstrip antennas are widely used in modern radio communication systems. They are preferred for antenna arrays in microware range (3-30 GHz) because of their small size, inexpensive and easy manufacture with printed-circuit technologies. Their main disadvantage is relatively narrow bandwidth typically about 5-6% [1,3,4,5], which is predetermined by resonant characteristic of radiating element. All basic characteristics of microstrip antennas depend on dielectric substrate's parameters [1-5]. This study describes mathematical models for analysis of microstrip antennas with rectangular patch. Simulative calculations based on these models are provided for determination of the bandwidth and directivity of microstrip antennas and their variation provoked by changes of dielectric substrate's parameter. It was explored the influence of dielectric substrate's height and permittivity over basic antenna characteristics.

II. METHODS OF TEORETICAL ANALYSIS FOR MICROSTRIP ANTENNAS

There are many theoretical methods of analysis for microstrip antennas, which are used depending on the specific purpose of examination. In general the antennas could be most properly described by using full-wave model. It includes finding the solutions of primarily electromagnetic field equations for near and far zone at particular defined conditions. In almost all cases calculations are based on Moment Method. This method gives very accurate results, can be applied for all types of electrodynamics calculations but it is very complicated, difficult to implement and usually provides less physical insight for the processes. That is the reason why in most cases are used other models, which also provides good and accurate results and are easier to implement. Most popular practical methods of analysis for microstrip antennas are transmission-line and cavity models [1,3,4,5]. They are relatively simple and give good physical interpretation of the electromagnetic processes in the antennas but less accurate results than full-wave model. In this section are presented main consideration of transmission-line and cavity model for microstrip antenna with rectangular patch (Fig.1).

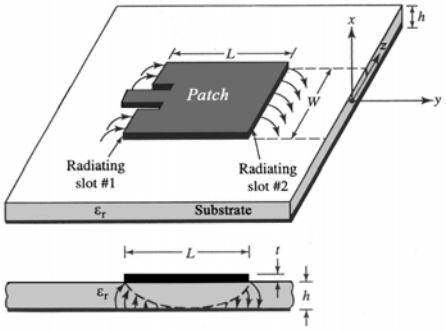


Fig.1. Microstrip antenna with rectangular patch

According the cavity model microstrip antenna is dielectric loaded cavity, which excite higher order resonances (Fig.2). In this model, the interior region of the patch is modeled as a cavity bounded by electric wall on the top and the bottom, and magnetic walls along the periphery. This is an approximate model which in principle leads to reactive input impedance and it does not radiate any power. This is the reason why there are introduced losses in the cavity. Results computed by this model are in good agreement with measurements [1],[5].

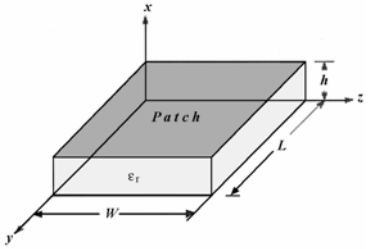


Fig.2 Cavity model – rectangular patch geometry

Analyses of the microstrip patch show that the type of electromagnetic field modes in the cavity is TM^{x}_{mnp}. The dominant mode is TM^{x}_{010} where x is direction normal to the patch according the geometry shown on Fig.2. The antenna dimensions are respectively W - width, L - Length, h - height.

[1] Nikola I. Dodov is with the Faculty of Communications and Communications Technologies, TU-Sofia, 1000 Sofia, Bulgaria. E-mail: ndodov@tu-sofia.bg

[2] Mario O. Krastev is with the Faculty of Communications and Communications Technologies, TU-Sofia, 1000 Sofia, Bulgaria. E-mail: mkrastev@tu-sofia.bg

Using The Field Equivalence Principle the four side slots of microstrip patch could be represented as radiating equivalent magnetic dipoles (Fig.3). The components of the electromagnetic fields, radiated from two of the slot separated by the length of the patch L≈λ/2, in a direction perpendicular to the ground plane add in phase and give a maximum radiation. So these slots are referred as radiating slot. The fields radiated by other two of the slots, which are separated by the width of the patch W are canceled along the normal plane because the equivalent magnetic current densities on these sides are with same amplitude but in opposite phase. So these two slots are called non-radiating. On the Fig.3 are shown the equivalent current densities and field distribution for radiating slot for dominant mode TM^{x}_{010}. One drawback of the cavity model is that it does not take into account fringing from the end of the patch.

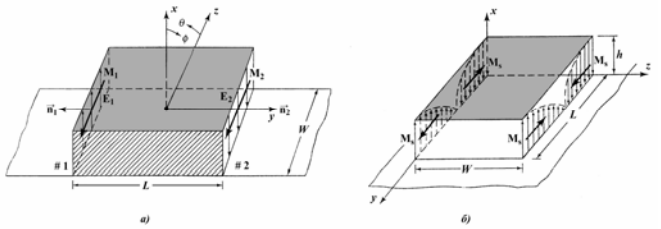


Fig. 3. Rectangular microstrip patch radiation and non-radiating slots at TM^{x}_{010} mode

Transmission Line model is the easiest and most popular of all other models. It gives relatively less accurate results, but very clears physical interpretation of the processes and is widely used for calculation of patch dimensions. According transmission line model microstip antenna is represented by two radiating slots separated by a low-impedance transmission line with length L. The radiating slots are equivalent magnetic dipoles, which form two elements antenna array. This model allows to work out good practical results for dependency between substrate's parameters (ε_r, h) and geometrical dimension of the patch (W, L). It also take into account fringing effects at the edges of the patch by introducing effective length L_{eff} and effective dielectric constant ε_{reff}. Transmission-line model is simple and with some consideration give nice practical results.

III. SHORT DESCRIPTION OF THE EXAMINATION

This study is done for microstrip rectangular patch antenna at central working frequency 10 GHz and different parameters of dielectric substrate. The intervals of variation for dielectric constant and height of substrate are as follow:

- 0,5 mm < h < 2 mm
- 2 < ε_r < 10

In these intervals are chosen 400 discrete equidistance points for substrate's height and permittivity. Therefore the steps of parameter's changes are:

$\Delta\varepsilon_r = 0{,}02$

$\Delta h = 0{,}004\ mm$

These steps are small enough and allow very clearly studying of the variation of antenna's directivity and bandwidth with the substrate's parameters changes. The delta steps are practically so small that reaches technological tolerance for dialectical material parameters. For each two discrete values of ε_r and h are calculated the antenna parameters. For these calculations are used theoretical equations based on models described in previous section.

For the purpose of examination are accepted copper patch metallization and very low loss dielectric materials with tangent of dielectric losses $tg\delta \le 9.10^{-4}$

Antenna's Directivity is defined as the ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions. If the direction is not specified it is implied the direction of maximal radiation intensity. The average radiation intensity is equal to the total power radiated by the antenna divided by 4π. Therefore the directivity of antenna is the ration of its radiation intensity in direction of the maximum radiation over that of an isotropic source. It could be mathematically expressed [1] as:

$$D = \frac{4\pi U_{\max}}{P_{rad}} \tag{1}$$

where

U_{max}, W/sr – maximum radiation intensity

P_{rad},W – total radiated power from the antenna

For a single slot total radiated power P_{rad} and maximum radiated intensity U_{max} is computed by using equivalence principle (according cavity model) and representing the slot as equivalent magnetic dipoles [1],[4],[5]:

$$P_{rad} = \frac{|V_0|^2}{2\pi Z_0}\int_0^{\pi}\left[\sin\left(\frac{\pi W}{\lambda_0}\cos\theta\right)\Big/\cos\theta\right]^2 \sin^3\theta d\theta \tag{2}$$

$$U_{\max} = \frac{|V_0|^2}{2\pi Z_0}\left(\frac{\pi W}{\lambda_0}\right) \tag{3}$$

where

V_0, Volts - equivalent voltage of the radiating slot

Z_0, Ω - characteristic (free space) impedance

Now using the transmission-line model the microstrip patch can be represented as two elements antenna array with distance between elements L≈λ/2. Consequently according antenna array theory the total directivity is:

$$D = \left(\frac{2\pi W}{\lambda_0}\right)^2 \frac{\pi}{I} \tag{4}$$

$$I = \int_0^{\pi}\int_0^{\pi}\left[\frac{\sin(0{,}5kW\cos\theta)}{\cos\theta}\right]^2 \sin^3\theta \times \cos^2(0{,}5kL_e\sin\theta\cos\phi)d\theta d\phi \tag{5}$$

The bandwidth of an antenna is defined as range of frequencies within which the performance of the antenna

conforms to specified requirements. For narrowband antennas bandwidth is expressed as a percentage of the frequency difference over the central working frequency. The antenna bandwidth is inversely proportional of its total quality factor Q. Therefore to define the bandwidth at first should be computed total quality factor.

Total quality factor is characteristic which represent the power transformation in antenna system. It is connected with losses mechanism. Typically there are radiations, conduction (ohmic), dielectric and surface wave losses. So the total quality factor Q is depended on all these losses and in general is written as [1],[3],[4],[5]:

$$\frac{1}{Q} = \frac{1}{Q_{rad}} + \frac{1}{Q_c} + \frac{1}{Q_d} + \frac{1}{Q_{sw}} \tag{6}$$

where different quality factors could be computed by using the following formulas [1,3]:

- Q_{rad} - quality factor due to radiation losses:

$$Q_{rad} = \frac{2\omega_0 \varepsilon_r W}{hG_{rad}}.K \tag{7}$$

- Q_c - quality factor due to conduction losses:

$$Q_c = h\sqrt{\pi f \mu \sigma} \tag{8}$$

- Q_d - quality factor due to dielectric losses:

$$Q_d = 1/tg\delta \tag{9}$$

- Q_{sw} - quality factor due to surface wave losses:

$$Q_{sw} = \frac{\omega_0 W_t}{P_{sur}} \tag{10}$$

In the formulas above there are used the following parameters:

σ, [S/m] – specific conductance of the patch;

$tg\delta$ – tangent of dielectric losses in substrate;

P_{sur}, [W] – power of surface waves;

ω_0,=$2\pi f_0$, [rad/s] – resonant frequency;

W_t, [W] – total energy stored in the cavity for one period, which for rectangular patch is:

$$W_t = \frac{1}{4}\varepsilon_0 \varepsilon_r V = \frac{1}{4}\varepsilon_0 \varepsilon_r hLW \tag{11}$$

G_{rad}, [S] - Radiation conductance of the antenna, which is computed by using expression for radiated power P_{rad} (2) and antenna array factor, according transmission line model [1],[3]:

$$G_{rad} = \frac{1}{30\pi^3}\int_0^{\pi}\int_0^{\pi}\left[\frac{\sin(0{,}5kW\cos\theta)}{\cos\theta}\right]^2 \times \sin^3\theta\cos^2(0{,}5kL_e\sin\theta\cos\phi)d\theta d\phi \tag{12}$$

The parameter K for a rectangular patch at the dominant mode is:

$$K = \iint_{Area}|E|^2 dA \Big/ \oint_{Perimeter}|E|^2 dl = 0{,}25L \tag{13}$$

For thin substrates ($h<<\lambda$) the losses due to surface wave are negligible and in most of cases are not taken into account. However for thicker substrates they need to be considered and the power of surface waves P_{sur} could be computed [3].

The bandwidth of the antenna Δf is inversely proportional to its total quality factor and can be defined as [1],[3],[4],[5]:

$$\Delta f = f_0/Q_t \tag{14}$$

This expression is approximate and is not useful for real purposes because does not take into account impedance matching between the antenna and feeder line. For practical calculations is used definition of the bandwidth where the VSWR (Voltage Standing Wave Ratio) at the feeder line is equal or less than desired maximum value. Therefore the antenna bandwidth is calculated by the formula:

$$\Delta f = f_0 \frac{VSWR-1}{Q_t\sqrt{VSWR}} \tag{15}$$

For practical usage the antenna bandwidth is measured at VSWR=2, which means that 90% of the transmitter power reached the antenna and the rest 10% reflected back. Therefore the bandwidth is measured on level $1/\sqrt{2}$ from the maximum:

$$\Delta f = \frac{f_0}{Q_t\sqrt{2}} \tag{16}$$

IV. RESULTS

The main purpose of this study is to find out numerical relations between substrate's parameters (height and permittivity) and one of the most important antenna characteristics - the directivity and the bandwidth. Major results of examination are shown on the figures below (Figs. 4-7). They illustrate the changes of bandwidth and directivity caused by dielectric substrate's parameters changes.

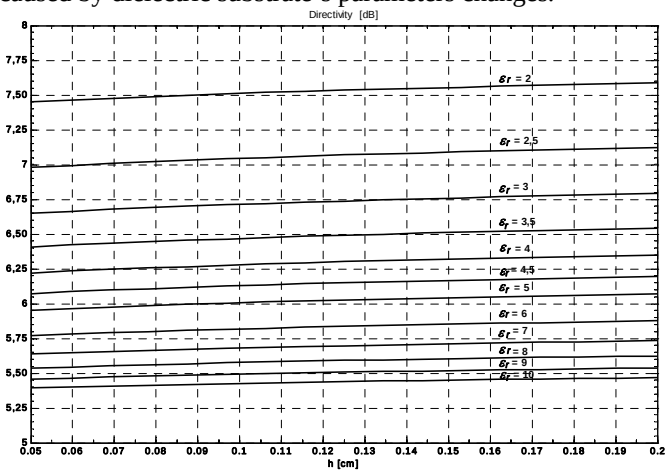


Fig. 4. Influence of substrate's height h over the Antenna Directivity

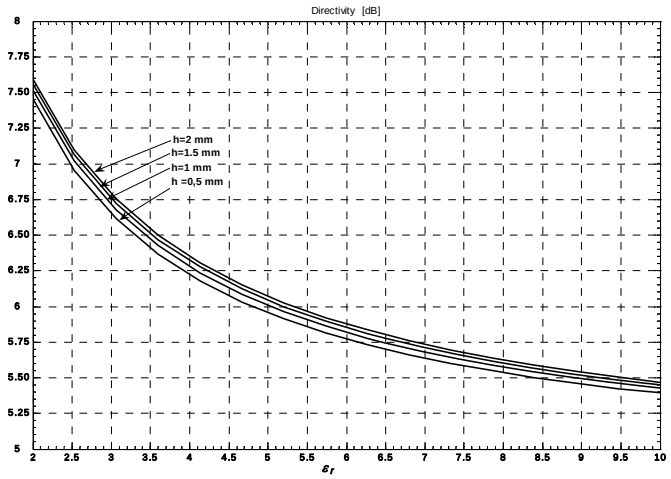


Fig. 5. Influence of substrate's permittivity ε_r over the Antenna Directivity

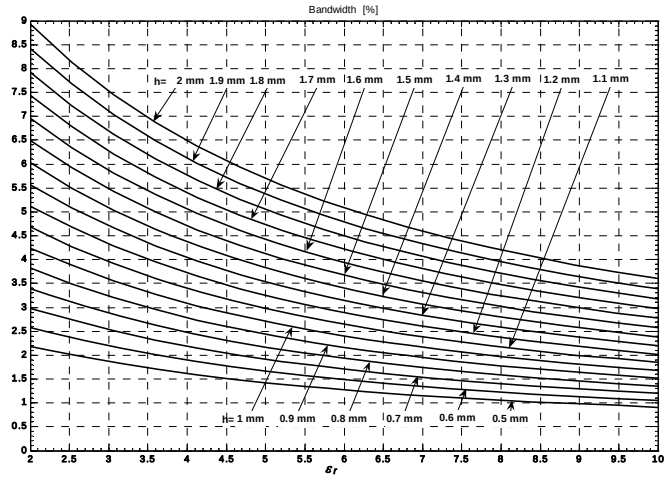


Fig. 6. Influence of substrate's permittivity ε_r over the Antenna Bandwidth

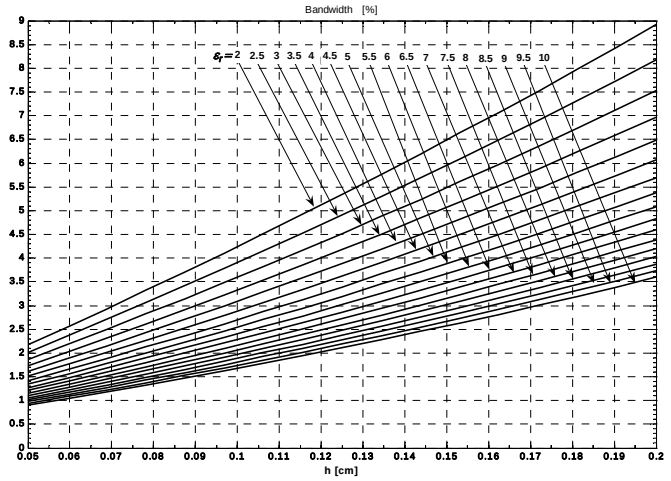


Fig. 7. Influence of substrate's height *h* over the Antenna Bandwidth

V. CONCLUSION

As a result of examination could be taken out many conclusions about dielectric substrate's parameters impact on microstrip antenna characteristics. It is possible to study out not only changes of bandwidth and directivity, but also changes of geometrical dimension, variation of fringing effects and quality factor. Analyses of the results lead to the following main conclusions:

- The antenna directivity is greater when materials with low value of dielectric constant ε_r (about 2÷3) are used. An increase of ε_r will concentrate more of the power in the patch and decrease the radiation, but at the same time leads to smaller size of the antenna. Variation of directivity in the examined interval $2<\varepsilon_r<10$ is about 25 %.
- Usage of thicker substrates leads to small increase of Directivity, which about 2% for whole explored interval.
- Directivity is to a large extent more dependent on dielectric material respectively ε_r than on substrate's height.
- Substrate's parameters influence over the bandwidth is more or less like over directivity. When the height is increased the bandwidth also increases with close to linear function. The change of bandwidth depending on substrate's height is relatively large - from 0,9% to 9% at different values of ε_r.
- The bandwidth is more sensitive to ε_r. When the dielelctric constant is increasing in diapason $2<\varepsilon_r<10$ the bandwidth is narrowed approximately with 50%.
- General conclusion is that maximum directivity and bandwidth are achieved with thin substrates and materials with low dielectric constant at the expense of bigger antenna size.
- Thicker substrates and low dielectric constant lead to extension of patch length (about 15%) and width (about 45%).
- The thicker substrates improve the bandwidth and directivity, but allow excitation of surface waves. Influence of surface waves could be neglected. For a single patch it leads to probable error of 1-2 %, but have to be taken into account at the antenna arrays.

This research make possible to be illustrated physical processes in the microstrip patches. The results of the examination allows making correct choose of materials for dielectric substrates with a view to achieve specific parameters of printed antennas and antenna arrays.

REFERENCES

[1] Balanis C.A., Antenna Theory (Analysis and Design), John Willey and Sons Inc., New York, 1998

[2] Dodov N.I., P.Zh.Petkov, The Surface Waves' Impact on the Coupling Effect in Microstrip Antennas, ICEST 2002, October 1-4, 2002, Niš, Yugoslavia

[3] Garg Ramesh, Prakash Bhartia, Inder Bahl, Apisak Ittpoboon, Microstrip Antenna Design Handbook, Artech House Inc., Boston, 2000

[4] Stutzman Warren, Gary Thiell, Antenna Theory and Design, John Willey and Sons Inc., New York, 1998

[5] Pozar, M.D., D.H.Schaubert, The Analysis and Design of Microstrip Antennas and Arrays, IEEE press, New York, 1996.

Modelling of Losses in a Microwave Applicator Using TLM Method

Jugoslav Joković [1], Bratislav Milovanović [1] and Tijana Dimitrijević [1]

Abstract - **In this paper, a microwave applicator based on loaded rectangular cavity is analyzed by using 3-D TLM software in order to investigate effects of the conducting and dielectric losses on EM field in the cavity.**

Keywords – **Microwave Applicator, Cavity, Losses, Waveguide, TLM**

I. INTRODUCTION

The utilization of microwaves in industry has led to the development of a number of microwave applicators in the processes of dielectric material heating and drying. A rectangular metallic cavity represents configuration very suitable for adequate modelling of some practical heating and drying applicators. The knowledge of the mode tuning behaviour under load and feed conditions has important significance and is helpful in designing these applicators. For that reason, a number of authors presented their researches of the metallic cavities, based on different approaches [1, 2].

When practical realization of microwave applicators is concerned, one of the most important issues is the resonant modes distribution inside the metallic cavity, in order to achieve equally material drying. Regarding applicator design process, the cavity walls and dielectric characteristics modelling has practical significance. As there is no analytical solution for the most cases of widely used loaded cavities, computational electromagnetic techniques emerge as an invaluable tool in the applicator design.

TLM (Transmission Line Modelling) method is a general, electromagnetically based numerical method that has been applied successfully to the wide range of problems. In that context, it has been applied to the modelling of metallic cavities loaded with dielectrics with different geometrical and electromagnetical properties [3, 4]. For establishing desired field distribution within the modelled cavity an impulse excitation of corresponding electromagnetic (EM) field component [4] or a real feed probe modelled by using TLM wire node [5] may be used.

This paper focuses on TLM modelling and analysis of how cavity walls and dielectric load losses influence EM field in the cavity.

[1]Jugoslav Joković, Bratislav Milovanović and Tijana Dimitrijević are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: [jugoslav, bata, tijana] @elfak.ni.ac.yu

II. TLM MODELLING PROCEDURE

In the TLM time-domain method, electromagnetic field strength in three dimensions, for a specified mode of oscillation in a cylindrical metallic cavity, is modelled by filling the field space with a network of link lines and exciting a particular field component [6].

Electromagnetic properties of mediums in the cavity are modelled by using a network of interconnected nodes (Fig. 1), a typical structure being the symmetrical condensed node (SCN) [7]. Each node describes a portion of the medium shaped like a cuboid or a slice of cake depending on the applied (rectangular/cylindrical) coordinate system (grid).

To operate at a higher time-step, a hybrid symmetrical condensed node (HSCN) [7] is used. An efficient computational algorithm of scattering properties, based on enforcing continuity of the electric and magnetic fields and conservation of charge and magnetic flux is implemented to speed up the simulation process.

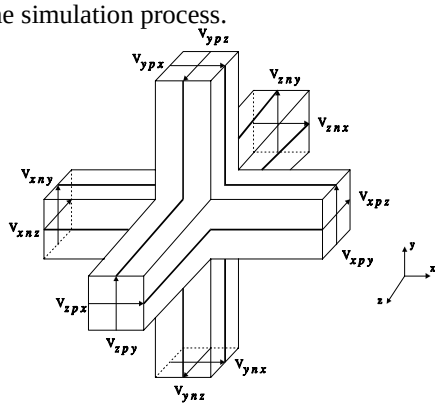


Fig. 1. Symmetrical condensed node

As in every numerical simulation, in TLM time-domain method it is also necessary to describe boundaries. External boundaries of arbitrary reflection coefficient ρ_w are modelled in TLM by terminating the link lines at the edge of the problem space with an appropriate load [7]. If the characteristic impedance of a link line differs from the intrinsic impedance of a medium, the equivalent link line reflection coefficient, ρ_{ij}, will be different from ρ_w. The link line reflection coefficient, ρ_{ij}, can be found by terminating the link line, of characteristic impedance Z_{ij}, with the same resistance:

$$\rho_{ij} = \frac{R - Z_{ij}}{R + Z_{ij}} = \frac{(1+\rho_w) - \hat{Z}_{ij}(1-\rho_w)}{(1+\rho_w) + \hat{Z}_{ij}(1-\rho_w)}, \qquad (1)$$

where a normalized characteristic impedance is introduced as $\hat{Z}_{ij} = Z_{ij} / Z_{ij}^{s}$.

If the external boundary represents an electric or magnetic wall we have $\rho_w = \rho_{ij}$. Otherwise, ρ_{ij} will depend on Z_{ij}. External boundaries modelling in TLM method, described by equation (1), will provide good results only if incident wave is perpendicular to the external boundary.

In many industrial processes, the water is often used as a cavity load. Losses in medium within a cavity can be incorporated in TLM model by introducing loss stubs into the scattering points that is nodes. The loss stubs may be viewed as infinitely long, or equivalently, as terminated (matched) by their own characteristic impedances. The matched stubs can be used to model both 'electrical' and 'magnetic' losses. In the HSCN, the presence of matched stubs is incorporated directly into the scattering matrix. Given the effective electrical conductivity σ_e, loss 'electrical' element for the 3-D time-domain TLM method is defined as [7]:

$$G_e = \sigma_e \, f(\Delta x, \Delta y, \Delta z), \tag{2}$$

where: Δx, Δy and Δz are dimensions of TLM node in the x, y and z directions respectively. Complex permittivity is related to effective electrical conductivity as:

$$\varepsilon^* = \varepsilon_0 \varepsilon_r^* = \varepsilon_0 \varepsilon_r' - j\sigma_e / \omega . \tag{3}$$

III. Numerical Results

Numerical results, illustrating the influence of the losses on resonant modes of the cavity, were obtained by modelling of the rectangular cavity with dimensions: a = 36 cm, b = 35 cm and h = 26 cm using TLM approach. The medium inside the cavity was modelled by uniform TLM network of dimension $x \times y \times z$=36×35×26 nodes ($\Delta x = \Delta y = \Delta z$ = 1 cm) where hybrid symmetrical condensed node (HSCN) was used.

In order to study effects of cavity walls and dielectric losses, numerical TLM results, representing E_z field component in TLM node (20, 20, 20), which were obtained in the case of impulse excitation of E_z field component applied to the nodes (1÷15, 1÷15, 1÷15), were considered. At the beginning, the analyzed cavity was filled with air, representing homogeneous lossless dielectric of relative permittivity ε_r = 1. The second stage of the analysis was accompanied with TLM modelling of fully air-filled cavity with hypothetically complex permittivity ε_r = 1-j0.01, representing homogeneous medium with losses. In both cases, reflection coefficent of the walls was varied, that is either cavity walls were presented by perfect conducted metal characterized by ρ = - 1 or ρ = - 0.99 was used to model losses in the cavity walls. The time-domain outputs for all considered cases are shown in Fig. 2.

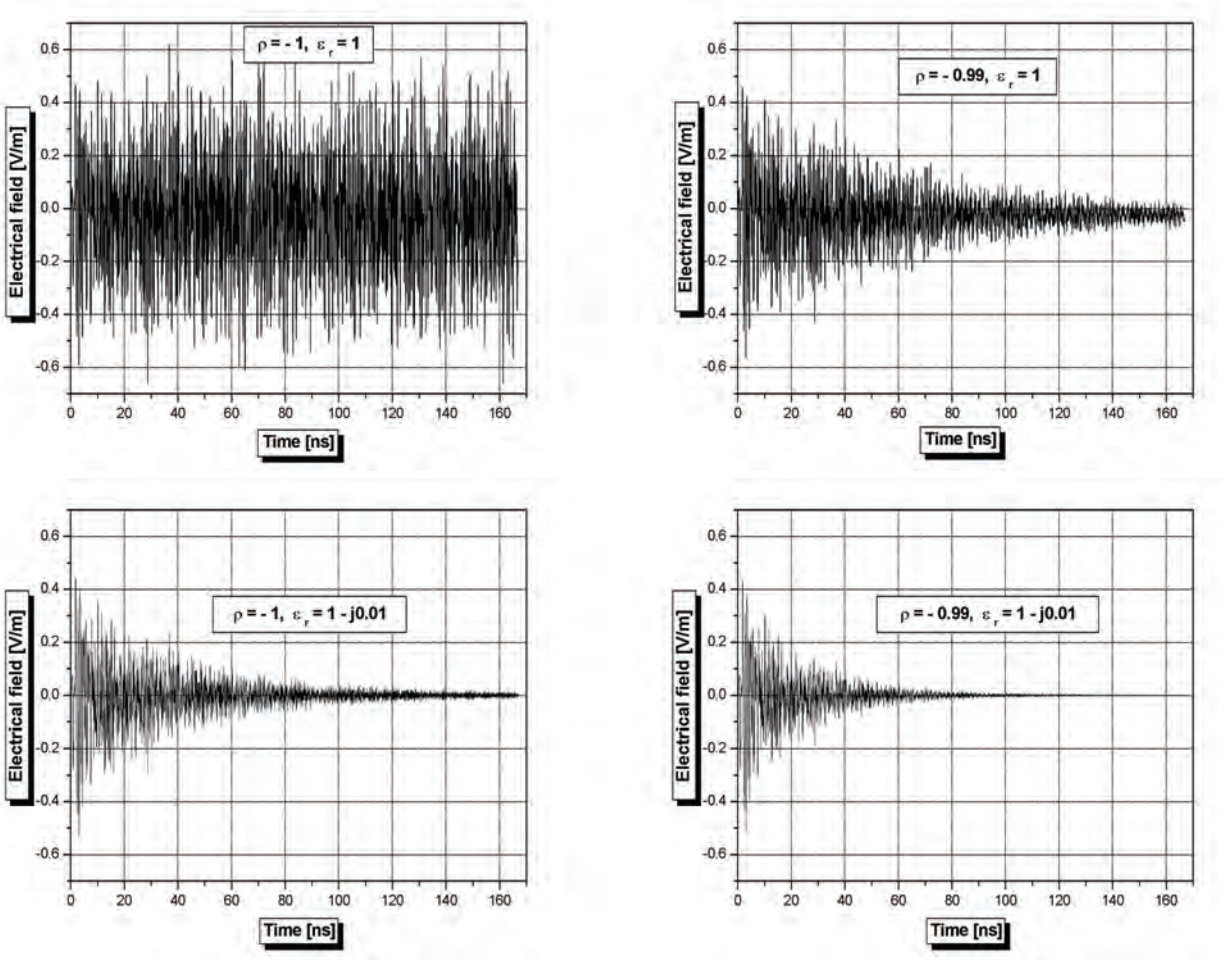


Fig. 2. Time-domain TLM results for air-filled cavity with different characteristics of walls and medium inside the cavity

When the lossless cavity is considered, the graph shows that envelope remains constant through period of time, whereas in the case when conducting losses as well as dielectric losses were taken into account an exponential decline of envelope is seen.

Corresponding frequency-domain results in the frequency range f = [500 ÷ 2000] MHz are shown in Fig. 3. It can be observed that amplitudes of the peaks, referred to as modes, are significantly reduced in the presence of conducting losses, but the same resonances can still be identified. Also, numerical results obtained for the lossless cavity shows greater oscillations of frequency-domain signal (as a result of time domain signal cut-off), compared to the cavity including conducting losses.

As far as homogeneous medium with losses (ε_r = 1 - j0.01) is concerned, presented graphs show that the presence of load losses has the same influence on the resonant modes as cavity walls reflection coefficients. On the other hand, when both cavity walls and load losses are incorporated, resonant modes peaks are additionally reduced, leading to difficulties concerning modes identification, especially when resonant frequencies are close.

Finally, partially loaded cavity with filling factor $t/h=0.2$ is analyzed. As a dielectric layer, placed on the cavity bottom, water at the temperature 20° is used. Relative dielectric constant of water ε_r = 81 - j10 is calculated from Debby's formula [8]. For more accurate modeling of this problem, a finer mesh within the homogeneous dielectric sample with losses and cells with arbitrary aspect ratio suitable for modelling of particular geometrical features, were applied. Corresponding numerical results in time and frequency domain are given in Figs 4. and 5, respectively. In the case of cavity without losses in the walls frequency-domain graph contains greater oscillations as a result of time domain signal cut-off. According to presented graphs conclusion can be derived that influence of the cavity walls losses on the resonant modes, in terms of frequencies and corresponding field level is not significant when the cavity contains good absorbing load, as water.

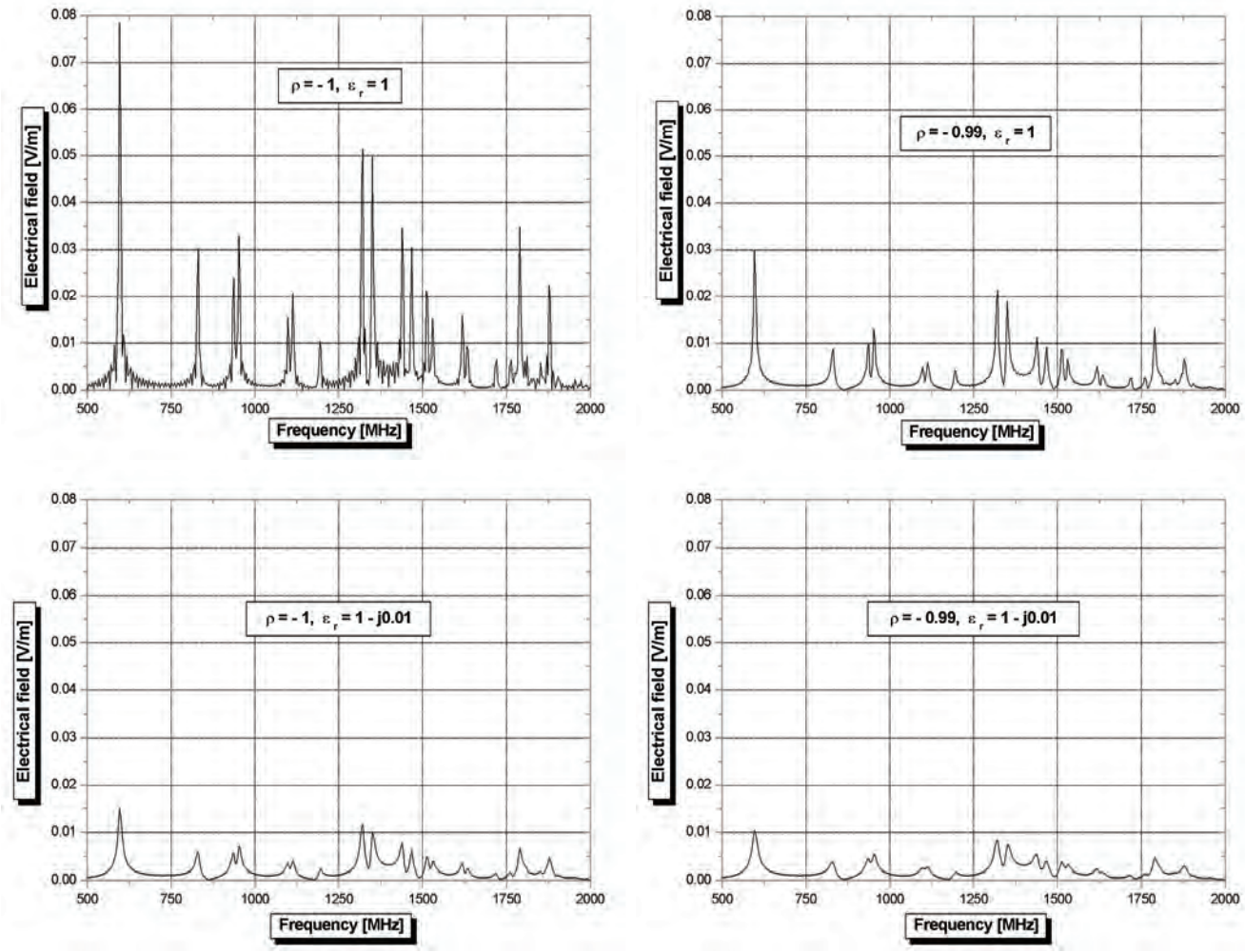


Fig. 3. Frequency- domain TLM results for air-filled cavity with different characteristics of walls and medium inside the cavity

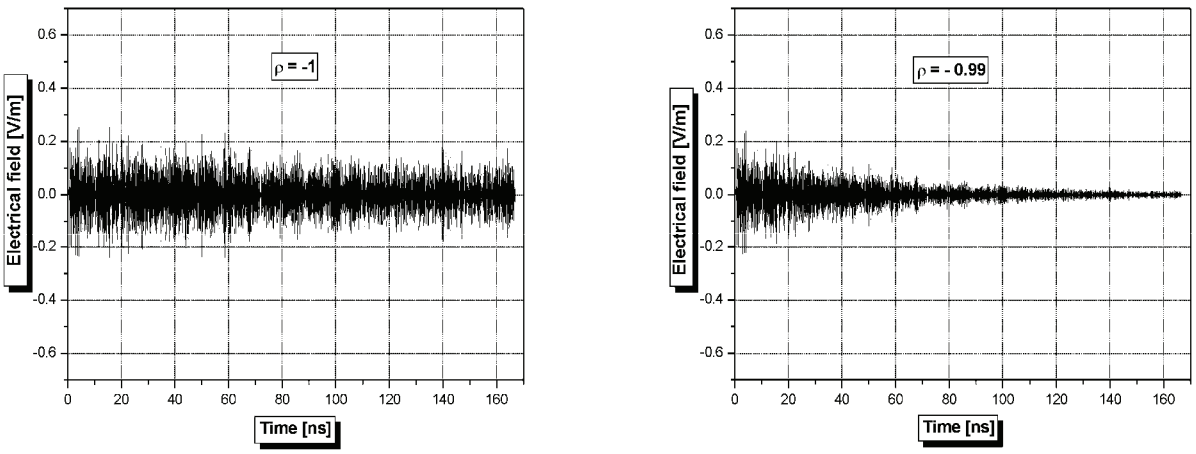


Fig. 4. Time-domain TLM results for partially loaded cavity with different reflection coefficients of walls

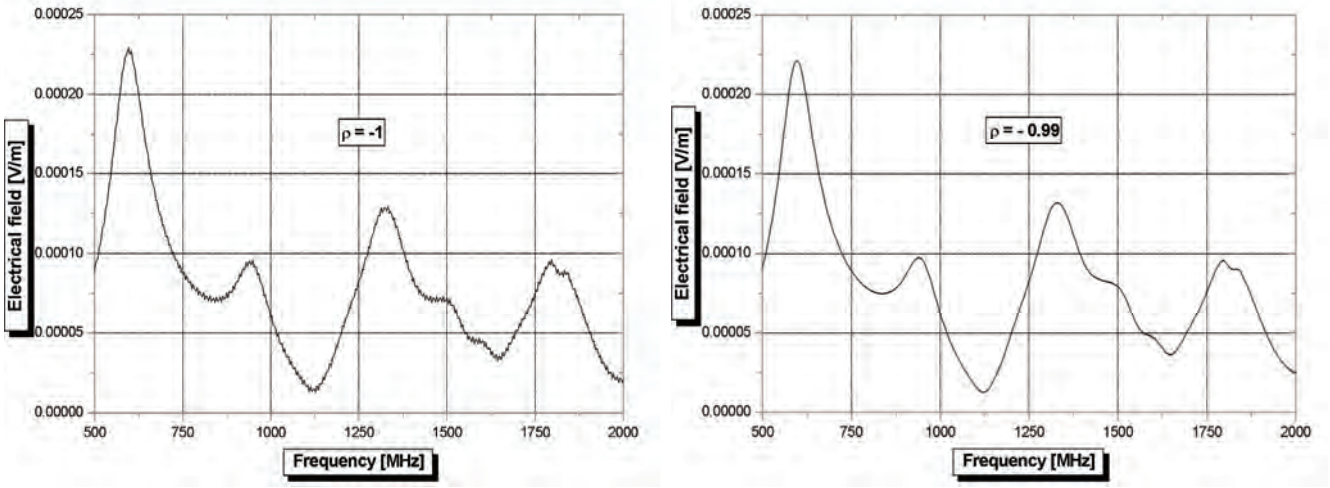


Fig. 5. Frequency-domain TLM results for partially loaded cavity with different reflection coefficients of walls

IV. Conclusion

This paper proposes TLM modelling of characteristics of microwave applicator based on conducting rectangular cavity loaded with dielectric layer. The cavity configurations with different reflection coefficients, representing losses in cavity walls, were analyzed in time and frequency domain by using 3-D TLM approach. Also, 3-D TLM software was applied to study effects of losses in dielectric load, represented by complex permitivity, in the cases of fully and partially loaded cavity for determining resonant modes in practical applications.

References

[1] A. Baysar, J.L. Kuester and S. El-Ghazaly, "Theoretical and Experimental Investigations of Resonance Frequencies in a Microwave Heated Fluidized Bed Reactor", IEEE MTT-S Digest, 1992, pp.1573-1576.

[2] M. Mladenović, A. Marinčić, B. Milovanović, "Mode Tuning of Microwave Resonator Loaded with Losses Multilayer Dielectric", Journal of Microwave Power and Electromagnetic Energy Vol 24, No. 3, 1989, 166-177.

[3] B. Milovanović, N. Dončov, V. Trenkić and V. Nikolić, "3-D TLM Modelling of the Circular Cylindrical Cavity Loaded by Lossy Dielectric Sample of Various Geometric Shapes", Proc. of the Third International Workshop on TLM Modelling-Theory and Applications, Nica, France 187-195, 1997

[4] B. Milovanovic, N. Doncov, A. Atanaskovic, "Tunnel Type Microwave Applicator Modelling Using TLM Method, Problems in Modern Applied Mathematics", A Series of Reference Books and Textbooks: Mathematics and Computers in Science and Engineering, WSES Press, pp.327-332, 2000.

[5] V.Trenkić, A.J.Wlodarczyk and R.A.Scaramuzza, "Modelling of Coupling between Transient Electromagnetic Field and Complex Wire Structures", International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, Vol.12, No.4, pp.257-273, 1999

[6] C. Christopoulos, The Transmission-Line Modelling Method, IEE/OUP Press, 1995.

[7] V. Trenkić, "The Development and Characterization of Advanced Nodes for TLM Method", Ph.D. Thesis, University of Nottingham, 1995.

[8] M.N.O.Sadiku, Refractive Index of Snow at Micro-Wave Frequencies, Applied Optics, Vol.24, No.4, pp. 572-575, 1985.

Improved Algorithm for Time-Domain Electromagnetic Field Calculation at Points outside the TLM Workspace

Nebojša Dončov and Bratislav Milovanović

Abstract – **In this paper, an algorithm for electromagnetic field calculation in the time-domain at points outside the defined TLM numerical workspace is presented. This algorithm is based on integration over Huygens surface surrounding all electromagnetic structures within the workspace. As this integration involves a significant amount of computation, a reduction of the integration frequency is suggested to improve the algorithm efficiency. This improvement incurs the cost of extra storage but has an advantage that this cost is fixed and does not increase with the number of output points. Accuracy and efficiency of the improved algorithm are illustrated on the appropriate example.**

Keywords – **EMC, TLM, Huygens surface, output outside workspace, integration frequency**

I. INTRODUCTION

Electromagnetic compatibility (EMC) is the branch of science and engineering concerned with the design and operation of equipment in a manner which makes them immune to certain amounts of electromagnetic (EM) interference, keeping, at the same time, equipment-generated interference within specific limits [1]. The scope of EMC is thus very wide as it encompasses virtually all equipment powered by electrical supplies. In recent years, the rapid increase in the use of radio communications, digital systems, fast processors and the introduction of new design practices have brought EMC to the forefront of advanced design.

How practical modern systems exceed in complexity anything that can be solved analytically or using approximate techniques, numerical computer-based models are the only alternative to study trends in EMC design and to understand the behaviour of systems. Differential numerical techniques in the time-domain, such as Finite-Difference Time-Domain (FD-TD) [2] and Transmission-Line Matrix (TLM) [3], are well established for solving a number of EMC problems over a wide frequency range. These methods, due to their characteristics and the development of powerful computer stations, offer a significant extension of the range of EMC problems that can be tackled.

However, there are still numerous practical EMC problems, where these techniques, even with the use of powerful computer stations are incapable to allow for their fast and correct modelling. One area of inefficient EMC numerical simulation represents equipment containing geometrically small but electrically important features (e.g. thin wires, slots, air-vents). For their description, extremely fine mesh is required which can result in a prohibitively large number of cells and large number of time-steps. In recent years, enhancements to TLM technique in the form of so called sub-cell or compact models have been developed to allow for an efficient simulation of these structures [4-6]. Prediction of level of interference generated by such equipment, as one of the key concerns in EMC design, is another area where numerical techniques struggle to allow calculation of EM field response at points located at distances exceeding the dimensions of radiating EM structures, by one order of magnitude or more. Again, large number of cells is required to cover that distance between equipment and outputs which leads to time consuming and inefficient EMC simulation even with the use of developed compact models.

The algorithm for efficient EM field calculation at points located at large distances from the radiating EM structures, without defining discrete numerical mesh up to these points, has been developed and verified in [7]. It was based on Love's equivalence principle [8,9] and required definition of so-called Huygens closed surface within the workspace surrounding all EM structures of analyzed EMC problem. However, integration over the Huygens surface at every time-step, to allow calculation of the time-domain EM field at points outside the numerical workspace, involved a significant amount of computation. The computational burden for a single output point proved acceptable but it can be easily made unacceptable by specifying many output points. Therefore, in this paper it is suggested to reduce the integration frequency to once every time-sample which will incur the extra memory storage but this cost does not increase with the number of output points. Improved algorithm is again implemented in TLM method and verified on the appropriate example. Compared to the conventional simulation, huge saving in simulation run-time can be achieved and maintained with the increase of output points.

II. LOVE'S EQUIVALENCE PRINCIPLE

The surface equivalence theorem is a principle by which actual sources, represented electrically by current densities J_1 and M_1, are replaced by equivalent sources J_s and M_s (Fig.1). The fictitious sources are said to be equivalent within a volume V_2, because they produce within that volume the same fields (E_1 and H_1) as the actual sources. According to this principle, the fields outside an imaginary closed surface S, i.e. in volume V_2, are obtained by placing, over the closed so-called Huygens surface, suitable electric and magnetic

Nebojša Dončov and Bratislav Milovanović are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia and Montenegro, E-mail: [doncov, bata]@elfak.ni.ac.yu

current densities that satisfy the boundary conditions. These equivalent current densities are given by Eqs.(1) and (2).

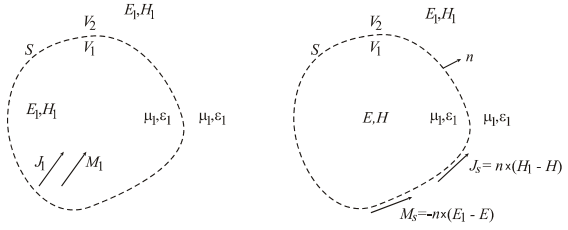


Fig.1 a) Actual and b) equivalent problem models

$$J_s = n \times (H_1 - H) \quad (1)$$

$$M_s = -n \times (E_1 - E) \quad (2)$$

The current densities are then selected so that the fields inside the closed surface (volume V_1), which is not the region of interest, are zero and outside the closed surface (volume V_2) are equal to the radiation produced by the actual sources:

$$J_s = n \times (H_1 - H)_{|H=0} = n \times H_1 \quad (3)$$

$$M_s = -n \times (E_1 - E)_{|E=0} = -n \times E_1 \quad (4)$$

Thus, this technique can be used to obtain the fields radiated outside a closed surface by sources enclosed within it. The formulation is exact but requires integration over the closed surface.

Since the current densities J_s and M_s radiate in an unbounded medium that is, have the medium (μ_1,ε_1) everywhere, they can be used in conjunction with the Eqs.(5) and (6) to find the fields everywhere:

$$E_1 = -j\frac{1}{\omega\mu_1\varepsilon_1}\nabla(\nabla \cdot A) - j\omega A - \frac{1}{\varepsilon_1}\nabla \times F \quad (5)$$

$$H_1 = -j\frac{1}{\omega\mu_1\varepsilon_1}\nabla(\nabla \cdot F) - j\omega F + \frac{1}{\mu_1}\nabla \times A \quad (6)$$

where A and F are vector magnetic and electric potential, respectively, that can be expressed through electric and magnetic surface currents as [8]:

$$A = \frac{\mu_1}{4\pi}\int_S J_s \frac{e^{-jkR}}{R} dS', \; F = \frac{\varepsilon_1}{4\pi}\int_S M_s \frac{e^{-jkR}}{R} dS' \quad (7)$$

III. Near-to-near Transform Algorithm

The contribution of an elemental patch of the Huygens surface to EM field at a point removed from it by the vector R can be found in the time-domain from Eqs.(5-7) as:

$$\Delta E(t + R/c) = \frac{\Delta S}{4\pi R \cdot R}\left(\Delta E_{rad}(t) + \Delta E_{ind}(t) + \Delta E_{stat}(t)\right) \quad (8)$$

$$\Delta H(t + R/c) = \frac{\Delta S}{4\pi R \cdot R}\left(\Delta H_{rad}(t) + \Delta H_{ind}(t) + \Delta H_{stat}(t)\right) \quad (9)$$

where ΔS is the area of the elemental patch determined by TLM node. As it can be seen the signal from each surface patch consists of a differential radiation term (ΔE_{rad} and ΔH_{rad}), a direct induction term (ΔE_{ind} and ΔH_{ind}) and an integral static field term (ΔE_{stat} and ΔH_{stat}) that can be expressed in rectangular coordinate system as:

$$\Delta E_{rad}(t) = \frac{|R|}{c}\frac{d}{dt}(Z_0((J_s \cdot R_{ort}) \cdot R_{ort} - J_s) - M_s \times R_{ort}) \quad (10)$$

$$\Delta E_{ind}(t) = Z_0(3(J_s \cdot R_{ort}) \cdot R_{ort} - J_s) - M_s \times R_{ort} \quad (11)$$

$$\Delta E_{stat}(t) = \frac{c}{|R|}\int Z_0(3(J_s \cdot R_{ort}) \cdot R_{ort} - J_s)dt \quad (12)$$

$$\Delta H_{rad}(t) = \frac{|R|}{c}\frac{d}{dt}(Y_0((M_s \cdot R_{ort}) \cdot R_{ort} - M_s) + J_s \times R_{ort}) \quad (13)$$

$$\Delta H_{ind}(t) = Y_0(3(M_s \cdot R_{ort}) \cdot R_{ort} - M_s) + J_s \times R_{ort} \quad (14)$$

$$\Delta H_{stat}(t) = \frac{c}{R}\int Y_0(3(M_s \cdot R_{ort}) \cdot R_{ort} - M_s)dt \quad (15)$$

where Z_0 is an intrinsic impedance of free space ($Y_0 = 1/Z_0$), R_{ort} *is* unit vector ($R_{ort} = R/|R|$) and J_s and M_s are the equivalent electric and magnetic current densities, respectively.

The time delay R/c in the near-to-near transform is introduced by means of a conveyor belt shown in Fig.2. Each output point is at the delivery end of its own conveyor which carries a time-domain signal towards it at the speed of light c. In order to introduce a delay R/c into a signal, the signal is dropped once every time-step Δt onto the conveyor at the distance R upstream of the output point.

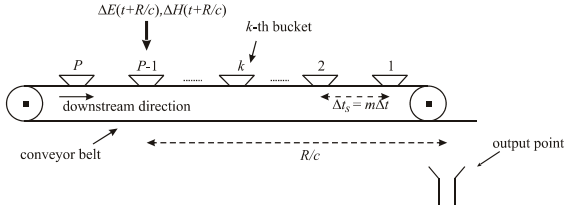


Fig.2 Conveyor belt for one output point

The conveyors are not continuous but consist of buckets spaced one time-sample, Δt_s, apart. Time-sample depends on highest frequency of interest for EM response analysis and in general can be expressed as integer number, m, of used time-step in TLM simulation, Δt ($\Delta t_s = m\Delta t$). If a signal is dropped onto the conveyor belt at time, $i\Delta t$, (i=1,2,...,I_{max}, where I_{max} is maximum number of iteration of simulation in the time-domain) when a bucket is not passing the loading point in question, then the signal falls between buckets and is lost. The conveyor belt (buckets on it) advances with time-step Δt towards output point (downstream direction).

The time delay R/c is introduced in the way that signal corresponding to the appropriate elemental patch of Huygens surface at distance R from the output point is dropped onto appropriate place at conveyor belt ($\Delta E(t+R/c)$ and $\Delta H(t+R/c)$, given by Eqs.(8-15)). The total length of a conveyor is determined by the maximum required delay, i.e. by the distance between the output point and the most remote patch of the Huygens surface and it can be calculated as:

$$P_{max} = 1 + \text{int}(R_{max}/c/\Delta t) \quad (16)$$

IV. Improving the Computational Efficiency

Because the integration over the Huygens surface involves a significant amount of computation, it may be worthwhile attempting to reduce the integration frequency from once

every time-step to once every time-sample. Even if the computational burden for a single output point proves acceptable, it can be easily be made unacceptable by specifying many output points. The integration frequency may be reduced in a variety ways. The approach described in this paper incurs the cost of extra storage, but has the advantage that this cost is fixed and does not increase with the number of output points.

At first, an additional conveyor belt is created for each elemental patch of the Huygens surface while keeping conveyor belt per output point as a main belt. These per-patch conveyor belts are very short, just one time-sample plus one time-step long, and have a bucket at every time-step. Each per-patch conveyor belt carries the equivalent surface currents J_s and M_s away from the patch, and it is therefore four real numbers wide.

Per-Time-Sample integration is performed as follows. For simplicity, assume that the integration is performed at those time-steps when a bucket of a main conveyor belt is level with its output point. At these time-steps there will be no buckets at any loading point corresponding to a fractional time-sample delay. In these circumstances, the delay is achieved by picking surface currents off the per-patch conveyor a distance downstream of the patch corresponding to the fractional part of the delay, doing necessary arithmetic, and then dropping the resulting values onto the main conveyor at a point corresponding to the integral part of the delay (where we have presupposed there to be a bucket present).

a) Differential radiation term

For the differential term, instead of dropping an up-down pulse onto the main conveyor belt, as suggested in [7], an up-down pulse is picked off the per-patch conveyor belt, and the resulting radiation term dropped straight into a single bucket of the main conveyor. When the desired delay n is not an exact half-integral number of time-steps, the same weightings are used as described in [7]:

$$\text{drop}[N\text{-}1] = (N + 1/2 - n)v(t)/\Delta t$$
$$\text{drop}[N] = 2(n - N)v(t)/\Delta t$$
$$\text{drop}[N\text{+}1] = (N - 1/2 - n)v(t)/\Delta t$$

where N = nint(n) is the nearest integer value.

The extra time-step allowed in the length of the per-patch conveyors simplifies the case of integral and near-integral time-sample delays. Without the extra length, it would be necessary to treat these cases specially. This would involve transferring one term from one end of the per-patch line to one bucket on the main conveyor belt and a second term from the other end of the per-patch line to the adjacent bucket on the main conveyor belt doubling the necessary computation.

b) Direct term

The value for the direct induction term is simply dropped, instead of main conveyor belt, onto the per-patch conveyor belt at the desired point. Again, for fractional time-steps, the drop weightings described in [7] are now used as pickup-weightings from the per-patch conveyor:

$$\text{drop}[N1] = (N2 - n)v(t)$$
$$\text{drop}[N2] = (n - N1)v(t)$$

where $N1$ and $N2$ are consecutive integers bracketing n.

c) Integral term

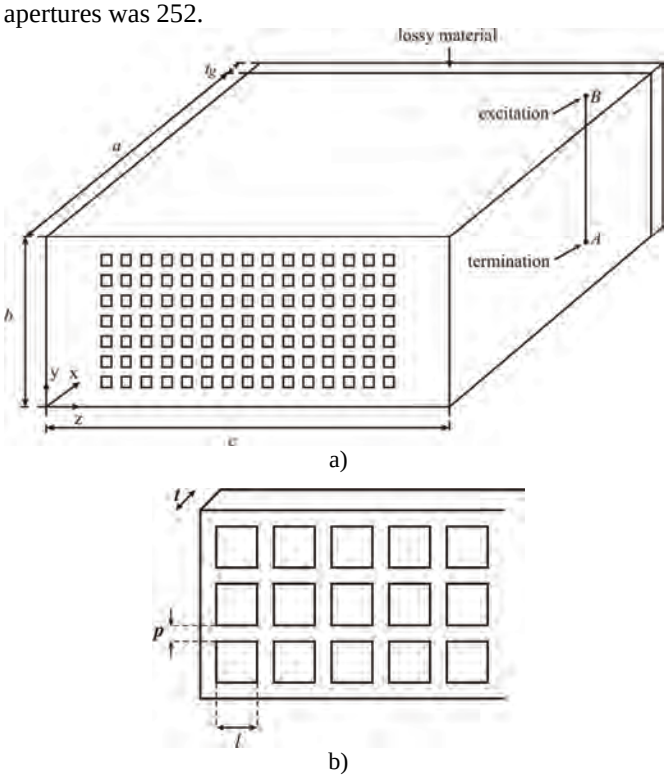
The main part of the integration is still performed on the main conveyor belt. It would be very inefficient to integrate surface currents from an individual patch, because of the presence of a D.C. field the integral would ramp up forever. The fractional-time-sample part of the integral is performed by picking up all values of the per-patch conveyor downstream of the desired fractional-delay point (excluding the extra time-step). The resulting term is dropped into the bucket of the main conveyor corresponding to the integral part of the delay. All buckets upstream from this one get the static field term resulting from the sum of all the values on the per-patch conveyor (whether upstream or downstream of the fractional pickup point, but excluding the extra time-step).

For fractional time-steps, the drop weightings specified in [7] is used as a pickup weighting:

$$\text{drop}[N] = (N + 1/2 - n)v(t)\Delta t$$

V. Numerical Analysis

The accuracy and efficiency of improved algorithm for calculating the time-domain EM field at points outside the TLM workspace are verified on the realistic EMC example in the form of rectangular shielding enclosure (Fig.3a). The enclosure was constructed of six pieces of t=0.635 cm thick aluminum with conductivity σ=3.54e7 S/m. The inside dimensions of the enclosure were a=50 cm, b=20 cm and c=40 cm. One of the enclosure walls in yz plane contains the square aperture array (Fig.3b). The aperture size was l=1 cm, with an edge-to-edge space p=0.5 cm. The number of apertures was 252.



a)

b)

Fig.3. a) Rectangular test enclosure with air-vent on front face, b) air-vent layout

The model was excited by a long-wire feed at the back of the box which was modelled by a simple voltage source V=1 mV with 50 Ohm resistance incorporated into a single cell at the feed point *B* (43 cm, 20 cm, 33 cm). The excitation had the center conductor of the probe extended to span the width of the cavity with a 0.16 cm diameter and terminated on the opposite cavity wall with termination of 47 Ohm at point *A* (43 cm, 0, 33 cm). In order to reduce Q-factor of enclosure, which can be extremely high for unloaded enclosure, lossy material of conductivity σ=2.27S/m and thicknes t_g=1 cm was used. The resistors were modelled as lumped elements. Compact wire [4] and air-vent model [6] were used to model wire conductor and array of square apertures, respectively, in order to avoid the need of TLM mesh with extremely high resolution. The choice of geometry, excitation and output was governed by experimental arrangements used in [10].

Huygens virtual surface, completely surrounding the TLM model of the enclosure, is located at the distance of one TLM node from the enclosure. The external boundaries of TLM mesh are placed around the enclosure at the distance greater than 30% of the largest dimension of the enclosure, which is still far away from the required output point. The results for far zone electric field at 3 m away from the face of enclosure containing air-vent, obtained by using the improved algorithm, are compared with the experimental results [10] and shown in Fig.4. An excellent agreement between these two results can be observed except at higher frequencies, probably due to neglecting the frequency dependant characteristics of the lossy material at the back of the box. In addition huge saving in the computer resources, compared to conventional TLM, is achieved with accuracy acceptable for most EMC applications.

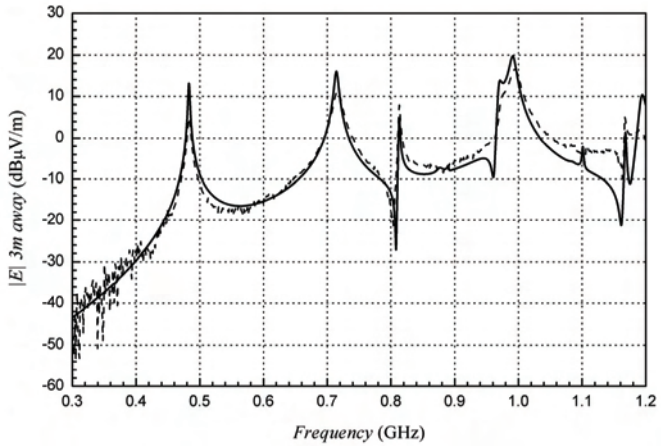


Fig.4 Radiated electric field at 3 m away from the enclosure: a) measurements (dotted line), b) improved algorithm (solid line)

VI. CONCLUSION

This paper describes an improved algorithm for calculating the time-domain EM field at points outside the TLM workspace. It is primarily indented for EMC applications. The new scheme is based again on integration over a Huygens surface within the defined workspace using the calculated equivalent electric and magnetic surface currents. However, integration frequency is now reduced from once every time-step as suggested previously to once every time-sample in order to maintain efficiency of proposed algorithm with the number of output points. The effect of time delay in signal propagation from elemental patches of Huygens surface to output points is realized through main conveyor belt per output point and additional belts per elemental patch of Huygens surface. The scheme is general and it can be implemented into other differential numerical techniques in the time-domain such as FD-TD method. The efficiency of this approach is illustrated on the appropriate examples.

If the number of time-steps per time-sample is high, then an intermediate sampling frequency must be used on the per-patch conveyor belts to avoid excessive memory usage. There need not be a whole number of intermediate time-samples per final time-sample. However, the transfer of the static field terms from the per-patch conveyor belt to the main conveyor belt must take into account what fraction of the final intermediate time-sample must be transferred to make up a complete final time-sample. This modification to allow efficient memory allocation, in the case of high time-sample, will be presented in future papers.

REFERENCES

[1] C. Christopoulos, *Principles and Techniques of Electromagnetic Compatibility*, CRC Press, 2000.

[2] C. Christopoulos, *The Transmission-Line Modelling Method*, Series on Electromagnetic Wave Theory, IEEE/OUP Press, 1995.

[3] K.S. Kunz, R.J. Luebbers, *The Finite Difference Time Domain for Electromagnetics*, CRC Press, 1993.

[4] A.J. Wlodarczyk, V. Trenkić, R. Scaramuzza, "*A Fully Integrated Multiconductor Model for TLM*", IEEE Transactions on Microwave Theory and Techniques, Vol.46, No.12, 1998, pp.2431-2437.

[5] V. Trenkić, R. Scaramuzza, "*Modelling of Arbitrary Slot Structures using Transmission Line Matrix (TLM) Method*", Proceedings of International Zurich Symposium on Electromagnetic Compatibility, Zurich, Switzerland, 2001, pp.393-396.

[6] N. Dončov, A.J. Wlodarczyk, V. Trenkić, R. Scaramuzza, "*Compact TLM Model for Air-vents*", Electronics Letters, Vol.38, No.16, 2002, pp.887-889.

[7] N. Dončov, B.Milovanović, "*Electromagnetic Field Calculation in the Time-Domain at Points outside the TLM Workspace*", Proceedings of 41st ICEST Conference, Sofia, Bulgaria, 2006, pp.37-40.

[8] E.C. Jordan, K.G. Balmain, *Electromagnetic Waves and Radiating Systems*, Prentice-Hall Inc., 1968.

[9] C.A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons, New York, 1989.

[10] M.Li, J.Nuebel, J.L.Drewniak, R.E. DuBroff, T.H. Hubing and T.P.Van Doren, "*EMI from Airflow Aperture Arrays in Shielding Enclosures – Experiments, FDTD, and MoM Modelling*", IEEE Transactions on Electromagnetic Compatibility, Vol.42, No.3, pp.265-275, 2000.

Strip Lines on the Cylindrical Dielectric Substrate

Žaklina J. Mančić[1], Saša S. Ilić[2] and Slavoljub R. Aleksić[3]

***Abstract* – The paper presents quasi TEM analysis of strip lines on the cylindrical substrate, when the strip is of a negligible thickness or has some final thickness. As a calculation method Equivalent Electrode Method was used, as well as the Finite Elements Method (FEM). The obtained results are shown in tables and graphs. It has been noted that the results correspond well using these two methods.**

***Keywords* – strip lines, quasi-TEM analysis, Equivalent Electrode Method (EEM), Finite Elements Method (FEM), Green functions.**

I. INTRODUCTION

Strip lines of usual rectangular geometry are widely used [19-20]. With them the characteristic impedance and relative dielectric permittivity depend on geometry and line dimensions. Using EEM, the strip line with cylindrical isotropic substrate, Fig.1, was calculated, and its characteristics were analysed.

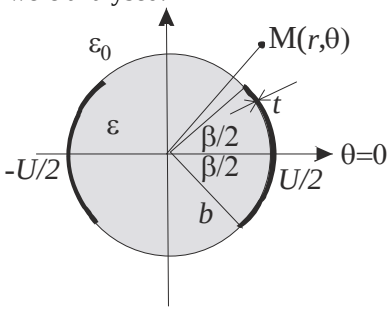


Fig. 1. Strip line on cylindrical substrate.

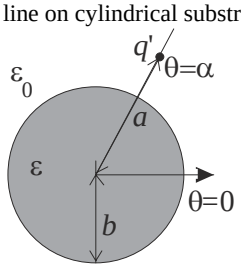


Fig. 2. Charge per unit length outside dielectric cylinder.

As EEM requires knowledge of Green function, Green function has been used [24] for charge per unit length in the proximity of dielectric cylinder, Fig.2:

$$\varphi=\begin{cases}\varphi_0-\dfrac{q'}{2\pi(\varepsilon+\varepsilon_0)}\ln\left[1-2\dfrac{r}{a}\cos(\theta-\alpha)+\left(\dfrac{r}{a}\right)^2\right], & r\le b\\ \varphi_0-\dfrac{q'}{4\pi\varepsilon}\ln\left[1-2\dfrac{r}{a}\cos(\theta-\alpha)+\left(\dfrac{r}{a}\right)^2\right]\\ -\dfrac{q'}{4\pi\varepsilon}\dfrac{\varepsilon_0-\varepsilon}{\varepsilon+\varepsilon_0}\ln\left[1-2\dfrac{b^2}{ar}\cos(\theta-\alpha)+\dfrac{b^4}{a^2r^2}\right], & r>b\end{cases} \quad (1)$$

where φ_0 is a constant dependent of the referential point choice. Equations (1) allow determination of expressions for electrical field components in a cylindrical coordinate system, for $r>b$:

$$E_r=\frac{1}{2\pi\varepsilon_0}\left(A+\frac{\varepsilon_0-\varepsilon}{\varepsilon_0+\varepsilon}B\right),\qquad E_\theta=\frac{1}{2\pi\varepsilon_0}\left(C+\frac{\varepsilon_0-\varepsilon}{\varepsilon_0+\varepsilon}D\right) \quad (2)$$

where:

$$A=\frac{r-a\cos(\theta-\alpha)}{a^2-2ar\cos(\theta-\alpha)+r^2},$$

$$B=\frac{arb^2\cos(\theta-\alpha)-b^4}{a^2r^3-2ar^2b^2\cos(\theta-\alpha)+b^4r},$$

$$C=\frac{a\sin(\theta-\alpha)}{a^2-2ar\cos(\theta-\alpha)+r^2},$$

$$D=\frac{b^2a\sin(\theta-\alpha)}{a^2r^2-2arb^2\cos(\theta-\alpha)+b^4}.$$

Relative error given in Table 1 was calculated using formula:

$$\delta[\%]=\frac{\left|Z_c^{EEM}-Z_c^{FEM}\right|}{Z_c^{FEM}}\cdot 100, \quad (3)$$

where characteristic impedance, as well as effective dielectric permittivity are defined in a standard way, that is, $Z_c=Z_{c0}\Big/\sqrt{\varepsilon_r^{eff}}$, $\varepsilon_r^{eff}=C'/C_0'$, where C_0' and Z_{c0} are parameters obtained for the case when $\varepsilon=\varepsilon_0$.

[1]Žaklina J. Mancic is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: zmancic@elfak.ni.ac.yu

[2]Saša Ilić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: silic@elfak.ni.ac.yu

[3]Slavoljub R. Aleksić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: as@elfak.ni.ac.yu

II. Results

Cylindrical strip line was analysed, of *b* radius with symetrically laid strip conductors, Fig.1. An angle at which the strip can be seen from the center of dielectric cylinder is β, while in the most general case the strip thickness is marked with *t.* The obtained results were compared with the results gained by using program package FEMM [13]. The results of comparison are shown in Table 1, for $\varepsilon_r = 2.2$, different ratio t/b and different values of the angle β.

TABLE I
COMPARISON OF RESULTS OBTAINED WITH EEM AND FEM TECHIQUE

$t/b = 0$				
β[deg]	30	60	90	120
$Z_C^{EEM}[\Omega]$	257.89	190.58	148.99	116.43
$Z_C^{FEM}[\Omega]$	256.50	189.76	148.59	115.94
δ[%]	0.54	0.43	0.27	0.42
ε_r^{eff} (EEM)	1.600	1.600	1.599	1.600
ε_r^{eff} (FEM)	1.600	1.599	1.601	1.601
$t/b = 0.01$				
β[deg]	30	60	90	120
$Z_C^{EEM}[\Omega]$	254.20	188.44	147.46	115.17
$Z_C^{FEM}[\Omega]$	253.34	187.96	147.23	115.01
δ[%]	0.34	0.26	0.16	0.14
ε_r^{eff} (EEM)	1.590	1.592	1.593	1.592
ε_r^{eff} (FEM)	1.591	1.593	1.593	1.593
$t/b = 0.05$				
β[deg]	30	60	90	120
$Z_C^{EEM}[\Omega]$	244.25	182.77	143.23	111.64
$Z_C^{FEM}[\Omega]$	243.91	182.42	142.91	111.35
δ[%]	0.14	0.19	0.22	0.26
ε_r^{eff} (EEM)	1.566	1.573	1.574	1.572
ε_r^{eff} (FEM)	1.567	1.574	1.575	1.573
$t/b = 0.1$				
β[deg]	30	60	90	120
$Z_C^{EEM}[\Omega]$	235.29	177.27	139.10	108.14
$Z_C^{FEM}[\Omega]$	235.15	177.01	138.85	107.91
δ[%]	0.06	0.15	0.18	0.21
ε_r^{eff} (EEM)	1.544	1.554	1.556	1.552
ε_r^{eff} (FEM)	1.546	1.555	1.557	1.553

Fig. 3 shows characteristic impedance in the function of angle β, where the relative dielectric permittivity is ε_r, and $t/b = 0$. Fig. 4 shows characteristic impedance in the function of strip thickness and cylinder radius ratio, for $\varepsilon_r = 2.2$ where β is parameter.

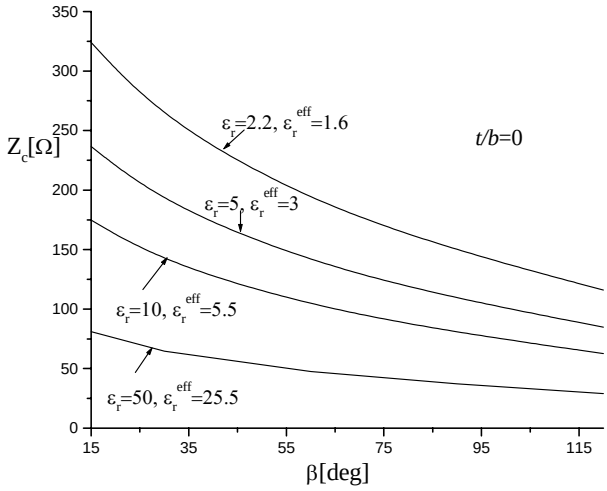


Fig.3. Characteristic impedance in the function of angle β.

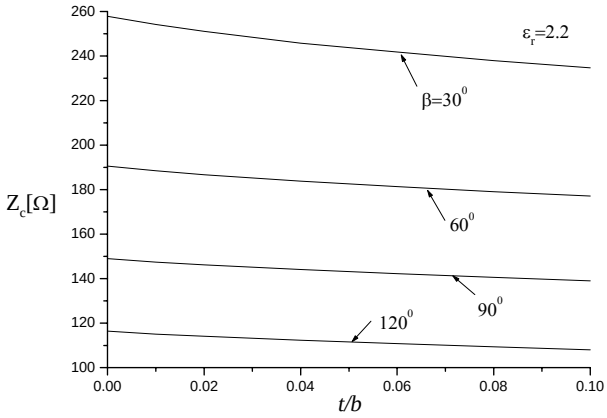


Fig.4. Characteristic impedance in the function of thickness strip, β is the parameter

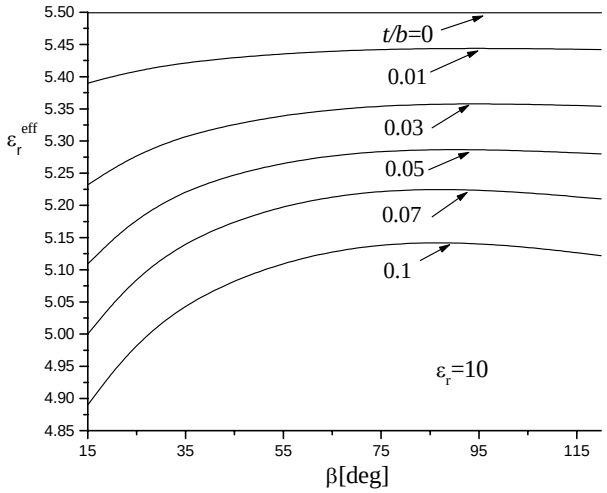


Fig.5. Relative effective dielectric permittivity in function of angle β, conducting strip thickness is the parameter.

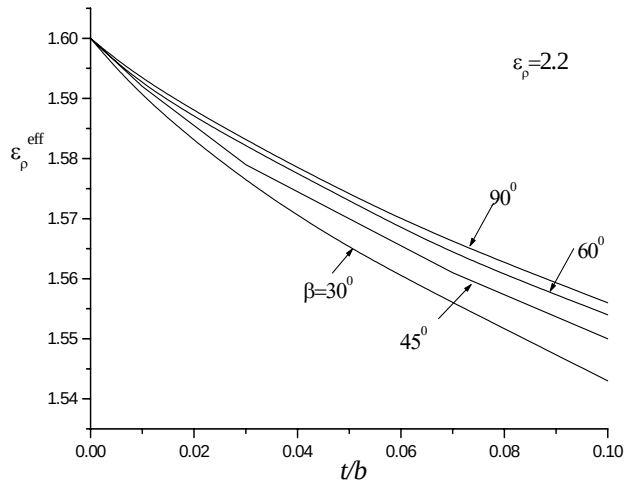


Fig. 6. Relative effective dielectric permittivity in the function of conductive strip thickness for various values of angle β.

The ratio of electrical field intensity in the function of distance from the cylinder center, E, and field intensity on the cylinder surface, E_0, for various directions $(r > b, \theta)$, where t/b is parameter, is shown in figures 8, 9 i 10. They illustrate that a large amount of energy is concentrated within dielectric substrate, and with the distance twice larger than cylinder radius, the field is negligible. A conclusion can be drawn from this, that with these lines it is desirable that relative dielectric substrate constant is as large as possible, so that the largest amount of energy is in it.

III. CONCLUSION

One type of strip lines has been calculated, both of negligible thickness and of final thickness, on a cylindrical substrate. The calculations have showed that in the case of disregarding strip thickness for the required dielectric cylinder constant, these lines have practically constantly effective dielectric constant, independent of line dimensions. This result is actually, the expected theoretical result, that the effective dielectric permittivity is actually equal to arithmetic mean of relative dielectric constant for air and relative dielectric substrate constant. This characteristic, as well as small dissipation of electrical field outside the line, could be an advantage over up to now used strip lines on the substrate of rectangular cross-section. Also, it should be said that the program developed in FORTRAN is much faster that FEMM program. Namely, the calculations were done on PENTIUM 3 (900 MHz) computer with 256 MB RAM, and the duration of calculating using EEM (600 electrodes) was about 30 seconds, while with 500000 final elements, it lasted for 8 minutes.

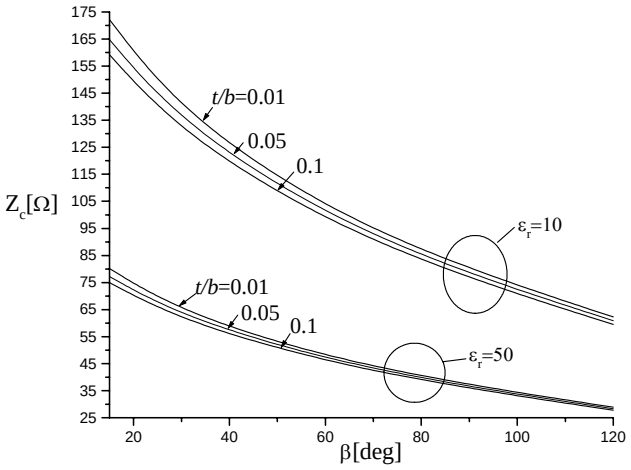


Fig.7. Characteristic impedance in the function of angle β for various thickness of conductive strips.

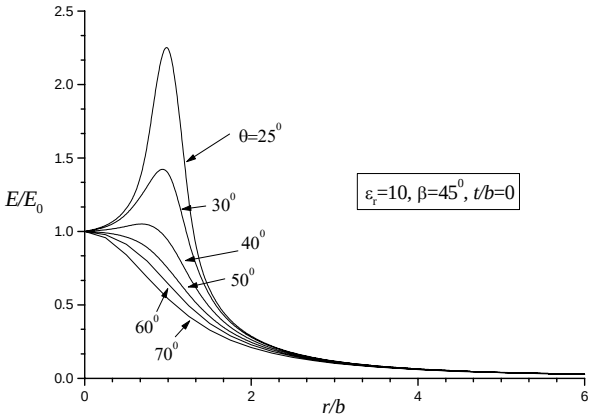


Fig.8. Field intensity in the function of distance for strips of negligible thickness and $\varepsilon_r = 10$.

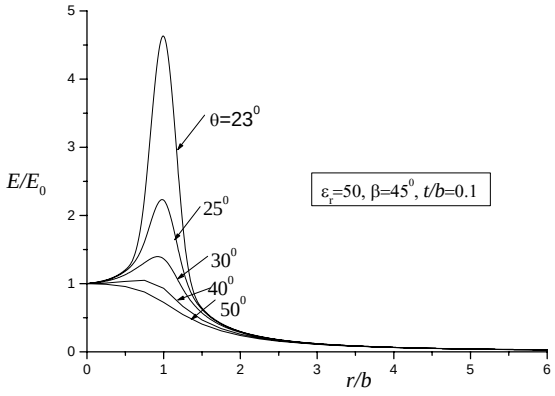


Fig.9. Field intensity in the function of distance for strips of final thickness and $\varepsilon_r = 50$.

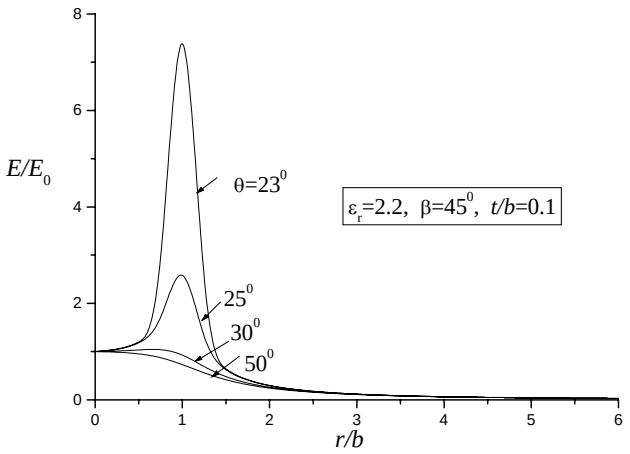


Fig. 10. Field intensity in the function of distance for final thickness strips and $\varepsilon_r = 2.2$.

REFERENCES

[1] Ž. J. Mančić, S. R. Aleksić, V.V. Petrović: **"***Comparison of FEM, MoM and EEM in Solving a Benchmark 2D Electrostatic Problem*", paper proposed for conference PES 2007, Niš.

[2] Nebojša B. Raičević, Saša S. Ilić.: One Hybrid Method Application on Anisotropic Strip Lines Determination, 23-rd Annual Review of Progress in Applied Computational Electromagnetics – ACES 2007 (on CD), March 19-23, Verona, Italy, pp. 8-13, 2007.

[3] Nebojša B. Raičević, Saša S. Ilić.: *Equivalent Electrodes Method Application on Anisotropic Striplines Calculations*, International PhD-Seminar, Computational Electromagnetics and Technical Appli-cations, 28.08.-01.09. 2006, Banja Luka, Bosnia and Herzcegovina, pp. 181-186, 2006.

[4] Saša S. Ilić.: *Izvođenje Grinovih Funkcija za Potencijal Podužnog Naelektrisanja u Ravnom i Cilindričnom Anizotropnom Ogledalu*, Zbornik radova konferencije ETRAN, Beograd, 6-8 jun, str. 243-246, Srbija, 2006.

[5] Nebojša B. Raičević, Saša S. Ilić.: *Primena metoda ekvivalentne elektrode na proračun mikrostripova sa anizotropnom podlogom*, Zbornik radova konferencije TELFOR 2005-CDROM (pel 7.9), 22-24 novembar, Beograd, Srbija i Crna Gora, 2005.

[6] Saša S. Ilić, *Slit Coaxial Line*.: Inernational PhD Seminar 'Numerical Field Computation and Optimization in Electrical Engineering', Proceedings of papers, Ohrid, Macedonia, 20-25 September, pp. 107-110, 2005.

[7] Saša S. Ilić, Dragutin M. Veličković.: *Simetrično spregnuti mikrostrip u oklopu sa prorezom*, Zbornik radova konferencije ETRAN, 6-10 jun, Čačak, str. 157-160, 2004.

[8] Dragutin M. Veličković, Saša S. Ilić.: *Shielded Strip Line*, Proceedings of Mediterranean Microwave Symposium (MMS'2003), 6-8 May, Cairo, Egypt, pp. 101-106, 2003.

[9] Dragutin M. Veličković, Saša S. Ilić.: *Shielded Slot Line of Two Dielectric Layers-Theoretical Analysis*, 6-th Internatinal Conference on Applied Electromagnetics PES 2003, 1-3 June, Niš, Serbia and Montenegro, pp. 297-300, 2003.

[10] Dragutin M. Veličković, Saša S. Ilić.: *Shielded Slot Line of Two Dielectric Layers-Numerical Results*, 6-th Internatinal Conference on Applied Electromagnetics PES 2003, 1-3 June, Niš, Serbia and Montenegro, pp. 301-304, 2003.

[11] Saša S. Ilić, Dragutin M. Veličković.: *Mikrostrip između dve provodne uzemljene ravni*, Zbornik radova konferencije ETRAN, Sveska 2, 8-13 jun, Herceg Novi, str. 219-222, 2003.

[12] Dragutin M. Veličković.: *Application of Equivalent Electrodes Method in Multilayer Media*, Proc. PES2001, pp. 183-188, 8th - 10th October 2001.

[13] David Meeker, *Finite Element Method Magnetics* (FEMM), Version 3.01, http://members.aol.com/dcm3c, 2001.

[14] Duncan, J. W.; Minerva, V. P.: *Bandwitch Balun Transformer*, Proc. IRE, vol.48, pp.156-164, Feb. 1960.

[15] Gunston, M. A. R.: *Microwave Transmission - Line Impedance Data*, Van Nostrand Reinhold Company LTD, New York, Cincinnati, Toronto, Melbourne, 1972.

[16] Veličković, D. M.: *Slit Cable Calculation*, EUROEM, Bordeaux, pp. THp-04-06, 30 May-June 4, 1994.

[17] Veličković, D. M.: *Calculation of Characteristic Impedance of Coax with Axial Slit*, RAILWAYS, No. 11-12, pp. 669-671, 1996.

[18] Veličković, D. M.: *TEM Analysis of Transmission Lines Using Equivalent Electrodes Method*, TELSIKS97, Nis/Yu, Vol. I, pp. 64-74, 8-10 October 1997.

[19] Veličković, D. M.; Mančić, Z. J.; Zulkić, D. G.: *Rectangular Coax with Axial Slit and with Rectangular or Circular Center Conductor*, EMC'98 ROMA, Roma, September 1998.

[20] Veličković, D. M.; Mančić, Z. J.; Zulkić, D. G.: *Axial Slit on Two Wire Line with Rectangular Shield*, ANALELE UNIVERSITATII DIN ORADEA, Fascicola ELECTROTEHNICA, Baile Felix, Romania, pp. 18-23, 30 mai - 1 juni, 1998.

[21] Veličković, D. M.: *Equivalent Electrodes Method*, SCIENTIFIC REVIEW, pp. 207-248, Belgrade, 1996.

[22] Veličković, D. M.; Ilić, S. S.; Zulkić, D. G.: *Computer Program TUNNEL*, Technical Report, Department of Theoretical Electrotechnics, Faculty of Electronic Eng., University of Nis/Yu, 1998.

[23] Veličković, D. M.: *General Numerical Program for Plan-parallel Electrostatic Fields Solving*, 2th International Conference on Electrostatics ELSTAT 1990, Wroclaw, Poland, 17-22 September 1990 (MATERIALS SCIENCE, Vol. XVI, No 4, pp. 89-94, 1990.).

[24] Veličković, D. M.: *Metodi za proračun elektrostatičkih polja*, knjiga prva, str. 142-190, Niš, 1982.

[25] Terry Edwards, *Foundations for microstrip cirquit design, John Wiley & Sons Ltd, England*, 1992.

The Linearization of Multichannel Amplifiers Connected in Cascade with Enhanced Efficiency

Natasa Males-Ilic[1], Aleksandar Atanaskovic[2] and Bratislav Milovanovic[3]

***Abstract* – The linearization of the chain of three amplifiers connected in cascade has been performed in this paper. The second harmonics of the fundamental signals (IM2) together with the fourth-order nonlinear signals (IM4) are fed into the third amplifier in cascade. The linearization scheme is arranged so that the second amplifying stage is exploited as the source of IM2 and IM4 signals whereas the IM2 signals are extracted at the output of the first amplifier. The third amplifier in cascade operates as harmonic-controlled amplifier at class-A driven with half-sinusoidal signal. This configuration enables high gain of class A amplifier combined with higher-drain efficiency. Consequently, power-aided efficiency is increased as well as the suppression of the intermodulation products (third and fifth order) of amplifying system.**

***Keywords* – Multichannel amplifier, Intermodulation distortion, Linearization technique, Second harmonics, Saturation, Power-aided efficiency.**

I. INTRODUCTION

The linearization of multichannel amplifiers in base stations of wireless communication systems has always been of concern. The basic concept of the linearization technique with the injection of the second harmonics (IM2 signals) gives good results in reduction of the third-order intermodulation products, IM3, [1]-[3], up to the certain power of the fundamental signals that depends on load impedance [2]. The linearization technique that uses the injection of IM2 signals has been extended in [4] by introducing the injection of the fourth-order nonlinearity of the fundamental signals (IM4) in order to suppress the fifth-order intermodulation products (IM5) of a single amplifier.

In practice, the transmit path of microwave signals is usually composed of a chain of cascaded amplifiers to achieve sufficient output power and signal gain. The linearization concept considered in [5] investigates linearization of three amplifiers connected in cascade. In the concept, IM2 signals are driven into the second amplifier in cascade whereas IM2 and IM4 signals are put together (IM2+IM4) into the third amplifier. The fundamental signals are extracted at the output of the first amplifier in cascade and put to the nonlinear components biased at the appropriate points to generate the IM2 and IM2+IM4 signals.

In order to avoid two additional nonlinear components in the circuit for linearization, a new concept that uses IM2 and IM2+IM4 signals, which are generated at the amplifier stages, is considered in this paper.

II. THEORETICAL ANALYSIS

The linearization concept proposed in [5] shows remarkable results in the reduction of IM3 and IM5 products of the amplifying system consisting of three amplifiers connected in cascade. In order to simplify the linearization circuit and lower the number of additional components required for the linearization, the new concept is considered in this paper. Instead of the nonlinear transistors that are biased at appropriate point to produce IM2 or IM2+IM4 signals, those signals are extracted at the outputs of amplifiers in cascade as shown in Fig. 1. IM2 signals produced at the output of the first amplifier in cascade are combined with the IM2 and IM4 signals turn up at the output of the second amplifier that is biased at class B or AB to provide the sufficient power level of those signals. Adjusted in amplitude and phase resulting IM2+IM4 signals are put at the input of the third amplifier to achieve both the suitable drive signal for the amplifier [6] and to decrease the intermodulation products. Appropriate harmonic termination of the third amplifier that corresponds to class-F amplifier [7] gives the increase of drain efficiency accompanied with the high gain of class-A amplifier.

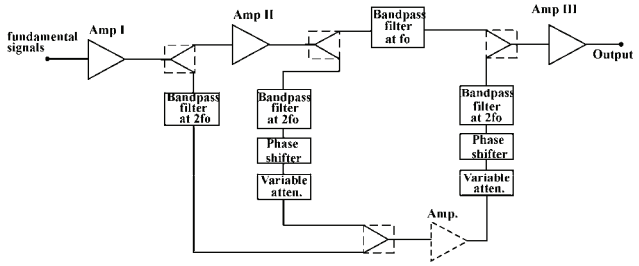


Fig. 1. Amplifying system with the additional circuit for linearization with IM2 and IM2+IM4 signals

Theoretical analysis of the proposed linearization approach is based on the nonlinearity of drain-source current expressed by a polynomial model [8], [9]. The expression for the nonlinearity of MESFET in amplifier circuit, under the assumption of neglecting the memory effect, is represented by eleven terms as given by (1). The drain-source current is dependent upon two control voltages: v_{gs}–voltage between

[1]Natasa Males-Ilic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: natasam@elfak.ni.ac.yu

[2]Aleksandar Atanaskovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: beli@elfak.ni.ac.yu

[3]Bratislav Milovanovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: bata@elfak.ni.ac.yu

gate and source and v_{ds}-voltage between drain and source of the transistor. The expression (1) connects the nonlinearity of the drain-source current, i_{ds}, in reference to voltage, v_{gs} that is represented by the coefficients $K_{10}^{(i)}$ to $K_{50}^{(i)}$. The higher order nonlinear terms $K_{40}^{(i)}$ and $K_{50}^{(i)}$ are included into the analysis according to the analysis performed in [10] that favors the terms of output current as function of v_{gs} up to the fifth-order.

The nonlinearity of drain-source current in terms of v_{ds} is expressed by the coefficients $K_{01}^{(i)}$ to $K_{03}^{(i)}$. In addition, the equation encompasses "mixing" terms $K_{11}^{(i)}$, $K_{12}^{(i)}$ and $K_{21}^{(i)}$.

$$\begin{aligned} i_{ds}(v_{gs}, v_{ds}) &= K_{10}^{(i)} v_{gs}(t) + K_{20}^{(i)} v_{gs}^2(t) + K_{30}^{(i)} v_{gs}^3(t) + \\ &+ K_{40}^{(i)} v_{gs}^4(t) + K_{50}^{(i)} v_{gs}^5(t) + \\ &+ K_{01}^{(i)} v_{ds}(t) + K_{02}^{(i)} v_{ds}^2(t) + K_{03}^{(i)} v_{ds}^3(t) + \\ &+ K_{11}^{(i)} v_{gs}(t) v_{ds}(t) + K_{21}^{(i)} v_{gs}^2(t) v_{ds}(t) + K_{12}^{(i)} v_{gs}(t) v_{ds}^2(t) + \ldots \end{aligned} \quad (1)$$

A carrier supplemented with a baseband spectrum $V_B(j\omega)$ can represent the spectrum of a digitally modulated fundamental signal, $V_{infund}(j\omega)$, as given below:

$$V_{infund.}(j\omega) = V_B(j\omega) \otimes \frac{1}{2}\delta(\omega \pm \omega_0) \quad (2)$$

The spectrum of the IM2 signals at the output of the first amplifier can be expressed as follows:

$$V_{IM2}^{(1)}(j\omega) = \left\{K_{20}^{(1)}\left[V_B(j\omega) \otimes V_B(j\omega)\right]\right\} \otimes \frac{1}{4}\left(\delta(\omega \pm 2\omega_0)\right) \quad (3)$$

The second amplifier produces the IM2 and IM4 signals which spectrum is expressed as:

$$\begin{aligned} &V_{(IM2+IM4)}^{(2)}(j\omega) = \\ &\left\{\left(K_{10}^{(1)2} K_{20}^{(2)} + K_{20}^{(1)} K_{10}^{(2)}\right) \rho_{23} e^{-j\varphi_{23}} \left[V_B(j\omega) \otimes V_B(j\omega)\right]\right\} \\ &\otimes \frac{1}{4}\left(\delta(\omega \pm 2\omega_0)\right) \\ &+ \left\{\left(K_{10}^{(1)2} K_{20}^{(1)} K_{10}^{(2)} + K_{10}^{(1)} K_{30}^{(1)} K_{20}^{(2)} + K_{10}^{(1)4} K_{40}^{(2)} + K_{40}^{(1)} K_{10}^{(2)}\right)\right. \\ &\left.\rho_{43} e^{-j\varphi_{43}} \left[V_B(j\omega) \otimes V_B(j\omega) \otimes V_B(j\omega) \otimes V_B(j\omega)\right]\right\} \\ &\otimes \frac{1}{4}\left(\delta(\omega \pm 2\omega_0)\right) \end{aligned} \quad (4)$$

Nonlinearity of the second order in the second amplifier (term $K_{20}^{(2)}$) mixes two linearly amplified fundamental signals of the first amplifier giving the first term in (4). Additionally, the second term stands for signals produced due to the second order nonlinear term in the first amplifier, $K_{20}^{(1)}$, which is linearly amplified by the second amplifier. Besides those IM2 signals, the expression (4) contains terms that relates to the IM4 signals (third to sixth terms). One can notice that they are products of nonlinear terms of first to fourth order, $K_{10}^{(i)}$ - $K_{40}^{(i)}$ i=1,2, in transfer characteristic of both first and second amplifiers. The amplitude and phase of IM2+IM4 signals from the second amplifier are adjusted to reach the required power level when combine with the IM2 signals from the output of the first amplifier. Signals obtained in this way are led to the input of the third amplifier. To approach the optimum power level of IM2 and IM4 signals for the linearization they should be set on adequate amplitude and phase throughout the injection path. Parameters ρ_{23} and φ_{23} represent amplitude and phase of the IM2 signal driven at the input of the third amplifier, whereas ρ_{43} and φ_{43} relate to the same parameters when IM4 signal is considered.

The drain-source current of the IM3 and IM5 products at the output of the third amplifier is given by (5), and (6) where the coefficients $K_{10}^{(1,2)}$ to $K_{50}^{(1,2)}$ are treated as dominant nonlinerities of the chain of the first and second amplifier in cascade. The coefficients $K_{10}^{(3)}$ to $K_{50}^{(3)}$, $K_{11}^{(3)}$ and $K_{21}^{(3)}$ characterize the third amplifier.

$$\begin{aligned} I_{ds}(j\omega)\big|_{IM3} \approx \left\{\left[\frac{3}{4} K_{30}^{(1,2)} K_{10}^{(3)} + \frac{3}{4} K_{10}^{(1,2)} K_{30}^{(3)} + \right.\right. \\ \left. + \frac{1}{4} K_{10}^{(1,2)} K_{20}^{(3)} \rho_{23} e^{-j\varphi_{23}} - \frac{1}{4} K_{10}^{(1,2)} K_{11}^{(3)} \rho_1 \rho_{23} e^{-j\varphi_{23}}\right] \\ \left. V_B(j\omega) \otimes V_B(j\omega) \otimes V_B(j\omega)\right\} \otimes \frac{1}{2}\delta(\omega \pm \omega_0) \end{aligned} \quad (5)$$

$$\begin{aligned} I_{ds}(j\omega)\big|_{IM5} \approx \left\{\frac{5}{8} K_{50}^{(1,2)} K_{10}^{(3)} + \right. \\ + K_{10}^{(1,2)}\left[\frac{5}{8} K_{50}^{(3)} + \frac{1}{4} K_{20}^{(3)} \rho_{43} e^{-j\varphi_{43}} - \frac{1}{4} K_{11}^{(3)} \rho_1 \rho_{43} e^{-j\varphi_{43}} + \right. \\ \left. + \frac{1}{8} K_{30}^{(3)} \rho_{23}^2 e^{-j2\varphi_{23}} - \frac{1}{8} K_{21}^{(3)} \rho_1 \rho_{23}^2 e^{-j2\varphi_{23}}\right] \\ \left. V_B(j\omega) \otimes V_B(j\omega) \otimes V_B(j\omega) \otimes V_B(j\omega) \otimes V_B(j\omega)\right\} \otimes \frac{1}{2}\delta(\omega \pm \omega_0) \end{aligned} \quad (6)$$

In (5) the first term relates to the IM3 products of the first and second amplifiers linearly transmitted by the third amplifier. The second term exists due to the cubic nonlinearity of the third amplifier that mingles fundamental signals. The term at the third place in the equation is made due to square nonlinearity of Amp. III, $K_{20}^{(3)}$, that mixes the fundamental signal with IM2 signal. The fourth term relates to the mixing product $K_{11}^{(3)}$ gained by interaction between IM2 signal at the input of the third amplifier and fundamental signal come out at its output. The amplitude of output voltage at fundamental signal frequency that is 180° out of phase in reference to the input signal is denoted as ρ_1.

According to this, it is possible to reduce spectral regrowth caused by third-order distortion of fundamental signal by choosing the appropriate amplitude and phase of the injected

IM2 signal (ρ_{23} and φ_{23}). Depending on the input signals' power and values of $K_{20}^{(3)}$ and $K_{11}^{(3)}$ coefficients, the fourth term may constrain the linearization effect, having the phase opposite to the third term in (5).

The first term in (6) represents IM5 products of the chained the first and second amplifiers, that are linearly amplified by the third amplifier in cascade. The second term expresses the drain-source current of the IM5 products that is formed due to the existence of fundamental signals and amplifier nonlinearity of the fifth-order, $K_{50}^{(3)}$. The third term is the mixing product between the fundamental signal and IM4 signal fed at the input of the third amplifier. Also, IM5 products are formed by $K_{11}^{(3)}$ term that stirs IM4 signals led at the input of the third amplifier and fundamental output signal. The IM5 products are also a function of $K_{30}^{(3)}$ mixing term made by a reaction between two IM2 signals and fundamental one. Mixing term which stands by $K_{21}^{(3)}$ in (6) is generated due to reaction between two IM2 signals at the third amplifier input and fundamental signal at its output. Since IM2 and IM4 signals that appear at the output of third amplifier are shortened (according to harmonic terminations in class-F amplifier) the other $K_{12}^{(3)}$ and $K_{21}^{(3)}$ terms do not exist.

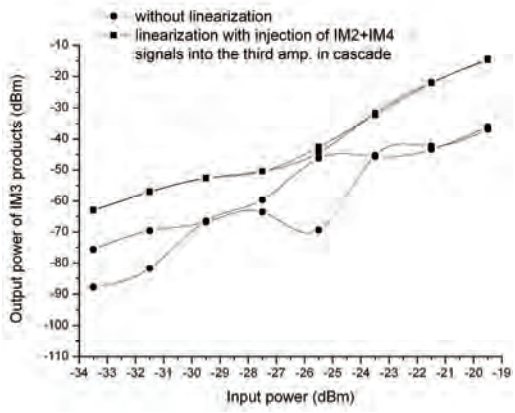
It is obvious that for the larger amplitudes of the fundamental signals the injected IM2 signals are supposed to have greater amplitudes as well according to (5). Since φ_{23} is to be equal to 180° to reduce IM3 products, the phase of the $K_{30}^{(3)}$ term in (6) is 360°. Accordingly, with the rise in amplitudes of IM2 signals, mixing $K_{30}^{(3)}$ term (the fifth term in (6)) starts increasing the power of IM5 spectrum. Due to the overlapping of IM3 and IM5 spectra the power augmentation in the range of IM3 spectrum is unavoidable. The influence of the fifth term to the power of IM3 and IM5 products can be cancelled by the sixth term.

Therefore, by adjusting the amplitude and phase of the appropriate IM4 signal the original IM5 product (the first and second terms) can be reduced. As is the case in (5), the raise in power levels of the IM2 and IM4 signals fed at the amplifier input are limited by $K_{11}^{(3)}$ term in (6) that is opposite in phase to $K_{20}^{(3)}$ term.

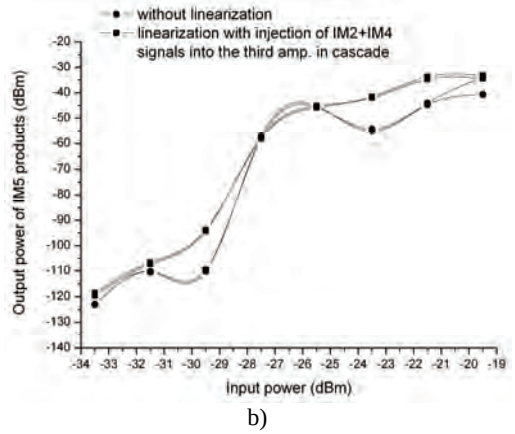
Proposed linearization concept can be applied to boost the drain efficiency of the third amplifier. This stage operates at class-A driven with half-sinusoidal signal that can be considered as the fundamental signal influenced by the second harmonics (IM2). By setting amplitude and phase of IM2 signals on appropriate values the shape of half-sinusoidal signal is control so that the third stage harmonic-controlled amplifier, [10], attains high-drain efficiency combined with high gain of class-A amplifier. Consequently, power-aided efficiency is increased in reference to class F-amplifier.

III. SIMULATED RESULTS FOR THE SECOND LINEARIZATION CONCEPT

The configuration of amplifying system consists of three amplifiers as depicted in [5]. However, in this paper the third amplifier is terminated at second harmonics (IM2 signals) and third harmonics as class-F amplifier but biased at class-A. Fig. 2 illustrates the reduction of IM3 products at frequencies 2.49 GHz and 2.52 GHz and IM5 products (2.48 GHz and 2.53 GHz) in case of two fundamental signals at the input of amplifying system, 2.5 GHz and 2.51 GHz. It follows from the figure that the suppression of IM3 products is not equal for both IM3 products in entire power range of input power observed. At the points where optimization is performed, (-29 dBm and -21 dBm), IM3 products is approximately 10 dB and 20 dB lower than before linearization, respectively. As far as IM5 products are concerned, the results are worse than those gained by the linearization concept described in [5] that also uses IM2 +IM4 signals. However, the linearization concept proposed here includes the linearization circuit of less complexity enabling satisfactory results for IM3 products in the large range of the input power going up to the higher power levels close to saturation point.



a)



b)

Fig. 2. Output power in terms of the input power before and after linearization for: (a) IM3 products; (b) IM5 products

In the case when the third stage is class-F amplifier the maximum negative gate voltage ascends to double of half-sinusoidal driven class-A harmonic control amplifier so that it

is necessary to rise the input power of fundamental signals in order to achieve the same output power. It is obvious that the class-A amplifier in the third stage provides the higher gain and output power than class-F amplifier; therefore, the power-aided efficiency (PAE) is improved in reference to the class-F case by 3 % observed at -19 dBm input power.

Additionally, the designed amplifying system has been tested for three sinusoidal fundamental signals at frequencies 2.5 GHz, 2.51 GHz and 2.522 GHz when the power of fundamental signals at the input of amplifier is –23 dBm that is 5 dB below 1-dB compression point of the third amplifier. The output spectra consisting of the fundamental signals, IM3 and IM5 products are compared in Fig. 3 for the cases before and after linearization. Various results are gained for different kinds of IM3 and IM5 signals. For example, IM3 products at frequencies $2\omega_i-\omega_j$ (the first kind) are lowered by 18 dB whereas they are approximately reduced by 15 dB at frequencies $\omega_i+\omega_j-\omega_k$ (the second kind) $i\neq j\neq k \in(1,2,3)$. It follows from Fig. 3 that all IM5 products are kept at the lower or nearly equal power level to the linearized IM3 products.

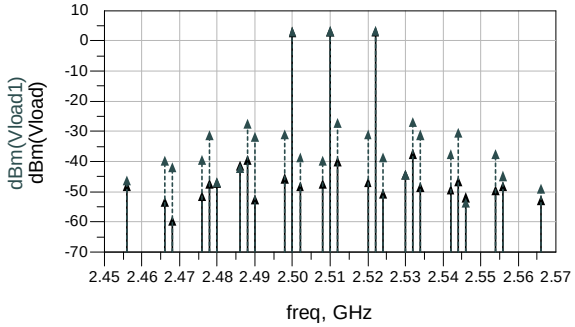


Fig. 3 Output spectrum before and after linearization for three fundamental sinusoidal signals

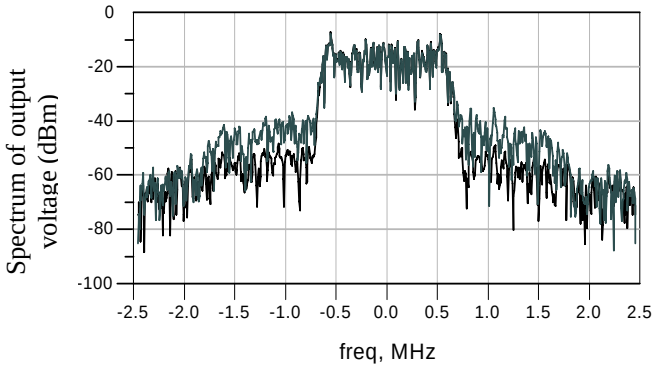


Fig. 4. Simulated spectrum of the output voltage for OQPSK digitally modulated signal before (grey line) and after linearization (black line) with the second concept for -19 dBm carrier input power

On the top of that, the amplifying system has been tested for OQPSK digitally modulated signal at carrier frequency 2.5 GHz and spectrum width 1.25 MHz. The results shown in Fig.4 compare the output spectra before and after the linearization for -19 dBm power levels of the fundamental signals. Improvement observed at ±900 kHz offset from the carrier over the 30 KHz frequency range is 11 dB whereas the ACPR at offset ±2.1 MHz becomes better for 4 dB.

IV. Conclusions

The new linearization concept of the amplifiers in cascade has been considered in order to simplify the linearization circuit. The linearization is based on the injection of the second harmonics (IM2 signals) and fourth-order nonlinear signals (IM4). IM2 and IM2+IM4 signals are extracted at the outputs of amplifiers in cascade and combined. When adjusted in amplitude and phase they are injected at the input of the third amplifier in cascade. It should be pointed out that the linearization circuit of less complexity reaches remarkable reduction of IM3 products for configuration of amplifying system which enables high power-aided efficiency. Injected IM2 and IM4 signals are adjusted not only to achieve the highest possible PAE but also to lower the IM3 and IM5 products, so the achieved results represents the compromise between the excellent results for both PAE and IM products simultaneously.

References

[1] Aitchison, C. S.; Mbabele, M.; Moazzam, M. R.; Budimir, D.; Ali, F.: Improvement of third order intermodulation products of RF and microwave amplifiers by injection. IEEE Trans. Microwave Theory and Techniques 49 (2001), 1148-1154.

[2] Kusunoki, S.; Kawakami, K.; Hatsugai, T.: Load-impedance and bias-network dependence of power amplifier with second harmonic injection. IEEE Trans. Microwave Theory and Techniques 52 (2004), 2169-2176.

[3] Males-Ilic, N.; Milovanovic, B.; Budimir, Dj.: Improvement in “Second Harmonics” Linearization Technique for Multichannel Amplifiers. Microwave and Optical Technology letters 38 (2003), 150-153.

[4] Males-Ilic, N.; Budimir, Dj.; Milovanovic, B.: Linearization technique for reducing third- and fifth-order intermodulation distortion products in multichannel amplifiers. Proc. International Conference EUMC 03, Munich, Germany, 7-9 October 2003, CD.

[5] Males-Ilic, N.; Milovanovic, B.: The linearization of multichannel amplifiers connected in cascade. Proc. International Conference MMS ’06, Genova, Italy, 21-23 September 2006, CD.

[6] Ingruber, B.; Wachutka, M.: High-Efficiency Harmonic-Control Amplifier. IEEE Trans. Microwave Theory and Techniques 46 (1998), 857-862.

[7] Raab, F.: Class-F Power Amplifiers with Maximally Flat Waveforms. IEEE Trans. Microwave Theory and Techniques 45 (1997), 2007-2012.

[8] Heiskanen, A.; Aikio, J.; Rahkonen, T.: A 5-th order Volterra study of a 30W LDMOS power amplifier. Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS), Bangkok, Thailand, 25-28 May 2003, 616-619.

[9] Aikio, J. P.; Rahkonen, T.: Detailed distortion analysis technique based on simulated large-signal voltage and current spectra. IEEE Trans. Microwave Theory and Techniques 53 (2005), 3057-3065.

[10] Colantonio, P.; Gianini, F.; Leuzzi, G.; Limiti, E.: High-Efficiency Low-IM Microwave PA Design. Microwave Symposium Digest, Phoenix, AZ, USA, 20-25 May 2001, Vol. 1, 511-514.

Software for Microwave Slotted Patch Antenna Design Based on Hybrid Empirical-Neural Model

Bratislav Milovanović, Zoran Stanković, Marija Milijić and Maja Sarevska

Abstract-**This paper presents a software *"MW Patch Antenna Design 2.0"* intended for modeling microwave slotted patch antennas. This software uses Hybrid Empirical-Neural Model (HEN) which, unlike the models based on a classical multi-layer perceptron (MLP) network, includes an existing partial knowledge about the resonant frequency behavior of patch antenna, yielding more accurate determination of the resonant frequencies f_r and the minimum value of S_{11} parameter (S_{11min}).**

Keywords-**Neural network, patch antenna, modeling, software**

I. INTRODUCTION

A microstrip patch antenna provides a great advantage over traditional antennas such as efficacy, compact physically realization and mechanical reliability. This narrowband, wide-beam antenna is lightweight, inexpensive and conformable. Due its numerous advantages, it is often used in wireless communication systems, mobile communications and many microwave applications [1]. Therefore, the tools for patch antenna modeling are dictated to be fast and failsafe.

The typical patch antenna is shown in Figure 1. It consists of a radiating patch of any planar geometry (e.g. circle, square, ellipse, ring and rectangle) on one side of a dielectric material substrate backed by a ground plane on the other side. The patch is generally made of conducting material such as copper or gold [2]. The radiating patch and the feed lines are usually photo etched on the dielectric substrate. A substrate with a low dielectric constant (ε_r) is used (typically ~ 2.5), but in loss critical applications, Alumina (ε_r =10) must be used [3].

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

Good alternative for overcoming all these problems is modeling of patch antennas using an artificial neural network model which can be fast and accurate as detailed EM simulation method [4,5,6]. But, the main disadvantage of MLP models is a need for providing a large set of training data, that could be difficult and time-consuming process [7].

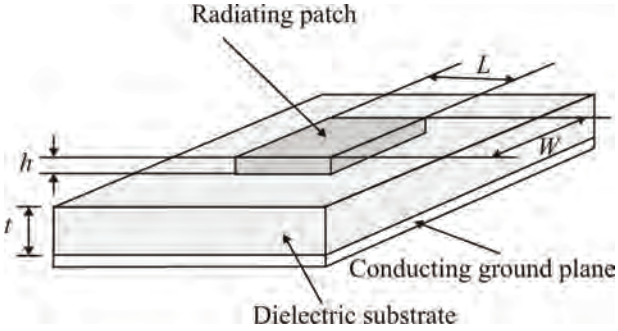


Figure 1. Typical patch antenna

Many earlier researches modeled only the resonant frequency of patch antennas [8,9,10], while they omitted modeling of patch antennas S_{11min} parameters which represents patch antenna losses. Hybrid Empirical-Neural (HEN) model of patch antennas, which incorporates knowledge about both the behavior of the resonant frequency and the behavior of S_{11min} parameters, needs smaller set of training data then MLP model for achieving satisfying model accuracy [11]. This HEN has six input parameters: patch antenna length L, patch antenna width W, depth of patch antenna slot l, width of patch antenna slot s, the resonant frequency f_r^e determined by approximate model and S^e_{11min} parameter determined by specially developed MLP model and it gives value of resonant frequency f_r and value of S_{11min} parameter of patch antenna. It is used for software "MW Patch Antenna Design 2.0" developing that calculates resonant frequency f_r and S_{11min} parameter of patch antenna.

II. KNOWLEDGE ABOUT PATCH ANTENNA RESONANT FREQUENCY

Slotted patch antenna, whose HEN model is presented in this paper, is shown in Fig. 2. Antenna is constructed of substrate with fallowing features: relative dielectric constant ε_r = 2.17, thickness h = 0.508 mm, conductor thickness T = 0.017 mm. Detailed earlier research [4,5,6] has shown that the resonant frequency f_r of patch antenna depends on patch antenna length L, patch antenna width W, depth of patch antenna slot l and width of patch antenna slot s for constant parameter w=0.5 mm:

The authors are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, Niš, Serbia and Montenegro, E-mail:batam@pogled.net, zoran@elfak.ni.ac.yu, marijam@elfak.ni.ac.yu,

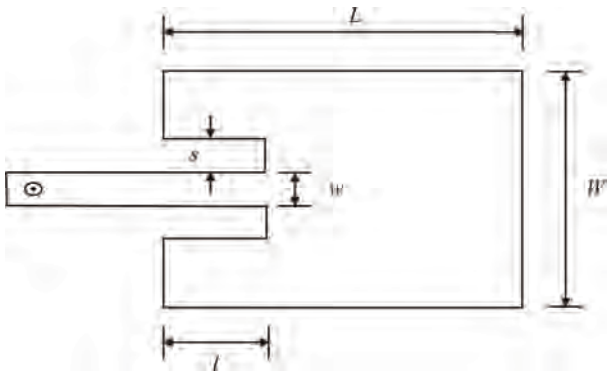


Fig. 2. Slotted patch antenna

$$f_r = f(L, W, l, s) \tag{1}$$

Also, the resonant frequency of a square-patch antenna in its dominant mode can be calculated using equation given by Wolf and Knoppik [8]:

$$f_r = \frac{c}{2L_{eff}\sqrt{\varepsilon_{dyn}}} \tag{2}$$

where c is the velocity of light in free-space, L_{eff} is the effective patch length:

$$L_{eff} = L + \Delta L \tag{3}$$

where ΔL is given as:

$$\Delta L = 0.412h\frac{(\varepsilon_{reff} + 0.3)(\frac{W}{h} + 0.264)}{(\varepsilon_{reff} - 0.258)(\frac{W}{h} + 0.8)} \tag{4}$$

ε_{dyn} is the dynamic dielectric constant considering the fringing field effects:

$$\varepsilon_{dyn} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left[1 + 12\frac{h}{W}\right]^{-1/2} \tag{5}$$

The resonant frequency f_r of a square-patch antenna given in analytical form (2) represents a partial semi-empirical knowledge from the problem domain implemented in the structure of HEN model whose architecture will be exposed further in this paper.

III. Hybrid Empirical-Neural Model of Patch Antenna

Hybrid Empirical-Neural Model of patch antenna appears as integration of the patch antenna approximate model as empirical knowledge holder, MLP network for calculating the value of S_{11min} parameter of patch antenna and MLP network. The basic idea in HEN model realization is that the empirical model and MLP network for calculating the value of S_{11min} parameter with corresponding connection to the neural

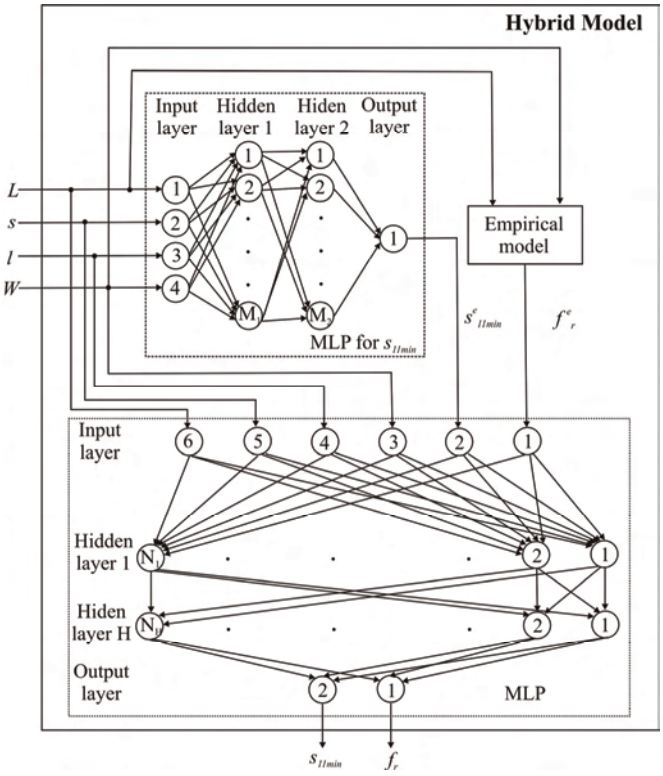


Fig. 3. Hybrid model of patch antenna

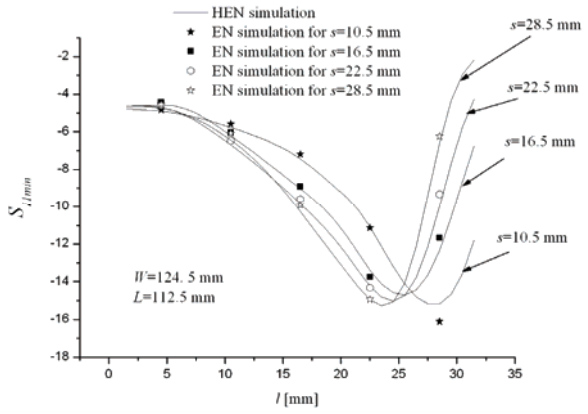
network provides higher generalization and extrapolation capabilities of the network [7]. This is achieved by presenting extra information about the problem at the input of the network. According to that, architecture of HEN model for slotted patch antenna design is presented in Fig.3. Approximate model determines the resonant frequency f_r^e using equation (2). MLP network for calculating the value of S_{11min} parameter has been trained to give approximate S^e_{11min} value. The output from the approximate model, f_r^e, and the output from MLP network for calculating the value of S_{11min} parameter, S^e_{11min}, are brought to MLP as additional inputs. According to this, MLP of the HEN model is given by $\mathbf{y}=\mathrm{y}(\mathbf{x},\mathbf{w})$, where $\mathbf{w}$ is a connection weight matrix among neurons [7,9], $\mathbf{x}=[L,W,l,s,\ f_r^e,S^e_{11min}]^T$ is the input vector, and output vector is $\mathbf{y}=[f_r,S_{11min}]$. During the training, network weight matrix $\mathbf{w}$ have to be adjusted in order to make the total main squared error $E(\mathbf{w})$ between the desired outputs and the actual outputs from MLP network, lower than the prescribed value E_c [7]. Adjustment of the network parameters is determined by the chosen training algorithm. General symbol for this HEN model is HEN*H-N1-…-N$_l$-…-N$_H$* where H is the total number of layers of neurons and N_i is the number of neurons in the i-th hidden layer. Activation functions of the hidden layers are sigmoid [7], while the output layer has linear activation function.

IV. Modeling Examples

HEN model is trained for calculation resonant frequency f_r and S_{11min} parameter of patch antenna. It is done in the wider range of input parameters: 52.5 mm ≤ L ≤ 112.5 mm, 4.5 mm

Table I TESTING RESULTS FOR HYBRID MODEL OF PATCH ANTENNA

Neural model	*WCE*	*ACE*	r^{PPM}
HEN4-4-2	17.28	2.05	0.97
HEN4-3-3	20.21	2.14	0.96
HEN4-2-2	21.20	2.09	0.97
HEN4-3-2	21.62	2.18	0.97
HEN4-2-4	22.02	2.14	0.96
HEN4-2-3	22.03	2.10	0.97
HEN4-3-4	22.47	2.24	0.96
HEN4-2-5	23.32	2.13	0.96



Slika 4 S_{11min} vs. parameters l

$\leq s \leq$ 28.5 mm, 4.5 mm $\leq l \leq$ 28.5 mm and 82.5 mm $\leq W \leq$ 112.5 mm. A training set of 375 samples has been obtained by electromagnetic simulation using HFSS 9.0 software.

Testing of the HEN model is done using testing data set (128 samples) that is not used in training process. The testing results for eight HEN models with the lowest average testing error (average test error (ACE [%]), worst-case error (WCE [%]) and Pearson Product-Moment correlation coefficient (r^{PPM})) are shown in TABLE I. Compared to the testing results of classical MLP model [6], it can see that the best HEN model is more accurate then the best MLP model, which vindicates using HEN model for patch antennas.

H4-4-2 model has the best testing results and it is used for checking obtained generalization level. First, it is used for representation parameter S_{11min} as function of antenna parameters l and these results have been compared with values obtained by EM simulation (Fig. 4.). As this figure shows, H4-4-2 model is correct and it can be used for modeling patch antenna S_{11min} parameter with acceptable accuracy.

This HEN architecture improves patch antenna modeling with very great speed of work. In Fig. 5., the dependence f_r on parameter L and W, when l and s parameters are constant, is shown. This dependence is shown using 900 values of f_r obtained by NN simulation for less then 1 seconds. If we use EM simulation to obtain the same number of f_r values, we will do it for 1 hour. Also, Fig. 6. represents S_{11min} as function of parameters s and l when W and L are constant values. If we use EM simulation to obtain 378000 values of S_{11min} parameter, we will finish it after 20 days of simulation. NN simulation obtains 378000 needed values of S_{11min} parameter

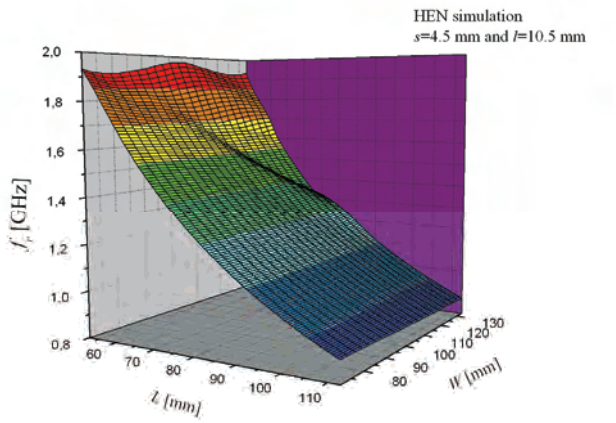


Fig. 5. f_r vs. parameters L and W

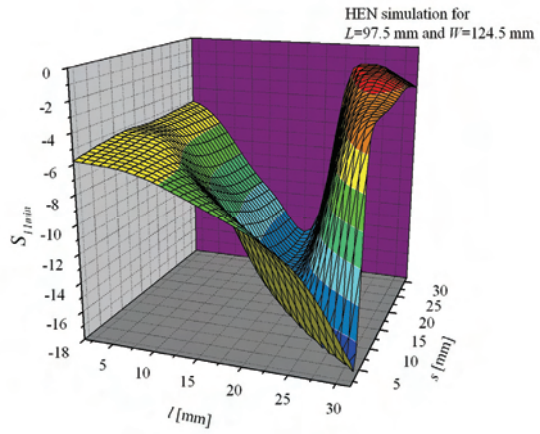


Fig. 6. S_{11min} vs. parameters l and s

for 2 second. For these reasons, NN simulation is better alternative in applications where simulation has to be finished in certain period of time.

V. MW PATCH ANTENNA DESIGN 2.0

Software module "MW Patch Antenna Design 2.0", whose code is written in Visual C++ programming language, uses HEN model for microwave slotted patch antennas design. Range of input parameters is limited by the range of parameters from training process. When user inputs parameters, the values of resonant frequency f_r and S_{11min} parameter are calculated for the patch antenna with these dimensions. This model works with great speed, e.g. the calculating resonant frequency f_r and S_{11min} parameter is done for less then 1 second. Also, it offers the possibility of optimization. It means that user do not have to input all antenna parameters (Fig. 7). The user can input one, two, three or no one parameters and this software module can calculate the values of resonant frequency f_r and antenna parameters when S_{11min} parameter is minimal, i.e. when antenna has minimal losses. Likewise, the range of resonant frequency f_r can be chosen. As a help to the user, the range of resonant frequency f_r is shown for the input antenna

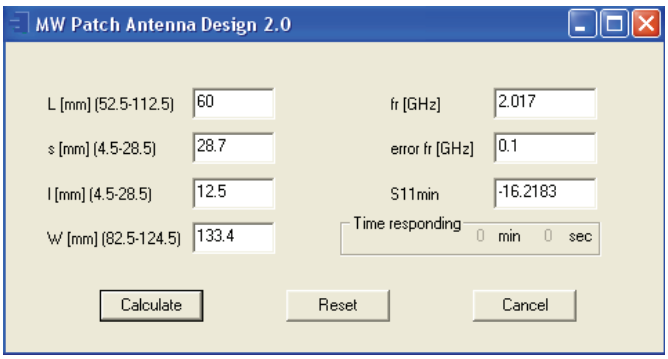


Fig. 7 Software MW Patch Antenna Design 2.0

parameters. If some antenna parameters are input, they are considered as constants and the other parameters values are looked for to obtain minimal value of S_{11min} parameter. The algorithm of this software is shown in Fig. 8. It calculates antennas parameters with accuracy of 10^{-1} mm while the accuracy of EM simulation is 100 mm. The optimization process has two phases. The first, which finds the approximate minimum of function S_{11min} =**g**(L,s,l,W), has accuracy of 100 mm. The second finds the real minimum of function **g** and its accuracy is 10^{-1} mm.

VI. CONCLUSION

EM simulation is the most used method for modeling of patch antenna. Despite its good accuracy, it has many limitations (the complicated, slow simulation process, high demands concerning the hardware resources necessary for its software implementation). Good alternative is to use neural network modeling. MLP model is easy to be developed, has high simulation speed and satisfying accuracy if its training set is large enough. But, in cases when the training set is too small for the acceptable MLP model accuracy, the HEN model retains accurate modeling by incorporating existing knowledge for problem domain while it keeps the same simulation speed. This can be main advantage of using HEN model instead MLP models for microwave patch antennas. This HEN model is used to develope software module "MW Patch Antenna Design 2.0", which is high qualitative, fast and accurate programme and which can satisfy the user's requests.

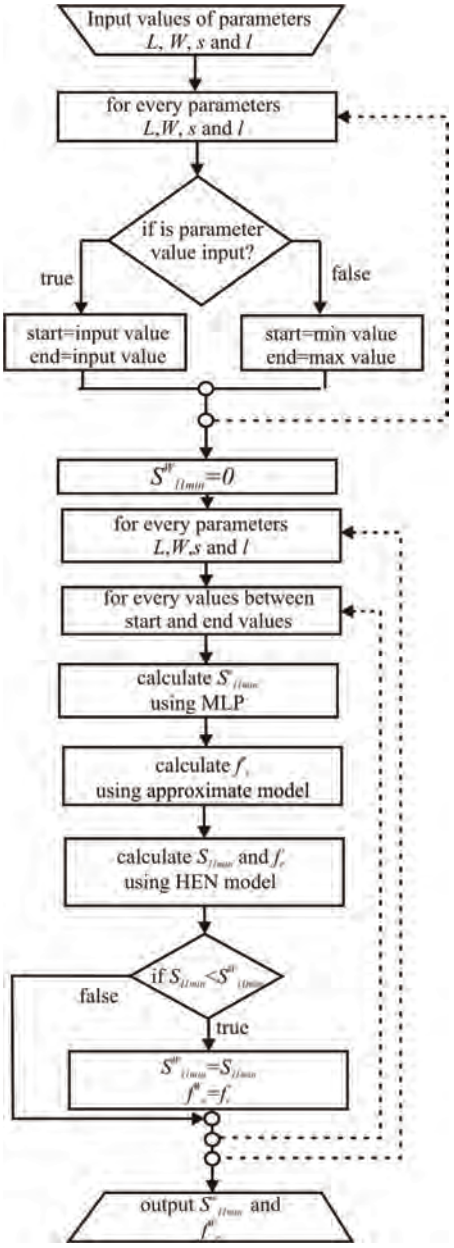


Fig. 8 The optimization process

REFERENCES

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

[2] etd.lib.fsu.edu/theses/available/etd-04102004-143656/unrestricted/ Chapter3.pdf

[3] www.odyseus.nildram.co.uk/Systems_And_Devices_Files/Patch_Antenna.pdf

[4] Marija Milijić, Zoran Stanković, Aleksandar Atanasković, "Modelovanje mikrotalasnih patch antena pomoću neuronskih mreža", ETRAN 2005, Budva, 2005

[5] Bratislav Milovanović, Marija Milijić, Aleksandar Atanasković, Zoran Stanković, "Modeling of Patch Antennas Using Neural Networks", TELSIK 2005 Konferencija, Niš, 2005

[6] Bratislav Milovanović, Zoran Stanković, Marija Milijić, "Efficient Neural Model of Microwave Patch Antennas", ICEST 2006 Konferencija, Sofia, Bulgaria, 2006

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

[8] Nurhan Turker, Filiz Gunes, Tulay Yildirim, "Artificial Neural Networks Applied to the Design of Microstrip Antennas", Microwave Review, No.1, Vol. 12, pp. 10-14, June 2006.

[9] R. K. Mishra, A. Patnaik, "Neural network-based CAD model for the design of square-patch antennas", Antennas and Propagation, IEEE Transactions, Vol. 46, No. 12, pp. 1890-1891, December 1998.

[10] R. K. Mishra, A. Patnaik, "Designing rectangular patch antenna using the neurospectral method", Antennas and Propagation, IEEE Transactions, Vol. 51, No. 8, pp. 1914-1921, Aug. 2003.

[11] Bratislav Milovanović, Zoran Stanković, Marija Milijić, "Hybrid Empirical-Neural Model of Microwave Slotted Patch Antennas", 8th Seminar on Neural Network Applications in Electrical Engineering, NEUREL-2006, Belgrade, Serbia, 2006

SESSION CSIT II

Computer Systems and Internet Technologies II

Simulation of Next Generation Network Applications

Ivaylo I. Atanasov[1]

***Abstract* – The paper presents a way for functional validation of an approach to service creation. The essence of the approach is synthesis of a new mark-up language used to describe service logic. An application that uses network functions accessed through open programming interfaces is designed. The application logic is described using the proposed mark-up language and the application behaviour is simulated by the use of network resource gateway simulator.**

***Keywords* – Mark-up languages for service creation, Parlay/OSA interfaces, network resource gateway simulator**

I. Introduction

Next generation networks (NGN) are expected to provide content rich and customized services. To be competitive network operators need to have not only the right technology, but the right tools to create these services efficiently.

An important ingredient of NGN service architecture is the concept of Application Programming Interfaces (APIs) that allows third parties to get in on developing services for telecommunication networks, with transportable suppliers independent skills [1]. A promising technology providing access to network function through open API is Parlay/OSA (Open service access).

One of the ways of implementing Parlay/OSA is by the use of mark-up languages. There exist some markup languages for service creation [2], but none of them supports the full range of network functions exposed by Parlay/OSA.

Service Logic Processing Language (SLPL) is a new mark-up language developed to meet challenges of NGN service creation [3]. The language supports the whole palette of network functions exposed by Parlay/OSA interfaces. To allow language usability an SLPL interpreter is developed. The SLPL functionality is validated by simulation.

First in the paper, the network resource simulator used is presented. Then an example that uses network functions accessed through Parlay/OSA interfaces is considered. The suggested mark-up language SLPL is presented in the context of the application logic description. Last, the process of simulation of the application functional behaviour is explained.

II. Network Resource Gateway Simulator

To validate the approach applicability the Ericsson Network Resource Gateway (NRG) simulator is used [4]. The NRG simulator emulates NRG node and its purpose is to test applications in absence of an actual NRG running on a live telecom network.

The NRG product is Service Capability Server, which provides network services to applications in a secure way and controls the network elements. Beneath the service capability server is the telecom network which uses a wide range of network protocols to implement the services in the network.

As it is shown in Fig. 1, the NRG provides a set of APIs independent from underlying network.

The NRG product includes Software Development Kit (SDK) also. The SDK offers software libraries to Java application developers, which simplify the development of applications that use NRG capabilities.

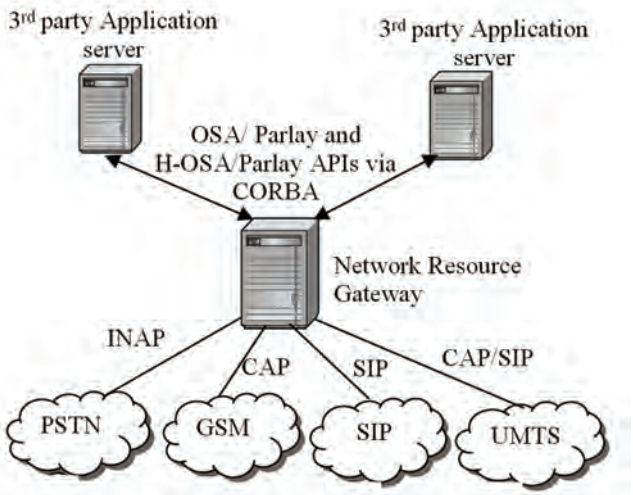


Fig. 1 Ericsson Network Resource Gateway

As it is shown in Fig. 2, by using the NRG the developers no longer need detailed knowledge of telecom network in order to use its services and with the help of SDK the need for detailed Common object request broker (CORBA) knowledge in order to use NRG is removed as well.

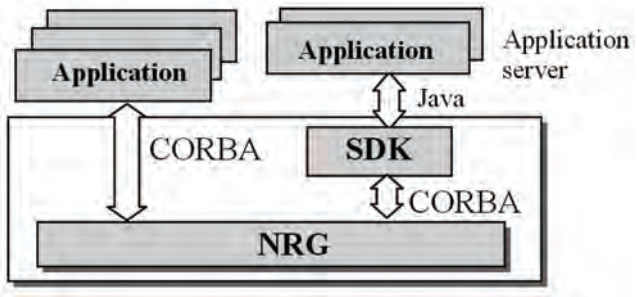


Fig. 2 An overview of NRG SDK

The NRG simulator mimics the behavior of an actual NRG node. It simulates not only the node but also the underlying

[1]Ivaylo I. Atanasov is with the Faculty of Communications and Communication Technologies, Technical University of Sofia, 1000 Sofia, Bulgaria, E-mail: iia@tu-sofia.bg

network resources (such as telecommunication network). It uses a graphical interface that allows a user to define the behavior of network resources.

The NRG SDK provides API for several services including framework, multi-party call control, user interaction, user location user status and messaging. Before an application can get access to a service, the application needs to authenticate itself towards the NRG using the framework API.

Most applications follow a similar design involving the following classes:

- Main – initializes, starts, stops and terminates the application
- Configuration – provides application with configuration data
- Graphical user interface (GUI) – provides a graphical user interface
- Feature – implements application logic
- YY_processors – use service managers to send requests to the NRG and to receive callback responses.

A feature (application) can have multiple processors.

In the next section an example of application that uses SDK APIs is considered.

III. Case Study

Let us consider a "Local entertainments" application that provides information about places of entertainment in a city. When a user dials the number of "Local entertainments" service, the application locates her position and asks the user about requested information for example, restaurants, discos or pubs. After the user enters her choice, the message is played containing information about the places of entertainments in the vicinity. The information about local places of entertainment is retrieved from a database.

The application uses the following classes available in NRG SDK:

- MPCCProcessor – a multi-party call control processor that uses service manager IpMPCCManager
- LocationProcessor – a processor used to determine the user position
- UIProcessor – a processor for user interaction that uses service manager IpUIManager
- IpMPCCManager – an interface for receiving results and notifications from IpMPCC interface
- IpUIManager – an interface for receiving results and notifications from IpUICall interface
- IpMPCC interface offering methods for multi-party call handling
- IpUserLocation interface offering methods for user location
- IpUICall interface offering methods for user interaction.

Further for the aim of the application two more classes are developed:

- DBProcessor – a database processor
- IpDataBase interface offering methods for database access.

The sequence of actions performed during application execution is shown in Fig. 3.

1 After the user dials the service number, IpMPCCManager notifies MPCCProcessor about the event.
2 The event is forwarded to the application.
3 The application requests user location.
4 The LocationProcessor sends a request for positioning to the user location service.
5 The user location service provides requested information.
6 The information about user location is forwarded to the application.
7 The application requests from the database identifications of the voice messages containing information about places of entertainment. The user coordinates are sent as parameter.
8 The DBProcessor calls DB_retrieve method to extract the requested data.
9 The result of database query is forwarded to the application. The result contains message IDs for places of entertainment.
10 The application starts dialogue with the user.
11 User interaction session is created.
12 The application requests from the UIProcessor to prompt the user about the requested entertainment type by entering digit.
13 UIProcessor requests from the user interaction service to prompt and collect information.
14 The user's choice is sent to the application.
15 The application sends to the UIProcessor the identification of the message to be sent.
16 UIProcessor requests from the user interaction service to play the message.
17 When the message ends this is reported to the UIProcessor.
18 The application requests to release user interaction session.
19 The UIProcessor frees the user interaction service.
20 The application deallocates the call-related resources.
21 The multi-party call control service is released.

IV. SLPL Description of Application Logic

The SLPL application logic description consists of definition and executive parts. In the definition part data types and methods are defined, and variables are declared.

The SLPL supports all data types defined in OSA specifications. Fig. 4 shows the definition of data type that represents the result returned by the database which contains the ID of the menu to be played and message IDs of the entertainment places in the vicinity.

Fig.5 shows variables used in the application script.

The application methods are also defined. When the user dials the "Local entertainments" service number the application is notified. The method "handleCall" handles incoming calls to the service, the method "locationReceived" is used to receive the geographical coordinates, and the method "locationTranslated" is used to get the database response in form of messages IDs corresponding to the geographical coordinates.

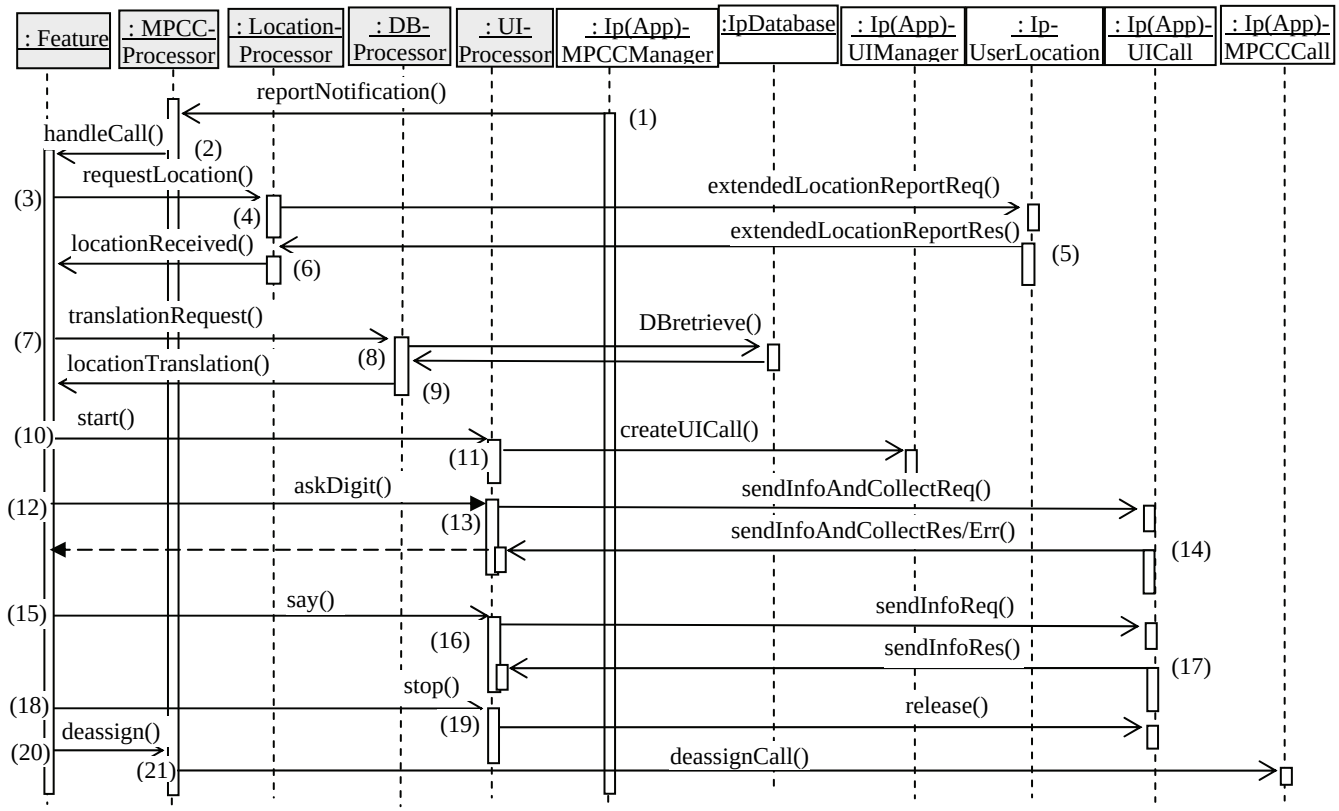


Fig. 3 Sequence diagram for "Local entertainments" applicationVariables of the types defined are declared.

```
<types>
    <structure name="PlaceDescription">
      <element name="menu" type="integer"/>
      <element name="Restaurants" type="integer"/>
      <element name="Discos" type="integer"/>
      <element name="Pubs" type="integer"/>
    </structure>
</types>
```

Fig. 4 An example of SLPL definition of structure type

```
<variables>
      <id name="theLongitude" type="float"/>
      <id name="theLatitude" type="float"/>
      <id name="thePlacesID" type="integer"/>
      <id name="theUICall" type="TpUICallIdentifier"/>
      <id name="thePlaces" type="PlaceDescription"/>
      <id name="digit" type="integer"/>
      <id name="menuID" type="integer"/>
</variables>
```

Fig. 5 An example of SLPL variable declaration

Fig.6 shows the definition of the method "handleCall" and the method "locationTranslated".

The executive part of the service logic script is built of statements. To request a service from the network-side of interface, the application logic has to invoke its methods. On method invocation its interface is specified and actual value of its argument is given. To allow synchronous communications the application has to wait for the result of network services. By invocation of application's methods, the network-side interfaces return the results of the requested service. "Case-statement" is used for multiple choices.

```
<method name="handleCall">
  <arguments>
    <argument name="aCall"
      type="TpMultiPartyCallIdentifier"/>
    <argument name="aLeg" type="TpCallLegIdentifier"/>
    <argument name="anOriginatingAddress"
      type="TpAddress"/>
  </arguments>
  <returns/>
  <body>
    <set refid="theCall" valref="aCall"/>
    <set refid="theLeg" valref="aLeg"/>
    <set refid="theOrigAddress"
      valref="anOriginatingAddress"/>
  </body>
</method>
<method name="locationTranslated">
  <arguments>
    <argument name="PlaceIDs" type="PlaceDescription"/>
  </arguments>
  <returns/>
  <body>
    <set refid="thePlaces" valref="PlaceIDs"/>
  </body>
</method>
```

Fig. 6 An example of SLPL method definition

First the application waits for call notifications. On call notification, the method "locationRequest" is invoked to receive user location. The application waits for user location data. On receiving user location, the method "translation-Request" is invoked to query database about message IDs for entertainments in the vicinity. The application waits for database response. To start interaction with the user the

method "start" is invoked. The method "askDigit" is invoked to prompt the user about her choice. The method "say" is invoked to play the message.

Fig. 7 shows the skeleton of the executive part of the script.

```
<execute>
  <wait/>
  <!—invoke method LocationProcessor.locationRequest -->
  <wait/><! – wait for user location data -->
  <!—invoke method "DBProcessor.translationRequest" -->
  <wait/> <! -- the DB result is received in variable thePlaces -->
  <set refid="theUICall">
    <invoke>
        <method name="itsUIProcessor.start">
          <arguments>
            <argument refid="theCall"/>
          </arguments>
        </method>
    </invoke>
  </set>
  <set refid="menuID" valref="thePlaces">
    <value><element name=”menu”/></value>
  </set>
  <set refid="digit">
    <invoke>
       <method name="itsUIProcessor.askDigit">
         <arguments>
           <argument name="aMsgID" refid="menuID"/>
         </arguments>
       </method>
    </invoke>
  </set>
  <case refid="digit">
    <on val="1">
      <set refid=”thePlaceID" valref="thePlaces">
        <value><element name="Restaurants"/></value>
      </set>
    </on>
    <on val="2">
      <set refid="thePlaceID" valref="thePlaces">
        <value><element name="Discos"/></value>
      </set>
    </on>
    <on val="3">
      <set refid="thePlaceID" valref="thePlaces">
        <value><element name="Pubs"/></value>
      </set>
    </on>
  </case>
  <!-- invoke method itsUIProcessor.say -->
  <!-- invoke method itsMPCCProcessor.deassign -->
</execute>
```

Fig. 7 Skeleton of executive part of SLPL script

V. Simulation of Application Logic

The SLPL interpreter is registered in the Framework of the NRG. When simulating the application behavior, the interpreter is in the role of an object of class Feature as depicted in Fig.3. The SLPL interpreter is supplied with the SLPL description of "Local entertainments" application.

When the "Local entertainments" application is started, the SLPL interpreter verifies the syntactical correctness of the description, generates an abstract syntax tree and makes the respective mappings of the abstract syntax tree onto the semantics i.e. calls corresponding Java methods. The calls of the interpreter are of methods exposed by the Ericsson NRG simulator.

In runtime on request of the user of the "Local entertainments" application, the current location is obtained and displayed in a map. When dialing the service number, an announcement is played asking for entering digit. After the user enters a digit, a message with requested information is played.

Fig. 8 shows a screenshot of service simulation

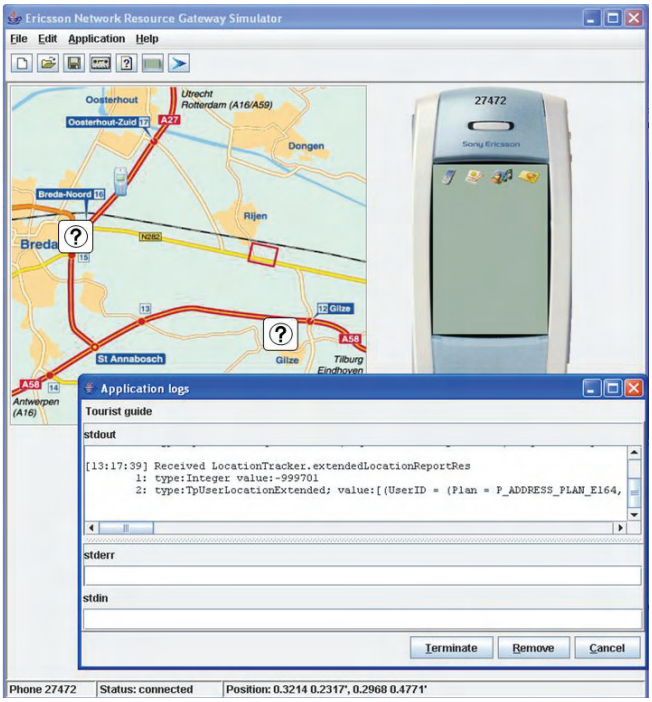


Fig. 8 A screenshot of "Local entertainments" application simulation

VI. Conclusion

In this paper an example of Parlay/OSA application is considered. The application logic is described using a new markup language SLPL. The SLPL functional applicability is verified by simulation.

References

[1] Bakker, J. Tweedie, D. and M. Umnehopa, “Evolving service Creation; New developments in Network Intelligence”, http://www.argreenhouse.com/papers/jlbakker/Bakker-telenor.pdf

[2] Bakker J-L, R. Jain, “Next Generation Service Creation Using XML Scripting Languages”, http://www.argreenhouse.com/papers/jlbakker/bakker-icc2002.pdf

[3] Atanasov I., E. Pencheva, “A Mark-up Approach to Add Value”, IJIT Enformatika, vol.3, Number 4, pp 267-276

[4] Ericsson Network Resource Gateway SDK (version R5A02) http://www.ericsson.com/mobilityworld/sub/open/technologies/parlay/tools/parlay_sdk

Pricing the Internet Services Using Extensive Game Solution

Vesna Radonjić[1] and Vladanka Aćimović Raspopović[2]

***Abstract* – In this paper we described the application of extensive game model. We considered the case of the economic interactions between local Internet Service Providers in a region which we modeled as an extensive game. We founded three Nash equilibrium points as the solutions of the game.**

***Keywords* – Pricing, Extensive game, Nash equilibrium.**

I. INTRODUCTION

The development of good pricing models for the Internet is a topic of current interest. Most Internet Service Providers (ISPs) currently employ flat rate pricing, i.e. user is charged a fixed amount per time unit, irrespective of usage [1]. Others, following the telephony model, price the time spent connected to the Internet (some dial-up services are priced this way). ISPs may also use percentile-based charging. Still others charge based on actual bytes transferred. In general, pricing schemes have to be defined and evaluated with respect to the heterogeneous technical, economic and social aspects.

Game theory, as a mathematical basis for the analysis of interactive decision-making processes, can be applied for solving various pricing problems in the current Internet. It is a collection of modeling tools that aid in the understanding of interactive decision problems.

In this paper we study how members of a group, who are in an identical position in the hierarchy, interact with each other. As an example, we consider the case of the economic interactions of local ISPs of a region with each other, which we model as an extensive game. The model of an extensive game defines the possible orders of the events. The players can make decisions during the game and they can react to other players' decisions. We focus on extensive game model as a possible solution for defining charges for exchanging traffic between local ISPs belonging to the same region.

The paper is organized in the following way. In Section 2 we briefly discuss the Internet as it looks like today and pricing issues in the current Internet. In Section 3 basic components of game theory and classification of games are presented. In Section 4 extensive game model is described and we give example of an extensive game in which we examine the case of the economic interactions of local ISPs of a region with each other. Conclusion is given in the Section 5.

[1] Vesna Radonjić is with the Faculty of Transport and Traffic Engeneering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: v.radonjic@sf.bg.ac.yu

[2] Vladanka Aćimović Raspopović is with the Faculty of Transport and Traffic Engeneering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: v.acimovic@sf.bg.ac.yu

II. PRICING THE INTERNET SERVICES

A. Internet model

The Internet is a heterogeneous body of privately owned infrastructure. Roughly speaking, it consists of two types of networks: 1) densely meshed networks in geographically localized regions which specialize in providing consumers with connection points to the network and 2) networks traversing large geographical distances which provide connectivity between the local networks [2]. All the networks are connected by means of an inter-operable protocol stack agreed upon by the Internet Engineering Task Force (IETF). Fig. 1 illustrates the Internet as it looks like today.

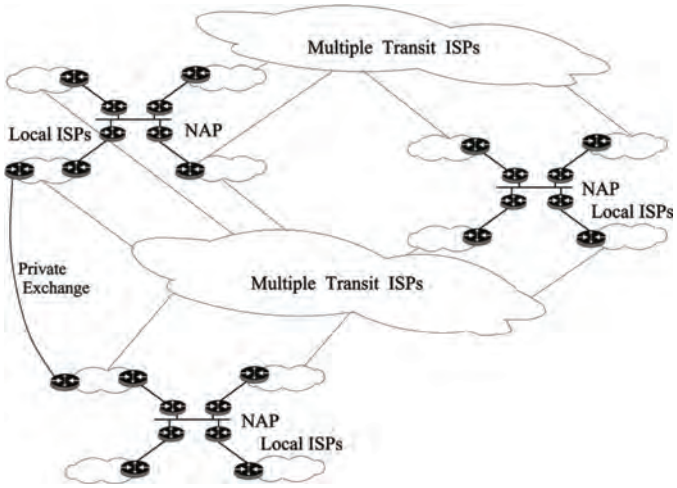


Fig. 1. Structure of the Internet consisting of local and transit ISPs[2]

There are local Internet Service Providers (ISPs) providing services in small regions, which compete for the same group of customers and transit ISPs which transfer data between such local groups. Local providers as well as transit providers establish a point of presence (shown as small "clouds" in Fig.1) at a Network Access Point (NAP) where traffic may be exchanged. Local ISPs in different regions have two options for traffic exchange—they can use the services provided by a transit ISP (shown as large transit „clouds") at a NAP, or can build their own means of private exchange (either at a NAP or independently). Recent maps of the Internet [3] indicate that private traffic exchanges are becoming popular.

B. Pricing issue

Pricing is one of the biggest challenges in in the current Internet. In general, pricing schemes have to be defined and

evaluated with respect to the heterogeneous technical, economic and social aspects. The main evaluation criteria encompass efficiency in the sence of maximizing utilities of customers and the provider, fairness and feasibility [4].

In today's Internet there is a need for shifting from simple charging schemes such as flat rate based or duration based charging towards the usage based charging [5] with different tariffs assigned to different service classes (DiffServ). However, providers usually require simple charging schemes which enable them to recover costs fairly and effectively allocate network resources.

Independently of pricing scheme the ISP uses, he needs to price and re-price its services from time to time. The need for re-pricing arises typically with changes in prices of competitors in the market or in its efforts at continual service differentiation through the periodic introduction of new services. The periodicities of price changes can be range from once or twice a day to several months.

Game theory is a good basis for the analysis of pricing problems in the current Internet.

III. Game Theory

A. Assumptions and Definitions

Game theory is a field of applied mathematics that describes and analyzes interactive decision making situations. It consists of a set of analytical tools that predict the outcome of complex interactions among rational players [6].

Basic components of a game are players, the possible actions of the players and consequences of the actions. The players are decision makers and their actions result in a consequence or outcome. The players 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. Rational players are assumed to maximize their payoff. If the game is not deterministic, the players maximize their expected payoff [6], [7]. 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 actions the players can take and what the consequences of the actions 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.

An outcome of a game is Pareto efficient, if there is no other outcome that would make all players better off. In implementation theory, the aim is typically to design a game that will end in a Pareto efficient outcome.

When a player makes a decision, he can use either a pure or a mixed strategy. If the actions of the player are deterministic, he is said to use a pure strategy. If probability distributions are defined to describe the actions of the player, a mixed strategy is used.

B. Classification of Games

Games can be classified into different categories according to their properties.

According to their focus, games can be divided into noncooperative and cooperative games. In noncooperative games, the actions of the single players are considered. In cooperative games the joint actions of groups are analyzed, i.e. what is the outcome if a group of players cooperate. In telecommunications, most game theoretic research has been conducted using noncooperative games, but there are also approaches using cooperative games.

According to their dynamics, games can be divided into strategic and extensive games. In strategic (or static) games, the players make their decisions simultaneously at the beginning of the game. While the game may last long and there can be probabilistic events, the players can not react to the events during the game. On the other hand, the model of an extensive game defines the possible orders of the events. The players can make decisions during the game and they can react to other players' decisions. Extensive games can be finite or infinite.

Games can be divided according to their payoff structures. A game is called zerosum game, if the sum of the utilities is constant in every outcome. Whatever is gained by one player, is lost by the other players. In telecommunications, the games are usually not zero-sum games.

Games can be divided on games with perfect and imperfect information. If the players are fully informed about each other's moves, the game has perfect information. Games with simultaneous moves have always imperfect information, thus only extensive games can have perfect information.

Games can also be divided on games with complete and incomplete information. In games with complete information the preferences of the players are common knowledge, i.e. all the players know all the utility functions. In a game of incomplete information, in contrast, at least one player is uncertain about another player's preferences.

In this paper we focus on extensive game model as a possible solution for defining charges for exchanging traffic between local ISPs belonging to the same region.

IV. Extensive Game

A. Basic model

The strategic game model is suitable for representing simple real life events such as auctions. A broader model is needed, when more complex interactions are occurring between the decision makers. Especially the possibility to react to the actions of the other players is essential in many applications. Extensive games eliminate the limitation of the simultaneous decisions, thus they make possible to model a wider range of real life situations.

For simplicity, the following formulation of extensive game does not allow simultaneous actions of the players, i.e. the game has perfect information.

Definition 1: An extensive game with perfect information has the following components.

- A set of players N,
- A set H of sequences (finite or infinite) of actions that satisfies the following three properties:
  - The empty sequence Ø is a member of H.
  - If $(a^k)_{k=1,\dots K} \in H$ (where K may be infinite) and $L < K$ then $(a^k)_{k=1,\dots L} \in H$.
  - If an infinite sequence $(a^k)_{k=1}^{\infty}$ satisfies $(a^k)_{k=1,\dots L} \in H$ for every positive integer L then $(a^k)_{k=1}^{\infty} \in H$.

  (Each member of H is a history; each component of a history is an action taken by a player.) A history $(a^k)_{k=1,\dots K} \in H$ is terminal if it is infinite or if there is no a^{K+1} such that $(a^k)_{k=1,\dots K+1} \in H$. The set of terminal histories is denoted Z.
- A function P that assigns to each nonterminal history (each member of $H \setminus Z$) a member of N. (P is the player function, $P(h)$ being the player who takes an action after the history h.)
- For each player $i \in N$ a utility function U_i on Z.

In strategic games, the behavior of the player is defined by the action the player takes. In order to define the player's behavior in an extensive game, more information is needed. A strategy describes the action of the player in every possible situation of the game.

Definition 2: A strategy of player $i \in N$ in an extensive game with perfect information. $\{H, N, P, (U_i)\}$ is a function that assigns an action in $A(h)$ to each nonterminal history $h \in H \setminus Z$ for which $P(h) = i$.

The solution of an extensive game is a Nash equilibrium from which no player has an incentive to deviate unilaterally.

Definition 3: A Nash equilibrium of an extensive game $\{H, N, P, (U_i)\}$ is a profile $a^* = (a_1^*, \dots a_N^*)$ of actions with the property that for every player $i \in N$ we have:

$$U_i(a^*) \geq U_i(a_1^*, \dots, a_{i-1}^*, a_i, a_{i+1}^*, \dots, a_N^*) \text{ for all } a_i \in A(h).$$

When a game is played, the rationality assumption will force the game into a Nash equilibrium outcome. If the outcome is not a Nash equilibrium, at least one player would gain a higher payoff by choosing another action. If there are multiple equilibriums, more information on the behavior of the players is needed to determine the outcome of the game. It is important to notice that while an equilibrium is a result of the optimization of the individual players, it does by no means imply that the result is "good" or globally optimum.

Extensive games with two players can be illustrated with matrices, as it is given in table 1. In this simple example, there are two cases:

1) Player 1 takes his first move choosing between 1 or 2. After observing player 1's decision, player 2 decides to choose 1 if the player 1's first move was 1 and 1.5 if the player 1's first move was 2.
2) Player 2 takes his first move choosing between 1 or 2. After that, player 1 decides to choose 0.5 if the player 2 played 1 and 1.5 if the player 2 played 2.

Table 1. Example extensive game with two players in matrix form

Strategies	Player 2	
Player 1	(1,1)	(2,1.5)
	(0.5,1)	(1.5,2)

The solution of the example game can be deducted easily. In the first case, player 1 prefers option that leads to higher utility, i.e. 2, which means that optimal strategy is (2,1.5). Respectively, in the second case player 2, as a rational player, chooses 2 (which is higher utility for him), hence the optimal strategy is (1.5,2). Clearly, the game has two Nash equilibrium points.

B. Example: Interaction among local ISPs

Next we consider the case of the economic interactions of local ISPs of a region with each other. Since they belong to a single group of the hierarchy, we observe this case as an interaction among equals. We study only intra-regional traffic and consider bilateral interactions of ISPs 1, 2 and 3 (Fig. 2.).

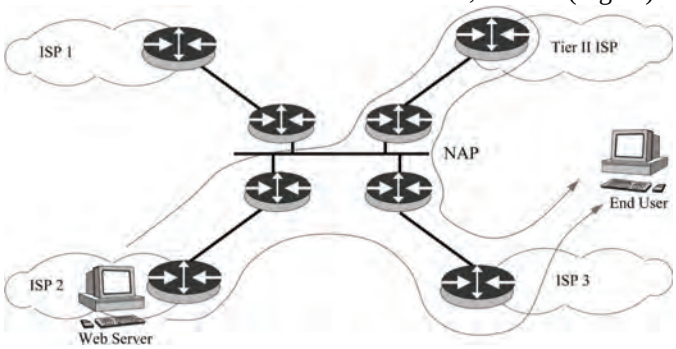


Figure 2. Local ISP structure

We suppose a small number of data carriers in a local region. This corresponds to the situation in the current Internet, where the number of data carriers is usually limited to a couple of ISPs providing a few QoS levels (DSL, cable/T1 and T3 seem to be most popular [2]). Local ISPs compete for consumers in a geographical region. They have to set prices for intra-regional and inter-regional[1] traffic. For intra-regional traffic, ISP i charges prices $\tilde{p}_i^r$ and p_i^r per unit traffic to websites and end-users, respectively. Each ISP must provide a guarantee of connectivity to all other users. This means that ISPs have to exchange traffic. They do this by means of a bilateral settlement, i.e., ISP i charges ISP j an access charge t_{ji}^r for termination of traffic, where ISPs i and j both belong to region r. The costs for all ISPs in the region assumed to be identical. All ISPs have identical fixed costs, which we take to be equal to zero (since it makes no difference to any of the results).

We also assume identical access charges for termination of traffic for all ISPs in the region (i.e. $t_{ji}^r = t_{ij}^r \ \forall i, j$ in region r) at the beginning of the game. We will find the optimum

[1] In this example, we will not examine charging for inter-regional traffic.

change of access charges, which would maximize the profit of individually rational ISPs.

Possible actions for any ISPi, $i = \overline{1,3}$ are:

- a_i^1 - not to change access charge to ISPj, $\forall j = \overline{1,3},\ j \neq i$
- a_i^2 - higher access charge to ISPj, $\forall j = \overline{1,3},\ j \neq i$ and
- a_i^3 - lower access charge to ISPj, $\forall j = \overline{1,3},\ j \neq i$.

No change of access charge gives ISP utility 0, higher access charge gives him utility 1 and lower access charge gives him utility -1. We assume that ISP1 takes his first move choosing between a_1^1, a_1^2 and a_1^3. After observing ISP1's decision, ISP2 and ISP3 decide to choose their actions, respectively. This example is ilustrated in table 2.

Table 2. Extensive game with three ISPs

a_1^1

	a_3^1	a_3^2	a_3^3
a_2^1	(0,0,0)	(0,0,1)	(0,0,-1)
a_2^2	(0,1,0)	(0,1,1)	-
a_2^3	(0,-1,0)	-	(0,-1,-1)

a_1^2

	a_3^1	a_3^2	a_3^3
a_2^1	(1,0,0)	(1,0,1)	-
a_2^2	-	(1,1,1)	-
a_2^3	-	-	-

a_1^3

	a_3^1	a_3^2	a_3^3
a_2^1	(-1,0,0)	-	(-1,0,-1)
a_2^2	-	-	-
a_2^3	(-1,-1,0)	-	(-1,-1,-1)

There are three possible cases for ISP2: 1) If ISP1 decides for action a_1^1, ISP2 choose between all three possible actions; 2) If ISP1 decides for a_1^2, ISP2 will eliminate third possibility, a_2^3 and 3) If ISP1 decides for a_1^3, ISP2 will eliminate second possibility a_2^2.

ISP3 plays by the following rules: 1) If ISP1 decides for a_1^1 and ISP2 decides for a_2^1, ISP3 will think over all three possible actions; 2) If ISP1 decides for a_1^1 and ISP2 decides for a_2^2, ISP3 will eliminate third possibility, a_3^3; 3) If ISP1 decides for a_1^1 and ISP2 choose a_2^3, ISP3 will eliminate second possibility a_3^2. 4) If ISP1 decides for a_1^2 and ISP2 choose a_2^1, ISP3 will eliminate third possibility, a_3^3; 5) If ISP1 decides for a_1^2 and ISP2 choose a_2^2, ISP3 will choose a_3^2 and 6) If ISP1 decides for action a_1^3, ISP3 will eliminate a_3^2 independently of what ISP2 choose.

There are three Nash equilibrium points in this game: (0,1,1), (1,1,1) and (-1,0,0).

V. Conclusion

This paper described the application of extensive game model as a mathematical basis for the analysis of various pricing problems in the current Internet.

Extensive game model is suitable for representing complex interactions, that occures between the decision makers. The model of an extensive game defines the possible orders of the events. The players can make decisions during the game and they can react to other players' decisions, which is essential in many applications. Extensive games eliminate the limitation of the simultaneous decisions, thus they make possible to model a wider range of real life situations. This model is a good basis for the analysis of pricing the Internet services.

We consider the case of the economic interactions of local ISPs of a region with each other. The problem of finding the optimum change of ISPs access charges for termination of traffic is modeled as an extensive game. We found three Nash equilibrium points as the solutions of the game.

Acknowledgments

This paper resulted from the researching project *Planning, Forecasting and Pricing in Telecommunication Networks* that is supported by Serbian Ministry of Science and Environmental Protection (project designation TP-6106A).

References

[1] A.Pras, B-J. van Beijnum, R. Sprenkels, R. Parhonoyi, "Internet Accounting", *IEEE Comm. Magazine*, vol. 39, no. 5, May 2001, pp. 108-113.

[2] S. Shakkottai, R. Srikant, "Economics of Network Pricing with Multiple ISPs", *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, December 2006, pp. 1233-1244.

[3] AS Internet Graph. 2003 [Online]. Available: http://www.caida.org

[4] V. Aćimović-Raspopović, M. Stojanović, "Accounting in QoS-enabled IP Networks", *Proceedings of POSTEL*, Belgrade, December 2005, pp.303-312.

[5] C. Courcoubetis et al., "A Study of Simple Usage-Based Charging Schemes for Broadband Networks", *Telecomm. Systems*, Vol. 15, No. 3-4, 2000, pp. 323-343.

[6] A. MacKenzie, L. DaSilva, *Game Theory for Wireless Engineers*, Morgan & Claypool, 2006.

[7] J. Leino, „Applications of Game Theory in Ad Hoc Networks", Master's thesis, Helsinki University of Technology, October, 2003.

A Hybrid Population Based Method Solving Convex Integer Optimization Problems

Vassil G. Guliashki[1]

***Abstract* – This paper presents a hybrid population based evolutionary programming method, designed to solve single objective convex integer optimization problems. The proposed method combines some ideas from scatter search and genetic algorithms, as well from the so-called ant systems and particle swarm optimization. The aim of the method is to explore in an efficient manner the whole feasible convex domain and to find out the global optimum of a multimodal objective function.**

***Keywords* – evolutionary programming, scatter search, genetic algorithms, convex integer optimization problems.**

I. INTRODUCTION

The integer programming problem considered in this study can be stated in the following form:

$$\textbf{Min } F(x) \quad (1)$$

$$\textbf{subject to: } g_i(x) \le 0; \quad i = 1,\ldots,m; \quad (2)$$

$$l_j \le x_j \le u_j; \quad j = 1,\ldots,n; \quad (3)$$

$$x \in \mathbf{Z}^n, \quad (4)$$

where x is an n-dimensional vector of variables x_j, $j = 1,\ldots,n$; which accept discrete values only. By l_j and u_j are denoted the bounds (lower and upper) of x_j, and $F(x)$ is the multimodal objective function. There is no necessary $F(x)$ to posses accessible to calculation derivatives in an explicit analytical form. The functions $g_i(x)$, $i = 1,\ldots,m$; are convex nonlinear functions and m is the number of nonlinear constraints **(2)**.

The convex integer problems (see [6, 16]) belong to the class of NP-hard optimization problems. There does not exist an exact algorithm, which can solve these problems in time, depending polynomially on the problem input data length or on the problem size. For this reason many efficient approximate evolutionary algorithms and metaheuristic methods have been created to find out the global optimum of such complex optimization problems. The most successful and efficient methods usually hybridize two or more metaheuristics.

In this study is proposed a new hybrid method, combining different ideas of such evolutionary techniques. It is organized to manage the system of constraints preserving the feasibility of the new obtained solutions. This is very important, as shown in [18]. The paper is organized as follows: Some common features of the population based methods are considered in Section II. The new hybrid population based method is described in Section III. An illustrative example of this new method is given in Section IV. Some conclusions are drawn in Section V.

[1]Vassil G. Guliashki is with the Institute of Information Technologies – BAS, "Acad. G. Bonchev" Str. Bl. 29A, 1113 Sofia, Bulgaria, E-mail: vggul@yahoo.com

II. COMMON FEATURES OF POPULATION BASED METHODS

The name „Adaptive Memory Programming" was proposed by Fred Glover in connection with the metaheuristic Tabu Search (TS) (see [10]). Many metaheuristics (i. e. methods designed to obtain a global optimum) can be classified as "adaptive memory methods" (see [19]) or "population based methods" (see [12]). The most familiar and powerful among them are Genetic Algorithms (GA) (see [11, 13]), Scatter Search (SS) (see [7, 9]), Tabu Search (TS) (see [8, 10]), Ant Systems (AS) (see [1, 2, 3]) and Particle Swarm Optimization (PSO) (see [5, 14, 15]).

The above mentioned metaheuristics possess the following common features (see [19]):

- They memorize solutions (or characteristics of solutions) in a population of individuals. Each individual is associated with a feasible solution of the problem at hand.
- They include a generating solutions search procedure, which uses the information stored in the memory.
- They apply some kind of local search method (a greedy improvement method, an elementary tabu search or simulated annealing) to improve the obtained solutions.

There are some shortcomings in GA. For example the mutation has unexpected results on the objective function value of an individual, i. e. it does not necessarily improve it (see [12]). Another crucial moment is the sufficient diversity of the population, i. e. the availability of diverse enough genetic material, which permits the exploration of the whole feasible domain and which would not restrict the search in a near part of the feasible domain around a local minimum (see [17]). Also GA are not designed for precisely locating the optimal solution(s) and the combination of genetic approach and of some form of local search is advisable in general, because it could improve the efficacy of GA (see [17, 19]).

The creation of successful global search methods is connected very often with the combination of two or more metaheuristics in hybrid methods. For example GA are combined with Tabu Search methods, or with a faster local search procedure, AS – with local search techniques (see [19]), GA – with clustering procedure (see[4]), SS – with TS or SS – with GA (see [9]).

All population based methods alternate during the search periods of self adaptation (the search process is intensified in some region of the search space) with periods of co-operation

(information collective gathered during the search process is used to direct further the search). The periods of self adaptation correspond to execution of mutation, improvement or local search procedure, and the periods of co-operation are connected with the selection, trace updating or generation of trial points. The major difficulty with the creation of global search methods is to overcome the premature convergence towards local optima. To obtain the global optimum of the problem at hand a diversification phases of the search process are necessary, so that new areas of feasible domain, that are remained still unexplored, can be investigated.

III. THE NEW SYSTEMATICALLY DIVERSIFIED SEARCH METHOD SDS

Considering the search process of a global optimum there is no sense to direct the search in the region of the best found so far (local optimal or near optimal) solution, because in the most cases it will not coincide with the global optimal solution. For this reason information sharing as in AS or in PSO will be necessary only after the whole feasible domain has been roughly explored. The exploration of the whole feasible domain means that there is a guaranteed systematically diversification of the search process. The basic idea of the proposed hybrid SDS (systematically diversified search) is to divide the feasible domain in sub-regions (cones having a common vertex) and to explore each of them applying systematically diversification of the search. Around the best found solution for each sub-region, a simple local search procedure is performed. Then the obtained best solutions from each sub-region form a final population, i. e. the information gathered during the search so far is shared like in AS or PSO. By means of this final population a few scatter search iterations are performed to obtain the global optimal solution. At the end a precise local search procedure is used in the neighborhood of the best obtained solution in order to locate precisely the global optimum.

Let the feasible domain be denoted by X. The method SDS starts from the Tchebicheff center of the feasible domain. The Tchebicheff center $x_{tch} \in X$ is the point located at the maximal Euclidean distance from the constraint surfaces. We assume that the Tchebicheff center is obtained by means of a method for solving convex problems with continuous variables. Then x_{tch} is rounded to the nearest integer point i_{tch}. The "pseudo-code" form of the SDS method is presented below:

Hybrid method SDS:

Round off the components of the Tchebicheff center x_{tch} rounded to their nearest integer values and use them to obtain the starting integer point i_{tch}.

Generate a regular simplex with n+1 vertices, using i_{tch} as one vertex. Generate the other vertices on the base of the elementary geometry in the following manner:

$$v^{(i)}_j = \begin{cases} i_{tch} + \varphi_1 & \text{if } j \neq i \\ i_{tch} + \varphi_2 & \text{if } j = i \end{cases} \quad (5)$$

$$\varphi_1 = \alpha.\left[\frac{\sqrt{(n+1)} + n - 1}{n\sqrt{2}} \right] \quad (6)$$

$$\varphi_2 = \alpha.\left[\frac{\sqrt{(n+1)} - 1}{n\sqrt{2}} \right] \quad (7)$$

Let i_{tch} be denoted as $v^{(0)}$. Round off each $v^{(j)}$, j = 1,…,n; to its nearest integer point. There are (n+1)combinations of n vertices, correspondingly for each facet of the simplex. The rays starting at the weight center of all simplex vertices cs and passing through the vertices belonging to each facet determine $K^{(i)}$ cones, i = 1,…, n+1; in the feasible domain.

Calculate the weight center of the simplex:

$$cs = \frac{\sum_{j=0}^{n} v^{(j)}}{n+1} \quad (8)$$

Round off each cs to its nearest integer point.

FOR i = 1,n+1; **DO**

Explore the cone $K^{(i)}$ as follows:

Create a population $P^{(i)}$ of (n+1) points, including the simplex vertices of the current simplex facet and their weight center cv, where

$$cv = \frac{\sum_{v^{(i)} \in K^{(i)}} v^{(i)}}{n} \quad (9)$$

Round off each cv to its nearest integer point.

Calculate the mean fitness value FV of the population $P^{(i)}$.

Let: $xold^{(j)} = v^{(j)}$, j = 1,…, n; and $xold^{(0)} = cv$.

Calculate $p^{(j)} = \gamma(xold^{(j)} - cs)$, j = 1,…, n; and $p^{(0)} = \gamma(cv - cs)$, where γ is a scale multiplier, tuned according to the concrete problem.

ITERATION

Calculate the points $xnew^{(j)} = xold^{(j)} + p^{(j)}$, j= 0,…, n; Round off each $xnew^{(j)}$ to its nearest integer point. In case there is a violated constraint from the system **(2)-(3)** reduce the corresponding $p^{(j)}$ as follows:

- If a constraint of type $\boldsymbol{x_k + a = 0}$ is violated, where $\boldsymbol{a}$ can have positive or negative value, then the corresponding component p_k of $p^{(j)}$, is used to reduce $p^{(j)}$:

$$p^{(j)} = \left| \frac{a - p_k}{p_k} \right| p^{(j)} \quad (10)$$

- If a constraint of type $\boldsymbol{g_i(x) \leq 0}$ is violated then $p^{(j)} = 0.8\ p^{(j)}$. If it is necessary repeat this reduction until the rounded off integer $xnew^{(j)}$ becomes feasible.

- If there are more than one constraints, violated by $p^{(j)}$, then chose the most reduced vector $p^{(j)}$, so that the rounded off integer $xnew^{(j)}$ becomes feasible.

Evaluate the objective function values $\boldsymbol{F(xnew^{(j)})}$, j = 0,…, n; In case $F(xnew^{(j)}) < FV$ then replace the point of P having worst (greatest) fitness value by the point $xnew^{(j)}$.

Calculate the mean fitness value FV of the updated population $P^{(i)}$.

Update $xold^{(j)} = xnew^{(j)}$, j = 1,…, n; In case some $p^{(j)}$ has been reduced at the current iteration then

$xold^{(0)} = \frac{\sum_{j=1}^{n} xold^{(j)}}{n}$; Round off $xold^{(0)}$ to its nearest

integer point;
else $xold^{(0)} = xnew^{(0)}$.
ENDofITERATION

Repeat the *ITERATION* until it is not possible to generate any new feasible points by means of $p^{(j)}$, $j = 0,\ldots, n$;

At the obtained end points along the cone rays try to make a step with $p^{(0)} = xold^{(0)} - cs$.

Around the best obtained point of *P* perform a simple local search by means of vectors *d*, having only one nonzero component: $d_j = \pm 1$.

ENDFOR

Create a final population *FP* of (*n*+1) points, including the best found points from the populations $P^{(i)}$, $i = 1,n+1$;. Calculate the mean fitness value *FFV* of the population *FP*.

For each two pints $x^{(1)}$ and $x^{(2)}$ of FP perform scatter search as follows:

$xnew = x^{(1)} + w(x^{(2)} - x^{(1)})$, where $w = \pm 1/\|x^{(2)} - x^{(1)}\|$.

In case the obtained points *xnew* have better objective function values than *FFV*, replace the worst points in *FP* by them.

Around the best obtained point of *FP* perform more precise local search by means of vectors *df*, having two nonzero components : $df_j = \pm 1$.

END

Performing consecutive steps from the weight center *cs* towards the boundaries of the feasible domain, no matter are they improving the objective function value or not, the search process in the SDS method is further diversified in different way, besides the separately exploration of each cone, generated in the feasible region.

IV. ILLUSTRATIVE EXAMPLE

Let us consider the following example. The objective function $\boldsymbol{F(x)}$ is given in a tabular form (see Table I), *x* is two-dimensional and there are simple constraints on each variable:

$$\boldsymbol{0 \le x_1 \le 10;}$$
$$\boldsymbol{0 \le x_2 \le 8;}$$

TABLE I. VALUES OF THE OBJECTIVE FUNCTION $F(x)$

x_2 \ x_1	0	1	2	3	4	5	6	7	8	9	10
0	11	11	10	11	11	10	10	9	7	6	8
1	7	8	10	11	12	13	12	10	7	3	7
2	6	4	7	10	15	14	13	11	9	7	9
3	7	8	9	10	13	12	14	13	10	9	10
4	8	9	12	11	13	12	17	16	11	12	13
5	16	14	12	11	12	11	10	14	12	13	14
6	15	14	12	12	12	10	8	9	8	11	15
7	13	12	11	10	11	9	8	4	6	10	13
8	10	12	11	11	10	9	7	0	7	9	11

Starting from the Tchebicheff center $i_{tch} = (5,5)$ the simplex with vertices (5,4), (6,7) and (8,5) is generated. The weight center of the simplex is $cs = (6,5)$.

The population $P^{(1)}$ includes the points (5,4), (8,5) and the middle point of the segment, determined by them – the point $cv = (7,5)$. The obtained mean fitness value $FV = 12.67$. For this problem $\gamma = 3$ is chosen, so that $p^{(0)} = (3,0)$, $p^{(1)} = (-3,-3)$, $p^{(2)} = (6,0)$; The last vector is reduced to $p^{(2)} = (3,0)$. Only two new points are obtained: (2,1) and (10,5). The point (2,1) enters in *P*, replacing the point (7,5). The new mean fitness value $FV = 11.33$. A reduction of $p^{(2)}$ is made, so that the point $xold^{(0)} = (6,3)$ with objective function value $F = 14$. The obtained point $xnew^{(0)} = (9,3)$ with $F = 9$. It enters in *P* replacing the point (5,4). The new mean fitness value $FV = 10.33$. Reducing $p^{(1)}$ to $p^{(1)} = (-1,-1)$ the point (1,0) is obtained. It has objective function value 11 and does not enter in *P*. The point $xold^{(0)} = (6,3)$. At the point (10,5) the step with $p^{(0)} = xold^{(0)} - cs = (6,3)-(6,5) = (0,-2)$ leads in point (10,3) with $F=10$. This point replaces the point (8,5) in *P* and $FV = 9.67$. A simple local search is performed around the point (9,3). The local optimal solution (9,1) is obtained with $F=3$.

The population $P^{(2)}$ includes the points (8,5), (6,7) and the middle point of the segment, determined by them – the point $cv = (7,6)$. The obtained mean fitness value $FV = 9.67$. The search vectors are: $p^{(0)} = (3,3)$, $p^{(1)} = (6,0)$, $p^{(2)} = (0,6)$; All the three vectors $p^{(0)}$, $p^{(1)}$ and $p^{(2)}$ are reduced and the obtained points are (9,8) with $F=9$, (10,5) with $F=14$ and (6,8) with $F=7$. Point (9,8) replaces the point (8,5) in P, and point (6,8) replaces the point (7,6). The new mean fitness value $FV = 8$. The calculated point $xold^{(0)} = (8,7)$ with $F=6$. It replaces the point (9,8). The new mean fitness value $FV = 7$. The vector $p^{(0)}$ is reduced to $p^{(0)} = (1,1)$ and the new point (9,8) is obtained, which has been already explored. A simple local search is performed around the point (8,7). The local optimal solution (7,8) is obtained with $F=0$.

The population $P^{(3)}$ includes the points (5,4), (6,7) and the middle point of the segment, determined by them – the point $cv = (6,6)$. The obtained mean fitness value $FV = 9.33$. The search vectors are: $p^{(0)} = (0,3)$, $p^{(1)} = (-3,-3)$, $p^{(2)} = (0,6)$; Both vectors $p^{(0)}$ and $p^{(2)}$ are reduced and the points (6,8), (2,1) and (6,8) are obtained. The point (6,8) has $F=7$ and the point (2,1) has $F=10$. Point (6,8) enters in *P*, replacing the point (5,4) and the new mean fitness value $FV = 7.67$. The calculated point

$xold^{(0)}$ = (4,5) with F=12. It does not enter in P. The vector $p^{(1)}$ is reduced to $p^{(1)}$ = (-1,-1) and the vector $p^{(2)}$ is reduced to $p^{(2)}$ = (0,0). The new points (1,0) with F=11 and (4,8) with F=10 are obtained. They do not enter in P. At the point (1,0) the step with $p^{(0)} = xold^{(0)} - cs$ = (4,8)-(6,5) = (-2,3) leads after reducing $p^{(0)}$ in point (0,2) with F=6. The last found point replaces point (6,7) in P and the new FV= 7. A simple local search is performed around the point (0,2). The local optimal solution (1,2) is obtained with F=4.

The points (9,1), (7,8) and (1,2) form the final population FP. The mean fitness value FFV = 2.33. After the scatter and the precise local search no better feasible solution is obtained. The found global optimal solution is x = (7,8) with F=0.

V. CONCLUSIONS

The presented new hybrid method SDS has the following good features and advantages:

- During the exploration of each sub-region the SDS method systematically diversifies the search process, avoiding in this manner the trap of local minima.
- The SDS method performs search in all defined sub-regions of the search space, so that the whole feasible domain is roughly explored.
- The formed final population contains diverse enough individuals, so that it is expected that the final search phase would lead to the global optimal solution.
- The applying of simple local search technique at the end of sub-regions exploration and of more precise local search at the end of the search process guarantees the good quality of the obtained final best solution.
- The SDS method can be efficient in comparison to other global search methods, because it explores only a small percent of all integer points in the feasible domain. This percent decreases with the increase of n and with growth of the feasible area.
- The populations used in SDS method have relatively small size, so that no great memory will be necessary for its implementation.
- The SDS method guarantees the feasibility of the obtained solutions.
- A great part of the integer points located on the rays forming each cone, which have been explored during the search in the corresponding sub-region, can be used during the exploration of the next sub-region. This may be used for creation of efficient program realizations of SDS method.

The SDS method will be tested on a set of test examples and may be further refined.

REFERENCES

[1] Colorni A., M. Dorigo and V. Maniezzo, "Positive Feedback as a Search Strategy", Department of Electronics, Politechnico di Milano, Italy, Working paper 91-16, 1991.

[2] Colorni A., M. Dorigo and V. Maniezzo, " Distributed Optimization by Ant Colonies", in Proceedings of the *First European Conference on Artificial Life (ECAL-91)*, F. J. Varela and P. Bourgine eds., The MIT Press: Cambridge, MA, 1991, pp. 134-142.

[3] Colorni A., M. Dorigo and V. Maniezzo, „An Investigation of Some Properties of an Ant Algorithm", in Parallel Problem Solving from Nature 2, R, Männer and B. Mandrieck eds., North-Holland: Amsterdam, 1992, pp. 509-520.

[4] Damavandi N. and Safavi-Naeini S., "A hybrid evolutionary programming method for circuit optimization", In Proc. *IEEE Transactions on circuits and systems – I: Regular papers*, vol. 52, No. 5, May 2005, pp. 902-910.

[5] Eberhart R. C. and J. Kennedy, "A new optimizer using particle swarm theory", Proceedings of the *Sixth International Symposium on Micromachine and Human Science*, Nagoya, Japan, pp. 39-43, 1995.

[6] Garey M. R. and Johnson D. S., "*Computers and Intractability: A Guide to the Theory of NP-Completeness*", W. H. Freeman, San Francisco 1979.

[7] Glover F., "Heuristics for integer programming using surrogate constraints", *Decision Sciences* 8, 1977, pp. 156-166.

[8] Glover F., "Future path for integer programming and links to artificial intelligence", *Computers & Operations Research* 13, 1986, pp. 533-549.

[9] Glover F., "Tabu Search for Nonlinear and Parametric Optimization (with Links to Genetic Algorithms)", *Discrete Applied Mathematics*, "Viewpoints on Optimization", 1991.

[10] Glover F., "Tabu search and adaptive memory programming – advances, applications and challenges", in *Advances in Metaheuristics, Optimization and Stochastic Modeling Technologies*, Barr, Helgason and Kennington eds.-, Kluwer: Boston, MA., 1997, pp. 1-75.

[11] Goldberg D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, Reading, Mass, 1989.

[12] Hertz A., Kobler D., „A Framework for Description of Population Based Methods", *Tutorials and Research Reviews*, 16-th European Conference on Operational Research Brussels, Belgium, 1998, pp. 48-59.

[13] Holland J. H., „*Adaptation in Natural and Artificial Systems*", The University of Michigan Press: Ann Arbor, MI, 1975, MIT Press, 1992.

[14] Kennedy J. and R. C. Eberhart, "Particle Swarm Optimization", Proceedings of *IEEE International Conference on Neural Networks*, Piscataway, N. J., 1995, pp. 1942-1948.

[15] Krusienski D. J. and Jenkins W. K. "Design and performance of adaptive systems, based on structured stochastic optimization strategies", *IEEE Circuits and Systems*, Vol.5, No.1, first quarter 2005, pp.8-20.

[16] NemhauserG. L. and Wolsey L. A. "*Integer and Combinatorial Optimization*", Wiley, New York. 1988.

[17] Pirlot M., "Heuristic Search Methods", Oper. Res. Designing Practical Solutions; Tutorial and Research Review Papers, Euro XIII/OR 36, The joint EURO/Oper. Res. Society Conference; University of Strathclyde, Glasgow, 19-22, VII. 1994. pp. 180-201.

[18] Reid D. J., „Enhanced Genetic Operators for the Resolution of Discrete Constrained Optimization Problems", *Computers & Operations Research*, Vol. 24, No 5, 1997, pp. 399-411.

[19] Taillard E. D., Gambardella L. M., Gendreau M. and J.-Y. Potvin, "Adaptive Memory Programming : A Unified View of Metaheuristics", *Tutorials and Research Reviews*, 16-th European Conference on Operational Research Brussels, Belgium, 1998, pp. 30-38.

Technical Aspects of Interconnect Management of Multimedia IP-based Networks

Aleksandar Tsenov[1] and Tashko Nikolov[2]

***Abstract* – In this paper some possible interconnection architectures regarding VoIP service are taken in consideration. The integration of the service functionality of Internet and PSTN/ISDN is a long process, which has to be carefully investigated in order to increase the operability of the service functions and to define appropriate architectures for the management of the QoS, the efficiency and the financial aspects of the network operation.**

***Keywords* – Interconnection, NGN, VoIP, Quality of services, Service integration**

I. Introduction

In this work some interconnection aspects are introduced. Especially the possible ways to interconnect between PSTN/ISDN and NGN are examine. On the base of versatile architectural decisions some scenarios for end-user connections are proposed and analysed. As a part of Multimedia services the VoIP service is taken in deeper consideration.

II. Interconnection Aspects

A. Quality of Services in NGN and NGI

This work deals with aspects of network interconnections between both NGN and NGI as well as between NGN and PTST/ISDN and between NGI and PSTN/ISDN. The delays in IP-based network elements of a connection must be added up over the continuous network elements in order to determine voice quality. It can be concluded that short delays over a connection link with high-quality and fast connections, particularly in NGN, make it interesting for a small ISP to tolerate increased delays in their own IP network. These incentives are particularly like to occur with regional ISPs using lower-quality und slower connections due to their low traffic concentration. The quality of the connection is constantly determined, however, by coding the voice signal with the lowest standard.

Therefore, voice services in IP-based networks can be given varying quality. Ensuring a quality comparable with PSTN/ISDN would require compliance with adjusted QoS parameter values like those for termination in PSTN/ISDN when terminating a voice connection from a PSTN/ISDN operator to an IP network operator with E.164 call number. Otherwise the voice quality deteriorates, either over the entire duration (if coding with reduced voice quality as G.711 is used) or occasionally, if time delays in the IP network jitter exceeding the permitted limit values. In these cases, the termination cannot be compared to the former PSTN/ISDN termination capacity from a QoS point of view. The question arises if and to what extent this affects former termination charges for mixed-type IP-PSTN/ISDN connections.

For a detailed analysis of the interconnection of several networks, here an existing interconnection system and its technical network foundations is described. To summarize, it can be said that the PSTN/ISDN and the former IP broadband network have a similar hierarchy, however, with a highly differing number of locations at the various hierarchical levels and totally different interfaces between the levels.

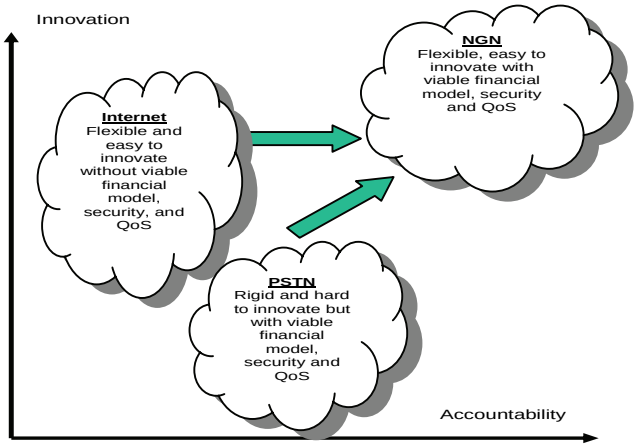


Fig. 1. Figure example [1]

On the Figure 1 one in the past year proposed interconnection architecture is shown [1].

The following can be drawn from this: The local traffic in PSTN/ISDN is transferred early to limit transit traffic in the exchanges of the core network. Contrary to this, IP routers in the core network are more heavily loaded with transit traffic since there is no local transfer in the access network (not least because of the dominant client-server connection structure of traditional IP services. With regard to traffic structures in PSTN/ISDN and for broadband services, it can be concluded that the traffic in PSTN/ISDN is mainly determined by dialog traffic between the terminal equipment from which a heavily meshed traffic structure is derived. A star-shaped traffic structure results for IP broadband traffic from the dominant

[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]Tashko Nikolov is with Telecom Department at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: tan@tu-sofia.bg

www traffic and other client/server applications, which is also the reason why traffic cannot be transferred locally, since the servers are typically installed at high-traffic locations in the core area of an IP network.

B. Interconnection Scenarios

Interconnection scenarios between national network operators and their NGN or PSTN/ISDN are below. A top-down implementation of the NGN (i.e. implementation originating in the core network) is assumed in this case with the expansion (in terms of the geographical extension) of the PSTN/ISDN being maintained in a first step. The following can be concluded:

- Connections between users from the PSTN/ISDN of various networks are routed as before via the appropriate external interconnection point.

Voice connections from a broadband user of the NGN to a PSTN/ISDN user of another network operator are routed as long as possible within the network of the "originating operator" with a view to as short and low-cost a termination as possible. The VoIP traffic originating in the NGN is routed by the network operator via their interconnection point (NGN to PSTN/ISDN) into their own PSTN/ISDN to then transfer the traffic as close to the B-subscriber as possible into the destination PSTN via the external PSTN interconnection point.

- If these are voice connections from a PSTN/ISDN user to another NGN broadband user of another operator, both PSTN networks must be interconnected again. The terminating network operator transfers the PSTN traffic to their NGN destination network to terminate it there.

It can also be drawn from the top-down approach for the implementation of the NGN that the voice traffic of major exchanges (trunk exchange, local exchange) is first integrated at the locations of the IP core network (NGN). PSTN/ISDN islands develop where traffic is routed among itself via internal PSTN-NGN interconnection points. It can therefore be described as an IP core network developing into an NGN. Since this development (due to the top down assumption) takes place at large locations first, the external interconnection point to another PSTN/ISDN installed there must be replaced by a Trunk Media Gateway (TMGW). The number of external interconnection points to other PSTN/ISDN, however, remains unchanged during this phase.

For medium-term development, it is assumed that other exchanges, their traffic and their PSTN users are integrated into the NGN concept. Assuming that an external interconnection to another PSTN/ISDN is always realized at an IP core network node by means of TMGW, the number of external interconnection points to another PSTN/ISDN is reduced converging to the locations of the IP core network nodes of an NGN in the long run. Based on even development of the PSTN/ISDN integration into the NGN with all major national PSTN/ISDN, this reduction is realized more or less simultaneously and a long-term migration takes place with external interconnection of PSTN/ISDN networks with each other being replaced by external interconnection between NGN networks. However, it cannot yet be identified whether this external interconnection for the termination of VoIP traffic is limited to the locations of IP network nodes or if interconnection at a lower level between the locations of the BAN is possible. Thus, the question concerning the number of external interconnection points between NGNs of various operators remains open from a technical and economic point of view.

In addition, it can be concluded that mainly regional or local PSTN/ISDN network operators are affected by the reduced number of interconnection points, in particular if the external interconnection points that they have been using so far together with their PSTN/ISDN networks, are discontinued by the national operators.

C. Existing Interconnect Regimes

Today exists a Interconnect Regime for PSTN/ISDN based on an Element Based Cost (EBC) model. It is proposed that the PSTN/ISDN is divided into two main parts – a backbone and access network. The backbone is further divided into two layers – Wide Transport Exchange Network (WTXN) and Regional Transport Exchange Network (RTXN). The RT Exchange contains the Interconnect Exchange points. On the access layer the Subscriber Exchanges (SE) are placed. The End-Users may be connected to the SE directly or via Concentrating Units (CU). The proposed architecture is shown on Figure 2.

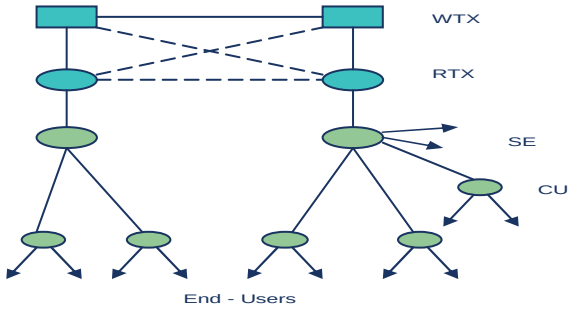


Fig. 2. Example of Interconnect Regime

The same layered architecture may be applied for

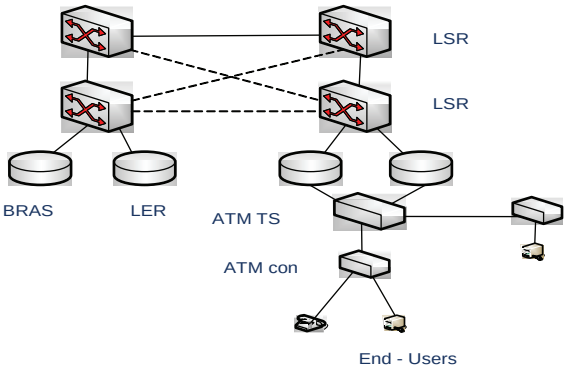


Fig. 3. Proposed Broadband IP – network architecture

structuring of Broadband networks and of Data networks as well. Nowadays such networks are based on the integration of

the ATM technology and the overlaid IP – Network. So arises the hierarchical structure of the Broadband IP – based network regarding to access point to stream-like services (Figure 3).

The following table proposes a short comparison between both architectures and their building elements. The correspondence between the elements allows the definitions of interconnect scenarios according end-to-end connections in such integrated structures.

Network	Layer	Equipment	
		PSTN/ISDN	IP BBN
Backbone	high	WTX	LSR
	low	RTX	LER/LSR
	interface	RTX	BRAS/ATM TS
Access	high	SE	ATM con
	low	CU	DSLAM

Table 1. Comparison between Network components

As shown in the table WTX corresponds to the Label Switch Router (LSR) in the MPLS – backbone network. Most interesting might be that the interface position differs in both networks. RTX is meanwhile part of the backbone and of the interface between two layers. In the IP Broadband Network (IP BBN) there are different elements for the corresponding functions – LER/LSR (Label Edge Router/LSR) and BRAS/ATM TS (Broadband Remote Access Router/ATM Traffic Selector). On the access level in the IP BBN the well known ATM Concentrators and DSLAM (DSL Access Multiplexer) are placed.

III. Interconnection Scenarios for Voice Services

Both architectures and the corresponding functionality on the same network layers are the base for the following scenarios propositions, which serve a purpose of the mentioned interconnection aspects in part II.

A. Interconnection between NGN and their precursor

NGN is actually the integration of PSTN and IP-Platform through one network operator. This integration process passes from top to bottom. That means – PSTN/ISDN networks have to operate for a relative long next period – together with the NGN. This is the reason why such interconnection has to be proposed and investigated (Figure 4).

Some important conclusions regarding the process mentioned above may be defined as follows:

- Connections between PSTN subscribers from different networks will remain unchanged – it will appear through the existing interconnection points on the PSTN level (the short dashed line).

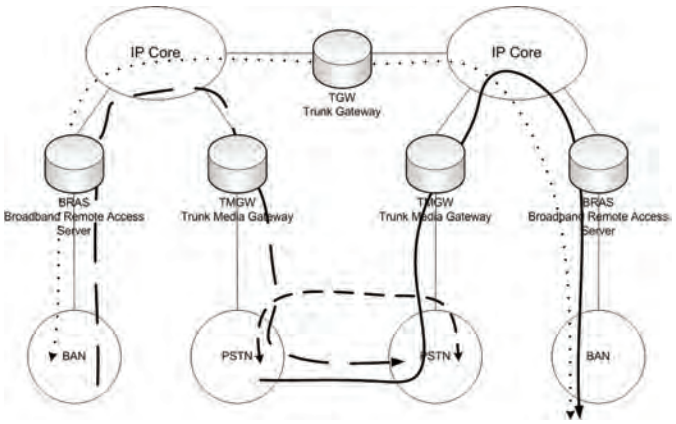


Fig. 4. Interconnection NGN – PSTN/ISDN

- Voice connections from a Broadband user (BBU) to a PSTN subscriber will be switched to the PSTN level of the same Network operator and then to the other PSTN network as described above (the long dashed line).
- Conversely – the connection from PSTN user to a BBU will be switched first on PSTN level and than to the BBU through BAN (Broadband Access Network) of the second Network operator (the thick line).
- Finally – a voice connection between BBU's goes to the higher level – the IP Core using well known VoIP protocols (the pointed line).

B. Interconnection between National Network Operators and Internet Service Providers

On the next figure the possible way for interconnection between IP – network operators on a national level and local ISP (Internet Service Providers) is shown. The assumptions is made that the national operator has already performed the integration to NGN and the ISP enhances the functionality of it own network for distribution of voice services. The same way as above different type of connections may be defined.

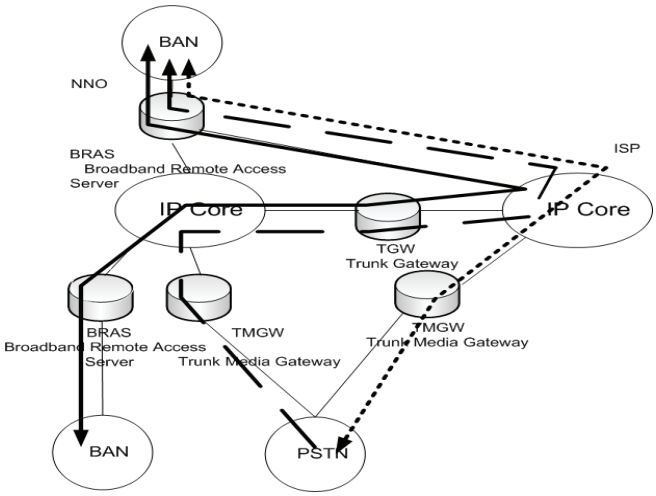


Fig. 5. Interconnection NNO - ISP

The considerations are made by the assumption that the ISP disposes of no or small broadband access infrastructure. So

the connections regarding traditional IP services go from the BBU of the ISP to BBU of the NNO through an external interconnection point – BRAS of NNO to LAS of ISP (the thick line).

Almost the same way are the voice connections from PSTN users to an ISP BBU. It goes either an internal TMGW (Trunk Media Gateway) to it's IP Core or through an external TMGW directly to the ISP's IP Core (the long dashed and the short dashed lines on the figure 4).

C. Interconnection between National Network Operator and Regional Network Operators

Finally the case of a local network operator or Regional Network Operator (RNO) will be considered. The assumption is made that the RNO of its own BAN disposes (figure 6).

The short dashed line shows a connection equal to this in the first case described.

Equal to the case mentioned are the connections between BBU's in both networks and the connection between a BB user in the network of the National network operator and a PSTN user from the network of the Regional network operator.

There is a difference regarding the opposite connection – a PSTN user in the NNO network builds a connection to BBU in the RNO network (the thick line on the figure).

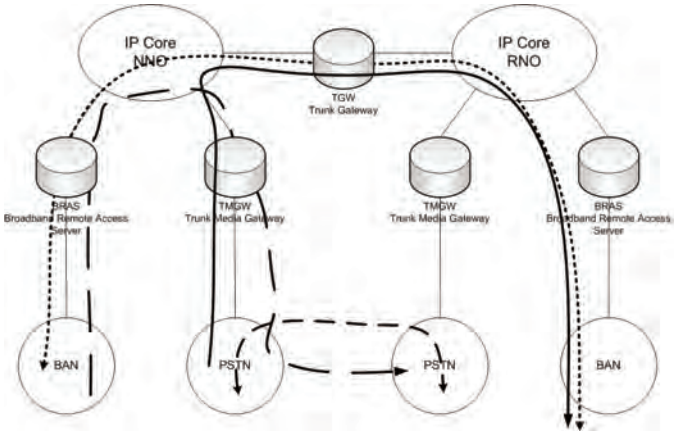


Fig. 6. Interconnection NNO - RNO

IV. CONCLUSIONS

The most important results of the analysis are summarized and initial conclusions for an interconnection system with and between IP-based networks drawn in the seventh and last chapter of the core report.

- Both national network operators and ISPs are striving towards an integrated voice and data offering with national operators generally establishing a centralized control platform - separate from the transport layer - into which the services from the PSDN/ISDN can be integrated. So-called soft switches are used for controlling the connections of the services. This new network concept is also called NGN. ISPs on the other hand improve their IP-infrastructures with peripheral intelligence in the form of SIP proxies and session border gateway controllers and media gateway controllers at network gateways and introduce additional protocols to ensure QoS parameter values; this method of network expansion is also called NGI. No major difference can be identified between the NGN and NGI concepts in terms of the transport layer.
- The central control layer of the NGN allows services offered by pure service providers to be integrated only in close technical and economic coordination with the NGN operator. On the other hand, the integration of separate services can be perceived more flexibly in the NGI.
- The three main QoS parameters (average delay, jitter of delay and packet loss rate) mean different things for the various services. Overdimensioning, prioritizing and capacity reservation have been introduced and examined as strategies for QoS realization with overdimensioning requiring the least additional QoS management effort and capacity reservation the highest.
- The service features of voice services (billing, connection availability, scaling, protocol stability, security and GoS/QoS) that are to be integrated into the NGN or NGI can be compared to those of the PSTN/ISDN (except for billing features). However, until these networks have been fully implemented and voice services having to be realized during this transitional period with hybrid infrastructures, restrictions in terms of scaling, protocol stability, security and GoS/QoS must be expected, particularly in the NGI concept.

ACKNOWLEDGEMENT

This work is made in connection to the **Project BY-TH 105/2005**.

REFERENCES

[1] Michel Bruyère, "Interconnection Issues Related To Next Generation Networks", Goulet TELECOM, Canto – 2006, Punta Cana, 18-21 July, 2006.

[2] NGN Management Focus Group, "NGN Management Specification Roadmap v3.1", November 2005.

[3] Donnelly E, "Evaluation of the state of the art in telecommunications management", SmartMan Consortium, 2000.

[4] ND1614:2006, "Management of general Connectivity of PSTN/ISDN Service Interconnect for UK NGN's" , Issue 1, NICC Document, December 2006.

[5] NGN Management Specifications, Annex 5 to the NGN Specification Roadmap v3.1, ITU, December 2005.

[6] Jia W, Bo Han, Ji Shen, Haohuan Fu, "Next Generation Networks Architecture and Layered End-to-End QoS Control", ISPA 2005, LNCS 3758, pp. 1055-1064, 2005.

[7] Markus J. S., "Framework for Interconnection of IP – based Networks – Accounting Systems and Interconnect Regimes in the USA and the UK", Study of the Federal Network Agency, Bad Honnef, March 2006

Computer System for Deductive Issue from Natural Language Text through Interpretation in Semantic Resolution

Iliyana Raeva, Rumen Raev and Dragan Stankovski

Abstract **- A system for Deductive Issue from Natural Language text written in English language is presented in the paper. The system solves some linguistics tasks in the logic area, through H- interpretation under Universal set U. An appropriate semantic type is chosen for the H-interpretation. As results from series known facts, the program gives an answer of the question if a given statement is false or true.**

Strings of statements are presented as a deductive system of knowledge used as initial data. The rules of predicate logic and principle resolution are applied.

The objective is to find an empty disjunct from the system. The programming language for the system is Delphi.

Key words: propositional calculus, logics, disjunctive system, semantic resolution, logical issue.

I. INTRODUCTION

The system presented in this work is a deductive model of decision making system. The problem to be solved is written as a formal deductive system. We confirm or refute the problem by using some axioms and the rules for issues from formal logics.

Decision making system models are mechanisms for management of complex objects. It is impossible to describe the models using only mathematical instruments. The objects themselves in that system evolve in the time, change their structures and functions. This leads to evolution of the management process itself. The goals of objects' management cannot be always expressed as a quantitative correlation. Logics – linguistics management models are applied for building such systems. A typical characteristic for retrieval of the used knowledge in the objective area is the availability of some logical tools. The knowledge is presented in a linguistic form.

The ideas developed in this work could be applied to management systems. An example is the operative-dispatcher management of power supply board system, sea and river ports, etc. One of the directions for increasing the dispatcher management's affectivity is creating a system for transferring information, incoming from the dispatcher to a computer.

Authors are with Department in Mathematics, University of Rousse; E-mail:iraeva@yahoo.com

II. ERBRAN'S PROCEDURE OF ISSUE. SEMANTIC RESOLUTION

The analysis and algorithms of this system are based upon the Erbran's procedure of issue based on the following theorem:

Theorem 1 A given set of disjuncts S is executive if and only if it takes value false in all interpretations in all areas of interpretation.

Due to the impossibility to examine all interpretations in all areas it is necessary to find a special area of interpretation. If S is impracticable in this area, then it will be impracticable in the other areas of interpretations [1]. Such area exists and it is called Universum of Erbran. The definition is the following:

Theorem 2 Let H_0 is a set of constants, appearing in the set of disujncts S.

- If there is not a constant in S, then an arbitrary constant "a" is included in H_0. т. е. H_0 = {a};

- For i=0, 1, 2 … H_{i+1} = $H_i \cup \{ f^n(t_1,t_2,...,t_n) \}$, where f^n is a n – positioned functional symbols, which occur in S a $t_1,t_{2,...,}t_n$ are elements of sets H_i. Then H= H_∞ named Universum na Erbran, for S , a H_i - its level i.

An effective modification of the principle of resolution is the semantic resolution. The cancels of disujncts which are tautologies and gluing of disjuncts going through the process of issue completely only slightly decrease the number of unduly disjuncts. Some cardinal ways are necessary to generate only significant disjuncts. The interpretation of predicate letters appears as a tool for generation of disjuncts.

III. SYSTEM DESCRIPTION

Entrance data

The program accepts at the entrance a file containing a logical task. A dialogue window appears on the screen through which the appropriate file must be chosen. The base format of the files for working the program is Special text *.stf, but it is possible to import text files *txt as well. In this away we can introduce a logical task using external text editor.

By using of a built text editor in application keyboard symbols are accepted to be introduced according to the Latin letters and punctuation signs. Mouse movements on the

windows of the applications and keyboards combinations of symbols corresponding to commands implementing in the program induced the program at the entrance as well.

Exiting data

On the screen the exit of the program displays a path of the decision and the result. The condition of the solving problem, the path of the decision and its result are printing on the text document.

As an option writing in a file is suggested of the edited logical task. The basic format used for keeping information is Special text *.stf, but it is given a possibility to export in a text file *txt as well. It allows the editing of the logical task text to be done by an exterior for the program text editor(Figure 1).

Description of the data flows

- Execute user's command - the process holds the choice of the user from the basic menu of the program and starts the other process carrying out the request.
- ***File manager*** - the process forms the text of the introduced by the user task as a text format (txt), keeps it in a text file or reads from a text file written in advanced text and supplies it in the text field of the program.
- ***Parse text*** - the process makes the text parse of the introduced by the user task. It defines the different logical facts and presents the text as formulas.
- ***Teach the program*** - the process is starting when it meets an unknown to the program word and requires from the user to define its type as a part of the speech.
- ***Resolution*** - the process transforms the described by logical formulas user's task in a disjunctive normal form. Then it applies the rule of the resolution to deny the goal (resolution refutation) and gives the result to the process show results.
- ***Show results*** - the process visualizes on the screen the path of the decision and its result. The text of the task, the decision path and the result can be printed by request of the user.
- ***Dialog*** - the process realizes the dialog with the user by giving a message on the screen and showing the possible answers.

We make the entering of the text of the logical task by text field in the frame of the main windows. The entered restrictions and spelling rules are observed. The goal we have to prove is entered in the last sentence.

IV. RESTRICTIONS

The text of the logical task, that the system has to solve, is entered only in English language. The relevant syntactic and logical rules are respected.

The using of the following type of sentences is allowed:

- simple sentences not containing conjunctions

Example: P → John works.

- Compound sentences including only one coordinative conjunction

Example: P&Q → John works and John smiles.

PVQ → John works or John smiles.

- Complex sentences with a conditional subordinate in any of the two parts of which there is either a simple or a compound coordinative sentence

Example: P&Q → R -If John works and John smiles then John is a boy.

P ↔ Q V R- John is a boy if and only if John smiles or John works.

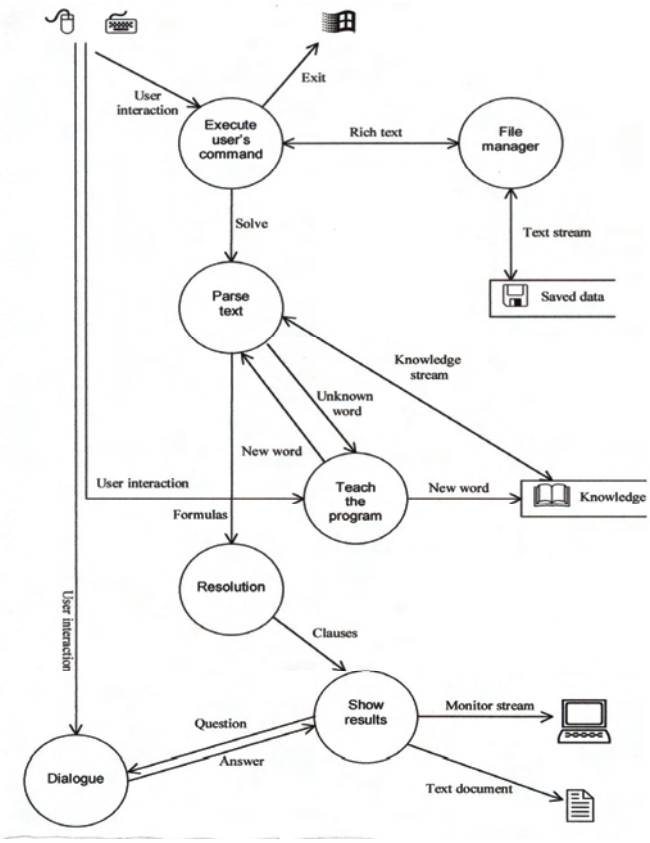


Figure 1 Diagram of the data flows and the processes in the program

Data streams and processes in the program

If a negative sentence is introduced the negation has to be written separately from the secondary or the modal verb.
Example: John does not work. John can not work.
The use of personal or possessive pronouns is not recommended. If the program text includes them
but yet, the definition of the action subject is needed.

Text example

Text of the logical task
John awakens. John brings a mop.
Mother is delighted if John awakens and cleans his room.
Ij John brings a mop then he clean his room.
Goal: Mother is delighted.

Meaning facts:
A - John awake
B - John brings a mop
C – John cleans John's room

D - Mother is delight

Path of the decision

(1)	pA	
(2)	pB	
(3)	nA*nC*pD	
(4)	nB*pC	
(5)	nD	
(6)	nA*nC	(5), (3)
(7)	nA*nB	(6), (4)
(8)	nC	(6), (1)
(9)	nA	(7), (2)
(10)	nB	(7), (1)
(11)	nil	(9), (1)

Decision results:

It is true that: “Mother is delighted.”

V. Conclusions and results

A system for deductive issue in a semantic resolution is realized from a text, written in natural language. The system could be used as a decision making system or as an expert system.

Explicit type logical tasks in logic and mathematical linguistic area can be solved effectively by using the system. In this sense the system can be used successfully as a learning system .

References

[1] Vagin V. N., Deduction and generalization in the Desition Making System, Moskow, 1988;

[2] Jim Woodcock, Martin Lomes, “ Software Engineering Mathematics GB 1991;

[3] Grishman Ralph, Computational Linguistics, Cambridge University Press, 1986;

[4] Michael Rosner, Roderick Johnson, Computational Linguistics and Formal Semantics, Cambridge University Press, 1992.

This page intentionally left blank.

Touch Interface Framework for GIS based C4I Systems

Aleksandar Milosavljević[1], Dejan Rančić[2], Aleksandar Dimitrijević[3] and Vladan Mihajlović[4]

***Abstract* – This paper presents *GinisUI* framework for rapid development of advanced graphical user interfaces with features such as control transparency, automatic arrangement and zooming. *GinisUI* is specially designed and implemented for appliance in GIS based C4I systems with touchscreen use. The framework relies on XML for specifying interface components and their layout. Because C4I systems combine GIS functionality with Virtual Reality (VR) their applications demand both 2- and 3-dimensional visualization. To enable seamless visual transition between these two modes, UI component visualization in *GinisUI* is implemented using both GDI and OpenGL drawing.**

***Keywords* – Touch Interface, Framework, GIS, C4I, XML.**

I. INTRODUCTION

A geographic information system (GIS) is special type of computer-based information system tailored to store, process, and manipulate geospatial data [1]. The ability of GIS to handle and process both location and attribute data distinguishes GIS from other information systems. It also establishes GIS as a technology important for a wide variety of applications [2]. For many years GIS has been considered to be difficult, expensive, and proprietary. The advent of graphical user interface (GUI), powerful and affordable hardware and software, and public digital data has broadened the range of GIS applications and brought GIS to mainstream use in the 1990s [2]. Despite the success of various GIS applications, their current use and usability is still constrained by overly complex user interfaces [3].

The acronym C4I is used to represent the following group of related military functions that enable the coordination of operations: Command, Control, Communication, Computers, and Intelligence [4]. *Command* and *Control* refers to the ability of the military commander to direct his forces. The addition of *Communications* to the grouping reflects the fact that communications is required to enable this coordination. In modern warfare, *computers* are also a key component, while *Intelligence* is the knowledge relevant to the coordination of forces.

One important capability that C4I systems provide commanders is situational awareness, i.e. information about the location and status of enemy and friendly forces [5]. Therefore, building a C4I system must rely on some GIS functionality. Certainly, the most important is mapping, positioning, map navigation, and overlay analysis using different layers of information to create a map.

Users of C4I systems are commanders at different levels of army command chain. Some lower levels of the command require embedment of C4I systems into combat vehicles. Such conditions require use of touchscreen for human interaction with the system.

To enable rapid development of an effective touchscreen graphical user interface for GIS based C4I systems we relied on a design and implementation of a GUI user interface framework.

Framework is a generic term for a powerful object-oriented reuse technique that typically emphasizes the reuse of design patterns and architectures [6]. There are two common definitions of an application framework [6]. The first defines an application framework as a reusable design of the entire, or part of a system represented by a set of abstract classes and the way their instances interact, while the second states that an application framework is the skeleton of an application that can be customized by an application developer. These definitions are complementary, not conflicting, since the former describes framework from the design perspective, whereas the latter describes it from the functional viewpoint.

The paper is organized as follows: Section 2 gives brief overview of user interface paradigms. Section 3 discusses design issues and the architecture of *GinisUI* framework. Section 4 presents GUIDL – XML language for describing *GinisUI* user interfaces. Section 5 considers presentation of a tank C4I system touch interface, and Section 6 summarizes the results achieved.

II. USER INTERFACES

A *user interface* is the aggregate of means by which people (the users) interact with a particular machine, device, computer program or other complex tool (the system) [7]. A *graphical user interface* (GUI) is a method of interacting with a computer through a metaphor of direct manipulation of graphical images and widgets in addition to text. *Touch interfaces* are graphical user interfaces using a touchscreen display as a combined input and output device.

[1]Aleksandar Milosavljević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: alexm@elfak.ni.ac.yu

[2]Dejan Rančić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: ranca@elfak.ni.ac.yu

[3]Aleksandar Dimitrijević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: aks@elfak.ni.ac.yu

[3]Vladan Mihajlović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: wlada@elfak.ni.ac.yu

Most modern general-purpose GUIs consists of graphical widgets such as windows, menus, radio buttons, check boxes, and icons, and employs a pointing device (such as a mouse, trackball, or touchscreen) in addition to a keyboard. Those aspects of GUIs can be emphasized by using the acronym WIMP (Windows, Icons, Menus, and Pointing device). GUIs that do not follow WIMP paradigm are most notably found in computer games. Another research direction in advancement of GUI is the Zooming User Interfaces (ZUI) [8].

The raising popularity of touch interfaces can be described by a fact that a touchscreen provides the simplest, most direct user interface with a computer [9]. Properly programmed, a touchscreen interface can be intuitive, requiring no learning curve for first time users. Touchscreens are also rugged enough to stand up to harsh environments where keyboards and mice can become damaged. Touchscreens ensure that no space is wasted because the input device is completely integrated into the monitor.

III. GINISUI FRAMEWORK

GinisUI is an application specific GUI framework specially designed and implemented for appliance in GIS based C4I systems with touchscreen use. General concepts underlying the design and the implementation of this application framework are already used in our GIS application framework *Ginis* [10].

GinisUI framework aimed domain is a C4I system embedded within combat vehicle and integrated with existing command, control, and communication subsystems. Harsh environment, and usually limited vehicle inner space require use of relatively small touchscreen interfacing devices.

Because these systems are GIS based, their basic purpose is to display layered map of surrounding area with a variety of geographically located information. So the problem was to develop a touch interface that fits on a relatively small screen already overcrowded with other information. Furthermore, pointing with a finger to a screen in the environment of a shaking combat vehicle requires buttons big enough to avoid mistakes.

Our solution to the problem was to develop a user interface that will enable:

- Transparency,
- Zooming and
- Automatic arrangement of controls.

Transparency is introduced to enable integration of user interface into a full-screen geographic map. This feature is combined with a zooming feature that represents enlargement of buttons when they get input (pointer) focus.

To further preserve an information view area, from the user interface, we included control showing/hiding and automatic arrangement. This feature enables design of a user interface that can automatically reconfigure to a current mode of an application.

GinisUI framework is designed and implemented to enable rapid development of user interfaces that fulfil identified requirements. To enable rapid development, the framework relies on XML for specifying interface components and their layout.

Logical architecture of *GinisUI* framework is based on relatively simple and known concepts (Fig. 1). Class `GUI` is the main class that acts as an interface between an application and the user interface (UI). It is derived from `GWindow` class, and it holds top-level UI controls (instances of `GControl` subclasses `GToolbar`, `GButton`, `GText`, and `GPicture`).

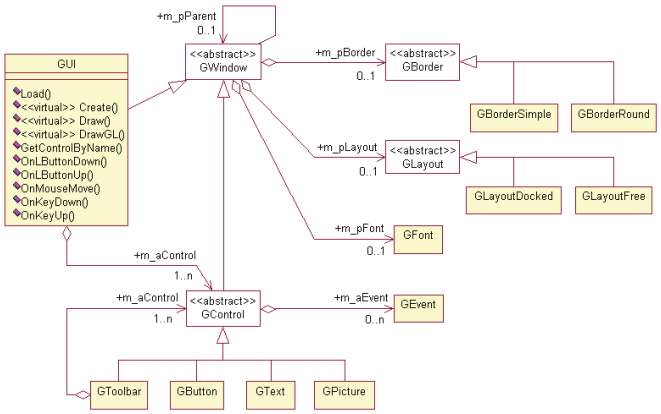


Fig. 1. Logical model of *GinisUI* framework

Characteristics of a UI control window are defined through `GWindow` class. General attributes of a window are background colour or background image and window transparency. Optional attributes are border, layout and font.

An implementation of different window border types is done through `GBorder` abstract class and their subclasses. Class `GBorderSimple` implements rectangular border type, while class `GBorderRound` is obliged for rounded window borders. Further extension of the framework with other border types can be done by implementation of appropriate `GBorder` subclasses.

Abstract class `GLayout` acts as a template for implementation of different window layouts. *GinisUI* framework currently enables docked (class `GLayoutDocked`) and free layouts (class `GLayoutFree`) of UI controls in corresponding parent windows. Controls can be docked on the top, bottom, left or right of the parent window, or they can be freely placed on some absolute position.

A style, size and colour of a text, that is to be displayed within `GText` control, are defined by `GFont` object attached to a control or some of its parent windows. This approach ensures range of possibilities from setting different font each control, to using single font for all controls in the UI.

As we already stated, `GWindow` part of a UI control object deals with display and arrangement of controls. Behaviour of a control (e.g. button) is implemented by event catching mechanism realized in `GControl` part of the object. Different event types (e.g. mouse button down, mouse move, keyboard button down, etc.) are defined by `GEvent` objects that can be attached to a UI control. When event happen, an application should call corresponding `GUI` interface methods (*`OnLButtonDown`*, *`OnMouseMove`*, etc.) that returns a pointer to a control that caught the event. An application that uses this framework, knowing the control name, should then handle that event.

Because C4I systems combine GIS functionality with Virtual Reality (VR) their applications demand both 2- and 3-dimensional visualization. To enable seamless visual transition between these two modes, UI component visualization in *GinisUI* is implemented using both GDI (`GUI` class `Draw` method) and OpenGL drawing (`GUI` class `DrawGL` method).

Described framework defines only basic building blocks needed for a user interface. Building a concrete user interface for some application additionally requires explicit definition what and how these blocks create the interface. For encoding of this definition we use XML. In order to constrain content of these XML documents, we specified appropriate XML language named *Ginis User Interface Definition Language* (GUIDL).

IV. XML LANGUAGE GUIDL

GUIDL represents XML language specified using *XML Schema Definition Language*. A valid GUIDL XML document represents definition of a user interfaces that can be realized using *GinisUI* framework.

The structure of GUIDL and *GinisUI* logical model are closely related. All classes shown in Fig. 1 have their corresponding XML elements in the language specification. To enable setup of instantiated object properties from XML definition, these classes implement virtual method `Create`.

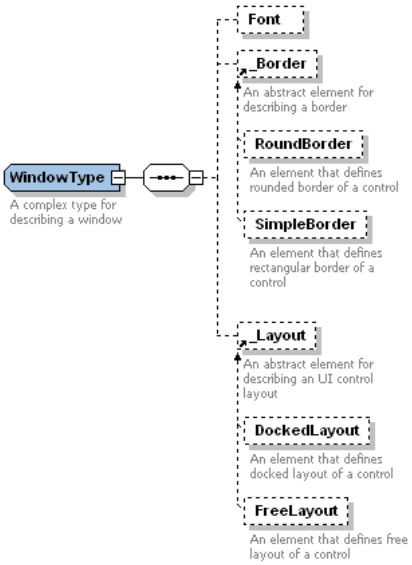


Fig. 2. Structure of an XML definition of a window

Fig. 2 shows structure of a complex XML element type `WindowType` that corresponds to `GWindow` class. Contained elements `Font`, `_Border` and `_Layout` are used to optionally setup font, border and layout properties of a control window. Elements `_Border` and `_Layout` are abstract and serves as placeholder for appropriate substitutes (e.g. `RoundBorder` and `DockedLayout`).

Complex type `ControlType`, shown in Fig. 3, is defined as an extension of `WindowType` and corresponds to `GControl` class. `ControlType` additionally defines events that control shall catch and different control titles for multilingual support.

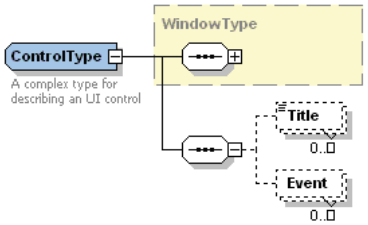


Fig. 3. Structure of an XML definition of a control

Finally, knowing structures of window and control element types, we can discuss structure of a root element `UIDef` (Fig. 4) that corresponds to `GUI` class. The root element inherits structure of `WindowType` that is used for specification of an UI background window. Additionally, it can contain several control elements defined through substitution group `_Control`.

Elements `Toolbar`, `Button`, `Text` and `Picture` (Fig. 4) are used to specify properties and structure of actual user interface controls. All these elements extend previously described `ControlType` (Fig. 3) with additional properties defined in corresponding classes `GToolbar`, `GButton`, `GText` and `GPicture` (Fig. 1). For example, `Toolbar` element can contain several control elements; `Picture` element has attribute defining filename of an image that will be shown and `Button` element can optionally contain images for normal and focused display.

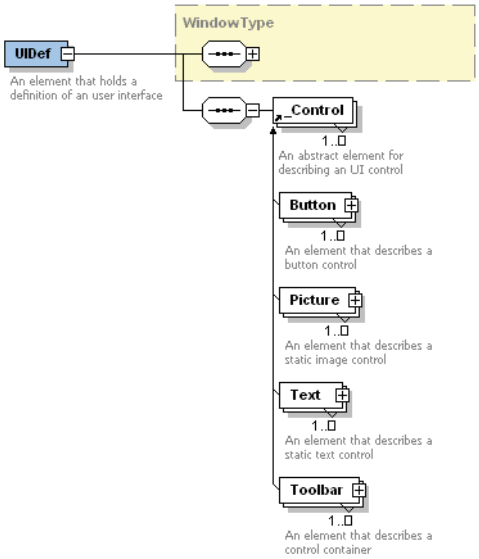


Fig. 4. Structure of an XML definition of a user interface

V. TOUCH INTERFACE FOR A TANK C4I SYSTEM

To illustrate features implemented in *GinisUI* framework, we present touch interface of a tank battlefield management system (*BMS*). *BMS* is in essence geographical information system extended with a command, control, communication, and virtual reality functions. *BMS* is equipped with a touchscreen for interfacing with a user.

BMS application can run in two modes: *combat* and *planning*; with two different views: *2D* and *3D*. Each mode with each view has different functions that user should be able to access through an interface.

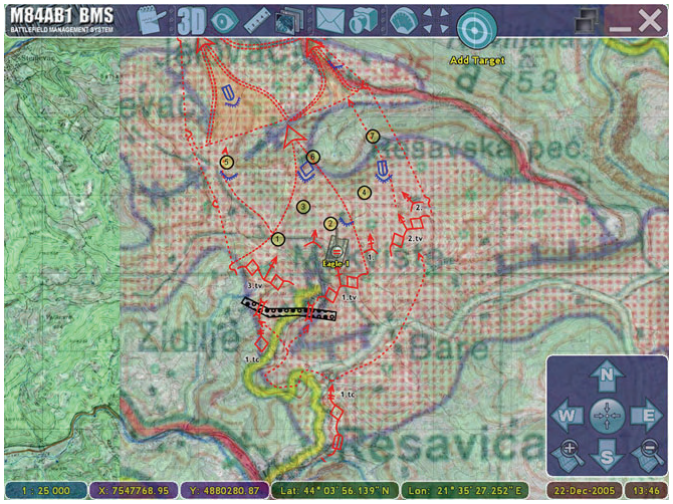


Fig. 5. BMS user interface in the 2D View

Fig. 6. BMS user interface in the 3D View

The user interface of *BMS* can be divided in always visible and visible on demand parts. Always visible parts are the main toolbar on top of the screen, status line at the bottom and navigation panel placed in the bottom-right corner. These interface elements are noticeable in Fig. 5 and 6. Depending on a mode or a view different controls can appear within these elements. The other, visible on demand part of the *BMS* user interface consists of various panels (Fig. 6), dropdown toolbars, message boxes, etc.

Fig. 6 shows *BMS* in the 3D view. *BMS* 3D view represents VR extension of GIS. It is used to display virtual terrain with some additional 3D objects (combat vehicles, trees, houses, buildings, etc.). This figure also shows 3D view specific compass panel and general system status panel.

User interface transition between 2D view (Fig. 5) which is drawn using GDI functions, and 3D view (Fig. 6) visualized using OpenGL, is seamless.

VI. Conclusion

We have discussed problems and identified requirements for a user interface of a GIS based C4I systems for combat vehicle embedment. Problems that we have faced were:

- Limited vehicle inner space – use of small touchscreen interfacing devices.
- Harsh environment – shaking of combat vehicles during use.
- Large amount of information to display – a geographic map or a virtual 3D terrain.

The solution that helps to reduce these problems relies on three key concepts:

- Transparency – enables integration of user interface with a background map or a virtual 3D terrain.
- Zooming – enables enlargement and additional information display of focused buttons.
- Automatic arrangement – combined with show/hide mechanism enables reconfiguration of a user interface to a current mode of an application.

Further, in the paper, we have introduced *GinisUI*, an application framework designed and implemented to enable rapid development of user interfaces based on these three concepts.

Described framework defines only basic building blocks needed for a user interface, while building a concrete user interface additionally requires explicit XML description what, and how these blocks create an interface. In addition to, and based on the framework, we specified XML language GUIDL to cope with these UI descriptions.

Finally, as an evaluation of our solution, we presented a case study concerning successful user interface realization of a C4I Battlefield Management System.

References

[1] Worboys, M., and Duckham, M., *GIS: A Computing Perspective, Second Edition*, CRC Press, Boca Raton, FL, 2004.
[2] Chang, K., *Introduction to Geographic Information Systems, Third Edition*, McGraw-Hill, New York, NY, 2005.
[3] Rauschert, I., Sharma, R., Fuhrmann, S., Brewer, I., and MacEachren, A., "Approaching a New Multimodal GIS-Interface", 2002, http://www.geovista.psu.edu.
[4] C4ISTAR, Wikipedia, http://en.wikipedia.org/wiki/C4I.
[5] What is C4I?, C4I.org, http://www.c4i.org/whatisc4i.html.
[6] Fayad, M., Johnson, R., and Schmidt, D., *Building Application Frameworks: Object-Oriented Foundation of Framework Design*, John Wiley & Sons, 1999.
[7] User Interface, Wikipedia, http://en.wikipedia.org.
[8] Bederson, B., and Meyer, J., "Implementing a Zooming User Interface: Experience Building Pad++", Software: Practice and Experience, Vol. 28, No. 10, pp. 1101-1135, 1998.
[9] Small, C. H., "Touchscreens Provide a Robust and Intuitive User Interface", TechOnLine, http://www.techonline.com.
[10] Milosavljević, A., Đorđević-Kajan, S., and Stoimenov, L., An Architecture for Open and Scalable WebGIS, 8th AGILE Conference on GIScience, pp. 629-634, Estoril, Portugal, 2005.
[11] Milosavljević, A., and Stoimenov, L., "Implementation Model of Ginis framework for Web-based GIS", ETRAN 2005, Conference Proceedings, pp. 35-38, Budva, Montenegro, 2005. (in Serbian)

Full-text Search Over WebDAV Repository

Vladica M. Ognjanović[1] and Milan Lj. Gocić[2]

***Abstract* – This paper introduces an advanced architecture of full text search over multiple WebDAV repositories. Proposed approach achieve integrated full text search including two important functionalities, namely the possibility to return results which only partially match the query and to rank results accordingly.**

In essence, our work is related with high performance indexing, i.e. processing the original data into highly efficient cross-reference lookup tables in order to facilitate rapid searching.

Experimental results from various tests show that execution time is within the limits of a few seconds making this solution applicable for most common users.

***Keywords* – WebDAV, full-text search, inverted index, metadata search, context search.**

I. INTRODUCTION

The amount of necessary information in a business world is growing at a prodigious rate. The users are becoming more dependent on search engines for locating relevant information.

In most cases, an issued query will result in hundreds of matching documents. In order to avoid flooding the users with a huge amount of results, the search engines present the results in batches of 10 to 20 relevant documents. The user then looks through the first batch of results and if the answer does not exist, he can potentially request to view the next batch. This process is repeated until the user either finds what he is looking for or gives up trying.

Architecture that we propose in this paper allows user to efficiently search over multiple WebDAV (Web-based Distributed Authoring and Versioning) [1] repositories. Currently, existing search support over WebDAV is limited to one repository at the time. It is also dependant on support of a server operating system indexing service (in our case windows indexing service). There are two types of search: search of a documents metadata and a context search. The metadata search includes retrieving some of the document properties like url, creation date, last modified date, length and others. The context search is used for fetching the documents based on their context and user request. This architecture introduces a multi-user system capable of handling the context search of multiple repositories with fast response, reviewing of results, partially matching the query and ranking results.

[1]Vladica M. Ognjanović is with Accordia Group, Kneginje Ljubice 1/1, 18000 Niš, Serbia, E-mail: vladicaognjanovic@gmail.com

[2]Milan Gocić is with Accordia Group, Kneginje Ljubice 1/1, 18000 Niš, Serbia, E-mail: mgocic@yahoo.com

II. ARCHITECTURE

The goal of this architecture is to simplify the search of multiple repositories. Users must register to the system and specify web folders of their interest. The system uses WebDAV as a front-end to collect documents and a file system as a back-end to store essential data of documents. Architecture of the solution is presented in Fig. 1.

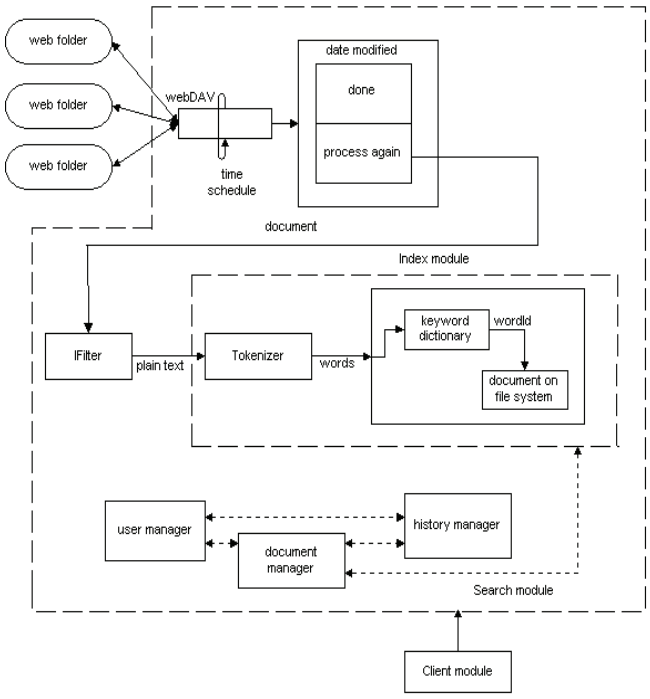


Fig. 1. Search service architecture

The presented search service architecture consists of five main modules: WebDAV module, IFilter module, indexer module, search module and client module.

A. WebDAV module

WebDAV [2-3] is a set of extensions (new headers and new methods) added to HyperText Transfer Protocol 1.1 to support collaborative authoring on the Web. While HTTP is a reading protocol, WebDAV is a writing protocol created by a working group of the Internet Engineering Task Force (IETF), which has defined extensions for six capabilities: overwrite protection, properties, name-space management, version management, advanced collections and access control.

The WebDAV extensions support the use of HTTP for interoperable publishing of a variety of content, providing a

common interface to many types of repositories and making the Web analogous to a large-grain, network-accessible file system. To this time, WebDAV is incorporated into most current operating systems and applications where it performs seamlessly. Since DAV is a single wire protocol like HTTP, it offers a more secure and faster method of file transfer than the one provided by the dual-channel File Transfer Protocol (FTP). Searching using metadata was one of the priorities in the development of DAV. Searching a context of remote web items over WebDAV is supported by operating system on the server in most cases is time consuming.

Collecting properties of web items is a primary task of the WebDAV module. A special accent is put on href, getlastmodified and iscollection properties. Metadata search process is done in three steps:

- Querying server for items metadata is done with a query like:

```
<?xml version="1.0"?>
<a:searchrequest xmlns:a="DAV:">
<a:sql>
        SELECT * FROM SCOPE('DEEP
        TRAVERSAL OF "<web folder
        ulr>'")
</a:sql>
</a:searchrequest>
```

  In this example all properties are requested from server. Explicit request of some properties can be done by modifing SELECT clause.

- Processing of a server reply is based on parsing XML. Server response contains many properties, but we will consider only the listed ones.

```
<?xml version = "1.0"?>
<a:multistatus xmlns:b = "… " xmlns:a = "DAV:">
<a:response>
  <a:href>
        <item url>
  </a:href>
    <a:propstat>
      <a:status>
          HTTP/1.1 200 OK
      </a:status>
      <a:prop>
                  …
        <a:getlastmodified
                  b:dt = "dateTime.rfc1123">
              Sat, 20 Jan 2007 17:14:02 GMT
        </a:getlastmodified>
        <a:iscollection b:dt = "boolean">
              0
        </a:iscollection>
                  …
      </a:prop>
    </a:propstat>
</a:response>
</a:multistatus>
```

  Documents and collections are identified by ***<itemurl>*** and the difference between is made by iscollection property.

- Based on all collected information on some item this module will either send document for a further processing or not. As shown in Fig. 1. this decision is based on the getlastmodfied field.

B. IFilter module

IFilter [4] is a COM component that uses installed IFilter providers to extract the text and to allow the indexer to read different file formats. The providers for the various formats are available from most vendors.

The IFilter interface scans documents for text and properties and extracts chunks of text from these documents, filtering out embedded formatting and retaining information about the position of the text, providing the foundation for building higher-level applications such as document indexers and application-independent viewers.

This interface is implemented to provide a filter for extracting information from a proprietary file format so that the text and properties can be included in the index.

IFilter module searches for a dll and ClassID of COM object responsible for filtering a specific file extension. Then it loads that dll, creates an appropriate COM object and returns a pointer to the IFilter instance which is used to read a plain text from a document.

C. Indexer module

The goal of storing an index is to optimize the speed and performance of finding relevant documents for a search query. Fast query evaluation makes use of an index: a data structure that maps terms to the documents that contain them. With an index, query processing can be restricted to documents that contain at least one of the query terms.

Search engine architectures vary in how indexing is performed and in index storage to meet the various design factors. Types of indices include: suffix trees, trees, inverted indices, citation indices, Ngram indices, term, document matrices.

Architecture presented in this paper is based on inverted indices. An inverted index [5-7] is an optimized structure that is built primarily for retrieval, with update being only a secondary consideration. The basic structure inverts the text so that instead of the view obtained from scanning documents where a document is found and then its terms are seen, an index is built that maps terms to. Instead of listing each document once, an inverted index lists each term in the collection only once and then shows a list of all the documents that contain the given term. Each document identifier is repeated for each term that is found in the document.

Index module architecture is given in Fig. 2. The core of index module is a keyword dictionary. Keyword dictionary is actualy a dictionary where key is a word and the value is a structure containing two integers and one boolean. Integers are wordId and rank, and boolean is reserved for all illegal

words and characters (like stop words, commas, etc.). The rank represents a frequency of word in user queries, and it is primarily used for search suggestion.

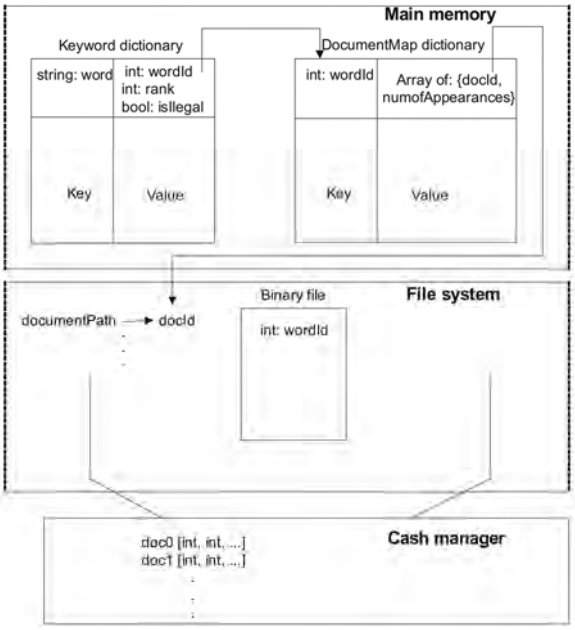


Fig. 2. Index module architecture

In close relation to keyword dictionary is a document map dictionary which is also represented as dictionary, with wordId as the key and array of structure of two integers as the value, the first integer is document id and the second is number of appearances. These two dictionaries reside in main memory. It is important to point out that every keyword and document have unique identifier.

When we get a word from a document, the first we check if a dictionary already contains that word, and if that is true we get the corresponding value in the documentMap dictionary and increase the number of appearances for the current document or add a new entry. Otherwise we create new word id and check if a word or a character is illegal. If illegal we just add it to dictionary, and if not we create a new entry in the documentMap dictionary.

Building index for some document is very simple. If a document is already processed we simply delete all previous information (this includes updating keyword and document map dictionary, and moving the appropriate file from file system to history). After extracting words from documents, we update keyword and document map dictionary as described. Every wordId is stored in the file system and these wordIds array represents documents in this architecture. The maintenance of documents on the file system is the main reason why all the words (legal and illegal) are kept in an one place.

Cash manager is utilized for quick accessing to frequently used documents and to managing the main memory occupied by cash.

Some of major factors in designing a search engine's architecture include:

- merge factors – merging is done by rebuilding index on document,
- storage techniques - file system is used as a back-end,
- index size - considering word encoding index size can be estimated up to 200 times smaller than original documents size,
- lookup speed – lookup time is within the limits of a few seconds,
- maintenance - rebuilding index makes a maintenance simplified, and
- fault tolerance - rebuilding index and periodical autosave of objects in main memory makes this approach reliable.

D. Search module

The primary task of search module is to manage all other modules, as shown in Fig. 1. User, document and history manager are additional parts of the search module. Document manager encapsulates document information, document owners and last index build time. History manager is used for versioning documents. For storage reasons it can be limited to last 10 versions of documents and the compression could be applied. User manager keeps track of user in system, his documents, authentication and user notification (notification of document updates by email).

Search module periodically iterate throw list of registered repositories and triggers WebDAV module. WebDAV module collects documents metadata, filters documents by iscollection and getlastmodified property and sends filtered documents to IFilter. IFilter extracts plain text from documents. The plain text is forwarded to tokenizer that is used for deriving words from text. Index module sends existing document index to history and rebuilds index on a current document.

Document searching involves:

- extracting tokens from user query,
- inquiring keyword dictionary for a list of document ids for each token,
- intersection of lists from keyword dictionary,
- ranking results based on tokens appearance in a document, and number of tokes found in document,
- returning rank results to user,
- results represent documents urls.

User can review a part of the document that contains tokens from search query. Namely, since all words from documents are encoded with corresponding ids, sliding window with the size of 50 words is used to parse document. Every window that contains ids of search tokens is decoded and returned to user.

E. Client module

Client module represents user interface, providing user registration and search service.

III. EXPERIMENTAL RESULTS

The prototype based on the proposed architecture is used for obtaining the experimental results; all tests were performed on four representative documents of different

types. The first and third documents are word document; the second is power point document; and the fourth is a PDF a document.

Index size in term of word count is given in Table 1.

TABLE I
WORD COUNT AND INDEX SIZE

Document number	Number of words per document (approximately)	Index size on disk [kB]
1	3500	14
2	2000	9
3	10000	39
4	34300	135

Original document size and index size ratio are represented in Table 2. Having in mind that each word is represented with four bytes the result can be estimated like in the following example: the expected index size of the third document with approximately 10000 words would be 10000 * 4B = 40 KB

TABLE II
ORIGINAL DOCUMENT SIZE AND INDEX SIZE RATIO

Document number	Original document size[kB]	Original document size / index size
1	3900	278
2	172	19
3	6300	161
4	145	1.07

If all words from document are present in keyword dictionary and if a document is already processed by tokenizer, the build index times would be as shown in Table 3.

TABLE III
DOCUMENT INDEXIG TIME

Document number	Build index time [s]
1	2.3
2	2.2
3	2.7
4	3.6

IV. CONCLUSION

In this paper we have proposed an architecture for full-text search over WebDAV repository.

This architecture improves efficiency of retrieval and enables efficient update of the inverted lists, which is especially important for swift acknowledgement of the document updates.

We plan to implement a complete system based on the proposed architecture that will validate the experimental results.

REFERENCES

[1] M. E. O'Shields, P. J. Lunsford II, "WebDAV: A Web-Writing Protocol and More", Journal of Industrial Technology, Volume 20, Number 2, 2004.

[2] E. J. Whitehead, M. Wiggins, "WebDAV: IETF Standard for Collaborative Authoring on the Web", IEEE Internet Computing, September/October 1998, pages 34- 40.

[3] WebDAV, Retrieved February 1, 2007, www.webdav.org

[4] IFilter, Retrieved April 10, 2007, http://msdn2.microsoft.com/en-us/library/ms691105.aspx

[5] A. Trotman, "Compressing inverted files", Kluwer Int. J. Inform., 2003., 5–19.

[6] E. M. Voorhees, D. K. Harman, "TREC: Experiment and evaluation in information retrieval", MIT Press, Cambridge, MA., 2005.

[7] J. Zobel, A. Moffat, "Inverted files for text search engines", ACM Computing Surveys (CSUR), Volume 38 , Issue 2 Article No. 6, ACM Press New York, NY, USA, 2006.

Models of Interconnect Management of Multimedia IP-based Networks

Tashko Ap. Nikolov

***Abstract* – The scope of network and service management related to interconnect relationships is limited compared to the need of network and service management within an administrative domain. However there are several crucial inter-domain interworking tasks that should be performed via a management interface, which will not be covered through a signalling or a transport plane interface. In this work some models for realization of interconnect management scenarios are present.**

***Keywords* – Interconnection, Multimedia services, IP – based networks, inter – domain networking, QoS.**

I. INTRODUCTION

Interconnection management tasks are important for efficient network operations and service delivery and assurance, with as little as possible interruptions, decreased time to repair, and efficient usage of resources. Management functionality supporting inter-domain monitoring and verifying of SLA (Service Level Agreement), efficient handling of faults, and interchange of charging data, all represents important functional areas applicable to inter-domain management.

II. ARCHITECTURAL ASPECTS OF INTERCONNECTIONS

A. Assumption and methodology

Capabilities and choices regarding the following contention may have great influence on the specific requirements for interconnect management operations and information flows. Capabilities with respect to the following need to be defined in order to specify the corresponding Interconnect Management requirements and capabilities needed.

- End-user services and QoS (Quality of Services) classes, their permanents, and objectives (SLA's),
- Network performance objectives,
- QoS and SLA measurement methodologies,
- QoS and SLA assurance methodologies,
- Transport plane capabilities and types of connectivity resources including CoS (Class of services), QoS, and resource reservation features, as perceived from an interconnection point of view,
- Constraints and assumptions regarding mapping of end user MMoIP (Multimedia over IP) flows (bearers) onto aggregate flows (tunnels),
- End-user naming scheme and "addressing" plan,
- Addressing scheme and addressing plan for MMoIP bearers (flows),
- Addressing scheme and address plan for tunnel end-points as well as for end-points of interconnect links,
- Routing information dissemination approach,
- Resource usage monitoring, charging and accounting schemes.

Several documents regarding management and OSS (Operation Support System) are applicable as sources for this work. Typically, the sources provide generic descriptions and requirements that need further investigation when analyzing and specifying what are applicable requirements and functionality in a given setting, such as for instance given interconnect architecture. Examples of such sources are: ITU-T M.3010 [1], ITU-T M.3200 [2], ITU-T M.3400 [3]etc.

Also some works in the past years include recommendations, example architectures and functionality definitions according the interconnection scenarios and interface specifications of interconnection management of Multimedia IP – based networks [5], [6].

B. Functional Layers of Management architecture

While in [6] and [7] the functional layers within the two planes are described in some details, the functional content of the management plane is not described in a corresponding detail (Fig. 1).

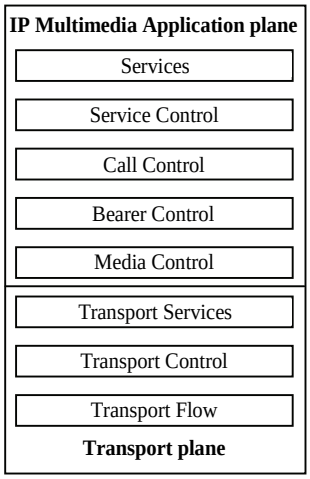


Fig. 1. Figure example

Tashko Nikolov is with Telecom Department at Technical University of Sofia, "Kliment Ohridsky" Blvd 8, 1756 Sofia, Bulgaria, E-mail: tan@tu-sofia.bg

In general, there are management functionalities associated with each of these functional layers. The complexity and management functionality content for each functional layer hi an interconnect setting will vary where some layers are more challenging from a management perspective than others. Several actors or administrative domains are likely to be involved when providing services. Typically, one makes a distinction between a service provider (SP) and a network operator (NO) where the later owns and operates the network within its domain. The NO can also play the role of a (end-user) SP, or rather be limited as a wholesale provider. However, more actors can be involved, and the exact delineation between an SP and an NO may vary from case to case.

C. Domain definitions and connectivity resources

The extent of an SP does not need to correspond with the extent of an NO. This is illustrated in the figure below.

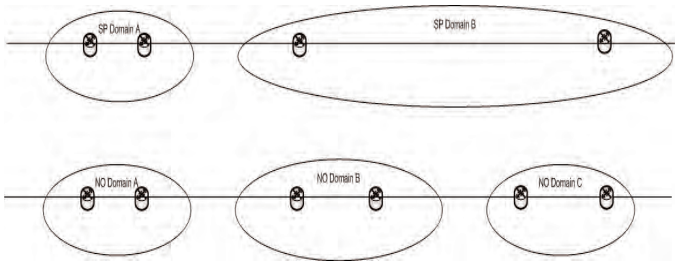


Fig. 2 NO and SP domains

In general, the new IP and MPLS (Multi Protocol Label Switching) packet-based technologies with associated QoS mechanisms such as for instance DiffServ (Differentiated Services) aware MPLS provide opportunities for more efficient network resource utilization than traditional TDM based solutions. However, interconnect and inter-domain network resource and QoS control and management are areas with several challenges.

D. Inter – office interconnect

The following will point out different types of connectivity entities (connection-less or connection-oriented) that are important to identify for subsequent identification of management capabilities needed. In order to get a conceptual view of these different transport plane network connectivity resources that are applicable to an interconnect relationship several figures are provided below.

Figure 3 identifies and illustrates three kinds of connectivity resources:

- An (end-to-end) end-user MMoIP flow (bearer). Such an end-user flow is most likely associated with an end-user application. The end-user flow can potentially be also an aggregate flow. However, this is transparent to the service providers and network operators. It is assumed that some end-user flows may request an SLA that requires monitoring of the individual end-user identifiable flow across multiple domains.

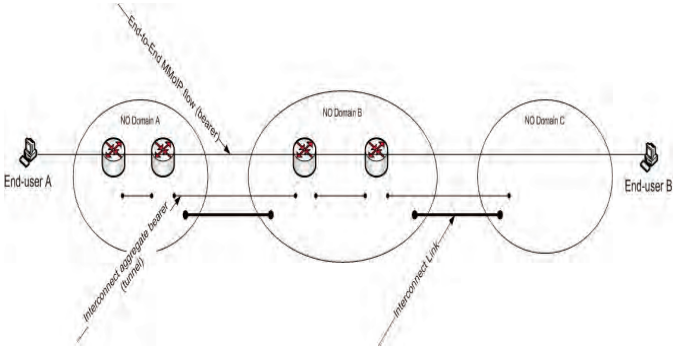


Fig. 3 Transport Plane (TP) – Kinds of connectivity resources – inter – office interconnect

- Tunnel(s) or aggregate bearer(s) that can support or carry several MMoIP end-user flows. Properties and features associated with such tunnels are further elaborated below. This is typically a preferred solution hi an inter-office case. (See main part of the document.)
- Interconnect link that extend between two neighbors NO domains and will typically aggregate several aggregate bearers (tunnels). It is assumed that this is connection oriented packet based or circuit based. Several hierarchical levels of interconnect links can be possible.

In order to avoid statically assignment of resources at a very fine granularity it is assumed that the bearer tunnels are MPLS packet-based. Such a tunnel can have various properties according to the needs of the interconnection relationship. Although MPLS is used for interconnect bearer tunnels. this still leave open to the network operator the choice of technology inside its domain. It is assumed that there is a set of service classes (CoS) available that are applicable to bearer tunnels. There must be an inter-domain coordinated way of mapping MMoIP flows onto tunnels to obey QoS properties. While in general admission control of MMoIP flows will be done per flow per session this will not result in per flow per session resource reservation on network elements (transport plane forwarding devices) providing interconnect network resources. Typically tins will be done on an aggregate basis.

There are several ways of dealing with resource reservation related to bearer tunnels. But in any case, bearer tunnels are an important means of SLA related monitoring and a durance. The following bearer tunnel resource reservation (i.e. some kind of bandwidth reservation) schemes are assumed as possible options:

- No per bearer tunnel resource reservation. Resource reservation can potentially be done at a more aggregate level, e.g. with respect to interconnect links.
- Resource reservation per tunnel (as agreed between interconnected domains) where the network operator maintains reservations at an aggregate level. Thus, no per tunnel resource reservation on network elements (transport plane forwarding devices) providing the tunnel interconnect network resources.
- Resource reservation per tunnel (as agreed between interconnected domains) where the resources are reserved also on network elements providing tunnel interconnect network resources.

The network operator can use (class based) DiffServ aware MPLS in the first two cases, while the last case per LSP (Label Switching Path) resource reservation is used. In any case, in addition to using MPLS LSPs to support SLA monitoring and assurance. MPLS LSPs are also used to ensure stable and predictable routing of IP flows. It is important to ensure that IP packet are directed to the expected interconnect border nodes where both NATing (Network Address Translation) and firewall functionality can be performed.

The number and kinds of bearer tunnels that one domain has interconnecting to other domains may vary. One domain may have only one tunnel to each of its neighbor network operators, which in turn provide interconnect transit services to the other domains. On the other hand one network operator may have several tunnels to each of its neighbors. Tunnels can go to different service providers being served by a neighbor network operator or tunnels can transit network operators to network operators with no interconnect link neighborship. These various kinds of tunnels are illustrated in Figure 4. In order to select different routs for the MMoIP flows they can be mapped onto different tunnels where the tunnels reach the same destination following different routes.

We also note that neighborship or interconnect relationship are relative terms. Both functional layers as shown in Figure 1 as well as the kind of connectivity resource involved will classify the interconnection relationship. Correspondingly, the devices involved in interconnection are of different kinds and may reside on various locations within an operator or provider domain.

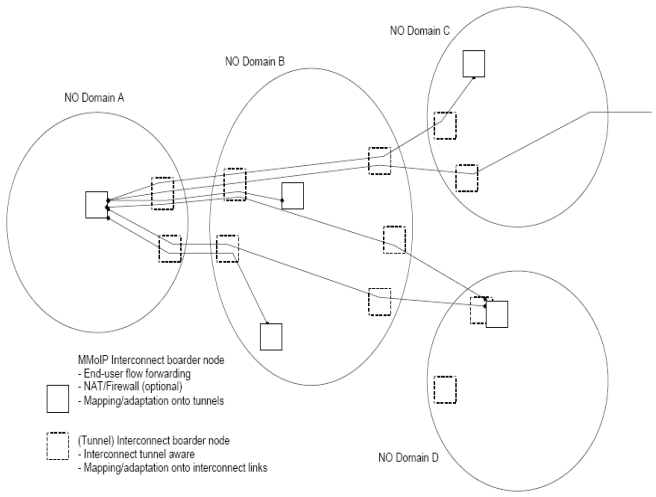


Fig. 4 Examples of aggregate bearers (bearer tunnels) – inter–office interconnect

It is expected that the tunnels as such are rather long lived, while their properties and associated bandwidth or resource reservation may vary depending different traffic patterns varying over time of day or over weeks or months. Several solutions for establishing such inter-domain tunnels exist. Typically one distinguishes setting up tunnels by signalling vs. by management or by a combination of the two ways. However, when looking at the "by management approach" we note that there is an emerging trend that the management and control of network resources and connection (tunnels) are not handled by traditional OSSs but by special purpose servers that can be more or less distributed according to scalability needs. In this sense one may use the notion of network resource control and management (NRCM) as a separate functional plane or layer. This functional plane is separated from the network elements and control plane components of the network elements. NRCM will typically provide functionality as identified belonging to the transport services and transport control functional layers. Policy-based configuration and management is an integral concept of such a functional plane.

Functionalities and capabilities of such a functional layer are crucial as these will enable and assure the QoS and service classification capabilities of the involved network domains provided that a consistent and coordinated mapping from MMoIP flows onto tunnels exist. In the following functionalities and solutions of NRCM will be covered to some extent both in this section and in subsequent sections.

The figure below illustrates the set up of a tunnel crossing two transit domains. It is assumed that NO domain B and C should not know where the tunnel terminates as this may provide undesirable knowledge about traffic patterns of other domains.

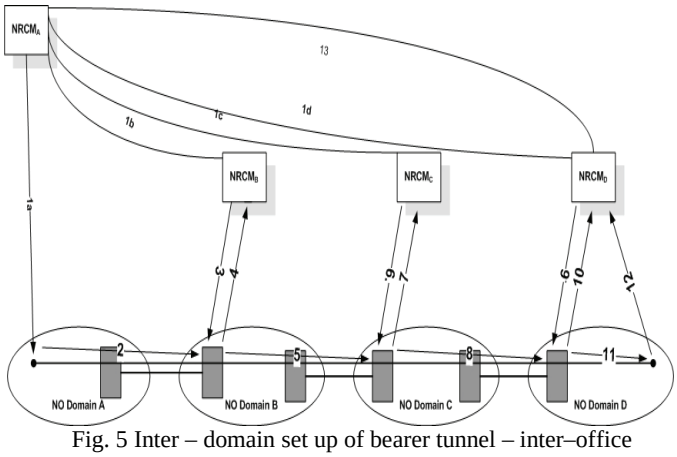


Fig. 5 Inter – domain set up of bearer tunnel – inter–office interconnect

This illustrates a setting where both signalling and "management" (both control plane and management plane) is involved in the set up. Often this is called soft-permanent connection set up. Taking a closer look at interaction (1b) and (3) the identifier provided by NO A is an abstract identifier Telling that the tunnel should go to domain C (or a particular abstract location in NO C). Then NRCMB must find an egress port leading to an appropriate ingress poit in NO C. This may require negotiation between $NRCM_B$ and $NRCM_C$. Whether resource reservation in the network elements is part of the set up should be optional.

To accommodate appropriate SLA monitoring and assurance as well as resilience and fault management there is a need to monitor and manage both the tunnel from an end-to-end perspective as well as the various intra-domain and inter-domain sections that the tunnel traverses. Several methods and mechanisms are candidates, for these purposes. It is a goal to select appropriate means while keeping the CAPEX and OPEX low.

III. MANAGEMENT FUNCTIONALITY

By analysing potential FCAPS functionality as identified in for instance M.3400 for each ICM functional layer, ICM functional requirements should be identified and appropriately structured. The next step is then to consider various alternative management entities and relationships between such entities, to reach a functional specification of the different interfaces. Protocol neutral information modeling may be used as a specification approach. Then, on a bilateral basis, the exact management protocol profile may be chosen from a set of standard protocol profiles.

Depending on the volume and complexity of management transactions, it is possible that some management tasks can be done manually. At a later stage, when the volume increases, automation by an OSS-OSS interface can be introduced.

There are several main management functional areas that could be applicable and need further investigation. In the following a brief (initial and preliminary) description of ICM functionality will be provided according to the structure identified in Figure 1. This structuring of the management capabilities needs further analysis and assessment. Some of the capabilities are applicable when considering multi-media services in general and not limited only to the voice service component.

It is assumed (below) that the MMoIP internetworking arena will be dynamic and evolving. It is important to support an evolutionary and step-by step approach where first, simple QoS-enabled connectivity services can be offered across domains. Then, more functionality and service features can be added step-by-step as experience is gained and the solutions mature. In order to allow such a vision in a cost-efficient manner it is important that management functionality exist to announce and discover changes to protocol or service capabilities, across the numerous interconnect relationships. This should be supported at each functional layer. It is a goal to allow different domains to develop their network and services differently, and one domain can have interconnect relationships to different other domains each supporting different capabilities.

Likewise, it is important to have efficient means of establishing a new interconnect relationship e.g. a new interconnect link, and efficiently updating or changing the configuration and knowledge of how features are interconnected. As interconnect topology and traffic patterns change it should for instance be possible to dynamically announce changes to routing policies. Routing policies are applicable at different layers, from routing of send requests, routing of call requests, routing of MMoIP flows, to routing of interconnect tunnels and links.

In Section II above it is discussed how bearer tunnels, in typically in the inter-office case, provide flexibility, efficient resource utilisation, and a means of ensuring SLA and QoS features. It is assumed (below) that CoS/QoS can be consistently mapped between the different functional layers and across domains. It is assumed that this is not a static mapping and different CoS/QoS schemes may be used internally within different domains, and CoS/QoS capabilities may evolve differently from domain to domain. Thus, this creates a need for efficiently supporting announcement and negotiation of CoS/QoS capabilities and mappings used for each (group of") interconnect relationship.

An alternative to this flexible scheme is to have static international agreements and no management system support for announcement and negotiation of service and CoS/QoS capabilities. A cost-benefit analysis may be needed to conclude on the preferred strategy. The functionality identified below assumes the dynamic approach.

However, in any case there is a need for some resource reservation scheme. As pointed out in Section II above, this can be done in several ways with different levels of granularity. Depending on the chosen scheme the management systems or Network Resource Control and Management systems will be involved in different ways.

SLA assurance is a crucial and challenging area. Procedures and international agreements must be developed to enable SLA assurance at various network levels and layers and corresponding SLA conformance reporting. Depending on the chosen scheme the management systems or Network Resource Control and Management systems will be involved in different ways.

IV. CONCLUSION

In this paper the approach represented in [5][6] is being followed further with recommendations according domains set up and interconnect scenarios for MMoIP networks interconnection. Some models are presented and future definitions are given. Also some management functionality requirements according interconnection of Multimedia networks are introduced.

REFERENCES

[1] ITU-T Recommendation M.3010. Principles for a Telecommunications management network, February 2000

[2] ITU-T Recommendation M.3200. TMN management services and telecommunications managed areas: overview. April 1997

[3] ITU-T Recommendation M.3341. Requirements for QoS/SLA management over the TMN X-interface for IP-based services. March 2003

[4] ITU-T Recommendation M.3400. TMN Management functions. February 2000

[5] Tsenov Al., “Interconnection of Multimedia service networks”, Proceedings of the 14-th International Scientific and Applied Science Conference - “Electronics - ET 2005”, pp. 138 - 143, Sozopol, Bulgaria, September 21 – 23, 2005

[6] Tsenov Al., “Management analysis for interconnection of Multimedia service networks”, Proceedings of the 14-th International Scientific and Applied Science Conference - “Electronics - ET 2005”, pp. 144 - 149, Sozopol, Bulgaria, September 21 – 23, 2005

SESSION TST III

Telecommunications Systems and Technology III

Pulse Shape Influence on Optical System Performance in the Presence of Interference, Dispersive and Nonlinear Effects

Mihajlo Č. Stefanović[1], Dragan Lj. Drača[2], Aleksandra S. Panajotović[3] and Daniela M. Milović[4]

Abstract - **Dispersion and nonlinear effects are most important limiting factors in optical transmission systems. The research subject of this paper is the common influence of dispersive and nonlinear effects on the optical system performance for different shapes of input pulse. Influence of inband interference on system performance is also studied in this paper. BER (Bit Error Probability) for IM-DD (Intensity Modulation – Direct Detection) telecommunication system is obtained for the worst case of interference influence.**

Keywords – **Chromatic dispersion, Kerr's nonlinearity, Inband Interference, Schrödinger equation, IM-DD system.**

I. INTRODUCTION

In the past, optical losses were the most important limiting factors in optical transmitting systems. Today, in the actual optical systems, dispersive and nonlinear effects restrict speed and transmission length of these systems. Highly nonlinear operating conditions or the interplay between the different linear and nonlinear effects can result in dramatic changes of the temporal and spectral properties of the pulse and thus lead to severe performance degradation.

Propagation of optical pulses in single-mode optical fibers is mainly influenced by the group velocity dispersion and the refractive index nonlinearity. The main contribution to the group velocity dispersion is represented by the parameter β_2, which leads in general to a broadening of the pulse shape. As a pulse widens, it can broaden enough to interfere with neighboring pulses (bits) on the fiber, leading to intersymbol interference. Dispersion thus limits the bit spacing and the maximum transmission rate on a fiber-optic channel. TOD (third-order dispersion) and FOD (forth-order dispersion) are higher-order effects originating from the wavelength dependence of the group velocity dispersion.

[1]Mihajlo C. Stefanovic is with Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia E-mail: misa@elfak.ni.ac.yu.

[2]Dragan LJ. Draca is with Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia E-mail: draca@elfak.ni.ac.yu

[3]Aleksandra S. Panajotovic is with Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: alexa@elfak.ni.ac.yu

[4]Daniela M. Milovic is with Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dacha@elfak.ni.ac.yu.

These dispersive effects can also distort ultrashort optical pulses in the linear as well as in nonlinear regime. In a dispersive medium, the index of refraction is a function of the wavelength. Thus, if the transmitted signal consists of more than one wavelength, certain wavelengths will propagate faster than other wavelengths. Since no laser can create a signal consisting of an exact single wavelength, chromatic (material) dispersion will occur in most systems [1], [2]. Second order dispersion has the strongest influence on pulse deformation and higher order dispersions are neglected in its presence. Single-mode optical fiber that works under anomalous dispersive regime is used as transmitting medium in this paper.

An optical signal power transmitted through the optical fiber, long hundreds of kilometers, is in the range of a few mW [3]. Influence of different sort of nonlinear effects had not to be neglected for such optical power values, especially the influence of Kerr's nonlinearity [1]. This nonlinear effect is consequence of refractive index which depends on the pulse intensity. Kerr's nonlinearity is able to decrease pulse deformation produced by chromatic dispersion in anomalous dispersive regime of the optical fiber.

Previous results have shown that shape of fundamental LP_{01} mode in single-mode fiber is similar to Gaussian shape. It is reason why many useful pulses in optical system has Gaussian shape [3]. It is more useful, in some cases, to use Sech shape of pulse [1], especially in nonlinear optical fiber.

The interference is one kind of disturbance that appears in optical telecommunication system and it can be inband or outband, i.e. it can be at the same or at different frequency in relation to a useful signal [4], [5]. Interference is always accumulated when the signal passes through the nodes and can't be eliminated by filters. Since inband interference can't be eliminated by optical filtering in a receiver, its influence on system performance is studied in this paper. This kind of interference can cause fluctuation of signal power which on the other hand can degrade system performances.

All results in this paper are determined by numerically solving Schrödinger equation using symmetrical split-step Fourier method.

II. RECEIVED PULSE

Pulse propagation along the nonlinear optical fiber can be described by Schrödinger equation [1]:

$$\frac{\partial U}{\partial z} = -i\frac{\operatorname{sgn}(\beta_2)}{2L_D}\frac{\partial^2 U}{\partial \tau^2} + \frac{i}{L_{NL}}|U|^2 U \tag{1}$$

where U is optical pulse envelope. β_2 is known as group-velocity dispersion parameter and defines dispersive regime of optical fiber. L_D and L_{NL} are dispersive and nonlinear length, respectively. In case when dispersive effects along with nonlinear effects affect pulse propagation, these lengths must be equal ($N^2 = L_D/L_{NL} = 1$).

An useful optical pulse at the beginning of optical fiber can be written as:

$$s(0,\tau) = U(0,\tau)\cos(\omega_r \tau) \tag{2}$$

$\omega_r = \omega T_0$ is normalized frequency and T_0 is pulse width. Inband interference is consequence of using nonideal devices in the optical system. It can appear anywhere along the fiber and it is time and phase shifted in regard to the useful optical pulse. The inband interference can be modeled as:

$$s_i(z_i,\tau) = U_i(z_i,\tau)\cos(\omega_r \tau + \varphi) \tag{3}$$

U_i is the inband interference envelope, φ is its phase shift and z_i is the place along the optical fiber where the interference appears and

The envelope and phase of the resulting signal at the place where interference appears are simply evaluated as [6]:

$$U_r(z_i,\tau) = \sqrt{U^2(z_i,\tau) + 2U(z_i,\tau)U_i(z_i,\tau)\cos\varphi + U_i^2(z_i,\tau)} \tag{4}$$

$$\psi(z_i,\tau) = arctg\frac{U_i(z_i,\tau)\sin\varphi}{U(z_i,\tau) + U_i(z_i,\tau)\cos\varphi} \tag{5}$$

and they represent new initial conditions for solving Schrödinger equation.

The time shape of received optical pulse can be obtained by solving Eq. (1) using symmetrical split-step Fourier method.

III. BIT ERROR PROBABILITY

IM-DD optical telecommunication systems are simple and pay off economically and these are the main reasons for their wide application. If we assume quantum noise to be Gaussian distributed also assume that both symbols (1 and 0) are equally probable, the conditional bit error probability of such system is [3], [7]:

$$P_{e/b} = \frac{1}{2}\left[\int_{V_p}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(y-\bar{y}_0)^2}{2\sigma_0^2}\right)dy + \int_{-\infty}^{V_p}\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(y-\bar{y}_1)^2}{2\sigma_1^2}\right)dy\right] \tag{6}$$

where $\bar{y}_0$, $\bar{y}_1$ are mean values of the resulting signals (4) and V_p is the optimal decision threshold:

$$V_p = \frac{\bar{y}_1'\sigma_0 + \bar{y}_0'\sigma_1}{\sigma_0 + \sigma_1} \tag{7}$$

$\bar{y}_0'$, $\bar{y}_1'$ are mean values of useful signal. The unconditional bit error probability is obtained by averaging relation (6), assuming that $p(b)$ is uniform probability density function of interference time shift:

$$BER = \int_{-\frac{1}{2B}}^{\frac{1}{2B}} P_{e/b}\,p(b)db \tag{8}$$

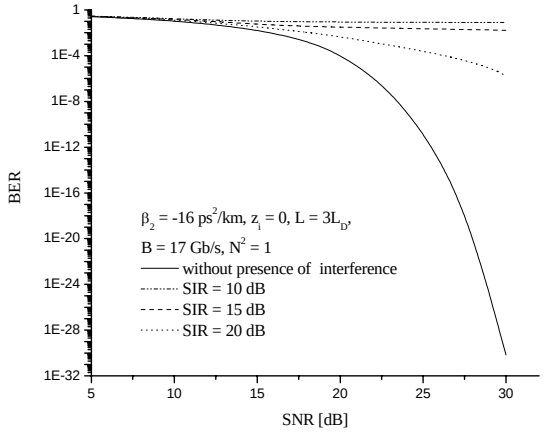
and B is bit rate.

Gaussian pulse is very often found as a useful optical pulse in optical telecommunication systems. Its envelope and envelope of inband interference can be modelled similarly as:

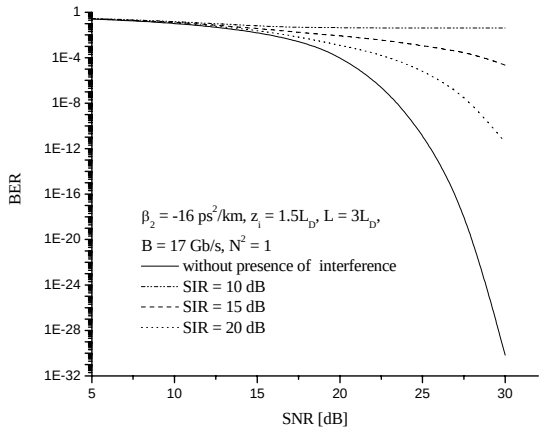
$$U(0,\tau) = a\exp(-\tau^2/2) \tag{9}$$

$$U_i(z_i,\tau) = a_i\exp(-(\tau-b)^2/2) \tag{10}$$

Value of parameter a depends on sending information. Parameter a_i shows normalized magnitude of interference and b is time shift of interference.



a)



b)

Fig. 1. The bit error probability as function of SNR (Gaussian pulse)

a) $z_i = 0$, *b*) $z_i = L/2$.

Bit error probabilities obtained using Eqs. (1), (4), (5), (6)-(8), (9) and (10) are presented on Fig. 1 (SIR – Signal-to Interference Ratio).

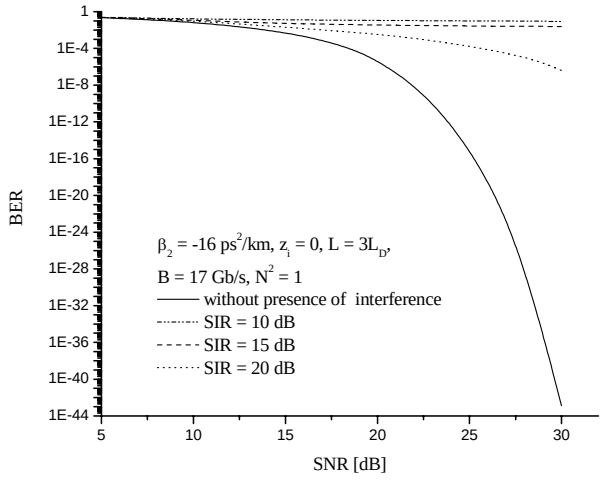
Scientific results showed that Kerr's nonlinearity can decrease pulse deformation induced by second order chromatic dispersion. If we want that Kerr's nonlinearity neutralize that negative influence of dispersive effects we should use pulse with Sech envelope as useful pulse:

$$U(0,\tau) = a\,\mathrm{sec}\,h(\tau) \tag{11}$$

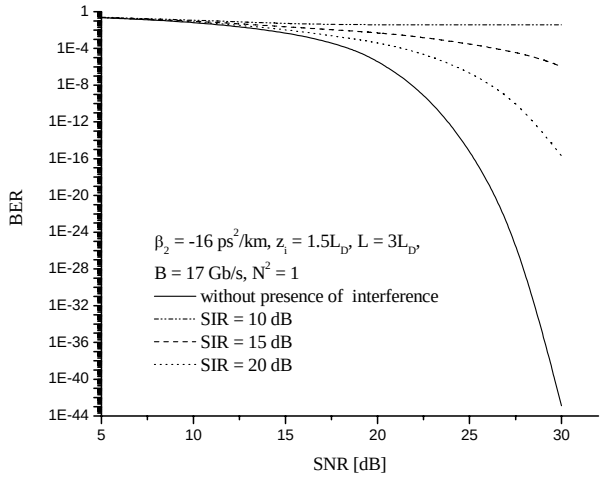
The envelope of inband interference has following shape:

$$U_i(z_i,\tau) = a_i\,\mathrm{sec}\,h(\tau - b) \tag{12}$$

Fig. 2 shows bit error probability calculated for this case.



a)



b)

Fig. 2. The bit error probability as function of SNR (Sech pulse)
a) $z_i = 0$, *b*) $z_i = L/2$.

It can be noticed from both figures that performances of optical systems are the worst when interference appears at the beginning of optical fiber than it appears along the fiber. There is a logical reason for such results. Condition for common influence of dispersive and nonlinear effects, i.e. $N^2 = 1$, is perturbed at the place when interference appears. From this place to the end of optical fiber influence of dispersive effects dominates in the fiber. This is reason for pulse broadening that induces such severe performances of optical system.

Solid lines on Figs. 1 and 2 show BER when Gaussian and Sech pulse propagate along the fiber. For SNR = 30 dB, BER for Sech pulse is 10^{12} times less than for Gaussian pulse. That shows better resistance of Sech pulse to common influence of dispersive and nonlinear effects in the presence of inband interference.

When inband interference appears at the beginning of optical fiber, system performance is better when useful pulse has Gaussian envelope. In the presence of such kind of interference, difference between resulting pulse and useful pulse at the receiver, at the decision time, is less than in case when useful pulse has Sech envelope shape.

Opposite situation happens when inband interference appears along the fiber. Sech pulse maintains its shape to the place when interference appears. Because of that this situation converge to case when Sech pulse propagates along shorter fiber in the presence of inband interference at the beginning of optical fiber. Shorter fiber means less influence of dispersive effects on pulse shape of resulting signal. This is reason for better performance of optical systems that use Sech pulse when the interference appears along the fiber.

In our previous paper we studied how pulse shape influence on system performance in the presence of interference and dispersive effects [8]. In that case better performances are obtained when interference appears at the beginning of optical fiber than somewhere in the middle of the fiber. It was shown in that paper that Sech pulse is more immune to interference influence than Gaussian pulse in both cases of the place of interference occurrence $z_i = 0$ and $z_i = L/2$.

IV. Conclusion

Influence of second order dispersion, Kerr's nonlinearity and inband interference on performance of IM-DD systems is studied in this paper. In optical telecommunication system Gaussian and Sech pulse are very often found as useful pulse. This is reason for using these pulses in our investigation. Calculated BER confirms fact that Kerr's nonlinearity enables propagation Sech pulse along the fiber without great pulse deformation. Inband interference can be very dangerous because it can't be eliminated by optical filtering at the receiver. We considered two cases. First, interference appears at the beginning of optical fiber. Second, interference appears in the middle of the fiber. Results obtained in this paper showed that Sech pulse is more resistant to influence of interference appearing along the fiber than Gaussian pulse. Situation is changed when interference appears at the

beginning of the optical fiber. For both useful pulse shapes, system has better performance when interference appears near to the receiver than to the laser.

REFERENCES

[1] G. A. Agrawal, *Nonlinear Fiber Optics*, Academic Pres INC, 1995.

[2] M. Cvijetić, *Digitalne svjetlovodne telekomunikacije*, Naučna knjiga, Beograd, 1989.

[3] J. M. Senior, *Optic Fiber Communications*, Prentice Hall, 1992.

[4] G. P.Agrawal, *Lightwave Technology - Telecommunication Systems*, Wiley, 2005.

[5] D. Milović, B. Stojanović, M. Milošević, "Analysis of interchannel crosstalk time-shift and optic power influence on useful signal propagation", ETRAN 2003, Conference Proceedings, 2, pp. 361-363, Herceg Novi, Serbia and Montenegro, 2003.

[6] M. Stefanović, D. Drača, P. Spalević, A. Panajotović, "The influence of crosstalk signal interference to signal propagation along the nonlinear and dispersive fiber", J. of Optic. Commun., vol. 26, no. 1, pp. 9-12, 2005.

[7] A. Marinčić, *Optičke telekomunikacije*, Univerzitet u Beogradu, Beograd, 1997.

[8] M. Stefanović, D. Drača, A. Panajotović, D. Milović, M. Petrović, "Influence of input pulse shape on optical system performance in the presence of interference and second order dispersion", INFOTEH 2007, Conference Proceedings, Jahorina, Bosnia and Hercegovina, 2007.

Fiber Nonlinearity Limitations in WDM CATV Systems

Lidia T. Jordanova[1] and Dobri M. Dobrev[2]

***Abstract* – The paper deals with some problems referring to fiber linearity effects on the parameters of the WDM CATV systems. Mathematical models of the scattering and Kerr effects in optical fibers are suggested taking into consideration both the parameters of the fiber chosen and the main characteristics of the system. Dependencies are given to determine the maximum launch power in a way to provide for acceptable worsening of the received information quality due to reduced carrier-to-noise ratio, intra-symbol distortion and cross-modulation.**

***Keywords* – WDM system, SBS, SRS, SPM, XPM, FWM**

I. INTRODUCTION

Wavelength division multiplexing technology (WDM) allows increasing the transport capacity of CATV systems with the number of wavelengths used. A wavelength in this technology has the meaning of an optical carrier modulated with either an analogue or a digital signal. The most challenging in WDM systems is to achieve simultaneously smaller channel separation as well as higher modulation rates (bit rates) of the optical carrier in order to increase the transmission capacity. These goals begin to be contradictory at a certain point, due to dispersion and nonlinear effects in the single-mode optical fiber.

In WDM systems the laser beams of strictly defined wavelengths λ_1, λ_2, … λ_n are modulated and then combined by an optical multiplexer in the same single mode optical fiber at the transmitting side. At the receiving side an optical demultiplexer splits the optical channels into the same wavelengths and directs them to the optical receivers. Modern optical fibers have low attenuation coefficients (about 0.2 dB/km to 0.3 dB/km) but if the optical signals are carried at a long distance, they must be amplified every 80 to 120 km. A standard erbium-doped fiber amplifier (EDFA) has a total output power of 17 dBm (50 mW) at the maximum. When the power of the transmitted signal turns out to be higher than a predefined threshold level, some nonlinear effects can appear in the optical fiber thus causing noise and distortion.

There are two types of nonlinear effects in the optical fibers: scattering effects and Kerr effects. Stimulated Raman scattering (SRS) and stimulated Brillouin scattering (SBS) are associated with the first type. They are due to non-elastic interaction between a pump wave of wavelength λ_p and the fiber core that transfers most of the pump energy into a Stokes light wave of wavelength $\lambda_s > \lambda_p$. Kerr effects occur because of the dependence of the refractive index *n* on the power density inside the fiber core. Three types of Kerr effects are considered to be important for modern optical communications: four-wave mixing (FWM), self-phase modulation (SPM) and cross-phase modulation (XPM). The effects here mentioned are due to elastic interactions with which no energy transfer from wave to fiber takes place i.e. the total energy of the incident waves is transferred to the emerging waves only.

II. LAUNCH POWER DETERMINATION WITH REGARDS TO SCATTERING EFFECTS

Though very similar in origin SRS and SBS differ due to the fact that optical phonons participate in SRS while acoustic phonons participate in SBS. A fundamental difference is that in optic fibers SBS occurs only in the backward direction (with respect to the pump) whereas SRS dominates in the forward direction. Growth of the Stokes power due to SRS is characterized by the relation [1]

$$P_s(L) = P_s^*(0)\exp\left[g_R(\omega_s) P_p(0) L_{\text{eff}} / A_{\text{eff}} - \alpha L \right], \quad (1)$$

where $P_s(L)$ is the Stokes power at the fiber output, $P_p(0)$ is the input pump power, $g_R(\omega_s)$ is the value of the Raman-gain coefficient at a Stokes frequency ω_s that is downshifted from the pump frequency by about 13.2 THz, L is the actual length of the fiber, L_{eff} is its effective length, α is the fiber attenuation coefficient and A_{eff} is the effective core area.

Magnitude $g_R(\omega_s)$ corresponds to the Raman gain peak whose typical value is 1.10^{-13} m/W at a pump wavelength of 1550 nm. It is well known that the Raman gain spectrum is broad enough to extend up to 40 THz with a broad dominant peak at ω_s. The following formulae are used to calculate L_{eff} and A_{eff} :

$$L_{\text{eff}} = \left(1 - e^{-\alpha L}\right)/\alpha, \; A_{\text{eff}} = \pi\left(MFD/2\right)^2 \quad (2)$$

where MFD is the mode field diameter of the fiber to be found in the manufacturer's data sheet.

Similarly to SRS, the Stokes power grows exponentially in the backward direction because of the Brillouin amplification due to SBS. That growth can be described by the same relation as (1) if $g_R(\omega_s)$ is replaced with the peak value of the Brillouin-gain coefficient $g_B(\omega_s)$. It should be pointed out that

[1]Lidia T. Jordanova is with the Faculty of Communications and Communications Technologies, 1756 Sofia, Bulgaria, e-mail: jordanova@tu-sofia.bg

[2]Dobri M. Dobrev is with the Faculty of Communications and Communications Technologies, 1756 Sofia, Bulgaria, e-mail: dobrev@tu-sofia.bg

Brillouin-gain spectrum is extremely narrow with a bandwidth of about 10 MHz. The peak value of Brillouin gain $g_B(\omega_s)$ is downshifted from the pump frequency by about 10 GHz, i.e., the Stokes shift in SBS is smaller (by order of three) if compared with SRS. For a narrow-bandwidth pump $g_B(\omega_s) \approx 5.10^{-11}$ m/W at a pump wavelength of 1550 nm, i.e. g_B is higher by more than two orders of magnitude if compared with g_R.

Both types of scattering effects reveal a threshold-like behavior, i.e. significant conversion of pump energy into Stokes energy occurs only when pump power exceeds a certain threshold level. The Raman threshold is defined as the input pump power at which Stokes power at the fiber output becomes equal to that of the pump, i.e.

$$P_s(L) = P_p(L) = P_p(0)\exp(-\alpha L). \quad (3)$$

The same formula applies to determine the Brillouin threshold if $P_s(L)$ is replaced with Stokes power at the fiber input $P_s(0)$.

The critical pump power required to reach the Raman threshold P_{RT} in a single mode fiber with $\alpha L >> 1$ can be calculated by the following formula:

$$P_{RT} = 16A_{\text{eff}} / g_R(\omega_s) L_{\text{eff}}. \quad (4)$$

The SBS power threshold P_{BT} is given by the equation

$$P_{BT} = 21A_{\text{eff}} / g_B(\omega_s) L_{\text{eff}}. \quad (5)$$

The typical value of the SBS threshold is less than 10 mW, while the SRS threshold is higher by about two orders and can reach 1 W.

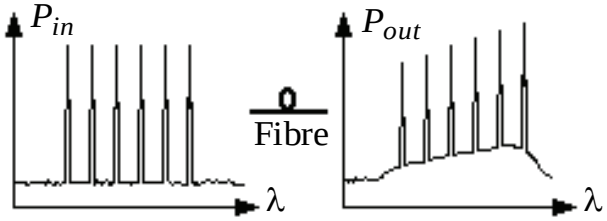


Fig. 1

The Raman-gain spectrum being very broad, SRS can cause problems in WDM systems and does not affect the parameters of the single-channel systems. Due to SRS an energy transfer from lower channels (shorter wavelengths) to higher channels (longer wavelengths) is observed, as shown in Fig. 1. This results in worsening the CNR in lower channels and limiting the transport capacity of CATV systems. The power penalty due to SRS is characterized by

$$PP_{SRS} = -10\lg(1 - P_k), \quad (6)$$

where P_k is the fraction of the power coupled from channel k to all other channels.

The following relation [2] can be used to calculate the total power P_{tot} of all WDM channels that is transferred to the fiber in order to provide 0.5 dB power penalty ($P_k < 0.1$):

$$P_{\text{tot}}(N-1)\Delta\lambda_s L_{\text{eff}} < 40\,000\,[\text{mW.nm.km}], \quad (7)$$

where N is the number of channels and $\Delta\lambda_s$ is the channel spacing.

III. LIMITATION OF THE INTRA-SYMBOL DISTORTION BY SPM AND XPM

SPM refers to the self-induced phase shift experienced by an optical field during its propagation along optical fibers. Its magnitude can be defined on the basis of the optical field's phase Φ according to the formula

$$\Phi = \Phi_o + \Phi_1 = (n_o + n_1 P/A_{\text{eff}}) k_o L_{\text{eff}}, \quad (8)$$

where n_o is the refractive index, n_1 is the nonlinear index coefficient, P/A_{eff} is the optical intensity inside the fiber, $k_o = 2\pi/\lambda$ and λ is the optical wavelength. SPM is responsible for broadening the pulses' spectrum and for producing the optical solitons in the fibers' anomalous-dispersion regime.

If the effect of group-velocity dispersion (GVD) on SPM is negligible then the intensity-dependent nonlinear phase shift at a point *z* arbitrarily chosen along the fiber at a moment *t* can be described by

$$\Phi_1(z,t) = |U(0,t)|^2 P_o \gamma L_{\text{eff}}, \quad (9)$$

where U(0,t) is the normalized optical field amplitude at z = 0, P_o is the peak power and $\gamma = 2\pi n_1/\lambda A_{\text{eff}}$ is the nonlinear propagation coefficient. Parameter γ can be calculated by the formula

$$\gamma = 2\pi n_1/\lambda A_{\text{eff}}, \quad (10)$$

where $n_1 \approx 3.2 \times 10^{-20}$ m^2/W.

Since Φ_1 is proportional to $|U(0,t)|^2$ its temporal variation is identical to that of the pulse intensity. The maximum phase shift occurs at the pulse center located at t = 0 and is given by

$$\Phi_{1\max} = \gamma P_o L_{\text{eff}}. \quad (11)$$

In order to avoid inadmissible intra-symbol distortion in NRZ digital systems the requirement $\Phi_{1\max} \le \pi/2$ must be fulfilled.

The SPM-induced spectral broadening is a consequence of the time dependence of Φ_1. A temporally varying phase implies that the instantaneous optical frequency differ across the pulse from its central value ω_o. The difference δω(t) is given by

$$\delta\omega(t) = -\frac{\partial\Phi_1}{\partial t} = -\gamma P_o L_{\text{eff}} \frac{\partial}{\partial t}|U(0,t)|^2. \quad (12)$$

The time dependence of δω can be viewed as a frequency chirp increasing in magnitude with the distance propagated. In other words, new frequency components are continuously generated as the pulse propagates down the fiber. These SPM-generated frequency components broaden the spectrum over its initial width at z = 0.

As shown through analysis, the temporal variation of the induced chirp δω is negative near the leading edge of the pulse (red shift) and becomes positive near the trailing edge

(blue shift). In other words, the result is a shift towards longer wavelengths at the leading edge of the pulse along with a shift to shorter wavelengths at the trailing edge. The SPM effect is illustrated in Fig. 2.

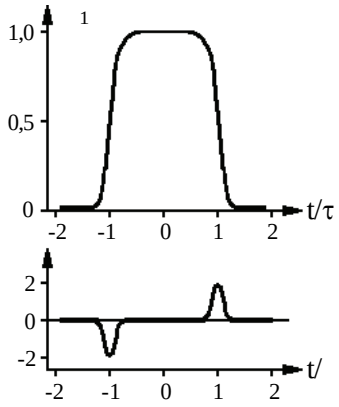


Fig. 2

In the CATV systems design the mutual influence of GVD and SPM effects on the pulse width must be taken into consideration. It is known that in the normal-dispersion regime (when GVD parameter is positive) the wavelength shift caused by GVD is the same as with SPM. Hence, in that case GVD leads to an enhanced rate of pulse broadening compared with that expected for SPM alone. The situation differs for pulses propagating in the anomalous-dispersion regime of the fiber (when GVD parameter is negative). In that case the spectrum is narrowing rather than broadening as expected by SPM in the absence of GVD. This behavior can be explained with the fact that the SPM-induced chirp is positive while the dispersion-induced chirp is negative, i.e. the two chirp contributions nearly cancel each other.

Cross-phase modulation appears when two or more waves propagate inside the fiber and interact between them in result of the nonlinearity of the refractive index produced by the total power inside the fiber. This effect is similar to SPM but the phase shift of one channel depends on the power of other channels. The XPM phase shift Φ_i associated with each channel (i = 1,2, …, N) can be estimated by adapting the formula used for SPM phase shift as follows

$$\Phi_i = \gamma L_{\text{eff}} \left[P_i + 2\sum_{n \neq i}^{N} P_n \right]. \qquad (13)$$

As seen from (13), XPM is always accompanied by SPM. If the optical fields are of equal intensity the XPM contribution to the nonlinear phase shift is twice as big as that of SPM. Like SPM, the XPM phase shift in a NRZ digital system becomes significant when $\Phi_i > \pi/2$.

In WDM systems both SPM and XPM can cause significant phase changes. When information is transmitted through amplitude modulation and is then incoherently demodulated, nonlinear phase changes are of little consequence. However, if coherent demodulation techniques are employed, such phase changes can limit the system performance.

Let's consider the case of amplitude-modulated coherent CATV system. If a phase sensitive (homodyne) detection technique is used, the phase Φ_i would vary from bit to bit depending on the bit pattern of the neighboring channels. In the worse case the XPM-induced phase shift of a given channel will be

$$\Phi_{i\max} = 2\gamma L_{\text{eff}} (N-1) P , \qquad (14)$$

where P is the power assumed to be the same in each channel. If we take $\Delta\Phi_{i\max} \leq 0.1$ as an acceptable value, the power in each channel is restricted to

$$P \leq 0.05 \left[\gamma L_{\text{eff}} (N-1) \right]^{-1}. \qquad (15)$$

IV. FWM EFFECT ON THE SYSTEM PARAMETERS

Four-wave mixing is the interaction between three transmitted channels of different frequencies f_i, f_j and f_k, producing a fourth product frequency

$$f_{ijk} = f_i + f_j - f_k . \qquad (16)$$

There are number of ways in which channels can combine to form a new channel according to the formula above. Figure 3 illustrates the possible frequency combinations that can be generated. With N-channel system the number M of unwanted signals known as ghost channels can be calculated by

$$M = 0.5\left(N^3 - N^2 \right). \qquad (17)$$

For example, a four-channel system would produce 24 ghost channels and an eight-channel system would produce 224 unwanted channels.

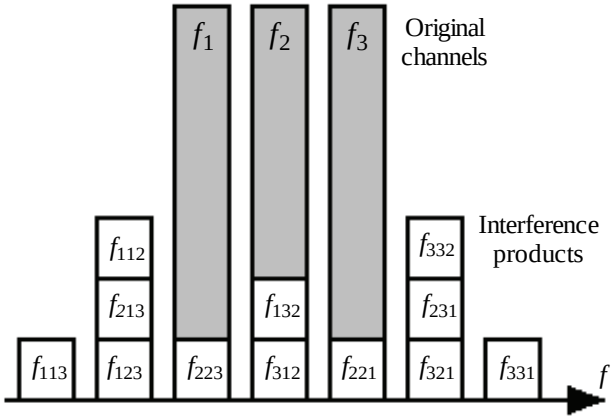


Fig. 3

FWM products reduce the energy in the transmitted channels, thus causing the carrier-to-noise ratio (CNR) to go down at the receiver input. In addition, if the resulting frequency product is within the bandwidth of the transmitted channel it will cause crosstalk at the receiver. As shown in Fig. 3, four of the products generated through FWM drop within the frequency band of the channels transmitted while the remaining six products are out of it and can be suppressed through optical filtering.

Although the equation (16) indicates the position of the potential FWM products it provides no information as to whether the product will be viable, i.e. if the process will be

efficient enough for the product to have significant power. The effect of FWM depends on the phase relationship between the interacting signals. That's why the efficiency of the FWM process is determined by the phase matching condition. Phase matching depends on the frequencies of the incident and resultant signals and the chromatic dispersion of the fiber.

The FWM phase mismatch $\Delta\beta = \beta_i + \beta_j - \beta_k - \beta_{ijk}$ can by expressed as [3]

$$\Delta\beta = \left(\frac{2\pi c}{\lambda^2}\right)\left(\frac{dD}{d\lambda}\right)\left(\frac{\lambda_i + \lambda_j}{2} - \lambda_o\right)\left(\lambda_i - \lambda_k\right)\left(\lambda_j - \lambda_k\right) \quad (18)$$

where λ_i, λ_j and λ_k are the three original wavelengths; λ_o is the zero-dispersion wavelength and $dD/d\lambda$ is the slope of the dispersion curve at $\lambda = 1550$ nm.

The following formula is used to determine FWM efficiency η:

$$\eta = \left(\frac{\alpha^2}{\alpha^2 + \Delta\beta^2}\right)\left(1 + \frac{4e^{-\alpha L}\sin^2\left(\Delta\beta L/2\right)}{\left(1 - e^{-\alpha L}\right)^2}\right) \mathrm{x} \quad (19)$$

$$\mathrm{x}\left(\frac{\sin\left(N\Delta\beta L/2\right)}{\sin\left(\Delta\beta L/2\right)}\right)^2,$$

where α is the attenuation constant, L is the actual fiber length, N is the number of transmitted channels and $\Delta\beta$ is as calculated above. Analysis shows that the effect of FWM is reduced whenever a dispersion is available and $\eta \to 0$ for a channel spacing of 100 GHz.

V. Optimization of the Maximum Power per Channel

Two types of fiber have been used to carry out the simulation studies – Fujikura Ltd. SMF and LEAF. The SMF chosen is designed to work within the wavelength range 1310-1600 nm and its parameters are as follows: $\alpha = 0.21$ dB/km at 1550 nm; MFD = 10.4 μm ($A_{eff} = 85\ \mu m^2$) and $D_{ch} = 17$ ps/nm.km at 1550 nm. The parameters of the chosen LEAF are as follows: $\alpha = 0.22$ dB/km, MFD = 9.6 μm ($A_{eff} = 72\ \mu m^2$), $D_{ch} = 2.0$ to 6.0 ps/nm.km at 1530 – 1565 nm.

LEAF refer to the non-zero dispersion shifted fibers (NZ-DSF), whose MFD is smaller than that of the standard SMF and makes them more susceptible to non-linear effects. LEAF offer system designers a higher power-handling capability, improved optical signal-to-noise ratio, longer amplifier spacing and maximum WDM channel plan flexibility as compared with NZ-DSF. It should be pointed out that fibers of a large A_{eff} provide a critical performance advantage – the ability to uniformly reduce all non-linear effects.

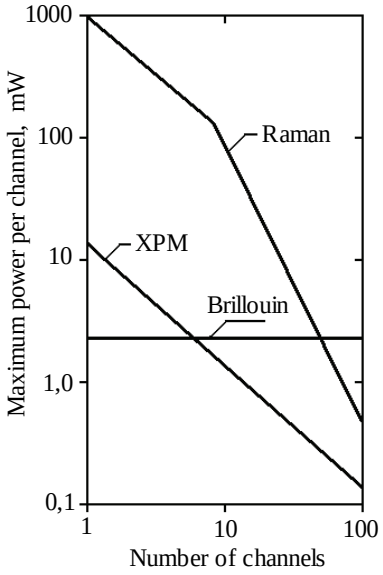


Fig. 4

The relation curves of the maximum power per channel of a WDM system investigated for various nonlinear effects are shown in Fig. 4. The experiment is based on the mathematical models here described and refers to a system applying an optical fiber of Fujikura Ltd LEAF type, the channel spacing being 200 GHz. The maximum power fed to the fiber in an 8-channel system is limited by a SBS power threshold of 3 dB, while limitations in a 16-channel system are due to the XPM effect.

VI. Conclusion

The dependencies here proposed enable the designers to determine the maximum launch power that provides the desired information quality if the fiber parameters and the number of transferred channels are given. The formulae have been developed into software products for the design of WDM CATV systems.

References

[1] Agrawal G. Nonlinear fiber optics. Academic press, Inc. University of Rochester, New Jork, 2001.

[2] Freeman R. Fiber-Optic Systems for telecommunications, John Wiley & Sons, Inc., New Jork, 2002.

[3] Baldwin T., S. Durand. IF Fiber Selection Criteria. ELVA Memorandum No. 32, Version 7, November, 2001

[4] Jordanova L., D. Dobrev. Influence of dispersion and non-linear effects in optical fiber on the parameters of CATV system. Proceedings of the XXXIX International Scientific Conference ICEST, Bitola, 2004, Volume 1, pp. 199 - 202.

[5] Fujikura Ltd. Optical fiber. Product information. 2003.

Analysis of Blocking and Clipping for TDM in Satellite Communication Systems

Sarhan M. Musa[1] and Nader F. Mir[2]

Abstract **– This paper presents the analysis of *time-division multiplexing* (TDM) applied to satellite communication systems. We evaluate multiplexing systems in which the number of input sources is greater than the number of available channels. We present analysis and simulation of blocking and clipping probabilities for TDM technology. For the blocking situation in synchronous TDM, we investigate the blocking probability and the average number of busy channels that can be delivered. For the clipping in statistical TDM, we examine the clipping probability and the expected number of busy channels that can be delivered.**

Keywords **– Time-division multiplexing (TDM), Satellite communications, Blocking and clipping probabilities.**

I. INTRODUCTION

The technology of satellite communications can support fixed and wireless data, voice, and video communications, the Internet connections, and enterprise networking. Satellite communication successfully uses continuously transmitted signal of Time-Division Multiplexing (TDM) as in the outbound (downlink) for an improvement of transmission quality instead of Orthogonal Frequency Division Multiplexing (OFDM), because of the linearity requirement on the power amplifier. TDMs are used in satellite networks for maximum transmission capacity of a high bandwidth line. Multiplexing allows many communication sources to transmit data over a single physical line.

With a typical time-division multiplexing, users take turns in a predefined fashion, each one periodically getting the entire bandwidth for a portion of the total scanning time. Given *n* inputs, time is divided into frames, and each frame is further subdivided into time slots, or channels. Each channel is allocated to one input. This type of multiplexing can be used only for digital data. Packets arrive on *c* lines, and the multiplexer scans them, forming a frame with *c* channels on its outgoing link. In practice, packet size is variable. Thus, to multiplex variable-sized packets, additional hardware is needed for efficient scanning and synchronization. TDM can be either *synchronous* or *statistical* and it has been used with other techniques as solution for satellite communications networks such as TDMA (Time division Multiple access), FDMA (Frequency division Multiple access), and PAMA (Pre-Assigned Multiple Access). There are a number of multiplexing methods used in satellite communication systems. One of the commonly used methods used in such systems is the TDM technology as in [1- 4]. Blocking and statistical probabilities are applied for TDM multiplexing as in [5-14]. In this paper we will analysis and simulate the blocking and statistical probabilities of TDM applied to satellite communications systems.

II. MULTIPLEXING MODELS AND RESULTS

Figure 1 shows a satellite communication system using TDM technology in interactive and sending/receiving different applications based on the importance of response time. The TDM used in the outbound link between the source (sender or host) and the user (receiver). A time-division multiplexing (TDM) system is a high-speed data stream scheme which acts at layer 1 (physical layer) of OSI model and at the layer 4 (network interface) of TCP/IP model. In TDMs Technology, users take turns in a predefined way, each one periodically getting the entire bandwidth for a portion of the total scanning time. The input sources, s, is divided into frames, and each frame is subdivided into time slots (channels), c, where each channel is allocated to one input as in Fig. 2. Packets arrive on s lines, and the multiplexer scans them, forming a frame with c channels on its outgoing link. There are two different types of TDMs to deal with the different ways in which channels of frames use could be allocated, Synchronous and statistical (asynchronous).

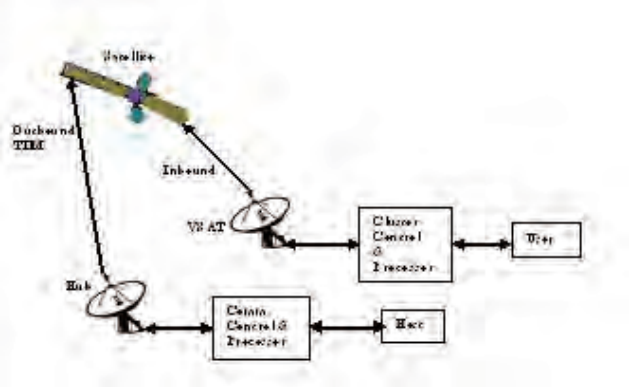


Fig. 1 Satellite communication System with TDM

[1]Sarhan M. Musa is with the Faculty of Engineering Technology Department, Prairie View A&M University, Texas, USA E-mail: smmusa@pvamu.edu

[2] Nader F. Mir is with the Faculty of Department of Electrical Engineering, San Jose State University, CA, USA.E-mail: nmir@sjsu.edu

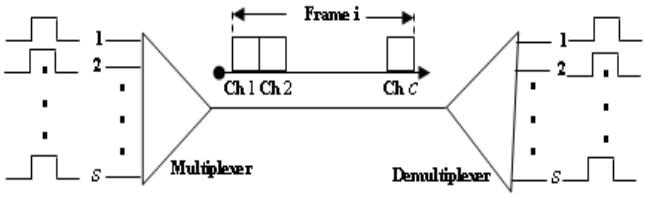


Fig. 2 A Time-division multiplexer (TDM) with s inputs and c channels

A. Synchronous TDM

In synchronous TDM, a frame is divided in a fixed–sized channels and channels are allocated to input sources in a fixed way. The QoS of synchronous TDM is base on how its transmission system is set up. For example, the multiplexer is inefficient when the number of users is greater than the available channels. This is true since the multiplexer scans all input sources lines without exceptions and the scanning time for each input source line (each connected to a user) is reallocated; as well as this time for a particular input source line is not altered by the system control. The scanner should stay on that input source line, whether or not there is data for scanning within that time slot. A synchronous TDM can also be programmed to produce same-sized frames, the lack of data in any channel potentially creates changes to average bit rate on the ongoing link.

To analyze a synchronous multiplexer, let t_a and t_d be the mean time for active input source and the mean time for idle input source respectively. Assuming that values of t_a and t_d are random and exponentially distributed. Also, consider a TDM with number of requesting input sources, s, is greater than available channels, c, where $s > c$, the TDM will react by blocking. The unassigned input sources are not transmitted and therefore remain inactive. The probability that an input source is active, ρ, can be obtained by $\rho = \dfrac{t_a}{t_a + t_d}$.

Let $P_s(j)$ be the probability of j different inputs out of s are active

$$P_s(j) = \binom{s}{j} \rho^j (1-\rho)^{s-j}, \qquad (1)$$

where $1 \le j \le s$. We know that $\sum_{j=0}^{c} p_s(j)$ can never be equal to 1 and in fact we must have $\sum_{j=0}^{s} p_s(j) = 1$. This can lead to normalization of $P_s(j)$ over c available channels. Thus, probability of j different output of c available channels $P_c(j)$ is given by

$$P_c(j) = \frac{\binom{s}{j}\left(\dfrac{\rho}{1-\rho}\right)^j}{\sum_{i=0}^{c}\binom{s}{i}\left(\dfrac{\rho}{1-\rho}\right)^i}, \qquad (2)$$

where $0 \le j \le c$ and $0 \le i \le c$. The blocking probability $P_s(c)$ can be obtained when $j = c$.

$$P_s(c) = \frac{\binom{s}{c}\left(\dfrac{t_a}{t_d}\right)^c}{\sum_{i=0}^{c}\binom{s}{i}\left(\dfrac{t_a}{t_d}\right)^i}, \qquad (3)$$

where in general $0 \le i \le c$.

Figure 3 shows the blocking probability for the fixed number of sources (s =10) and different numbers of channels ($c = 2$, $c = 5$, and $c = 10$). The blocking probability clearly rises with the increased utilization, ρ, for all three cases; and also it is higher when a fewer number of channels, c, is available.

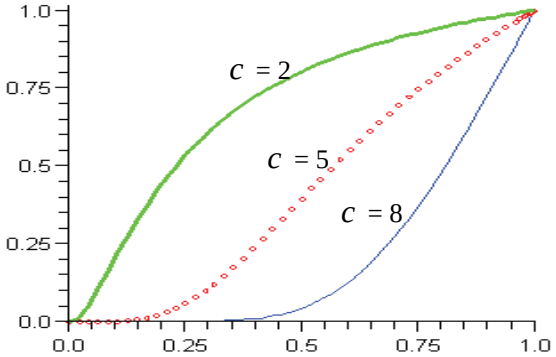


Fig. 3 Comparison of blocking probability, $P_s(c)$ vs. probability of active input source, ρ, with $0 \le \rho \le 1$, for s =10, c = 2, 5, and 8.

Then, we can calculate the expected number of busy channels for the multiplexer, $E_c(b)$, by

$$E_c(b) = \frac{\sum_{j=1}^{c} j\binom{s}{j}\left(\dfrac{t_a}{t_d}\right)^j}{\sum_{i=0}^{c}\binom{s}{i}\left(\dfrac{t_a}{t_d}\right)^i}, \qquad (4)$$

where $1 \le j \le c$, $0 \le i \le c$ and $\left(\dfrac{\rho}{1-\rho}\right) = \left(\dfrac{t_a}{t_d}\right)$.

Figure 4 shows the expected number of busy (unavailable) channels for fixed number of sources ($s = 10$) with different numbers of channels (*C=2*, *C=5*, and *C=8*). The expected number of busy channels of for *C=5* is higher than the one for *C=2* channels and lower than *C=8.*

.

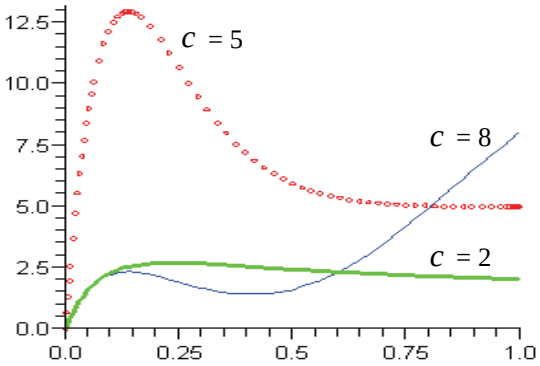


Fig. 4 Comparison of expected number of busy available channels, $E_c(b)$, vs. probability of active input source, ρ, where $0 \le \rho \le 1$ for $s = 10$, c = 2, 5, and 8

B. Statistical TDM

Statistical TDM method has high efficiency because a frame's time slots are dynamically allocated, based on demand and it removes all the empty slots on a frame. But, it is difficult to give a grantee QoS, because the requirement that additional overhead be attached to each outgoing channel. This additional data is needed because each channel must carry information about which input source line it belongs to. The frame length is available not only because of different channel sizes but also because of the possible absence of some channels.

We consider that t_a and t_d as random and exponentially distributed. Also, consider a TDM with number of requesting input sources, s, is greater than available channels, c, where $s > c$, the TDM will reacts by clipping; the unassigned input sources are partially transmitted.

If more than c inputs channels are active, we can dynamically choose c out of s active sources and temporarily block other sources. In this temporariy blocking, the source is forced to clip or lose data for a short period of time, where the amount of data lost depends on t_a, t_d, s, and c, but the source may return to a scanning scenario if a channel becomes free. This method maximizes the use of the common transmission line and offers a method of using the multiplexer bandwidth in silence mode.

The clipping probability, $P_i(l)$, or the probability that an idle source finds at least c channels busy at the time it becomes active, can be calculated by considering all s sources minus 1 (the examining source)

$$P_i(l) = \sum_{i=c}^{s-1} \binom{s-1}{i} \rho^i (1-\rho)^{s-1-i}, \qquad (5)$$

where $c \le i \le s-1$.

Figure 5 shows the clipping probability for fixed number of sources (s =10) and different number of channels (c = 2, c = 5, and c = 10). The clipping probability of 5 channels has the highest clipping probability compared to 2 and 8 channels.

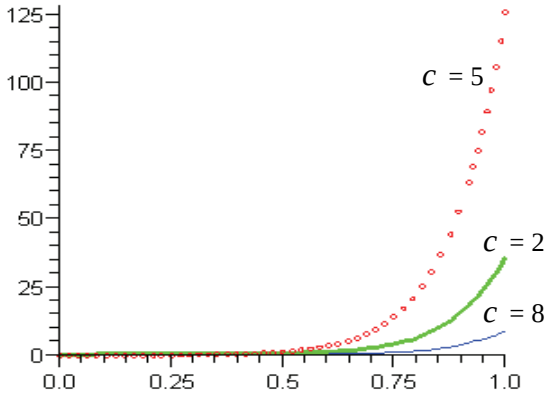


Fig. 5 Comparison of clipping probability, $P_i(l)$, vs. probability of active input source, ρ, where $0 \le \rho \le 1$, for s =10, c = 2, 5 and 8

Clearly, the average number of used channels, $A_c(u)$, is given by

$$A_c(u) = \frac{\sum_{j=1}^{c} j \binom{s}{j} \left(\frac{t_a}{t_d}\right)^j}{\sum_{i=0}^{c} \binom{s}{i} \left(\frac{t_a}{t_d}\right)^i}, \qquad (6)$$

where $0 \le i \le c$, and $1 \le j \le c$. Figure 6 shows the average number of used channels for fixed number of sources (s =10) and different number of channels (c = 2, c = 5, and c = 10). The average number of used channels of 8 channels has the highest average number of used channels compared to the ones for 2 and 5 channels.

Thus, the average number of busy channels, $A_c(b)$, is expressed by

$$A_c(b) = \frac{\sum_{j=1}^{c} j \binom{s}{j} \left(\frac{t_a}{t_d}\right)^j}{\sum_{i=0}^{c} \binom{s}{i} \left(\frac{t_a}{t_d}\right)^i} + c \sum_{j=c+1}^{s} \binom{s}{j} \rho^j (1-\rho)^{s-j}, \qquad (7)$$

where $0 \le i \le c, 1 \le j \le c$, and $c+1 \le j \le s$.

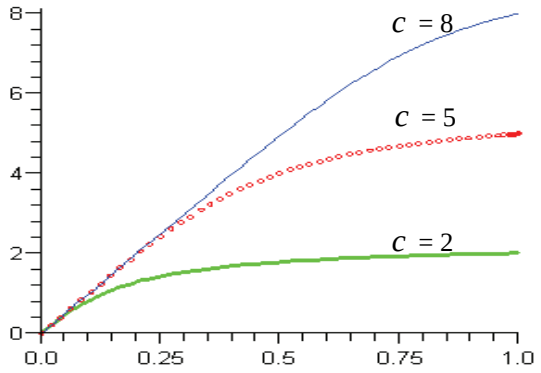


Fig. 6 Comparison of average number of used channels, $A_c(u)$,vs. probability of active input source, ρ , where $0 \le \rho \le 1$, for s =10, c = 2, 5, and 8

Figure 7 shows the average number of busy channels for fixed number of sources (s =10) and different numbers of channels (c = 2, c = 5, and c = 8). The average number of busy channels for c = 8 has the highest value compared to the ones for 2 and 5 channels.

III. CONCLUSION

This paper presented the analysis of Time Division Multiplexing (TDM) applied to satellite communications system when input sources are greater than available channels. The analysis and simulation of blocking and clipping for TDM was successfully investigated. For the blocking in synchronous TDM, we investigated the blocking probability and its average number of busy channels that could be delivered. For the Clipping in statistical TDM, we examined the clipping probability and its expected number of busy channels that could be delivered.

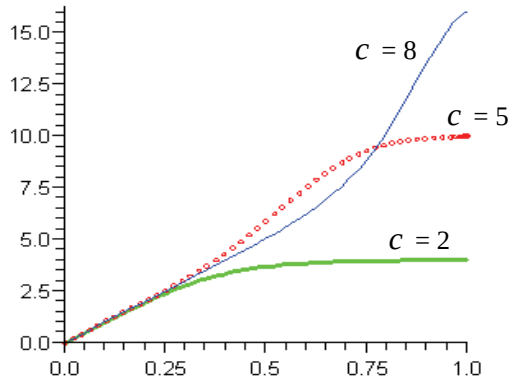


Fig. 7 Comparison of average number of busy channels, $A_c(b)$, vs. probability of active input source, ρ , where $0 \le \rho \le 1$, for s =10, c = 2, 5, and 8.

REFERENCES

[1] O. R. Ramires, "Use of TDM/PAMA satellite based communication as an option for the aeronautical communication", *IEEE National Aerospace and Electronics Conference*, pp. 683-686, Oct. 2000.

[2] Y. Chaehag, "TDM framing for gap filler operation in satellite digital multimedia broadcasting System A", *IEEE 59th Vehicular Technology Conference*, Vol. 5, pp. 2782 – 2786, May 2004.

[3] K. Zhang, and S. Hryckiewicz," An integrated approach for IP networking over the wideband Gapfiller satellite", *IEEE Military Communications Conference*, pp. 1556-1561, 2004.

[4] K. Eng, and A. Acampora ,"Fundamental Conditions Governing TDM Switching Assignments in Terrestrial and Satellite Networks", *IEEE Transactionson Communications*,Vol. 35, No. 7, pp. 755-761, Jul 1987.

[5] W. K. Lai, Y. J. Jin, H. W. Chen, and C. Y. Pan, "Channel assignment for initial and handoff calls to improve the call-completion probability", *IEEE Transactions on Vehicular Technology*, Vol. 52, Issue 4, pp. 876 – 890, July 2003.

[6] J. W. Chong; B. C. Jung;, and D. K. Sung," Statistical multiplexing-based hybrid FH-OFDMA system for OFDM-based UWB indoor radio access networks", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 54, No. 4, pp. 1793-1801, June 2006.

[7] A.I. Elwalid, and D.Mitra,"Statistical multiplexing with loss priorities in rate-based congestion control of high-speed networks", *IEEE Transactions on Communications*, Vol. 42, No.11, pp. 2989-3002, Nov. 1994.

[8] M. Balakrishnan, R. Cohen, E. Fert, and G. Keesman,"Benefits of statistical multiplexing in multi- program broadcasting", *International Broadcasting Convention*, pp. 560-565, Sept. 1997.

[9] J. Liebeherr, S. Patek, and E. Yilmaz, "Tradeoffs in designing networks with end-to-end statistical QoS guarantees", *Eighth International Workshop on Quality of Service*, pp. 221-230, June 2000.

[10] Z. L. Zhang, J. Kurose, J.D Salehi, and D. Towsley, "Smoothing, statistical multiplexing, and call admission control for stored video", *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 6, pp. 1148-1166, Aug. 1997.

[11] W. C. La and S. q. Li, "Statistical multiplexing and buffer sharing in multimedia high-speed networks: a frequency-domain perspective", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 3, pp. 382-396, June 1997.

[12] A. Iera, A. Molinaro, S. Marano, and D. Mignolo, "Statistical multiplexing of heterogeneous traffic classes in ATM-satellite based networks", IEEE *Wireless Communications and Networking Conference*, vol.1, pp. 174-178, Sept. 1999.

[13] L. Tancevski and I. Andonovic,"Block multiplexing codes for incoherent asynchronous all-optical CDMA using ladder network correlators", *IEE Optoelectronics*, Vol. 142, No. 3, pp. 125-131, June 1995.

[14] M. Zafer, and E. Modiano, "Blocking probability and channel assignment in wireless networks", *IEEE Transactions on Wireless Communications*, Vol. 5, No. 4, pp. 869-879, April 2006.

Influence of Optical Fiber Nonlinear Effects in HFC Television Networks with WDM Multiplexing

Krasen Angelov[1], Kiril Koitchev[2] and Stanimir Sadinov[3]

***Abstract* – The modern hybrid cable TV networks use for a backbone network high speed fiber-optical rings. Different nonlinear modulation effects have serious influence on the transmitted signals in the optical fiber as well as nonlinear effects due to lightwave scattering. These effects are directly related to the power of the transmitted signal and therefore they cause limitation in the launched in the optical fiber top permissible power. The nonlinear effects are main limiting factor for the transmission rate of digital information on optical fibers. This paper will deal with the analysis of the optical fiber nonlinear effects and an evaluation of their influence on the transmitted signal.**

***Keywords* – SBS, SRS, SPM, FWM, HFC**

I. INTRODUCTION

The fiber optical cables as a transmission environment are very widely applied as a backbone network for signal transmission in hybrid fiber-coaxial (HFC) cable television networks. The main purpose is transmitting the signals with carrier-to-noise ratio (*CNR*) bigger than *46dB*, composite second order distortions (*CSO*) not worse than *-62dBm*, and composite third order distortions (*CTB*) not worse than *-65dBm*.

In wavelength division multiplexing (WDM) system a multiplexing is realized when transmitting a few optical carrier wavelength in one optical flow and demultiplexing when receiving the flow. The signal gain on the transmission channel is realized by EDFA optical amplifier. The gain factor of about *20* to *25 dB* of the amplifier allows losses compensation input by the passive elements of the multiplexer and demultiplexer, including many other devices [1, 2, 3].

In order to achieve longer transmission distances or to provide more fiber splits, higher optical power needs to be launched into the optical fiber. In this situation the nonlinear interactions between the optical signal and the optical system start making sense.

The optical fiber nonlinear effects could be divided into two main groups:

– nonlinear effects related to light scattering – stimulated Brillouin scattering (SBS) and stimulated Raman scattering (SRS);

– nonlinear effects related with variation of the reflection index- they are a result of the dependence of refraction index of optical power intensity – self- phase modulation (SPM), the cross-phase modulation (XPM) and the four wave mixing (FWM).

The main problem with the nonlinear distortions is that they are always related to the power level of the transmitted signal, therefore they cannot be removed at the point of signal receiving without its parameters being changed.

In this paper an analysis of the optical fiber nonlinear effects will be made and an evaluation of their influence on the characteristics of the transmission medium and signals transmission.

II. NONLINEAR EFFECTS INFLUENCE RELATED WITH LIGHT SCATTERING

A. Stimulated Brillouin Scattering Effect

The stimulated Brillouin scattering (SBS) effect converts the transmitted optical signal in the fiber to a backward-scattered signal, and thus limits the maximum optical power that can be launched into the single-mode fiber (SMF). The SBS process generates a backward-propagating light beam at a lower frequency. Above the so-called SBS power threshold, the optical power of the backward-scattered signal increases exponentially, leading to significant optical power losses. The injected optical power level at which this increase happens, is defined as SBS threshold. Based on Smith's condition [4], the SBS power threshold is defined when the backward-scattered power is equal to the injected power, which for a uniform fiber is given by [5]

$$P_{SBS} = \frac{21\alpha A_{eff}}{\left(1-e^{\alpha L}\right) g_B}\left(1+\frac{\Delta\nu_L}{\Delta\nu_B}\right), \qquad (1)$$

where A_{eff} is the effective fiber core area, α – fiber loss coefficient, L – total fiber length, g_B – peak Brillouin gain and is given by

$$g_B = \frac{2\pi K n^7 p_{12}^2}{c\lambda^2 \rho_0 V_a \Delta\nu_B}, \qquad (2)$$

where n is the fiber refraction index, ρ – material density, V_a – acoustic velocity, p_{12} – elasto-optic coefficient, K – accounts for the optical field polarization, $\Delta\nu_L$ – laser

[1]Krasen K. Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@mail.bg

[2]Kiril R. Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

[3]Stanimir M. Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

linewidth, $\Delta\nu_B$ – spontaneous Brillouin linewidth of the fiber. For silica-based fibers, the Brillouin gain bandwidth is about *20MHz* at *1550nm*.

From Eq. (1) is viewed that the SBS threshold depends on the fiber length where the dependence is conversely proportional to the effective fiber length L_{eff} ($L_{eff}=(1-e^{\alpha L})/\alpha$) and exponential to the total fiber length L (Fig.1).

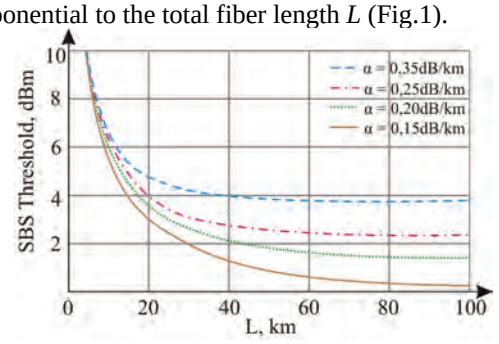


Fig. 1. The dependence of the SBS threshold level of the total fiber length for different fiber loss coefficient

Eqs. (1) and (2) indicate that the SBS threshold is directly proportional to $\Delta\nu_B$. For a standard single-mode fiber, $g_B=2.10^{-9}cm/W$ and the SBS threshold without modulation is approximately *6dBm* (*4mW*) at *1550nm* and *9dBm* (*8mW*) at *1310nm*. In addition to the dependence of the SBS threshold on the fiber type and uniformity along its length, the SBS threshold depends on the specific optical system requirements such as modulation format, symbol rate, and modulation type (i.e., direct or external). At the same time it does not depend on the WDM channel number.

As injected optical power is increased above *+6dBm*, the SBS scattered power dramatically picks up and the *CNR* rapidly degrades. The SBS also includes *CSO* and *CTB* distortion degradations in externally modulated *1550nm* laser transmitters, which are typically used due to the chromatic dispersion degradation. The *CSO* and *CTB* distortions can be approximated by [6]

$$CSO_{(dB)} = 10\lg\left\{N_{CSO}\left[\frac{(1-\sigma)m}{4\sigma}\right]^2\right\}, \quad (3)$$

$$CTB_{(dB)} = 10\lg\left\{N_{CTB}\left[\frac{3(1-\sigma)m^2}{16\sigma}\right]^2\right\}, \quad (4)$$

where $\sigma=(1-R_{BS}/0{,}85)/2$ is the fractional transmission coefficient for the single-mode fiber that was empirically estimated from the ratio of the backward-scattering power to the injected power (R_{BS}), m – modulation index, N_{CSO} and N_{CTB} are the *CSO* and *CTB* product counts.

To maintain the *CSO* distortion below *-62dBc*, the backward-scattered power should be less than *-20dB*, corresponding to a SBS threshold power of about *14dBm*, to achieve the necessary phase-modulation index.

The usual signal levels in CATV at wavelength *1550nm* often cause SBS effect because usually they range *8 - 14 dBm*, i.e. over the SBS threshold.

One of the most used methods for SBS effect reduction (i.c. increase of SBS threshold) is to apply a low frequency dithering of the optical frequency of the laser transmitter by modulating the laser bias current resulting in a broader effective laser line-width. The SBS threshold power increase can be obtained from

$$\Delta P_{SBS}^{th}(\Delta\nu_D) = 10\lg\left(1+\frac{\Delta\nu_D}{\Delta\nu_B}\right). \quad (5)$$

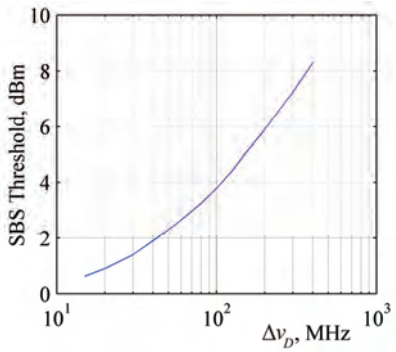


Fig. 2. SBS threshold versus the time average effective laser linewidth $\Delta\nu_D$

Fig. 2 shows the calculated from Eq. (5) SBS threshold increase versus the time average effective laser linewidth $\Delta\nu_D$. This method can easily provide an SBS threshold above *+20dBm* by allowing a frequency excursion above *1GHz*, and requires only a low-power laser current modulation.

It is necessary to note that the SBS threshold depends on the number of the EDFA optical amplifiers used; as well as knowing that with the increase of their number the power threshold decreases. For a system with optical amplifiers the SBS threshold power will be

$$P_{SBS\,N,[dBm]} = P_{SBS,[dBm]} - 10\lg N. \quad (6)$$

B. Stimulated Raman scattering effect

The stimulated Raman scattering (SRS) effect makes considerably lesser problems compared to the stimulated Brillouin scattering effect. This effect causes signal degrading only when the optical power level is high. In WDM systems the influence of this kind of light scattering is composed by lightwave scattering and its specter mixing in the long-wave area, where a redistribution of the optical power from short-wave to long-wave channels is realized.

The SRS threshold is much bigger (with about *3* levels) compared to SBS. The minimum SRS threshold power is given by [4]:

$$P_{SRS} \approx \frac{16\alpha K A_{eff}}{(1-e^{\alpha L})g_R}, \quad (7)$$

where A_{eff} is the effective fiber core area, α – fiber loss coefficient, L – total fiber length, g_R – peak Raman gain, зависещо от честотата на излъчване на сигнала и от свойствата на материала, K – accounts for the optical field polarization.

The dependence of the SRS threshold level on the fiber length for fibers with different coefficient of optical losses is shown on Fig.3.

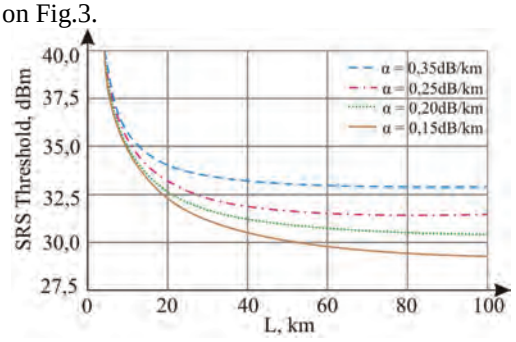


Fig. 3. The dependence of the SRS threshold level on the fiber length for fibers with different coefficient of optical losses

The power redistribution from short-waves channels to long-waves a result of the SRS effect leads to signal-to-noise ratio decrease for short-waves channels. As a result the total system's transmission capacity is limited, recognizing limitation of the total WDM channel number, channel spacing along wavelength, total system length and average launched power.

III. NONLINEAR EFFECTS INFLUENCE RELATED WITH VARIATION OF THE REFLECTION INDEX

A. Self-Phase Modulation Effect

The self-phase-modulation (SPM) effect occurs due to the interaction between the fiber's chromatic dispersion and the modulated optical spectrum of the propagating signal in the fiber. *CSO* and *CTB* distortions are generated when high-power (below the Brillouin threshold) optical signals are transmitted through a dispersive nonlinear optical fiber [7, 8 3]. The SPM effect named since the lightwave signal propagating through the fiber modulates its own phase.

The self-phase modulation leads to transmitted impulse expansion and temporary signal expansion or compression. In WDM systems with very small channel spacing, the spectral expansion caused by SPM influence, make result in interference between neighboring channels.

The analytical decision for the resulting SPM *CSO* distortions has the following view [9]

$$CSO_{SPM}(\Omega) = N_{CSO}\left[\frac{mP_0}{2A_{eff}}\left(\frac{L-L_{eff}}{\alpha}\right)\beta_2 k n_2 \Omega^2\right]^2, \quad (8)$$

where N_{CSO} is the *CSO* product count, m – optical modulation index per channel, P_0 – average optical power, A_{eff} – effective fiber core area, L – total fiber length, L_{eff} – effective interaction length, α – fiber loss coefficient, $ß_2$=–$D\lambda^2/2\pi c$, D – dispersion coefficient of the fiber, n_2 – nonlinear refraction index of the fiber, Ω^2 – subcarrier frequency.

Since the resulting *CTB* distortion caused by SPM was more than *30dB* smaller than the *CSO* distortion, it is neglected.

B. Cross Phase-Modulation Effect

The cross phase- modulation (EXP) is quite similar to SPM but is completely typical for WDM systems [9]. XPM is a result of a refraction index variation of the fiber when lightwave intensity increases.

XPM effect increases the nonlinear phase shift with *2N*, where *N* is the number of the working optical channels in the optical fiber. At this situation a bigger dependence on fiber dispersion compared to SPM is viewed.

A XPM decrease could be achieved by using optical fibers with large effective fiber core area, and, if it is possible the level of the channel optical power to be decreased.

C. Four Wave Mixing Effect

The four wave mixing (FWM) occurs only in multichannel light-wave systems and by its character it is similar to *CTB* distortions. It can completely destroy the WDM system.

Every time when three or more signals are propagated along the fiber, a four wave mixing might be expected. These three lightwave signals ω_i, ω_j, ω_k, generate a forth signal ω_{ijk}, obeying the ratio:

$$\omega_{ijk} = \omega_i + \omega_j + \omega_k . \quad (9)$$

The FWM level is sensible to the following system characteristics:

- increase of the optical power in the channel;
- increase of channel number;
- channel spacing decrease.

For the WDM system with channel number *N*, the total number of originating as a result of the FWM activity frequencies is

$$N_\Sigma = \frac{N^2(N-1)}{2}. \quad (10)$$

For example for a four channel WDM system the number of the additional channels is *24*, and for eight channel system *– 224*.

The FWM level significantly decreases in systems with channel spacing *100MHz*.

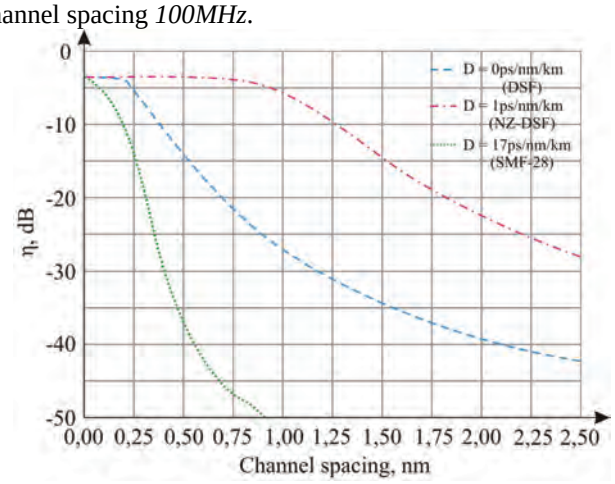


Fig. 4. FWM efficiency according channel spacing with different amounts of dispersion coefficient *D* of the fiber

In Fig.4 the FWM efficiency according channel spacing with different amounts of dispersion *D* coefficient of the fiber is shown.

The power level of the forth wave depends on multiple factors and can be written as

$$P_4(L) = \eta\left(\frac{1024\pi^6}{n^4\lambda^2c^2}\right)\left(\frac{L_{eff}}{A_{eff}}\right)^2 (DX_3)^2 P_1P_2P_3e^{(-\alpha L)}, \quad (11)$$

where η is the four-wave mixing efficiency, L – total fiber length, L_{eff} – effective interaction length, λ – wavelength, c – light speed, DX_3 – degenerating factor depending on channel spacing and chromatic dispersion of the fiber, α – fiber loss coefficient. The four-wave mixing efficiency η could be deducted as

$$\eta = \frac{P_4(L,k)}{P_4(L,k=0)} = \frac{\alpha^2}{\alpha^2+k^2}\left[1+\frac{4e^{-\alpha L}\sin^2(kL)}{\left(1-e^{-\alpha L}\right)^2}\right], \quad (12)$$

where k is a coefficient depending on the channel spacing and the condition of polarization sequence.

FWS decreases with the decrease of the absolute value of the chromatic dispersion. Moreover when the composite harmonics come to the working channel frequencies, a parametric interference originate, which might bring to an increase or to a decrease of the working impulse amplitude depending on the phase ratio among the working signal and the side-bands signals. The parametric losses cause close of the eye-diagram at the receiver outlet, resulting in worsening the BER level.

IV. Conclusion

The various nonlinear effects in an optical fiber limit the upper bound of the injected optical power in the optical part of HFC access network, which sets a limit to signal-to-noise ratio, and therefore a maximum transmission capacity of the optical system.

One of the possibilities for decreasing SBS and SRS influence is by decreasing the optical power of the channel below their threshold levels. But it is not a good decision, because it needs a using additional EDFA amplifier which worsens the threshold levels. Also the carrier-to-noise ratio is decreased.

The mostly used method for effects influence decrease is increasing the specter of the laser linewidth. Therefore laser sources with external modulation or continues wave are used, which also save the RIN level in its acceptable limits (*-155dB\Hz*).

The nonlinear effects mostly depend on the type of the optical fiber used. The threshold levels SBS and SRS could be changed depending on the type of the optical fiber. For DSF fibers (G.653) the SBS threshold is several times smaller than for systems with standard single-mode fibers (G.652). It could be concluded that this is true for all nonlinear effects.

At WDM systems it is suitable to use NZ-DSF fiber (G.655). In this fiber a limited chromatic dispersion is kept in working range. This leads to a decrease of the undesired influence of other nonlinear effects – self-phase modulation and cross-phase modulation. This type of fiber also reduces the effect of the four-wave mixing.

Acknowledgement

This paper has been sponsored under the auspices of the "Planning of Multimedia Telecommunications Networks with Traffic and QoS Management" project – a part of the "SCIENTIFIC INVESTIGATIONS" Fund, contract ВУ-ТН-105/2005.

References

[1] K. Angelov, K. Koitchev, S. Sadinov, An Investigation of Noise Influences in Optical Transmitters and Receivers in Cable TV Networks, ICEST 2006, Proceedings of Papers, pp.102-105, Sofia, Bulgaria, 2006.

[2] K. Koitchev, K. Angelov, S. Sadinov, An Effectiveness Investigation of Erbium-Doped Fiber Amplifiers for Cable TV Networks in presence of Noise, ICEST 2006, Proceedings of Papers, pp.106-108, Sofia, Bulgaria, 2006.

[3] R. Freeman, Fiber-Optic Systems for Telecommunications, John Wiley & Sons, New York, 2002.

[4] R. G. Smith, Optical Power Handling Capacity of Low-Loss Optical Fibers as Determined by Stimulated Raman and Brillouin Scattering, Applied Optics, vol.11, pp.2489-2494, 1972.

[5] X. Mao, R. W. Tkach, A. R. Charaplyvy, R. M. Jopson, and R. M. Derosier, Stimulated Brillouin Threshold Dependence an Fiber Type and Uniformity, IEEE Photonics Technology Letters, vol.4, pp.66-69, 1992.

[6] M. R. Philips, K. L. Sweeney, Distortion by Stimulated Brillouin Scattering Effect in Analog Video Lightwave Systems, OSA TOPS, System Technologies, vol.12, pp.182-185, 1997.

[7] D. A. Atlas, Fiber Induced Distortion and Phase Noise to Intensity Noise Conversion in Externally Modulated CATV Systems, NCTA Technical Papers, pp.289-293, 1996.

[8] C. Desem, Composite Second Order Distortion at 1550nm, IEEE Electronics Letters, vol.30, pp.2055-2056, 1994.

[9] M. C. Wu, C. H. Wang, W. I. Way, CSO Distortions Due to the Combined Effects of Self- and External-Phase Modulations in Long Distance at 1550nm AM-CATV Systems, IEEE Photonics Technology Letters, vol.11, pp.718-720, 1999.

Adaptive Packet Selection Algorithm for Bluetooth Data Packets

Radosveta Sokullu[1] and Engin Karatepe[1]

***Abstract* –In this paper we present a Matlab-Simulink model for performance evaluation of Bluetooth DHx data packets. The simulation model developed takes into account channel quality, distance and interference in the Bluetooth frequency band. We propose a new adaptive algorithm for packet selection based on the number of retransmissions until successful packet reception under current environment conditions. The results obtained using the simulation model have proved the increase in performance provided by this algorithm**

***Keywords* – Modeling Bluetooth ACL packets, adaptive packet selection, Matlab-Simulink**

I. INTRODUCTION

Even though Bluetooth is already a well established short range wireless technology due to its very low price, easy implementation, versatility and great user convenience it provides, increasing its throughput and performance is still an open research issue. Simulation models play a great role in this process because they allow multifaceted parameter and performance evaluation, testing and optimization because they allow for reducing the price and risks of hardware implementation. Research groups have dealt with simulating voice packets, investigating performance under different channel conditions and suggested algorithms for choosing best performing packet types.

In [2] the authors have tested Bluetooth performance for different BER and according to the results provided estimation for most suitable packet type selection in a certain BER range. In [3] the authors suggest a dynamic selection scheme based on measurements for SNR. In [4] a detailed OPNET Bluetooth model is developed and the mutual interference of Bluetooth and 802.11b system is evaluated. In [6] behavior of different packets is examined a link management scheme is suggested based on a set of rules analytically derived from channel quality measurements. In another article [9] the authors give an analytical expression for the throughput as a function of the number of frequency channels used for frequency hopping and the duration of the packet types. [7] suggests two schemes for dynamic transmission scheduling based on classifying the frequencies as "good" or "bad" depending on the registered packet loss, called BIAS (Bluetooth Interference Aware Scheduling) and AFH (Adaptive Frequency Hopping). Even though they give quite good results they are quite complex to realize. Another drawback of most models is the fact they concentrate either on the physical or link layer only. Matlab-Simulink is an integrated simulating environment which allows both very precise physical layer description as well as system level description. This is also supported by the fact that the new version of Matlab 7.0 has included and example of simulating Bluetooth SCL packets. In a previous work [8] we have developed and tested simulation models for the DMx packets. In this work we provide models for the DHx packets and incorporate all data packet types in a Matlab-Simulink model.Further on we simulate the performance of a new, simple to implement adaptive packet selection algorithm based on counting the number of retransmissions required until a packet is received successfully under different environment conditions including both interference and variable distance between the devices.

From here on the paper is organized as follows. In the next part we present a brief review of Bluetooth and the structure of different packet types. In part 3 we discuss our simulation model and in part 4 the suggested adaptive packet selection algorithm. Following that we present simulation results and conclude the paper.

II. BLUETOOTH ACL PACKETS STRUCTURE

Bluetooth is a short distance (10 to 100 m) wireless technology in the ISM (2.40-248 GHz) band based on FHSS. Communication is carried out according to a master slave model. A master and max of 7 active slaves form a piconet using TDD and transmission is divided into time slots of 625 μs duration. Information, both voice and data is carried in packets. The Bluetooth standard specifies different structure for voice, data, and control packets. Data packets are divided into DMx and DHx types. The DM packets carry control and data information protected by FEC and were envisaged for noisy channels, while the DHx type packets carry more data without FEC. All packets are divided into tree subtypes – DM1/DH1, DM3/DH3 and DM5/DH5 depending on the number of time slots they occupy. After the master addresses a slave in one slot the active slave has to respond in the next slot with either a 1-slot (DM1/DH1), 3-slot (DM3/DH3) or 5-slot (DM5/DH5). The selection of the suitable packet to be used is very important and highly depends on the channel and interference at a certain moment in time. DM1 carry least information but are suitable for worst conditions, while DH5 packets carry most information which is least protected

[1]Radosveta Sokullu and Engin Karatepe are faculty members of the EEED, Faculty of Engineering, Ege University, Izmir, Turkey, E-mail: radosveta.sokullu@ege.edu.tr; engine.karatepe@ege.edu.tr

against errors. Selecting the most suitable packet type is a major factor in ensuring highest throughput under specific channel conditions.

As in general the practically achievable data rates are quite low (despite the theoretical 1 Mbps a practical max is about 700 kbps) implementing a suitable packet type selection algorithm is of utmost importance. Such an algorithm has to be easy to implement and simple to operate in order to meet the performance requirements. The decision of our simulating tools, models and environment has been governed by the need to precisely reflect the specifics of both physical and data link layer functions. In all previous work known to us either one or other aspect has been favored depending on the specifics of the simulation tool used.

A. Structure of the ACL packet types

The Asynchronous Connectionless Link (ACL) is a point-to-point connection between a master and active slaves in the piconet. An automatic Repeat Request (ARQ) is applied to ACL packets and they are repeated until a positive ACK is received. The positive ACK is piggy-backed in the header if the returned packet. The ARQN is set to 1 or 0 depending on whether or not the previous packet was successfully received. Packets consist of three major parts: a 72-bit Access Code, a 54-bit Header and a variable length Payload, which has is own header data part and CRC. The first bit from the packet that is transmitted is the LSB. In our simulation this is represented by the first bit of the Access code. The payload field varies depending on the packet type from 0 to a max of 2745 bits. The difference between the user information carrying possibilities are clearly seen if we compare the payload fields: for DM1 the data payload is max 17 bytes, compared to DH1 – 27 bytes; for DM3, which occupies 3 slots, it has a value 121 bytes compared to 183 for DH3; and for DM5 – 224 versus 339 for the DH3 packet type.

Achievable throughput in kbps for different packet types are given in TABLE 1 and an example of the DH5 packet structure and comparison with the DM5 packet type is given in Fig..1 As defined in the Bluetooth standard, [1], the DH1 payload itself consists of 1 byte header, 27 bytes data and 2 bytes CRC. Accordingly the DH3 and DH5 payloads carry 183 and 339 bytes data, accompanied by a 2 byte header and 2 byte CRC.

B. Simulation model description

We have based our simulation model on the IEEE 802.15.1 and the Bluetooth standard ver.1.2 and realized it using Matlab ver.7.0. The model consists of 4 major building blocks: master, slave, channel and interference. Fig. 2 gives the main, top level window of the simulation program.

In general each Bluetooth device can function either as a master or a slave depending on the creation of the piconet.

TABLE I

ACL PACKET TYPES USING ARQ

Packet Type	Duration (slots)	Payload data length (Bytes)	Hamming FEC code	Peak throughput
DM1	1	17	Yes	108.8 kbps
DH1	1	27	No	172.8 kbps
DM3	3	121	Yes	387.2 kbps
DH3	3	183	No	585.6 kbps
DM5	5	224	Yes	477.9 kbps
DH5	5	339	No	723.2 kbps

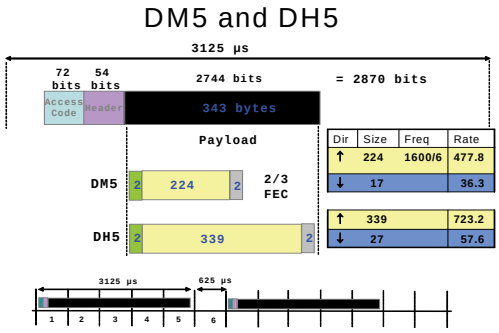


Fig. 1: Structure and comparison of the DM5 and DH5 packet types

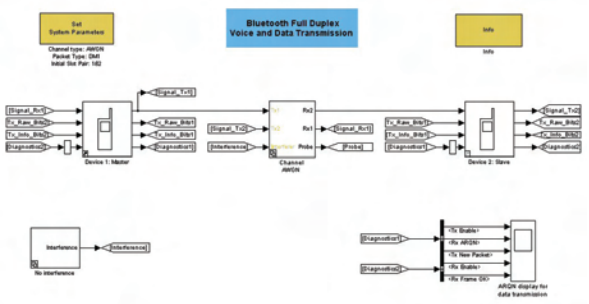


Fig.s 2: Main program window

The detailed functional structure of our master/slave simulation model is presented in Fig. 3. The Bluetooth hopping pattern is implemented using a pseudo-random generator.

The transmitter block shown in Fig. 3 realizes the functions of formatting the data payload, including its header and CRC and incorporating it with the required Access Code and Packet Header. Another major function incorporated here is the selection of a suitable packet type. The decision mechanism has to be embedded in the transmitter part. We assume asymmetric transmission. Besides this the transmitter has an encoder part, while the receiver has a filter part responsible for decoding the incoming data based on the understanding of the packet type. The receiver part contains also the performance metrics evaluation block.

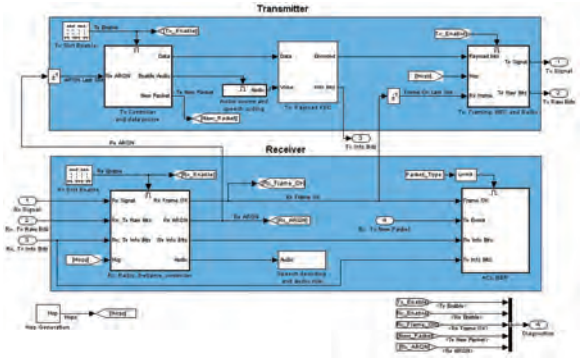


Fig. 3: Structure of the simulation model of the master/slave

Referring to the OSI Data Link Layer model the functions of data error and flow control are realized using Simulink. An example of the realized data flow model and error correction mechanism is given in Fig. 4.

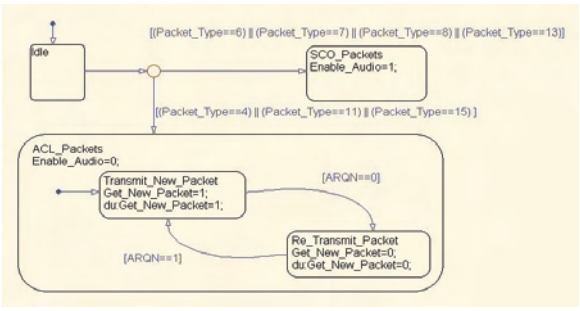


Fig. 4: ACL packets data flow control diagram

As seen in Fig. 4 this is an asynchronous type transmission and an ARQN=1 requests a new packet to be transmitted while in the opposite case the packet has to be retransmitted. In more detail, at the receiver side the first operation is checking the received packet for errors in the CRC. If the output is a Frame_OK the rest of the packet parts are checked and finally the packet is accepted. In the opposite case the packet is discarded and the ARQN is set to 0. A more detailed receiver side control flowchart is represented in Fig. 5. It should be noted that there are differences in the master and slave decision mechanism which for simplicity are not included here. For the channel and interference block is concerned we have considered an AWGN channel and 802.11b interference as well as the effect of distance between the master slave pair and interference sources.

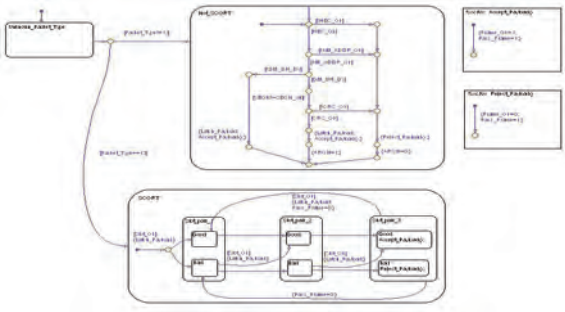


Fig. 5: Details of ACL packets data flow control

C. Simulation results

In Fig. 6 we present results for the throughput as a function of EsNo achieved with different DH1, DH3 and DH5 packet types for a noiseless channel using model presented in detail in the previous subsection. It has been verified against the analytical model from [3].

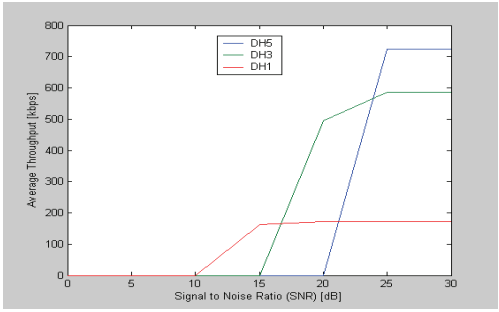


Fig. 6: Throughput against Es/No for DH1, DH3 and DH5

III. Adaptive Packet Selection Algorithm

Main objective of the ACL packets in the Bluetooth technology is carrying asynchronous data with min error and max throughput independent of the environment with a limited predetermined power levels. Frequency hopping allows for continuous change of the transmission frequency which leads to reducing effects of interference in certain bands. In some work elaborate frequency selection schemes have been proposed to increase mitigate the interference. [4,6]. The disadvantage of these schemes is their complexity in realization and the fact that they require major changes in the radio layer which is undesired by chip producers. Including FEC is an additional measure for ensuring min error rate. But it has also the disadvantage of reducing the user data rate. As a measure, specific to Bluetooth different packet types have defined with the aim to select the most suitable one for the given transmission conditions. The standard has left open the details of how these packets will be selected. Most of the suggested schemes are based on measurements and calculations using the LQI which means that they still require changes in the lower layer of the stack.

Based on the simulation model we have developed we have investigated the relation between the number of re-

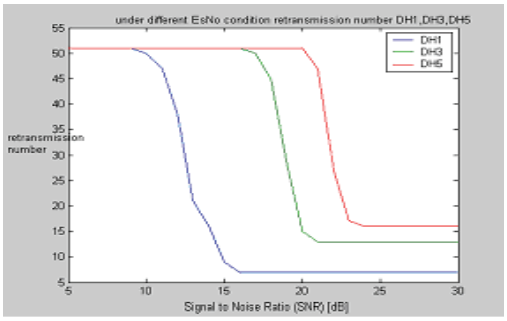


Fig. 7: Required number of retransmissions for different DH packet types.

transmission required until the successful reception of a packet in different noise conditions. The results are given in Fig. 7.

As it can be seen there are distinctive thresholds for different packets and different EsNo. Based on this we have simulated an algorithm for adaptively changing the packet type when a threshold is reached. The decision mechanism is incorporated in the sender side. The state diagram is presented in Fig. 8.

Fig. 2. Figure example

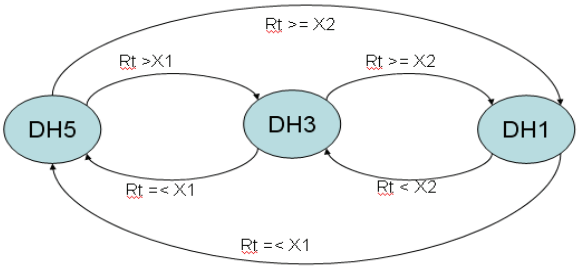


Fig. 8: State diagram of the suggested algorithm

The decision mechanism is based on two parameters X1 and X2 which reflect the number of retransmissions recorded by the sender. Results (Fig. 9) show gradual reduction of the average throughput with worsening channel conditions and easy adaptability to changes independent of the factors causing them.

In another work [10] work we have implemented this algorithm using only changes in the software stack and preliminary results have supported the simulation results presented above. In our experimental setup we have also observed that allowing fragmentation and encapsulation of large used frames using simultaneously two packet types, for example DH3 and DH5 increases the throughput with some10% more.

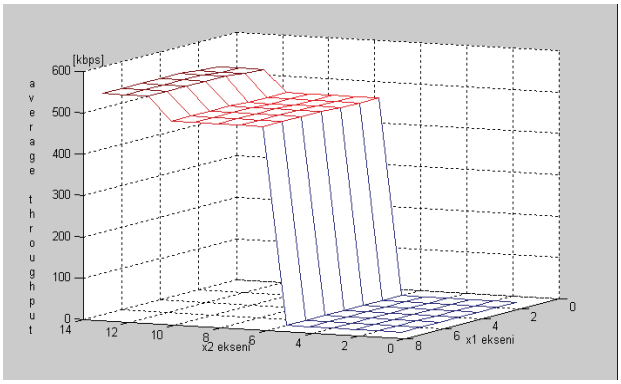


Fig. 9: Results of throughput changes as a function of Es/No using adaptive packet selection.

IV. . CONCLUSION

In this paper we describe the model developed for simulating Bluetooth ACL data packets and link management functions using Matlab. We discuss its details and advantages compared to other existing simulation models. Further on we describe and show simulation results for an adaptive selection algorithm for ACL packets. It is fast and simple to realize in software without requiring major changes in the Bluetooth radio. We believe that our work is a contribution to the efforts of the research community to increase with min possible changes in the hardware and firmware of Bluetooth chips the data throughput under changing environmental conditions.

ACKNOWLEDGEMENT

The authors want to express their thanks to all the members of the EgeBlue research group at the Dept. of EEE for their valuable support and contributions.

REFERENCES

[1] Bluetooth SIG, *"Bluetooth Specification" ver.1.0 p. Core and ver 1.2 p.Profiles*, Dec. 1999.

[2] Dogan K., Gurel G., Kamci A.K. and Korpeoglu I. 2004, *TUBITAK research projects EEEAG-101E025, EEAG-101E048, and EEAG-103E014*, Bilkent University, Dept. of EEE

[3] Valenti MC, Robert M., "Custom Coding, Adaptive Rate Control and Distributed Detection for Bluetooth", 2002, In *VTC Proceedings*, 918-922

[4] Golmie N., Van Dyke R.E.,Tonnerre A. and Rebala O., "Interference Evaluation of Bluetooth and IEEE 802.11b Systems", 2003, *Wireless Networks*, 201-21

[5] Valenti M., Robert M., Reed J., "On the Throughput of Bluetooth Data Transmissions", 2002, In the *Proceedings of the Wireless Communications and Networking Conference*, 119-123

[6] Ju M-C., Park C-H., Hong D-K., Youn K-J., Cho J.W., 2002 "Link Management Scheme of Bluetooth Based on Channel Quality Estimation", *Electronics Letters*, Vol 38, No, 15, 789-790

[7] Golmie N., Rebala O., "Bluetooth Adaptive Techniques to Mitigate Interference", 2003, In the Proceedings of GLOBECOM"03, Vol.1, pp.405-409

[8] Sokullu R, Özdemir Ö, Küçük T, Karatepe E, "Simulating DM1, DM3 ve DM5 Bluetooth Packet Types in Matlab", 2005, In the *Proceedings of the ITUSEM Symposium*, pp 109-114, 17-19 Nov. 2005

[9] Floren F., Stranne A., Etfors O., and Molin B-A.,"Throughput of Strongly Interferening Slow Frequency Hopping Networks", 2004, *IEEE Transactions on Communications*, July Vol.52, No.7, pp. 1152-1159

[10] Sokullu. R., Poyraz U., "CQDDR Based Bluetooth Throughput Improvement", 2006, *In the Proceedings of ELECO 2006*, pp. 116-121,

The Analysis of Rician PDF Integral Properties from the DSRC System Viewpoint

Aleksandra Mitić[1], Dejan Blagojević[2], Dimitrije Stefanović[3] and Veljkovic Stanimir[4]

***Abstract* – In this paper some statistical properties of DSRC structures, presented through the Rician fading channel model have been investigated. The approach to envelope like singular solution of the PDE, concerning the empirical conditions of PDF determining, enable to a better view on the physical process and nature of fading. The existence of singular solution could be used in the optimisation process of the *ITS* in sense of long-range prediction of the mobile fading envelope and hardware optimisation**

***Keywords* – DSCR. The Rician fading channel models, PDF, envelope, distribution, singular solutions.**

I. INTRODUCTION

The numerous statistical models of the wireless communications influenced by propagation character of the space have been developed [1-4] To provide valid performance of signal that are close to those in real situations, it is necessary to set up adequate channel model of these systems by the characteristic of the above mention statistical models. The signal envelopes and the signal phase at the receiver are the important parameters affected on the system performance. The PDF (probability distribution function) of these parameters has the different shape concerning the channel conditions, so the various types of fading could exists [5-8]. With these models not only average power level but also the maximum fluctuation of received signal can be predicted.

There are numerous services of ITS (intelligent traffic system) such as advanced toll collection, automatic vehicle control, and mobile computing and internet service are considered to be most important service in the future traffic development. Also, the *ITS* is in focus throughout the world as possible and unique measures to provide advanced and intelligent service [9-12]

The *ITS* services have been based on the DSCR (dedicated short range communication) telecommunication structure. However, along the dominant *LOS (line of sight)* component, the other numerous multipath signals are expected to arrive to antenna, so the Rician fading channel model is adequate for the *DSRC* system. Since the implement cost and high level of complexity of the user terminal is very critical point in the *ITS* services, statistical models and their characteristics, play important role.

The start point in these paper have been the numerical analyze of Rician PDF's, like particular solutions of corresponding differential equation for *K=const*., and *z=const*. in the case when we analyzed these distribution through the forms of curves family. On the other side, the envelope of curves family shows that initial differential equation has singular solution. The existence of singular solution could be used during the analysis of different conditions in the information chain with or without combiner. By the fact, that combined signal mostly treated as a deterministic process, the conclusion about singular solution could be use in the optimization process of ITS in sense of long-range prediction of the mobile fading envelope and hardware optimization in designed process of OBE (on board equipments) and RSE (road side equipments) units.

II. ANALYSIS:

The Rician fading channel models consider that the dominant wave can be a phasor sum of dominant signals, e.g. the LOS component and sum of multipath signals. This combined signal is then mostly treated as a deterministic (fully predictable) process. The wave components can also be subject to shadow attenuation.

The channel transfer function can be expressed as impulse response function $h(t)$:

$$h(t) = \delta(t) + \alpha e^{j\phi\delta(t)} \qquad [1]$$

The first item of the right side means LOS component and second presents the sum of multipath signal with amplitude α and phase ϕ. The amplitude of the multipath components has Rayleigh distribution with $E[\alpha^2] = 2\sigma^2$.

Let A^2 denote the power of LOS component, $2\sigma^2$ is a power of multipath components, then the Rician *K*-factor [3] *i.e.* the ratio of the LOS component signal power to the reflected interference power can be expressed as:

$$K = \frac{A^2}{E[\alpha^2]} \qquad [2]$$

[1]Aleksandra. Mitic is with the Faculty of electronic engineering Aleksandra Medevdeva 20 18000 Nish Serbia, E-mail: alekmi@elfak.ni.ac.yu

[2]Dejan I. Blagojevic is with the The High Technical School, Aleksandra Medevdeva 20 18000 Nish Serbia, E-mail: blagojevicd70@yahoo.com

[3]Dimitrije C. Stefanovic is with the Faculty of Faculty of electronic engineering Aleksandra Medevdeva 20 18000 Nish Serbia, E-mail vule@elfak.ni.ac.yu

[4]Veljkovic Stanimir is with the Faculty of electronic engineering Aleksandra Medevdeva 20 18000 Nish Serbia, E-mail

The starting assumptions about dominant line-of-sight component and a set of reflected waves, modeled as a phasor addition of signals with random amplitude and phase, lead to Gaussian Inphase and Quadrature component, and corresponding Rician amplitude. When the *LOS* signal component power tends to zero subsequently *K* tends to 0, the Rician PDF approaches to Rayleigh PDF. For large *A* the Rice distribution approaches a shifted Gaussian distribution centered at $z=\sigma$.The PDF dependence of the Rician factor *K* [3] could be expressed like:

$$p_z(z) = \frac{2z(K+1)}{\Omega} e^{-K-\frac{(K+1)z^2}{\Omega}} I_0\left(2z\sqrt{\frac{K(K+1)}{\Omega}}\right) \qquad [3]$$

where is Ω, the mean power of envelope, and ***z*** is signal level.

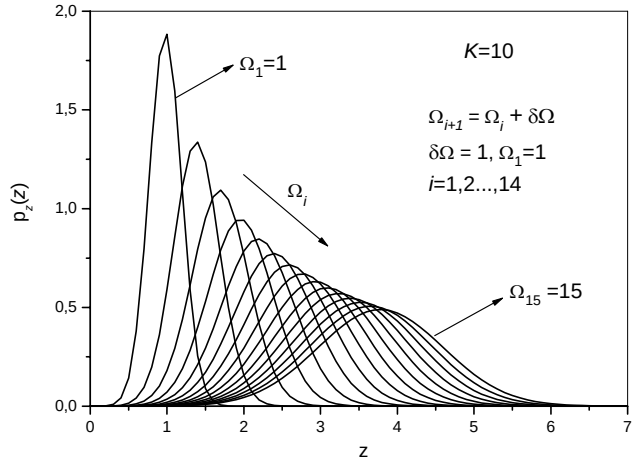


Fig. 1 The PDF dependance from z with parameter Ω, and Rician *K*-factor constant

We analyzed a PDF depenadance Fig 1. from the *z* variable with Ω like parameter and for constant value of Rician *K*-factor ./K=10, Ω=1÷15/. The PDF maximum decreasing with increasing the Ω. Also, it is obviouse that these distributions could be analyzed through the family of curves with Ω like the parameter of the family. We presented this family in log-log coordinates too Fig 2., and analysed the locations of the PDF's maximum for the different values of z in both systems.

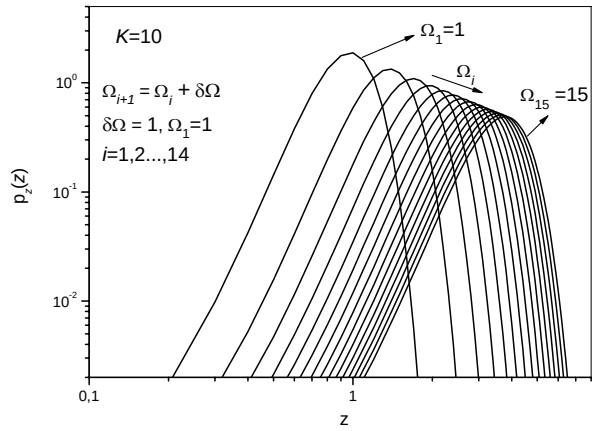


Fig. 2. The PDF dependence from z with Ω parameter and constant Rician *K*-factor in log-log coordinates

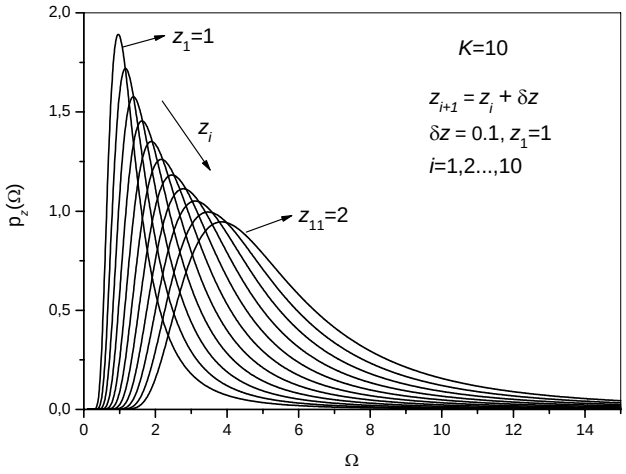


Fig. 3 The PDF dependance from Ω with parameter z, and K=10

We analyzed a PDF depenadance Fig 3. from the Ω variable with *z* .like parameter and for constat value of Rician *K*-factor ./K=10, z=1÷2/. The PDF maximum decreasing with increasing the *z*. Also it si obviouse that these distributions presents the family of curves with *z* like the parameter of the family. We presented this family in log-log coordinates too Fig 4, and analysed the locations of the PDF's maximim of the different values of Ω in both systems.

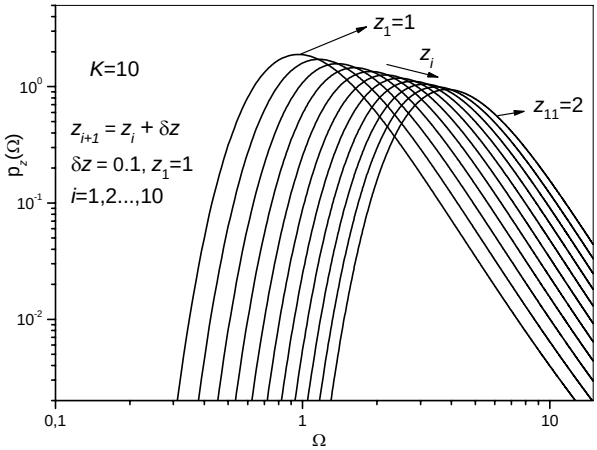


Fig. 4. The PDF dependance from z with parameter Ω, and and constant Rician *K*-factor in log-log coordinates

III. DISCUSSION

The maximum of PDF's are located on the envelope of the curve family. These envelopes could be approximated by exponential function Figs 1 and 3, or straight line in log-log coordinates Figs 2 and 4. The envelope existences enable to treat the PDF's like particular solutions of some differential equation, corresponding to a specific value of the equation free parameters. Also, numerical analysis Figs 2 and 4, confirms that beside the particular solutions, exist the singular solutions too, which present the above mention envelope of the analyzed curves family. We determined the envelope equations by graph-analytical methods. The envelope equations in the log-log coordinate for the parameters of the

curves families, Figs 2 and 4, have been approximated as (4) and (5) respectively:

$$\log(\max p_z(z)) = k_1 \log z + n_1 \quad (4)$$

$$\log(\max p_z(\Omega)) = k_2 \log \Omega + n_2 \quad (5)$$

The importance of the Rician distribution according the Shannon theory [13], made us to analyze the integral properties of Rician PDF from parameters, where are PDF treated, like we already mention, as particular solutions of first order PDE (partial differential equation) with the parameters of distribution like variables. Also, it means that the PDF could be analyzed like particular integral of differential equation for specific values of parameters. On the other side the curves family envelope shows that initial differential equation has singular solution, which existence could be used during the analysis of different conditions in the information chain with or without combiner.

The dependency of Ricean PDF versus z and Ω is presented by three-dimensional graph, as shown on Figure 5.

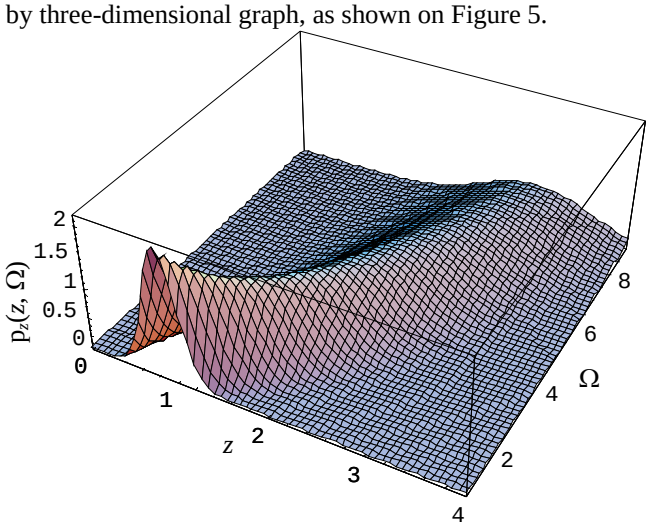


Fig. 5. The 3D view of Rician PDF

So the approach to envelope like singular solution of the PDF, concerning the empirical conditions of PDF determining, enable to a better view on the physical process and nature of fading. The fading analysis from the physical viewpoint needs a forming of particular differential equations with information entropy, which roles lie in relating conditions in one region to those in another, through a recursive process which allows us to progressively work out the solution from one place to another, a little bit at a time. The framework is prescribed by the differential equation, but the information which is to be relayed from one place to another is contained in the boundary conditions.

These remarks could be used in modeling of fading process, and the designed process of OBE and RSE hardware by providing very high date transfer rates in circumstances where minimizing latency in communication and isolating relatively small communication zones. The designed of the hardware for *OBE* and *RSE* units and many physical process and phenomena in high speed integrate circuits such as a peak noise voltage, varying load impedance, etc., based on one this approach, will result in the optimal relations between the system performance and design parameters in the various communications conditions as well as highway traffic conditions

IV. Conclusion

The importance of the Rician distribution according the Shannon theory, made us to analyze the integral properties of Rician PDF from parameters, where are PDF treated, like we already mention, as particular solutions of first order PDF (partial differential equation) with the parameters of distribution like variables. We determined the envelope equations by graph-analytical methods. The envelope equations in the log-log coordinates for the parameters of the These remarks could be used in modeling of fading process, and the designed process of OBE and RSE hardware by providing very high date transfer rates This approach could be use in the analysis of the other physical processes involved in the radio diffusion, optoelectronic and physics generally.

References

[1] T.Aulin., “A modified model for the fading signal at mobile radio channel”, IEEE Trans.Veh.Techn., Vol. 28, No.3, August 1979

[2] C. Loo., “A statistical models for fading channels with applications to digital transmission”, IEEE Transactions on vehicular technology, Vol. 34, No.3, pp 122-127, August 1985

[3] D.Drajic, Uvod u statisticku teoriju telekomunikacija, Akademska misoa, Beograd, 2003

[4] H.Sharam, H.Homayoun, “A propagation model for micro cellular mobile and personal radiocommunications.” Proc. IEEE PIMRC., vol.12, sept.1995, pp. 392-396

[5] M.K, Simon, M.S.,Alouni, “Digital telecommunication over fading channels”, A Unified approach to performance analysis, John Wiley&Sons, Inc.2000

[6] J.D., Parsons, “The mobile propagation channel”, John Wiley&Sons, Inc.2000

[7] C.Loo, J.S.Butterworth., “Land mobile satellite channel measurement and modeling”, Proceeding of IEEE, Vol. 86, No. 7, pp. 1442-1463, July 1998

[8] M. Patzold, F.laue., “Level-crossing rate and average duration of fades of deterministic simulation model for Rice fading channels”, IEEE Trans.Veh.Techn., Vol. 48, No.4, pp 1211-1129, July 1999

[9] M. Yasunaga, “ITS research and Development Activities in TAO, Japan, “Proc. of telecommunication Workshop of ITS,” May 2000.

[10] D. Kwon, Y.Hahm,I.Jeong, and D.Im, “CDMA Mobile System Test –bed and Field test“, ETRI Journal, vol. 19,no.3, Oct.1997, pp.259-280

[11] C. Wietfeld. “Performance evaluation of vehicles-roadside communication systems in Shadowing&Multipath fading environment“, IEEE Trans.Veh.Techn., vol VT-45,1995,pp.947-952

[12] Byung-Seub Lee ,et al. “Performance evaluation of the physical layer of the DSRC operating in 5.8 GHz frequency band”. ETRI Journal, vol 23. no 3, sept. 2001, pp 121-128

[13] C. E. Shannon, “A matematical theory of information“The Bell System Technical Journal, Vol,27,pp 379-423, July, 1948

This page intentionally left blank.

Some Integral Properties of Nakagami-*m* distribution

Hana Popovic[1], Aleksandra Mitic[2], Ivan Stefanovic[3] and Dimitrije Stefanovic[4]

***Abstract* – In this paper Nakagami fading channel model is described and some statistical characteristics of Nakagami-*m* distribution are analyzed. For analytical and numerical evaluation of system performance, the Nakagami probability density functions are analyzed like particular solutions of corresponding differential equation. The existence of singular solution is considered and analyzed under different conditions.**

***Keywords* – The Nakagami fading channel model, probability density function, Nakagami distribution, singular solutions.**

I. INTRODUCTION

Mobile communication is burdened with particular propagation aspects that make reliable wireless communication more difficult than fixed communication. The propagation characteristics change from place to place and, if the terminal moves, from time to time. The mobile radio channel is usually evaluated from statistical propagation models: no specific terrain data is considered and channel parameters are modeled as stochastic variables. Two mutually independent, multiplicative propagation phenomena can usually be distinguished: multipath fading and large-scale path loss (shadowing). As the result of multipath reception, the mobile antenna receives a large number of reflected and scattered waves. Because of wave cancellation effects, the instantaneous received power seen by a moving antenna becomes a random variable. Multipath propagation causes rapid fluctuations of the phase and amplitude of the signal if the vehicle moves over a distance in the order of a wave length or more. Multipath fading thus has a small-scale effect. The rapid fluctuations of the instantaneous received power due to multipath effects are usually described with Rayleigh, Rician or Nakagami-m model [1].

The Rician model is useful for modeling mobile wireless communication systems when the transmitted signal can travel to the receiver along a dominant line-of-sight (LOS) or direct path. The reflected paths those the signal travels are modeled with multipath Rayleigh fading channel. Rayleigh model is used in cases when direct LOS doesn't exist. Besides Rayleigh and Rician models, the distribution of the signal amplitude and power can be well described with Nakagami-*m* probability density function (pdf). Rician and Nakagami fading models are two generalizations of the model for Rayleigh fading. In the literature, the Nakagami model is often used for analytical simplicity in cases where Rician fading would be a more appropriate model. The Nakagami distribution matches some empirical data better than other models, but it is not always an appropriate approximation for Rician fading model. It has an essentially different behavior for deep fades, such that results on outage probabilities or error rates can differ by orders of magnitude [6].

The main point in this paper is the numerical analysis of Nakagami-m pdfs, as particular solutions of corresponding differential equation, for one varying parameter, while the others are set to constant values. On the other side, the envelope of family of pdf curves shows that the initial differential equation has the singular solution. The existence of singular solution could be used during the analysis of system performance.

II. NAKAGAMI FADING CHANNEL

The Nakagami fading model was initially proposed because it matched empirical results for short wave ionospheric propagation. The Nakagami distribution is described by pdf:

$$p_z(z) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m z^{2m-1} e^{-\frac{m}{\Omega}z^2}, \; z > 0, m \geq \frac{1}{2} \qquad (1)$$

where z is the received signal level, $\Gamma()$ is the gamma function, and m is an integer shape factor called fading figure. It is defined by [4]:

$$m = \frac{E^2[z]}{Var[z^2]} \qquad (2)$$

while Ω is defined as:

$$\Omega = E[z^2] \qquad (3)$$

The expressions for Rician fading are less convenient, mainly due to the occurrence of a Bessel function in the Rician pdf of received signal amplitude. Approximations by Nakagami distribution, with the simpler mathematical expressions, have become very popular, but it is not always an appropriate model.

The Rician and Nakagami models behave approximately equivalently near their main value. This observation has been used in many recent papers to advocate the Nakagami model as an approximation for situations where a Rician model would be more appropriate [10].

[1]Hana Z. Popovic is with High School of Electrical Engineering, Vojvode Stepe 283, 11000 Beograd, Serbia, E-mail: hanap@eunet.yu

[2] Aleksandra Mitic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: alekmi@elfak.ni.ac.yu

[3] Ivan D.Stefanovic. is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vule@elfak.ni.ac.yu

Dimitrije C.Stefanovic. is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vule@elfak.ni.ac.yu

In the special case that m=1, Rayleigh fading is recovered, while for larger m the spread of the signal strength is less, and the pdf converges to a delta function for increasing m. When using the Nakagami model to approximate the pdf of the power of a Rician fading channel, matching the first and second moments of the Rician and Nakagami pdfs gives:

$$K = \sqrt{m^2 - m} \Big/ \left(m - \sqrt{m^2 - m} \right) \qquad (4)$$

where is K Rician factor, defined as the ratio of LOS signal power to the random path signal power. The results are strikingly different for m>1. As the relation between K and m was based merely on the first and second moments, it is likely to be most accurate for values close to mean. The probability of deep fades differs for these two models, so an approximation the pdf of a Rician fading by a Nakagami pdf can be highly inaccurate.

III. NUMERICAL RESUTLTS

The graphic representation of numerical analysis of Nakagami pdf is given on following figures.

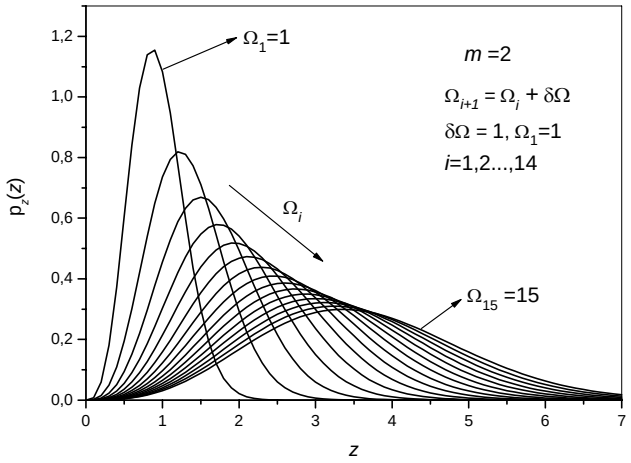


Fig. 1. The Nakagami pdf versus signal level z, for case m=2 with Ω taking values from 1 to 15

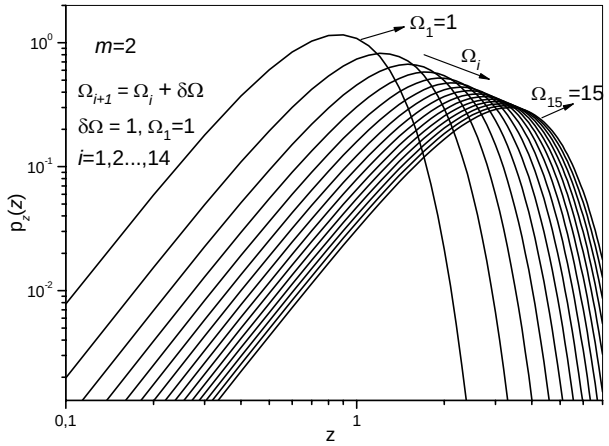


Fig. 2. The Nakagami pdf versus signal level z, in logarithmic scale, for case m=2 with Ω taking values from 1 to 15

It can be seen on Fig. 1. that with the increase of Ω the maximums of pdfs achieve smaller values for larger values of z. Also, the width of the curves is getting larger for larger values of Ω.

The same dependency is shown on Fig. 2. in logarithmic scale, where it can be seen that envelope of pdf curves family is a straight line.

The existence of the envelope of the family of pdf curves enables the pdfs to be regarded as particular solutions of some differential equation in further analysis.

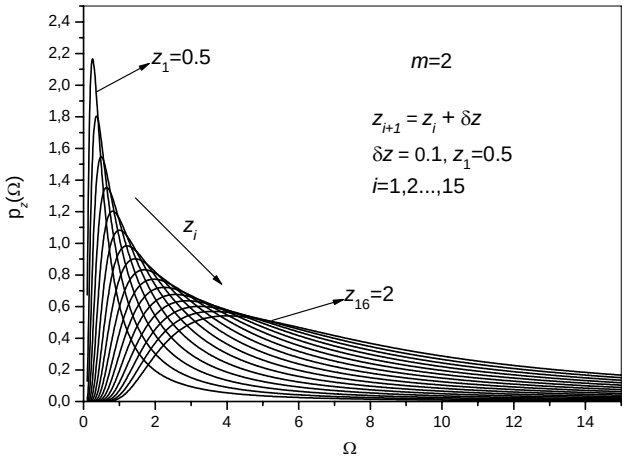


Fig. 3. The Nakagami pdf versus Ω, for case m=2 with z taking values from 0.5 to 2

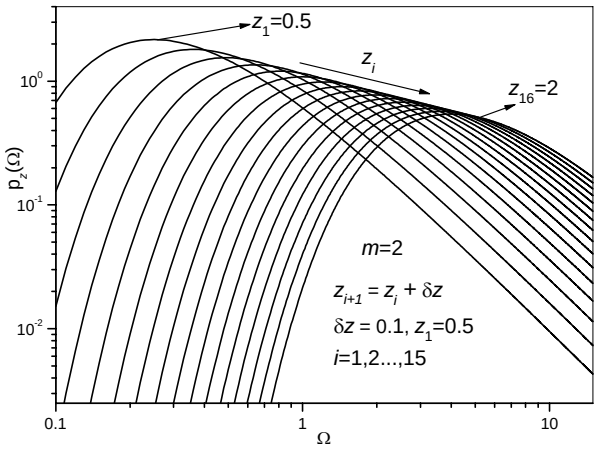


Fig. 4. The Nakagami pdf versusΩ, in logarithmic scale, for case m=2 with z taking values from 0.5 to 2

The Figs 3. and 4. show the families of pdf curves where z is taken as parameter. It can be seen that similar conclusions exist as in previous case. With the increase of z the maximums of pdfs achieve smaller values for larger values of Ω, and also, the width of the curves is getting larger for larger values of z. In logarithmic scale, it can be seen that envelope of pdf curves family is a straight line.

Analytical and numerical analysis shows that as well as particular solutions to the aforementioned differential equation, the singular solutions also exist, and they represent the envelopes of the analyzed families of pdf curves. Since these solutions can not be found through analysis of the initial

differential equation, they were found by grapho-analythical means. As the result, the solutions were found to be exponential functions, or in logarithmic scale, straight lines.

The dependency of Nakagami pdf versus z and Ω is presented by three-dimensional graph, as shown on Fig. 5.

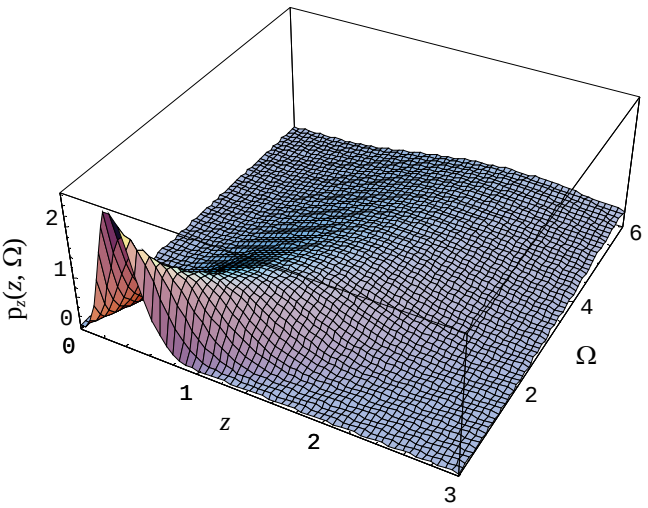


Fig. 5. A three-dimensional graph showing the dependency of Nakagami pdf versus z and Ω

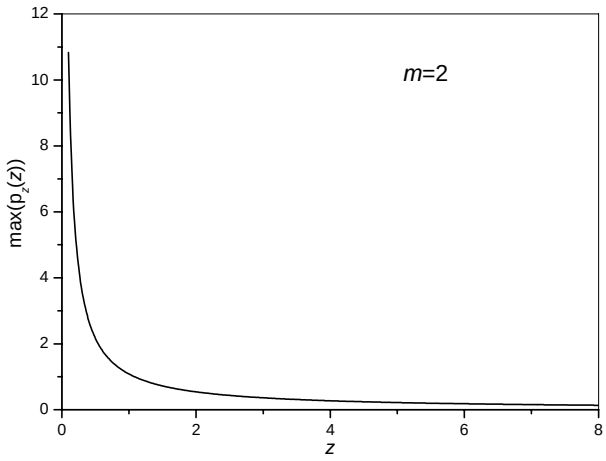


Fig. 6. The max Nakagami pdf versus z, for case m=2

The dependency of pdfs maximums versus z is shown on Fig. 6. That dependency can be described by equation:

$$\log(\max p_z(z)) = k \log z + n \tag{5}$$

In logarithmic scale, it is presented by a straight line. All of pdfs maximums rest on a straight line, defined by Eq. (5), where k=-1 and n=0.0345.

The dependency of pdfs maximums versus Ω is shown on Fig. 7. That dependency is also described by Eq. (5), but with different parameters: k=-0.5 and n=0.064. In logarithmic scale it is also presented by a straight line, which means that all of pdfs maximums rest on a straight line.

It can be shown that pdfs maximums versus z rest on envelope of pdf curves family given on Fig. 3. and the pdfs maximums versus Ω rest on envelope of pdf curves family given on Fig. 1.

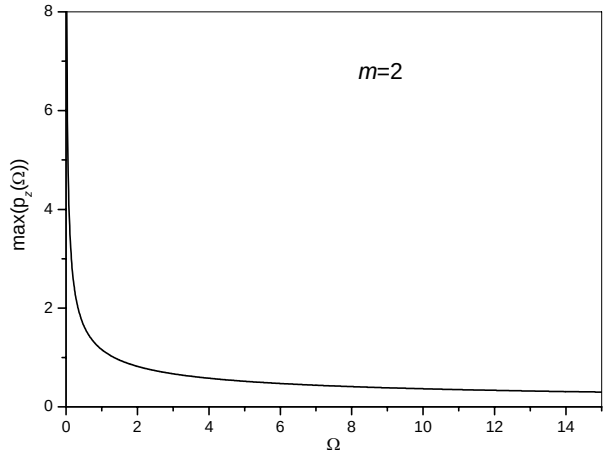


Fig. 7. The max Nakagami pdf versus Ω, for case m=2

Analytical and numerical analysis of integral properties of pdfs is very important for predicting the system performance. The theoretical analysis of performance (like Bit-Error-Rate and outage probability) for various modulation techniques, with the different pre-detection and post-detection diversity combining techniques, is based on analysis of properties of pdf.

This approach could also be used in the analysis of the other physical processes involved in the radio-diffusion, optoelectronic and physics generally.

IV. Conclusion

An advantage of the Nakagami distribution is that it can be reduced to the Rayleigh distribution and can model fading conditions more severe or less severe than those in Rayleigh case. In some situations, Nakagami model can be used as an approximation for Rician fading model. Analytical and numerical analysis of Nakagami pdf that describes the received signal envelope, is very important for predicting the performance of a modulation or coding scheme and is essential for the efficient evaluation and validation of system design [8].

Analytical and numerical analysis of Nakagami-m pdfs, as particular solutions of corresponding differential equation, for one varying parameter, while the others are set to constant values, shows the existence of singular solution, which could be used during the analysis of system performance.

References

[1] J. Proakis, “Digital Communications”, 3rd Ed., McGraw-Hill, 1999

[2] M. K. Simon, M. S. Alouni, “Digital Communications over Fading Channels: A Unified approach to performance analysis”, John Wiley, New York, 2000

[3] N. L. Johnson, S. Kotz, N. Balakrishnan, “Continuous Univariate Distributions”, Vol.1, New York: Wiley, 1994

[4] M. Nakagami, “The m-distribution, a General Formula of Intensity Distribution of Rapid Fading in Statistical

Methods in Radio Wave Propagation", W. G. Hoffman, Ed.Oxford, England: Pergamon, 1960

[5] M. C. Jeruchim, P. Balaban, K. S. Shanmugan, "Simulations of Communication Systems", New York: Plenum Press, 1992

[6] J. D. Parsons, "The Mobile Propagation Channel", John Wiley&Sons, Inc.2000

[7] R. H. Clarke, "A Statistical Theory of Mobile Radio Reception", Bell Systems Technical Journal, vol. 47, 1968, pp. 957-1000

[8] W. Y. C. Lee, "Mobile Cellular Communications" McGraw-Hill Book Co., New York, 1989

[9] R. Price, P. E. Green, "A Communication Technique for Multipath Channels", Proceedings of the IRE, March 1958, pp. 555-70

[10] N. Youssef, T. Munakata, M. Takeda, "Fade Statistics in Nakagami Fading Environments", Proc. IEEE 4th Int. Symp. Spread Spectrum Techniques & Applications, Mainz, Germany, pp. 1244-47, 1996

LAN Traffic Shaping Analyses in Metropolitan Area Networks

Rossitza. Iv. Goleva[1], Mariya At. Goleva[2], Dimitar K. Atamian[3] and Kostadin At. Golev[4]

Abstract - **This paper represents LAN traffic shaping analyses in routers that apply the three QoS management techniques – IntServ, DiffServ and RSVP. The results show queue management and packet stream shaping based on simulation of the mostly demanded services – VoIP, LAN emulation and transaction exchange.**

Keywords - **Packet network, IP, Quality of Service, shaping.**

I. INTRODUCTION

IP networks are becoming new complex communication facility nowadays. They are enough easy for implementation and with high demand of Quality of Service (QoS) support. Last 10 years there were strong talks concerning QoS level but the price of its support was very high. Recently most of the routers and gateways that are not among the cheapest production set support such functionality. In this paper we analyze the traffic shaping effect of the three mostly used techniques IntServ, DiffServ, RSVP. The analyses are made on the basis of the three popular services – VoIP, LAN emulation, transaction exchange [1], [2], [6], [7]. The shaping effect is estimated on all services. The presented results are mostly on the LAN traffic that is second low priority traffic in comparison to the real time traffic from VoiP and video over IP. The model uses queues and priorities specific for the IntServ, DiffSerf, RSVP. This fractional shaping phenomenon is important in small to medium Metropolitan Area Networks (MAN) that grows rapidly and new customers appear almost every day.

II. TRAFFIC SOURCES

Three types of traffic sources are assumed in an example metropolitan area network – Voice over IP, LAN emulation and transaction exchange. LAN traffic is lower priority in comparison to the VoIP traffic and with higher priority in comparison to the transaction traffic. The there services are mixed together with some assumptions. In Voice over IP (VoIP) service silence and talk intervals are exponentially distributed with equal mean values [1]. In some papers on-off model is proposed [6].

The limits for waiting times are calculated under consideration of end-to-end delay for every service [5]. Servicing times per packets are fixed on 100 Mbps line interface. Table I represents all the parameters for traffic sources in the model.

TABLE I
TRAFFIC SOURCES PARAMETERS

Parameter	VoIP	LAN	Transac-tions
Pear rate, packets per second	10	164	0
Mean call/ session duration, sec	180	20	10
Mean duration between calls/sessions, sec	360	10	15
Mean talk/ silence duration, sec	20	5	2
Distribution of call/series duration	Exp.	Exp.	Exp.
Maximal waiting time, sec	0.00072	0.6	1
Maximal number of waiting packets	210	1804	2
Traffic sources	5000	500	1500
Priorities	High	Medium	Low
Packet length, bytes	800	800	800

LAN emulation is specific with its sessions. Sessions are established for any Internet connections. Packet rate is higher in comparison to the VoIP. Session duration is low. The traffic source is behaving as on-off model with exponential duration of the silence and transmission intervals [5]. Transaction exchange is specific with few packets exchange. The service is not time demanding.

Number of traffic sources is taken from the typical image in a business area. Packets are taken to be long. In VoIP traffic 800 bytes carry up to 80 milliseconds voice. This means that quality voice can be transmitted only in the area using up to 2-3 hops. More precise investigation can be done with 200 bytes or less voice packets.

III. INTEGRATED SERVICES

Integrated Services (IntServ) is a complex technique often called protocol that ensures Quality of Service in IP networks. It is applied usually in access routers or gateways tried to serve packets from different services in a different ways depending on the quality requirements. IntServ classifies services into three main classes depending on the traffic requirements as elastic, tolerant real-time and intolerant real-time [2].

[1]Rossitza Iv. Goleva is with the Department of Telecommunications, Technical University of Sofia, Bulgaria, Kl. Ohridski blvd. 8, Sofia. 1756, Bulgaria, E-mail: rig@tu-sofia.bg

[2]Mariya At. Goleva is with Department of Communication Networks, University of Bremen, Germany, E-mail: mgoleva@gmail.com

[3]Dimitar K. Atamian is with Department of Telecommunications, Technical University of Sofia, Bulgaria, Kl. Ohridski blvd. 8, Sofia. 1756, Bulgaria, E-mail: dka@tu-sofia.bg

[4]Kostadin At. Golev is with Bianor Ltd., Bulgaria, E-mail: kotseto@gmail.com

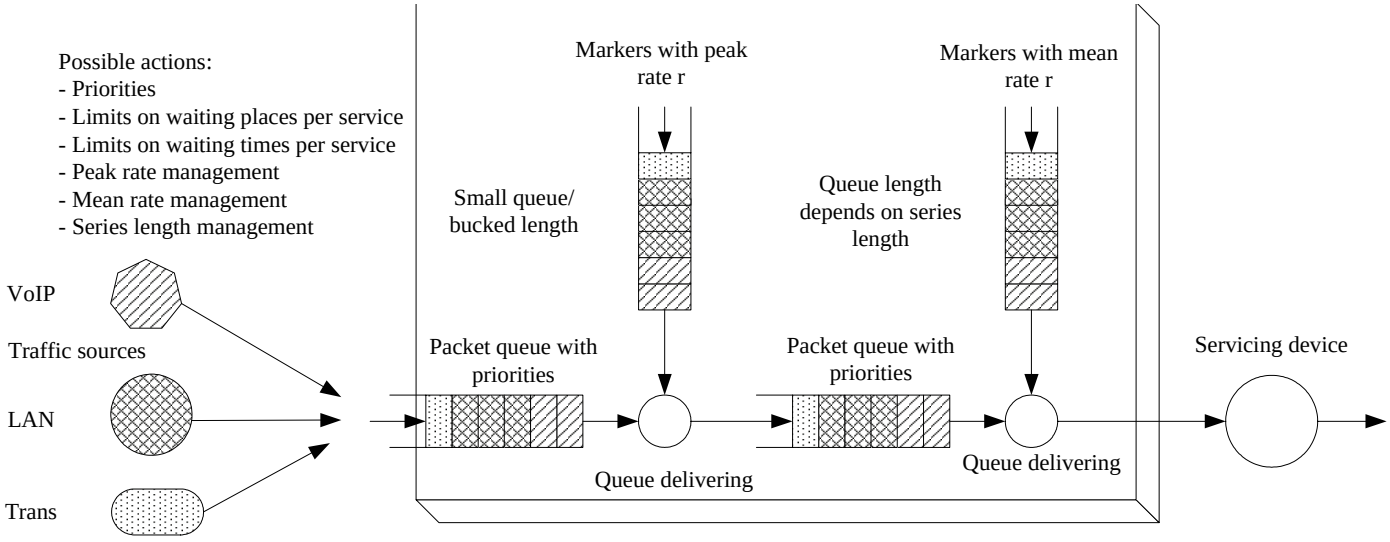


Fig. 1. Black box IntServ model approximation

Elastic applications are served in a “best effort” discipline [8]. They are served without any guarantee of quality level like transaction exchange. Tolerant real-time applications are delay sensitive and usually require high bandwidth. Token bucket model with peak rate control is a proper model for such traffic. LAN emulation is usually modeled this way. Some authors propose token bucket with series length and mean rate control for more accuracy (Fig. 1). Many authors propose the two token buckets to be connected in a cascade [7]. Intolerant real-time applications require low delays and almost guaranteed bandwidth. The model with two cascaded token bucket is compulsory for such traffic [1]. VoIP service is intolerant to the quality degradation service. IntServ simulation model is based on two cascaded token buckets that bound peak rate, series length and mean rate of the traffic (Fig. 1) [9]. The model from Fig. 1 is quite complicated for simulation. Due to this reason it is approximated as a black box that changes the characteristics of the data into output data in specific for IntServ way. As a result after approximation and few calculations it is easy to derive simpler model with one FIFO queue, priorities, fixed rate at the output and different limits for waiting times in the queue. This is the model that has been simulated further. Table II represents main data for model behavior.

IV. Differentiated Services

Differentiated Services (DiffServ) is another quality management technique that is more applicable for core networks. Due to its nature DiffServ applies its rules on aggregated traffic. After appropriate marking of the aggregated packets they are gathered in the way that is defined for their class. There are three main types of services we try to highlight in this paper [6]: premium service with low delay, low loss, guaranteed bandwidth like VoIP; assured service with less requirements to the delay and loss in comparison to the premium service like LAN emulation; olympic service with no time requirements at all like transaction exchange.

The model from Fig. 1 is simplified with DiffServ procedure in mind. Simple FIFO queue with priorities and limits for waiting times per service is derived. System parameters are shown on Table II.

V. RSVP

Resource Reservation Protocol (RSVP) is a technique useful for delay sensitive traffic like VoIP. Three types of services are identified for RSVP like: wildcard filter with maximal requirements for given interface like LAN emulation; shared explicit with maximal requirements for the interface taking into account called address (transaction exchange is modeled as shared explicit service); fixed filter with full reservation for quality sensitive services like VoIP (Table II).

TABLE II
Model characteristics.

Parameter	IntServ	DiffServ	RSVP
Queue length, packets	2016	1840	1840
VoIP queue length fraction, packets	210	200	200
LAN queue length fraction, packets	1804	1640	1640
Transaction queue length fraction, packets	2	2	2
Maximal waiting time for VoIP, sec	0,000716	0,0303	0,07508
Maximal waiting time for LAN, sec	0,6	0,27876	0,69
Maximal waiting time for transactions, sec	1	1	1
Priority for VoIP	Highest	Highest	Highest
Priority for LAN	Medium	Medium	Medium
Priority for transactions	Low	Low	Low

VI. RESULTS

Simulation is performed on C++ language. The pseudo exponential pseudo deterministic characteristics of the traffic sources are reached after usage of combination between many random generators [3], [4], [5]. The queue behavior is complex due to the priorities and limits on waiting times. Many parameters have been derived from the model like time and space loss probabilities, probabilities to wait for different types of traffic, probability distribution functions and probability density functions of the packets intervals, queue lengths, waiting times at many interface points in the model like output of the traffic sources, input and output of the queue. Statistical accuracy of the derived results is proven by Student criterion. IntServ, DiffServ and RSVP have different way to gather with packets and this influences the way they drop packets and shape them. On Fig. 2, Fig. 3 and Fig 4 probability density function of packet intervals at the input and output of the queue is shown for all services. It is interesting for shaping estimation. The effect of fast servicing in RSVP can be seen from Fig. 2. The delay variation of the packet intervals is becoming smoother and tends to constant value. Similar result is visible for IntServ on Fig 3. On Fig 4 shaping of the IntServ and DiffServ is seen again.

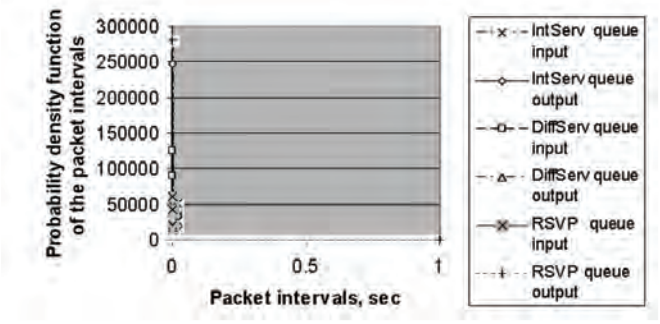


Fig. 2. Delay variation reduction in RSVP

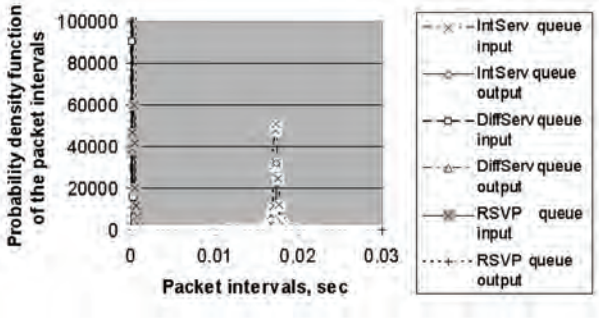


Fig 3. Delay variation reduction in IntServ

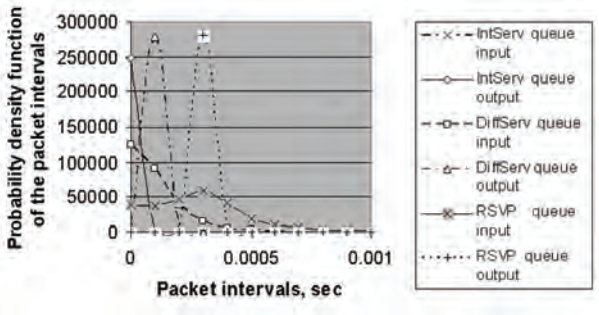


Fig 4. Delay variation reduction in IntServ and DiffServ

Fig. 5 and Fig. 6 represent probability density function of the packet intervals at input and output of the queue only for LAN traffic. The effect of low delay is visible for IntServ and DiffServ. The queue is full and the overall effect of traffic acceleration is compensated by bigger traffic delays for anothre type of traffic. This is visible from Fig. 7. The deterministic nature of the traffic is still kept.

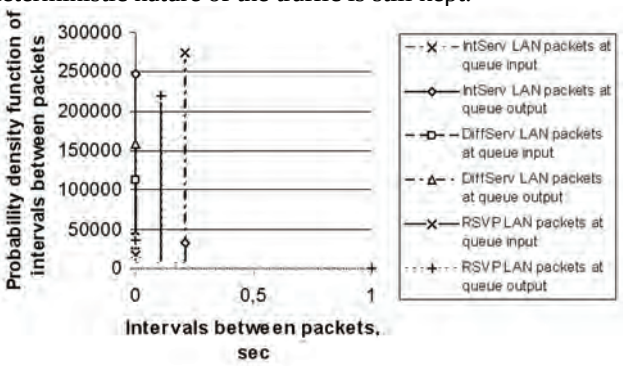


Fig. 5. Probability density function of queue intervals at queue input and output for IntServ, DiffServ and RSVP for LAN traffic.

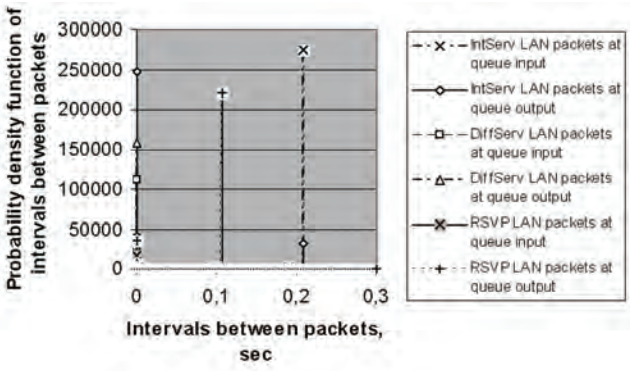


Fig. 6. Probability density function of queue intervals at queue input and output for IntServ, DiffServ and RSVP for LAN traffic.

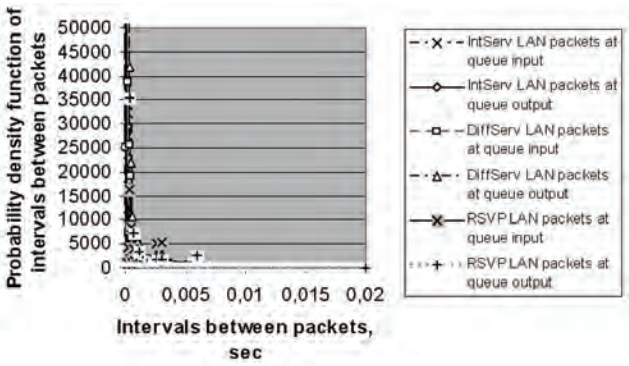


Fig. 7. Probability density function of queue intervals at queue input and output for IntServ, DiffServ and RSVP for LAN traffic.

The mean values for the intervals of LAN traffic at queue input and output for IntServ, DiffServ and RSVP model for different services are shown on Table III, Table IV and Table V. The dispersion and mean squire deviation are high for the rare traffic events. This means that special attention is needed to be paid to the more precise estimation of the status of the queues if necessary. The three QoS techniques distribute the queue resource in a different way. This is the reason to see different mean values. The acceleration effect visible from the graphs is explained more precisely by these tables. The acceleration effect for given traffic (e.g. VoIP on Table

IV) is due to the delay efferct for another traffic. The overall results are always with delay.

TABLE III
INTERVALS BETWEEN PACKETS IN QUEUE INPUT AND OITPUT FOR INTSERV

Traffic source	Mean value	Disperssion	Mean squire deviation
Overall at queue input	0.01726 sec	0	0.00678
VoIP at queue input	0.01732 sec	0	0.00055
LAN at queue input	0.2067 sec	0.00005	0.01255
Transactions at queue input	0.00459 sec	0	0
Overall at queue output	0.01732 sec	0	0
VoIP at queue output	0.01733 sec	0	0.00096
LAN at queue output	0.2076 sec	0	0.00231
Transactions at queue output	0.01937 sec	0	0

TABLE IV
INTERVALS BETWEEN PACKETS IN DIFFSERV

Traffic source	Mean value	Disperssion	Mean squire deviation
Overall at queue input	0.00014 sec	0	0.02628
VoIP at queue input	0.00038 sec	0	0.49381
LAN at queue input	0.00053 sec	0.00005	0.20726
Transactions at queue input	0.00494 sec	0	0.0628
Overall at queue output	0.00015 sec	0	0.02037
VoIP at queue output	0.00037 sec	0	0.43854
LAN at queue output	0.00055 sec	0.00005	0.23368
Transactions at queue output	0.01609 sec	0	0.02573

VII. CONCLUSION

In this paper we show probability density functions of the packet intervals at the queue input and queue output as well as probability density function of queue length per service type. These results demonstrate the specific characteristics of the queue as a packet shaper in three QoS management algorithms IntServ, DiffServ, RSVP. The deterministic nature of the packets streams suppress shaping and increase losses. The statistical multiplexing effect is very limited due to the deterministic streams.

TABLE V
INTERVALS BETWEEN PACKETS IN RSVP

Traffic source	Mean value	Disperssion	Mean squire deviation
Overall at queue input	0.00035 sec	0	0.01637
VoIP at queue input	0.00054 sec	0	0.47737
LAN at queue input	0.08488 sec	0.00173	0.18633
Transactions at queue input	0.00396 sec	0	0
Overall at queue output	0.00038 sec	0	0
VoIP at queue output	0.00054 sec	0	0.38552
LAN at queue output	0.08555 sec	0.00164	0.17985
Transactions at queue output	0.00608 sec	0	0

The results demonstrate the capability DiffServ to support excellent resource management and utilization. The authors refine the simulation model with more traffic sources and more precise generation of the packets from these sources based on the observation of the real traffic. MMPP or geometric/ Weibull distributions are also considered. Limits criteria for queue management is also under investigation.

ACKNOWLEDGEMENTS

This paper is sponsored by the Ministry of Education and Research of the Republic of Bulgaria in the framework of project No 105 “Multimedia Packet Switching Networks Planning with Quality of Service and Traffic Management”.

REFERENCES

[1] Jha, S., M. Hassan, “Engineering Internet QoS”, Artech Hose, 2002.
[2] Janevski, T., “Traffic Analysis and Design of Wireless IP Networks”, Artech House, 2003.
[3] Kleinrock, Leonard, “Queueing Systems”, Volumes I and II, John Wiley and Sons, 1976.
[4] Iversen, V., “Teletraffic Engineering Handbook”, ITU-D, 2005.
[5] Lavenberg, Stephen S., Editor, “Computer Performance Modeling Handbook”, Academic Press, 1983, ISBN 0-12-438720-9.
[6] Pitts, J., J. Schormans, “Introduction to IP and ATM Design and Performnce”, John Wiley&Sons, Ltd., 2000.
[7] Ralsanen, V., “Implementing Service Quality in IP Networks”, John Wiley & Sons, Ltd., 2003.
[8] Tanenbaum, Andrew S., “Computer Networks, Second Edition”, Prentice-Hall International, Inc., 1989, ISBN 0-13-166836-6.

SESSION RCMTA III

Radio Communications, Microwave Technique and Antennas III

Generating of Basic Wave Digital Elements for Modeling of Two-dimensional Planar Structures

Biljana P. Stošić and Miodrag V. Gmitrović

Abstract - In this paper, the main attention is focus on generating basic two-dimensional (2D) wave digital elements (WDE) and their use for modeling of 2D microwave structures. A reference 2D analog structure is separated into segments. One segment containing four inductors and one capacitor is modeled by frequency-dependent WDE. This four-port WDE is used as base for generating three-port and two-port WDE. Wave digital structure equivalent to the analog structure represents cascade connection of basic WDE and two-port adaptors used for impedance matching. Obtained digital structure is solved in the time domain. The suggested procedure with use of basic 2D WDE is implemented in MATLAB environment. An application example, proving the response accuracy of the new technique, is given.

Keywords – wave digital elements, wave digital structure, microwave circuits, microstrip circuits

I. INTRODUCTION

Microwave planar structures have many applications in practice. Because of this, a great attention is given to their analysis. They are two-dimensional (2D) structures and can be modeled by 2D LC circuits [1-4]. A computer program FAMIL (Frequency Analysis of Microwave Lines) [4] or more known programs GENESYS (RF and Microwave Design Software) and ADS (Advanced Design Software), can be used for analyzing these circuits in the frequency domain.

Standard wave digital elements (WDE), such as delay, adder, multiplier and adaptor, are used in design of wave digital filters [5-8]. A wave digital filter can be designed by direct translation of reference analog filter when WDE are used. In paper [9], a procedure for transmission line modeling using WDE is presented. LC transmission line, observed as one-dimensional (1D) structure, is modeled in two ways: by using unit wave digital element and by cascade connection of T-wave digital elements. In paper [10], a microwave planar lowpass structure is modeled by 1D WDE. Transmission lines are modeled by unit WDE. A basic idea for modeling of a single microwave line by using 2D WDE is presented in paper [11].

In this paper, the main attention is focus on generating what would now be called *basic 2D wave digital elements*. When compared with a 1D wave digital realization, a general 2D wave digital realization approach results in a higher implementation effort. A complex planar structure is to be separated into structure segments first. Then, each structure segment is divided into multiport segments. Three types of multiport segments occur in the structure: four-port, three-port and two-port segments. A single four-port segment of analog 2D circuit is modeled by four-port network containing inductors in series branches and a capacitor in parallel branch. A bilinear frequency transformation is used for modeling of an analog segment by equivalent symmetrical four-port WDE. This four-port symmetrical WDE represents start element used for obtaining all basic 2D WDE needed for modeling of a complex planar structure. Basic WDE and two-port adaptors for impedances matching are used for obtaining wave digital structure that is equivalent to the reference 2D analog structure. Obtained digital structure is solved in the time domain.

The suggested procedure is built-in in MATLAB environment. The frequency response is found by use of standard MATLAB programs such as dlinmod.m, dimpulse.m and fft.m. An application example, proving the response accuracy of the suggested procedure by use of basic WDE, is given. The analysis results are then compared with the analysis results of equivalent analog circuit obtained by computer program FAMIL.

II. BASIC 2D WAVE DIGITAL ELEMENTS

A planar microwave structure divided into segments is depicted in Figure 1. Three types of multiport segments occur in any complex 2D structure: four-port central (Central – Cen), three-port edge (LeftEdge – LE, RightEdge – RE, UpEdge – UE and DownEdge – DE) and two-port corner (DownLeftCorner – DLC, DownRightCorner – DRC, UpLeftCorner – ULC and UpRightCorner – URC). In order to generating basic 2D wave digital networks corresponding to the multiport segments, a four-port segment is analyzed first. An analog structure of the four-port segment is shown in Figure 2a. Each segment is modeled by inductors in the series branches and capacitor in the parallel branch.

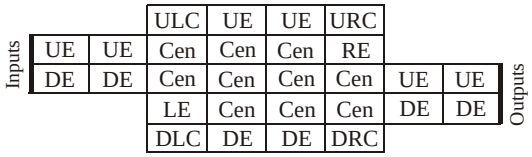


Fig.1. Segmented 2D structure.

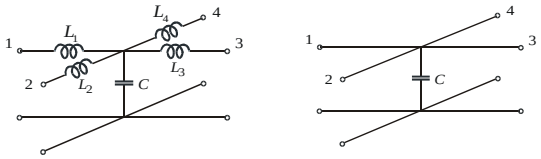


Fig.2. (a) An equivalent LC analog network of four-port segment, and (b) a four-port C analog network.

Biljana P. Stošić and Miodrag V. Gmitrović are with the Faculty of Electronic Engineering, Department of Telecommunications, Aleksandra Medvedeva 14, Niš, Serbia,

E-mails: bilja@elfak.ni.ac.yu, gmitrovic@elfak.ni.ac.yu

The inductances and capacitance can be obtained by expressions given in [3-4]. Wave digital network of the analog LC network shown in Figure 2a, is obtained starting from the four-port network containing capacitor in parallel branch, Figure 2b.

The five-port parallel adaptor with dependent port five is used for modeling capacitor in parallel branch. In this way the four-port wave digital network of the capacitor in parallel branch (WC_Cen), shown in Figure 3, is obtained [1-3]. The WC_Cen is symmetrical because of $L_1 = L_3$ and $L_2 = L_4$ for homogenous structures. This network corresponds to centrally placed segments. For this wave digital network, the equations for wave variables are

$$A_5 = z^{-1}B_5, \tag{1}$$

$$A_S = \alpha_1(2A_5 - A_1 - A_3) + \alpha_2(2A_5 - A_2 - A_4), \tag{2}$$

$$B_5 = A_5 - A_S, \tag{3}$$

$$B_k = B_5 + A_5 - A_k, \quad k = 1,2,\cdots,4, \tag{4}$$

where A_k and B_k are incident and reflected wave variables. The multiplier coefficients are calculated by expressions

$$\alpha_1 = \frac{2GL}{2GL + 2GLt + GC} \tag{5}$$

and

$$\alpha_2 = \frac{2GLt}{2GL + 2GLt + GC}, \tag{6}$$

where GL and GLt are port conductances.

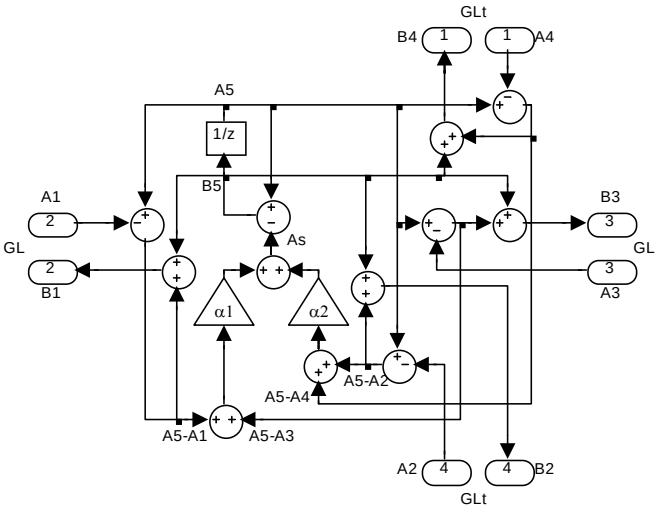


Fig.3. Symmetrical four-port wave digital network of capacitor for centrally placed segments (WC_Cen).

TABLE I

PORT SIGNAL ORDERING FOR WAVE DIGITAL EDGE ELEMENTS AND MULTIPLIER COEFFICIENTS

1) Left	2) Right	3) Up	4) Down
B4 A4 GLt	B4 A4 GLt	B1 A1 GL	B1 A1 GL
B3 A3 GL	B1 A1 GL	B2 A2 GLt	B4 A4 GLt
B2 A2 GLt	B2 A2 GLt	B3 A3 GL	B3 A3 GL
$\alpha1 = 2GL/SG$ $\alpha2 = 2GLt/SG$ $SG = GL + 2GLt + GC$		$\alpha1 = 2GLt/SG$ $\alpha2 = 2GL/SG$ $SG = 2GL + GLt + GC$	

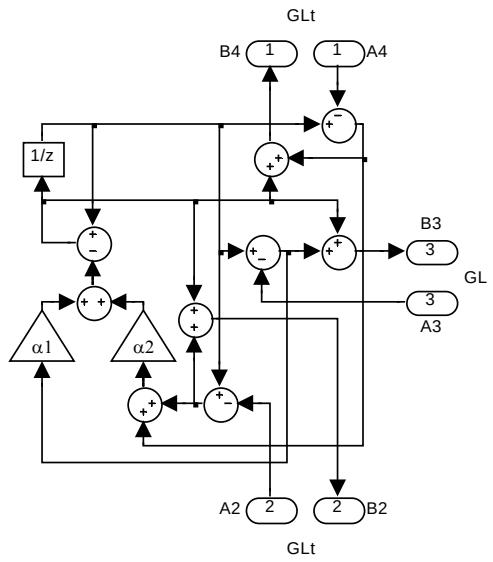


Fig.4. Three-port wave digital network of capacitor for segments edging structure at the left side (WC_LE).

The four-port wave digital network of WC_Cen segment is start point for obtaining wave digital networks of all other segments (three-port and two-port segments).

There are four different segments edging the structure: LE, RE, UE and DE. Ordering of port signals for edge WDE and relations for multiplier coefficients are given in Table I. Three-port wave digital network of capacitor for segment edging structure at the left side (WC_LE) is shown in Figure4.

Also, there are four different two-port segments placed at the structure corners: DLC, DRC, ULC and URC. Ordering of port signals for corner WDE and relations for multiplier coefficients are given in Table II. Two-port wave digital network of capacitor for segment placed at the structure down left corner (WC_DLC) is shown in Figure 5.

TABLE II

PORT SIGNAL ORDERING FOR WAVE DIGITAL CORNER ELEMENTS AND MULTIPLIER COEFFICIENTS

1) Down Left	2) Down Right	3) Up Left	4) Up Right
B4 A4 GLt	B4 A4 GLt	B2 A2 GLt	B2 A2 GLt
B3 A3 GL	B1 A1 GL	B3 A3 GL	B1 A1 GL
$\alpha1 = 2GL/SG$	$\alpha2 = 2GLt/SG$	$SG = GLt + GL + GC$	

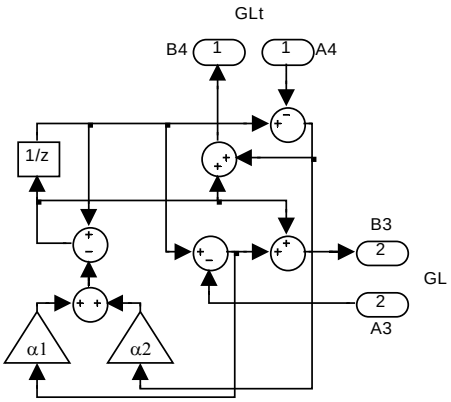


Fig.5. Two-port wave digital network of capacitor for segment placed at the structure down left corner (WC_DLC).

The inductors in the series branches are modeled by unit wave digital elements (UE) [6]. For symmetrical networks, the port resistances of UE corresponding to the inductors L_1 and L_3 are $R_1 = L_1$, and the port resistances of UE corresponding to the inductors L_2 and L_4 are $R_2 = L_2$. Each inductor is represented by a delay element. In this paper, one delay is placed at input port (A1) and the three other delays at output ports (B2, B3, B4), Figure 6a. Some of the wave digital elements corresponding to the analog *LC* networks for different types of segments such as central (WLC_Cen), edge (WLC_LE) and corner (WLC_DLC), are shown in Figure 6.

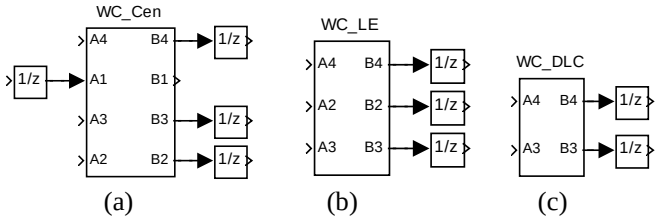


Fig.6. Wave digital elements of LC analog network for:
(a) central segments (WLC_Cen),
(b) edge segments (WLC_LE) and
(c) corner segments (WLC_DLC).

III. MICROWAVE T-JUNCTION AND ITS MODELING BY 2D WDE

A microstrip T-junction is shown in Figure 7. This homogenous planar structure is separated into three structure segments first. Structure segments, Line1 and Line 3, are divided into 3x5 segments, and Line2 into 22x10 segments. Structure lines are then modeled by cascade connection of corresponding WDE. Wave digital structure (WDS) obtained by modeling analog structure with WDE is depicted in Figure 8. Line1 is cascade connection of five SubNetwork3 shown in Figure 9. Adaptors (ADP-In, ADP-L12, ADP-L23 and ADP-Out) are used for port resistance matching. The coefficients of two-port frequency-independent adaptors, α_S and α_L, are given in [6].

In practice, a great number of WDE can be used and complex network obtained. A complex WDS can be formed by multiplication of basic 2D WDE. The WDE are formed in Simulink toolbox of MATLAB environment.

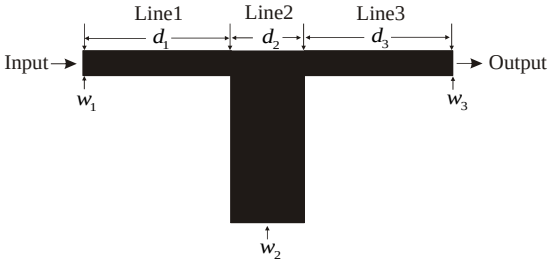


Fig.7. Microstrip T-junction

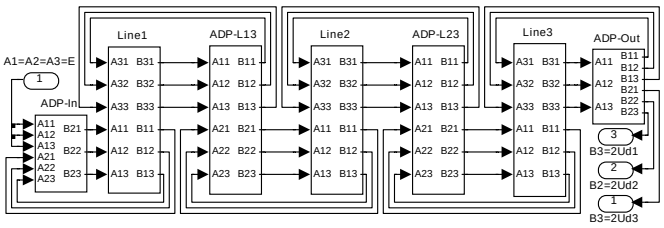


Fig.8. Wave Digital Structure (WDS).

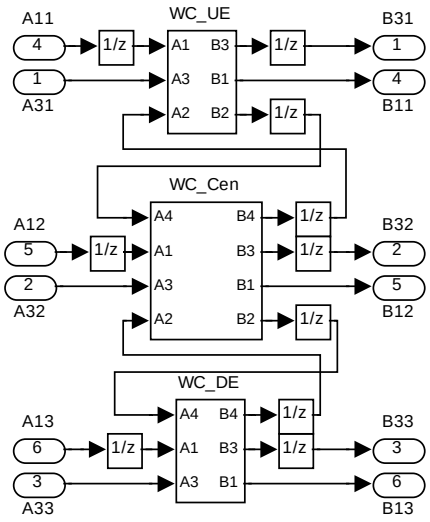


Fig.9. SubNetwork3 in Line1.

IV. APPLICATION EXAMPLE RESULTS

For observed T-junction, substrate material is Polyguide of relative dielectric constant $\varepsilon_r = 2.32$ and high $h = 1.58\,mm$.

The widths of the cascade-connected lines are: $w_1 = w_3 = 4.71\,mm$ and $w_2 = 34.71\,mm$. Their lengths are: $d_1 = d_3 = 30\,mm$ and $d_2 = 15.76\,mm$. The resistances of $50\,\Omega$ are connected to the input and output ports.

Formed WDS, verifying the suggested procedure and proposed basic WDE, is depicted in Figure 8. Comparison of WDS 2D simulation results for three sampling frequency values is depicted in Figure 10. It can be concluded that WDS 2D response accuracy is better for higher values of sampling frequency. For this T-junction, increasing of sampling frequency above $750\,\mathrm{GHz}$, influences simulation results a very little. Because of this, Figure 11 shows the comparison of $|S_{21}|$ simulation results of WDS 2D procedure for $F_{oj} = 750\,\mathrm{GHz}$ and program FAMIL.

The multiplier coefficients of the two-port adaptors at the source side (ADP-In) are $\alpha_S = -0.9094$, and at the load side are $\alpha_L = 0.9094$. For block ADP-L12, the coefficients are $\alpha_A(1) = 0.4226$, and for block ADP-L23 $\alpha_A(2) = -0.4226$. Each WDE has two multipliers and their coefficients are given in Table III.

TABLE III

MULTIPLIER COEFFICIENTS OF THE CENTRAL AND CORNER WDE

	WC_Cen		WC_C	
	α_1	α_2	α_1	α_2
Line1 & Line3	0.0022	0.0321	0.0022	0.0326
Line2	0.0278	0.0277	0.0286	0.0285

MULTIPLIER COEFFICIENTS OF THE EDGE WDE

	WC_E			
	WC_LE & WC_RE		WC_UE & WC_DE	
	α_1	α_2	α_1	α_2
Line1 & Line3	0.0022	0.0321	0.0022	0.0326
Line2	0.0282	0.0281	0.0282	0.0281

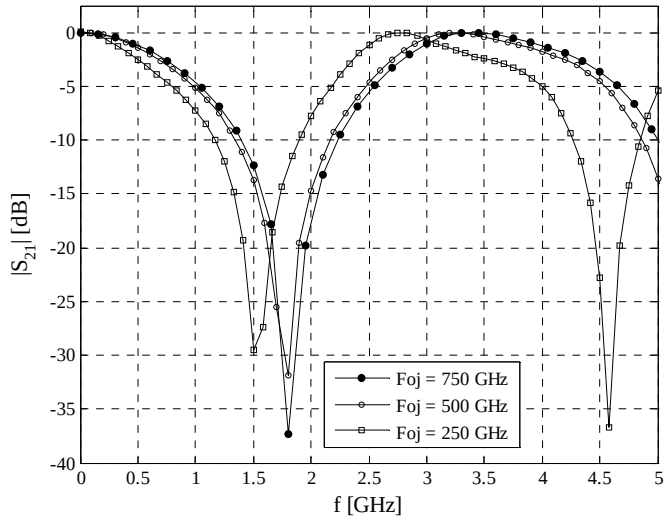


Fig.10. Comparison of WDS 2D simulation results for different sampling frequency values.

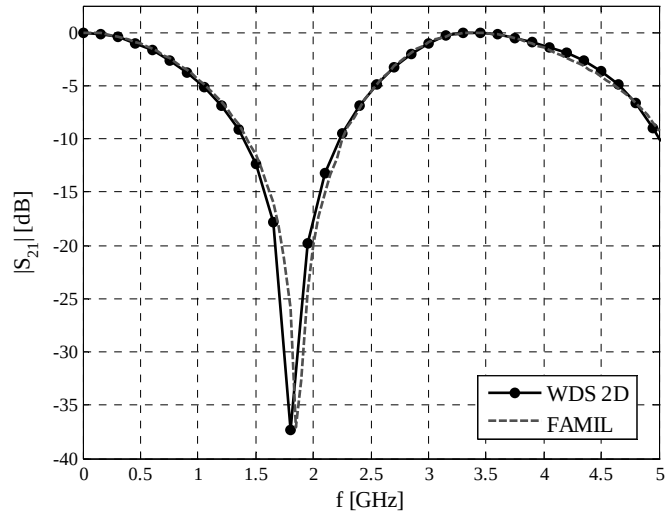


Fig.11. Comparison of $|S_{21}|$ [dB] simulation results for $F_{oj} = 750\,\text{GHz}$.

V. CONCLUSION

The procedure for modeling of 2D microwave structures by WDE is described. The attention is given to generating the basic 2D WDE. A general four-port analog 2D segment is modeled by symmetrical four-port digital network. In this paper, basic WDE corresponding to segments placed at different position in the structure, are given. According to the segment positions, there are three types of WDE: WDE for centrally placed segments (WLC_Cen), WDE for segments edging structure (WLC_ E) and WDE for segments at the structure corners (WLC_C). Also, a procedure for forming complex WDS equivalent to the reference analog structure is described shortly. In WDS, two-port adaptors are used for port resistance matching. Obtained WDS is analyzed in time domain, and its frequency response is obtained by use of fast Fourier transformation. In order to solve the problem, a program for MATLAB environment is written. The MATLAB standard programs from Simulink toolbox are used. WDS, all basic WDE, and two-port adaptors are drawing directly in Simulink toolbox and save as 'model_name'.mdl files.

An application example, proving the modeling of microwave structure by 2D WDE, is given. The correct choice of the sampling frequency leads to good agreement of the WDS 2D simulation results and analog circuit solution.

ACKNOWLEDGMENT

This paper is supported by the Ministry of Science and Environmental Protection of Serbia.

REFERENCES

[1] B.P. Stojanović and M.V. Gmitrović, "Analysis of 2D Lumped Circuit by Equivalent Thevenin Source Method", 14th Intern. Conf. MIKON 2002, May 20-22, Gdansk, Poland, pp. 549-552.

[2] M.V. Gmitrović and B.P. Stojanović, "Analysis of Cascade-Connected Planar Transmission Lines by ETS Method", XXXVII Intern. Scientific Conf. ICEST 2002, October 1-4, Niš, Yugoslavia, Volume 1, pp. 317-320.

[3] B.P. Stojanović and M.V. Gmitrović, "Planar Transmission Line Analysis", XLVI Conference ETRAN, Republic of Srpska, Banja Vrućica-Teslić, 2002, Volume II, pp. 241-244.

[4] B.P. Stošić, *Frequency Analysis of Planar Microwave Circuits by Equivalent Thevenin Source Method*, Master thesis, Faculty of Electronic Engineering,University of Niš,Niš, 2003 (in Serbian).

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

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

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

[8] Lj. Milić and M. Đurić, *Recursive Digital Filters*, Belgrade, 1982 (in Serbian).

[9] M. Gmitrović, B. Stošić and A. Đorđević, "Modeling of 1D Structures by using Wave Digital Elements", L Conference ETRAN, Serbia, Belgrade, June 6-8, 2006, pp. 317-320.

[10] M.V. Gmitrović and B.P. Stošić, "Analysis of Planar Structures Modeled by Wave 1D Digital Elements", 14. Telecommunication Forum – TELFOR 2006, Serbia, Belgrade, November 21-23, 2006, pp. 418-421.

[11] B. Stošić, M. Gmitrović and M. Veljković, "Modeling of 2D Structures by using Wave Digital Elements", L Conference ETRAN, Serbia, Belgrade, June 6-8, 2006, pp. 321-324.

Analysis of the Class E Amplifier with Load Variation

Iliya V. Nedelchev[1] and Iliya N. Nemigenchev[2]

***Abstract* – The operation is presented for class E power amplifier with output impedance converter. The ZVS conditions are specified for achieving high efficiency via minimizing the switching losses. The results for 100W, 1.8MHz are presented. Plots for $R = f(R_L), X_S = f(R_L), P_0 = f(R_L)$ and waveforms from experimental work at different load resistance are presented. According to the results, a wide area of variation of the load in ZVS mode can be achieved.**

***Keywords* – Power Amplifier Class E, High efficiency, Zero-Voltage Switching (ZVS).**

I. INTRODUCTION

The popularity of Resonant Switching Power Amplifiers (RSPS) as high frequency power sources increased in the last years due their advantages - high efficiency , high operation frequency and low cost. RSPS major application fields are dc-dc converters, SMPS (Switching Mode Power Supply), electronic ballasts, induction heating and telecom. Nearly any application suffers from load limitations for a guarantied safe and effective operation. Complicated control loops and protection circuits are required allowing maintaining efficiency in wide range load deviation. Improved circuit of class E amplifier with impedance converter can be used instead.

II. ANALYSIS

A. Assumptions

Consider Fig1 the amplifier class E consists of frequency source, bidirectional switch (MOSFET and anti-parallel diode), DC inductor for a constant current, parallel capacitor, series resonant LC circuit and load. The switch current can be bidirectional, but the Drain voltage can be only positive.

The analysis is based on the main assumptions [1]:

-The transistor has been replaced with ideal switch - zero forward resistance and voltage drop, open circuit in nonconductive state, instant switching.

-Ideal components C_1, C_2, L, i.e., they are linear, lossless and do not have parasitic resonances. The shunt capacitance C_1 includes the transistor output capacitance, the winding capacitance of the DC inductor and wiring capacitance. The transistor output capacitance is independent of the Drain to Source voltage.

-The DC inductor is an ideal current source – I_{dc}

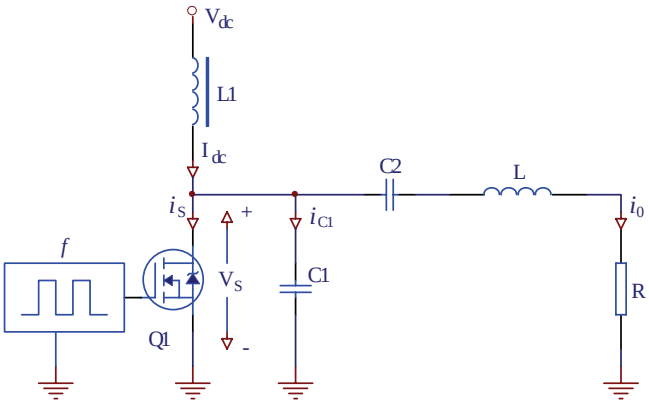


Fig. 1. Class E tuned power amplifier.

B. Parameters

The main amplifier parameters are calculated as follow. Series resonance frequency and Q - factor of L, C2, R, during transistor ON state are

$$\omega_{01} = \frac{1}{\sqrt{LC_2}}; \quad Q_1 = \frac{\omega_{01}L}{R} = \frac{1}{\omega_{01}C_2R} \qquad (1)$$

where $\omega_{01} = 2\pi f_{01}$.

Series resonance frequency and Q - factor of L, C1, C2, R, during transistorOFF state are

$$\omega_{02} = \frac{1}{\sqrt{\frac{LC_1C_2}{C_1+C_2}}}; \quad Q_2 = \frac{\omega_{02}L}{R} = \frac{C_1+C_2}{\omega_{02}C_1C_2R} \qquad (2)$$

where $\omega_{02} = 2\pi f_{02}$.

The operating frequency f differs from both resonant frequencies f_{01} and f_{02}. Therefore, it is convenient to introduce the following ratios of frequencies:

$$A_1 = \frac{f_{01}}{f}; \qquad A_2 = \frac{f_{02}}{f} \qquad (3)$$

From (1) to (3) the relationship (4) can be obtained

[1]Iliya V. Nedelchev is with the Technical University, Str. H. Dimitar 4, 5300 Gabrovo, Bulgaria, E-mail: ilned@tugab.bg

[2]Iliya N. Nemigenchev is with the Technical University, Str. H. Dimitar 4, 5300 Gabrovo, Bulgaria, E-mail: nemig@tugab.bg.

$$\frac{Q_1}{Q_2}=\frac{\omega_{01}}{\omega_{02}}=\frac{A_1}{A_2}=\sqrt{\frac{C_1}{C_1+C_2}} \tag{4}$$

The loaded Q-factor for the operating frequency can be expressed as

$$Q_L=\frac{\omega L}{R}=\frac{Q_1}{A_1}=\frac{Q_2}{A_2} \tag{5}$$

To maintain the efficiency the amplifier must be operated at 50% Duty Cycle and loaded Q-factor must be at least 5. Assuming above conditions the components can be calculated as [4]

$$R=0.5249\frac{V_{dc}}{P_0};\quad C_1=\frac{0.2067}{\omega R} \tag{6}$$

$$C_2=\frac{0.2269}{\omega R};\qquad L=\frac{5.673R}{\omega} \tag{7}$$

In a case of load changes the impedance converter on Fig. 2 can be used. The parallel combination of X_P and R_L (Fig. 2b) can be converted into its series-equivalent circuit $X_S - R$ (Fig. 2a). This conversion leads to the basic circuit of the class E amplifier. Using the equivalent two-terminal networks method, the relationships among the component of the two circuits at the operating frequency *f* are [5]

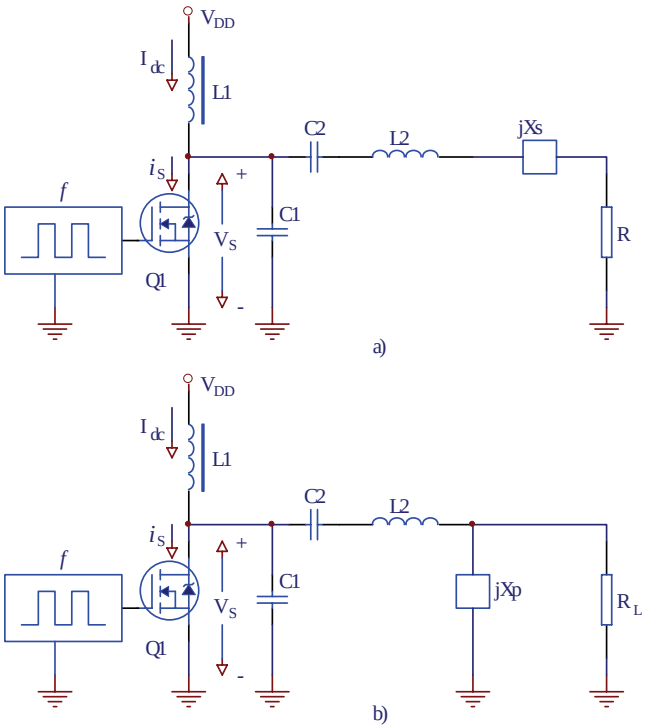


Fig. 2. Class E amplifier: a) Basic circuit. b) Circuit with impedance transforming network.

$$R=\frac{R_L}{1+\left(\frac{R_L}{X_p}\right)^2}=\frac{R_L}{1+q^2} \tag{8}$$

$$X_S=\frac{X_P}{1+\left(\frac{X_P}{R_L}\right)^2}=\frac{X_P}{1+\frac{1}{q^2}} \tag{9}$$

where $q=\frac{R_L}{X_P}$.

As the load resistance R_L from short to open circuit, the input resistance R first increases from zero to its maximum value $R_{max}=X_P/2$ and then decreases to zero. In the same case, the equivalent series input reactance X_S changes from 0 to X_P. In the case of inductive impedance transformation $X_S=\omega L_S$ and $X_P=\omega L_P$, L_S changes from 0 to L_P, so the overall inductance in the series resonance circuit is $L=L_2+L_S$ and increases with R_L. At the same time the series resonance frequency $f_{01}=1/2\pi\sqrt{C_2(L_2+L_S)}$ decreases with R_L. The ratio $A_1=f_{01}/f$ also decreases with R_L for a fixed operating frequency *f*.

B. Design Equations

Input parameters for class E power amplifier design are V_{dc} - the supply voltage, *f* - operating frequency, P_0- output power required. High efficiency can be achieved via ZVS adding $Q_1>5$ and 50% duty cycle to the above requirements.

$$R_{max}=0.5249\frac{V_{dc}^2}{P_0};\qquad C_1=\frac{0.2067}{\omega R_{max}} \tag{10}$$

$$C_2=\frac{0.2269}{\omega R_{max}};\qquad L=5.673\frac{R_{max}}{\omega} \tag{11}$$

From (8),

$$q_{min}=\sqrt{\frac{R_{L\min}}{R_{max}}-1};\quad X_P=\frac{R_{L\min}}{q_{min}} \tag{12}$$

can be defined, where $L_p=\frac{X_P}{\omega}$.

Using (9)

$$X_{S\min}=\frac{X_P}{1+\frac{1}{q_{min}^2}};\quad L_{S\min}=\frac{X_{S\min}}{\omega} \tag{13}$$

which leads to

$$L_2 = L - L_{S\min} \quad . \tag{14}$$

To achieve minimal losses at any R_L, $q = 1$ is required as $R_{L\min}$. In this case the design point is for maximal value of the input resistance R and optimal operation mode.

$$X_P = R_{L\min} = 2R_{\max}; \quad X_{S\min} = \frac{X_P}{2} = R_{\max} \tag{15}$$

L_1 inductance affects the main amplifier parameters. This dependence can be ignored if the DC inductor has a value bigger then the minimal inductance $L_{1\min}$ [3]

$$L_{1\min} = \left(\frac{\pi^2}{4} + 1\right)\frac{R_{\max}}{f} = 3.5\frac{R_{\max}}{f}. \tag{16}$$

The power consumption and maximum DC current are

$$P_{DD} = \frac{P_0}{\eta}; \quad I_{dc\max} = \frac{P_{dc}}{V_{dc}}. \tag{17}$$

The normalized pick transistor current and voltage are $I_{SM}/I_{dc} = 2.783$ and $V_{SM}/V_{dc} = 3.610$, respectively

$$I_{SM} = 2.783 I_{dc\max}, \; V_{SM} = 3.610 V_{DD} \tag{18}$$

The current amplitude in the serial resonant branch L_2–C_2 is

$$I_{0m} = \sqrt{\frac{2P_0}{R_{\max}}} \tag{19}$$

Giving a pick voltages across C_2 and L_2

$$V_{C2p} = X_{C2} I_{0m}; \quad V_{L2p} = X_{L2} I_{om} \tag{20}$$

III. Experimental Rezults

Using the above methodology P_0 = 100W, f = 1.8 MHz, class E amplifier has been designed with V_{dc} = 30V power supply having efficiency η = 94% for optimal load R_{Lopt}= 8 Ω. The components value are L_1=20 μH, L_2 = 1.7 μH, L_3= 0.7 μH, C_1= 4.6 nF, C_2= 5 nF, V_{dc}= 30 V, D = 0.5, Q_1=5.

The impedance converter characteristics R and X_S versus R_L are plotted on Fig. 3. As mentioned before the input resistance R of the impedance converter changes from 0 to $R_{\max} = X_P/2$ when R_L changes from 0 to infinity. In the same time X_S changes from 0 to X_P. Defining $R_{opt} = R_{\max} = X_P/2$ equivalent to q = 1, $X_S = X_P/2$ ZVS can be achieved for any value of R_L.

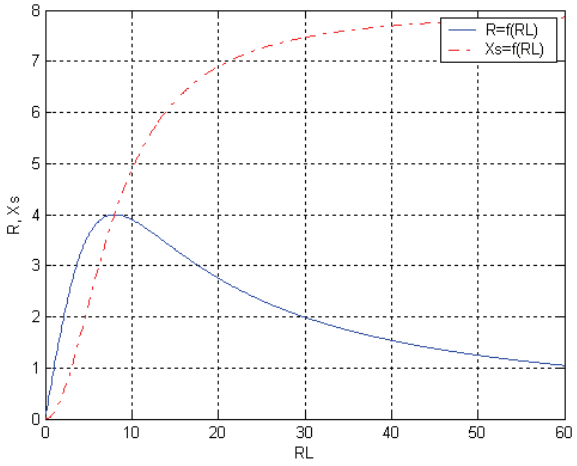


Fig. 3. Characteristics of impedance inverter

The measured characteristic of output power Po from load resistance R_L at constant frequency f = 1.8 MHz is plotted on Fig.4 where Po decreases with the increase of R_L.

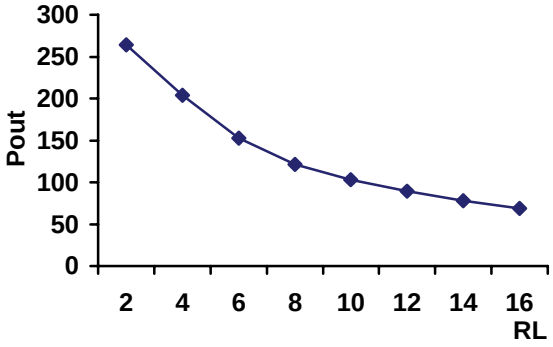
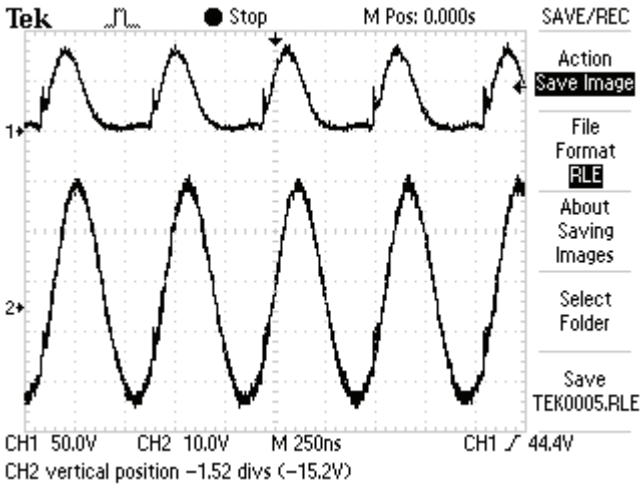
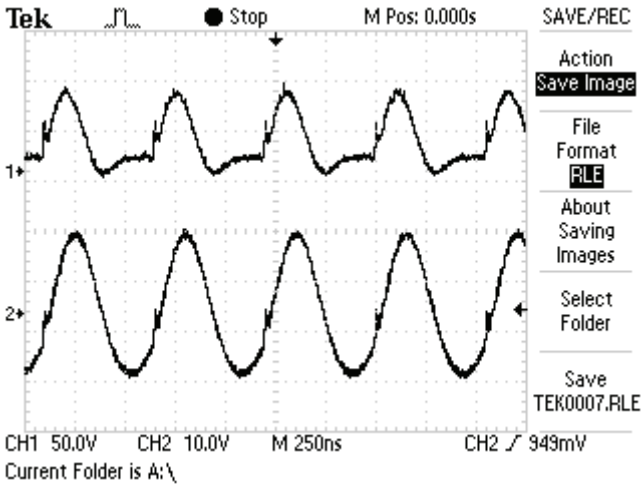


Fig. 4. Characteristics of P_0 versus R_L

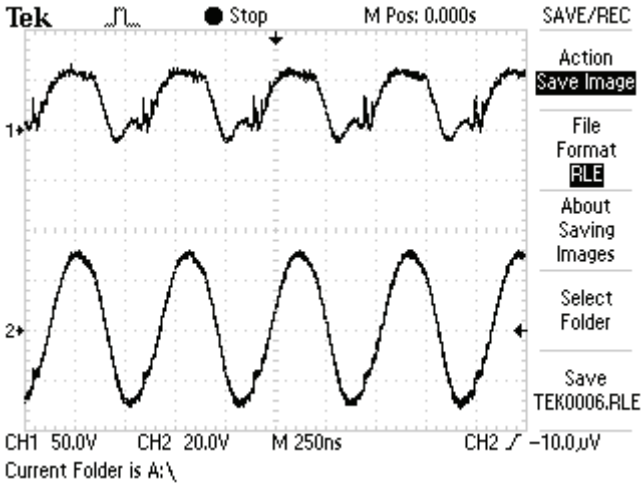
Fig. 5 shows the waveforms of Drain and output voltage at constant operating frequency f = 1.8 MHz and various of the load resistance R_L for ± 50 % .



a)



b)



c)

Fig. 5. Voltage waveforms in the class E amplifier with an inductance: a) – 50 % R_{Lopt}, b) $R_L = R_{Lopt}$, c) + 50 % R_{Lopt}.

The oscillograms show that using the impedance converter ZVS mode can be achieved in class E amplifier for a wide deviation of the load resistance, low losses and high efficiency.

IV. CONCLUSION

The class E amplifier with integrated impedance converter can maintain ZVS for a wide load changes, fixed operating frequency, at low losses and high efficiency. Two cases of operation may be distinguished:

-If the amplifier is designed so that optimum operation for q=1, ZVS can be achieved for changing load resistance from zero to infinity.

-If the amplifier is designed so that operation occurs at q>1, ZVS can be achieved for the load resistance changing from a minimum load resistance R_{Lmin} to infinity.

The series equivalent inductance L_S is a function of load resistance R_L, it increases with R_L. Consequently, the resonant frequency $f_{01} = 1/\left[2\pi\sqrt{C_2(L_2 + L_s)}\right]$ automatically decreases with R_L and the ratio $A_1 = f_{01}/f$ also decreases with R_L for a fixed operating frequency *f*. Therefore, the circuit automatically adjusts its operating conditions to provide ZVS over a wide range of the load resistance R_L.

Additional features have to be implemented in order to protect the amplifier operating in a load short circuit conditions because of the high load power. The load open circuit condition is not dangerous for the amplifier class E with an inductive impedance converter.

REFERENCES

[1] F. Raab, “Idealized operation of the Class E tuned power amplifier”, IEEE Transactions on Circuits and Systems, vol. cas-24, pp. 725-735, December 1977.

[2] H. Krauss, C. Bostian and F. Raab, *Solid State Radio Engineering*, New York, John Wiley & Sons, Inc., 1980.

[3] M. Kazimierczuk, “Collector Amplitude Modulation of the Class R Tuned Power Amplifier”, IEEE Transactions on Circuits and Systems, vol. cas-31, No. 6, pp.543-549, June1984.

[4] M. Kazimierczuk, K. Puczko, "Exact Analysis of Class E Tuned Power Amplifier at any Q and Switch Duty Cycle", IEEE Transactions on Circuits and Systems, vol. cas-34, No. 2, pp.149-159, February 1987.

[5] M. Kazimierczuk, X. T. Bui, “Class E Amplifier Operating from a Short Circuit to an Open Circuit”, Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON’89), Dayton, OH, vol. 1, pp. 240-245, May 22-26 1989.

[6] N. Sokal, A. Sokal, “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

[7] S. Cripps, *RF Power Amplifier for Wireless Communications*, Artech House Boston London, 1999.

Polar Diagram of Multiple Driver Loudspeaker Systems

Ekaterinoslav S. Sirakov

***Abstract* –In this work are present theoretical analysis polar characteristics and horizontal or vertical off–axis frequency responses of loudspeaker system. Discussed are the mathematical dependencies for generating of sound pressure from loudspeakers operating in a given frequency range, specified by crossover network.**

***Keywords* – Loudspeaker system, multiple driver, crossover network, polar diagram, Horizontal and Vertical off–axis frequency responses.**

I. THEORETICAL FUNDAMENTALS FOR DEFINITION OF SOUND FIELD

Sound field may be defined using the Huygen – Fresnel equation, known in acoustics as the Relay integral.

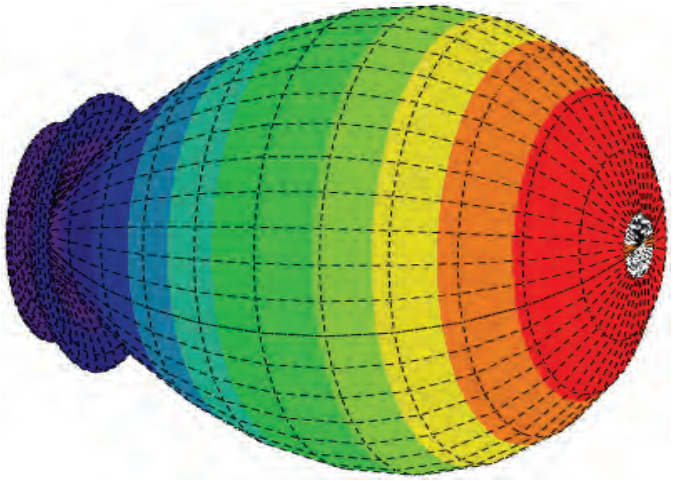


Fig.1. 3D polar diagram of a circular piston with diameter d=30cm for 2.5 kHz frequency, Eq. (1)

For establishment of sound pressure in a far-field region by a loudspeaker installed in the system we may use with success the mathematical expressions for circular piston surface situated on infinite absolutely solid screen.

For circular piston located in infinite screen, for the far-field region using the Frounhofer's approximation we have the following relative simple formula:

$$D(\theta)=\frac{1}{\pi.a^2}\int_0^a \rho \int_0^{2.\pi} \exp(-j.k.\rho.\sin(\theta).\cos(\gamma))\,d\gamma\,d\rho$$

where: a is the radius of circular piston, (1)

$\kappa=\frac{\varpi}{c}$ is the wave-number,

c is the speed of sound.

Illustrated on fig. 1 is a 3D polar diagram for Relay's integral Eq. (1), in the far-field region for a circular piston with diameter $d=2.a=30cm$ for frequency *2.55 kHz.*

Illustrated on fig. 2 is the frequency response of the sound pressure created by a circular piston with diameter *d=30cm* for three values of the angle θ. The first zero is at $\theta=\pi/2$ *(sin(θ)=1)* and is for frequency 1.4 kHz. The second zero is for frequency 2.55 kHz, represented by 3D polar diagram on fig. 1. By reducing the value towards emission axis, the frequency increases, as illustrated on fig. 2 for $\theta=45^0$, the first zero is at 2 kHz frequency.

$$D_{SPL}(f,\theta)=90+20.\log|D(f,\theta)| \quad (2)$$

$$D(f,\theta)=\frac{1}{\pi.a^2}\int_0^a \rho \int_0^{2.\pi} \exp(-j.k.\rho.\sin(\theta).\cos(\gamma))\,d\gamma\,d\rho$$

$$k=\frac{2.\pi.f}{c}$$

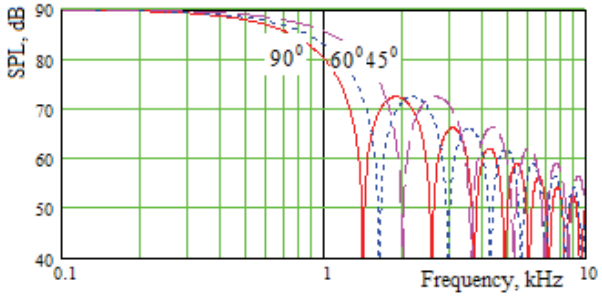


Fig. 2. Frequency response for sound pressure created by circular piston at parameter angle (90^0, 60^0 and 45^0), Eq. (2)

The solution of Riley's integral for circular piston, installed on infinite solid screen, [1], [3] and [4] is:

$$D(\theta)=\frac{2.J_1(k.a.\sin(\theta))}{k.a.\sin(\theta)}, \quad (3)$$

$J_1(k.a.\sin\gamma)$ is a first-order Bessel function.

[1]Ekaterinoslav S. Sirakov is with the Department of Radio engineering, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria, E-mail: katio@mail.bg, katiosirakov@abv.bg

II. Definition of Sound Pressure with Two Sources

For two point sources in phase of equal amplitude, sound pressure may be defined with the expression [1], [2]:

$$D_2(f,\theta)=\frac{1}{2}.\left(e^{-\frac{j.\pi.\frac{d}{2}.f}{c}.\sin(\theta)}+e^{\frac{j.\pi.\frac{d}{2}.f}{c}.\sin(\theta)}\right) \quad (4)$$

To establish a model for the off-axis frequency response of two loudspeakers of the same type operating in a phase of the same amplitude in vertical plane, used with success may be multiplication theorem (on the diagram for circular piston, Eqs. (1) or (3) and the diagram for two point source, Eq. (4)).

$$D_{2_LS}(f,\theta)=90+20.\log\left|D_2(f,\theta).D(f,\theta)\right| \quad (5)$$

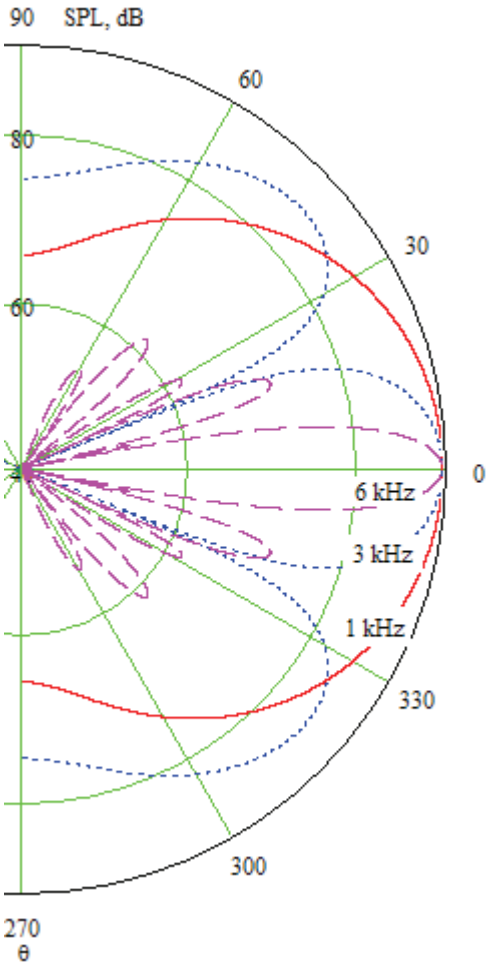


Fig.3. Vertical polar response of two source for frequencies 1, 3 and 6 kHz and distance between a acoustic centers 30 cm and equal diameter of circular piston 30 cm for both drivers, Eq. (5).

III. Model Frequency Response of Sound Pressure Created Loudspeaker

The frequency characteristic of every single loudspeaker may be presented with the band-pass function [2], [4] and [9]:

$$K(s)=K_1(s).K_2(s) \quad (6)$$

$$K_1(s)=\frac{s^2}{s^2+\frac{2.\pi.f_0}{Q}.s+(2.\pi.f_0)^2} \quad [2] \quad (7)$$

$$K_{21}(s)=\frac{1}{1+s} \quad [9]$$

$$K_{22}(s)=\frac{1}{s^2+\frac{2.\pi.f_h}{Q_h}.s+(2.\pi.f_h)^2} \quad [4] \quad (8)$$

$$s=j.\varpi=j.2.\pi.f$$

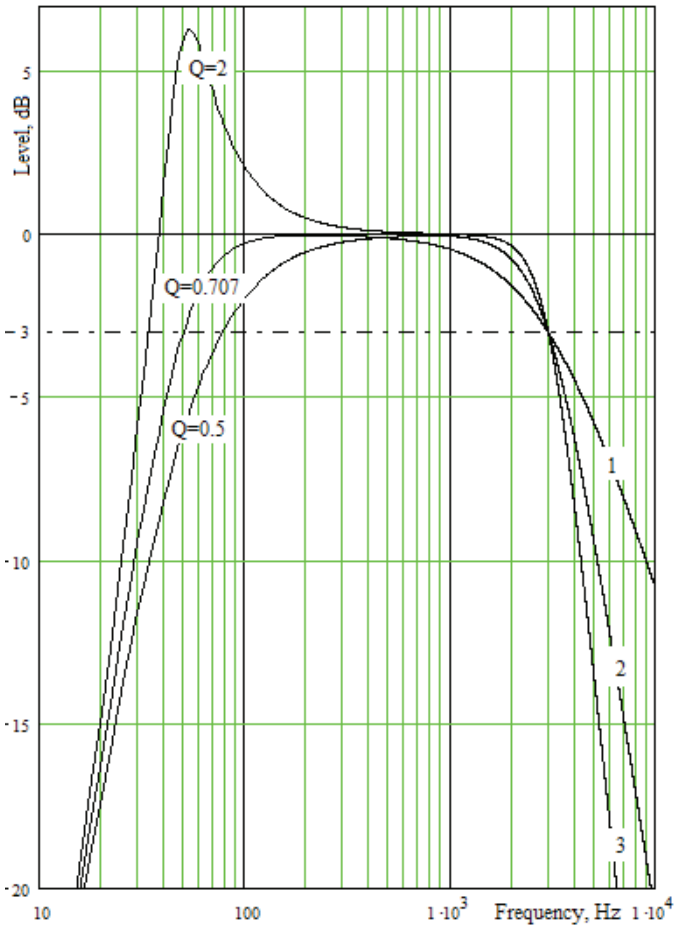


Fig.4. Model for frequency response of sound pressure created by loudspeaker, Eqs. (6), (7) and (8)

IV. TARGET FUNCTIONS APPROACH OF CROSSOVER (FILTERS) FOR TWO FREQUENCY BANDS

Frequency separation for loudspeaker systems is achieved by passive (for high signal level), active (for low signal level) and digital (for digital loudspeaker systems) separation circuits (crossover) or in a mixed mode.

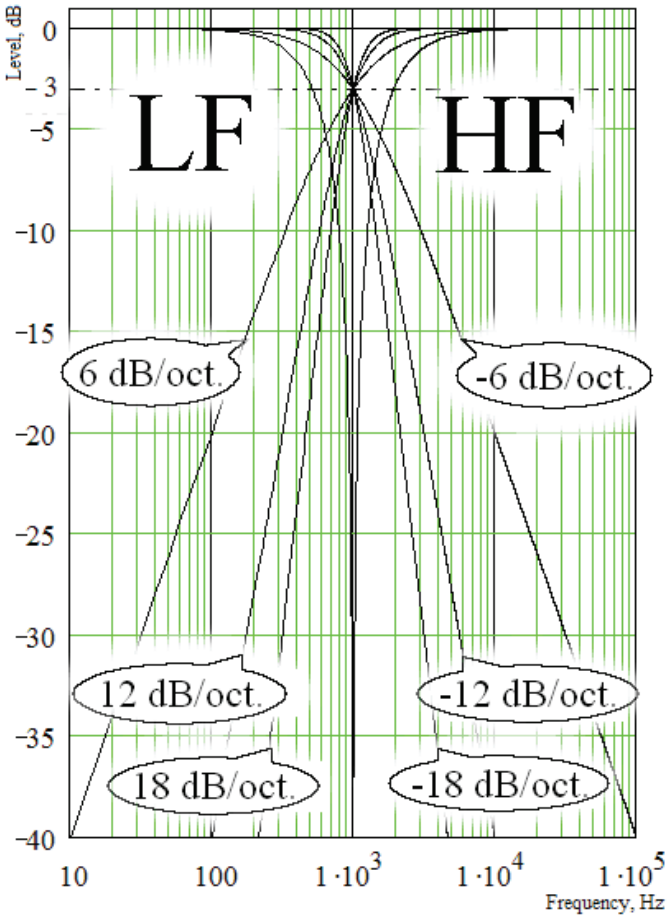


Fig.5. Two-way crossover response (6, 12 and 18 dB/octave rolloffs), Eqs. (9), (10) and (11)

All crossover networks (filters) may be described by their target functions approach. Selected for filters is Linkwitz-Riley, Bessel, Butterworth or different polynomial approximation. The rolloff of filters defined by polynomial approximates, is: 6 dB/oct. for first order, 12 dB/oct. for second order, 18 dB/oct. for third order etc.

The selection of filter sequence and type as well their polynomial approximation depend on a number of factors and is subject to compromises in actual realization.

For symmetrical filters of 1st order Butterworth type, the channels are in phase, and crossover networks (filters) in passive realization has constant input resistance [2]:

$$T_{L_1}(s)=\frac{1}{1+s},\ T_{H_1}(s)=\frac{s}{1+s},\ T_{L_1}(s)+T_{H_1}(s)=1 \qquad (9)$$

For Butterworth of the second order target functions approach:

$$T_{L_2}(s)=\frac{s/\sqrt{2}+1}{1+\sqrt{2}.s+s^2} \qquad T_{H_2}(s)=\frac{s^2+s/\sqrt{2}}{1+\sqrt{2}.s+s^2} \qquad (10)$$

$$T_{L_2}(s)+T_{H_2}(s)=1$$

For third order:

$$T_{L_3}(s)=\frac{2.s+1}{1+2.s+2.s^2+s^3}$$

$$T_{H_3}(s)=\frac{s^3+2.s}{1+2.s+2.s^2+s^3} \qquad (11)$$

$$T_{L_3}(s)+T_{H_3}(s)=1$$

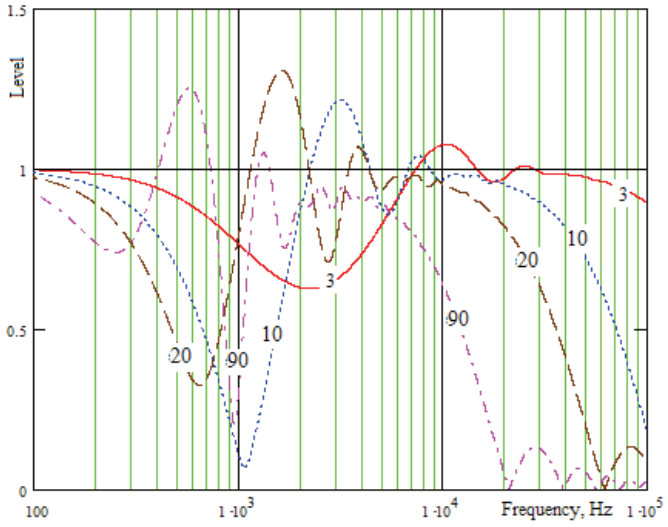


Fig.6. Frequency characteristic of sound pressure from loudspeaker system with two frequency bands at angle θ= 3°, 10°, 20° and 90°, Eq. (12).

Illustrated on fig. 6 is frequency response of sound pressure from loudspeaker system with two frequency bands at angle $\theta=3^0$, 10^0, 20^0 and 90^0 and crossover with Butterworth polynomial approximation of first order. Low frequency loudspeaker effective circular piston diameter is $D_{LF}=20cm$, and for high frequency loudspeaker $D_{HF}=2cm$ respectively, with minimum distance between both loudspeaker centers is $d=15cm$.

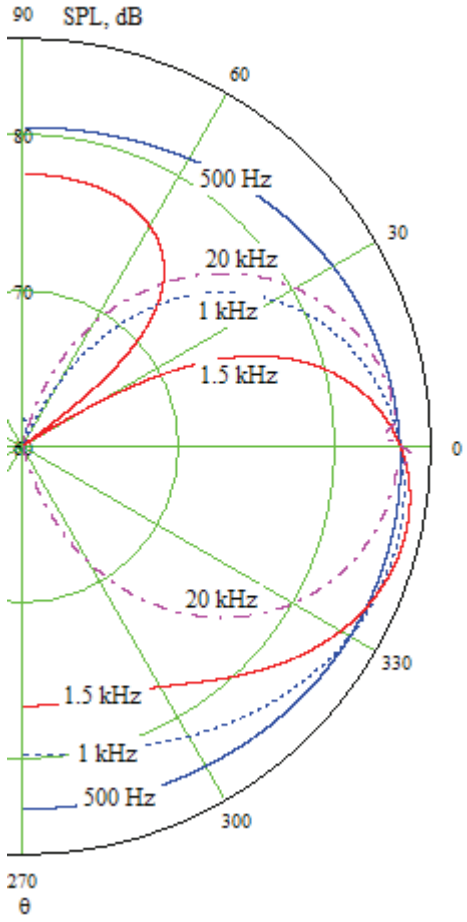


Fig.7. SPL polar response of loudspeaker system with two frequency bands for frequencies 0.5, 1, 1.5 and 20 kHz, Eq. (12).

Illustrated on fig. 7. is the vertical polar plot of two-way loudspeaker system for frequencies 0.5, 1, 1.5 and 20 kHz.

$$D_{V_dB}(f,\theta) = 90 + 20.\log\left|D_V(f,\theta)\right| \tag{12}$$

$$D_V(f,\theta) = \frac{1}{2}.\left\{ \begin{array}{l} D_L(f,\theta).\left[T_{L_1}(f).e^{-\frac{j.\pi.\frac{d}{2}.f}{c}.\sin(\theta)} \right] + \\ D_H(f,\theta).\left[T_{H_1}(f).e^{\frac{j.\pi.\frac{d}{2}.f}{c}.\sin(\theta)} \right] \end{array} \right\} \tag{13}$$

Driver piston diameter is $d_L=2.a_L=20cm$ for low frequency loudspeaker (for $D_L(f,\theta)$, Eqs. (1), (2) or (3)) and $d_H=2.a_H=2cm$ for high frequency loudspeaker (for $D_H(f,\theta)$, Eqs. (1), (2) or (3)), with distance between acoustic centers d=20 cm

Crossover frequency is $f_0=3kHz$. Butterworth approximation of crossover network of the first order is, Eq. (9).

V. Conclusion

The components of a single loudspeaker system: separation circuits (crossover) for respective frequency bands, usually two or three, geometry of loudspeaker arrangement – one or two for respective band, frequency characteristic of every single loudspeaker, influence the general polar characteristics of the system.

The presentation of loudspeaker as a circular piston Eq. (1) facilitates the investigation of frequency characteristic irregularity depending on angle parameter diameter (fig. 2), and the selection of upper limit frequency.

With two loudspeakers for the same frequency range according to classical geometry (one over the other normally at specified distance) or in the so called D'Appolito geometry aimed to enhance the generated sound pressure will sharply narrow the vertical polar response (fig. 3).

Used successfully for measuring the frequency characteristics of specific loudspeaker may be the mathematical dependencies Eq. (6) taking into account the specification T/S parameters (fig. 4) or actually measured characteristic values [Martin J. King, 6].

Of considerable importance is the selection of a specific separation circuit (crossover) type and filter sequence, not only for the generated sound pressure along emission axis, but for different angles too (Eq. (13), fig. 6) as well as the common diagram, depending on frequency, especially for frequencies close to separation (Eq. (12), fig. 7).

The results obtained in practice may be used for theoretical analysis, and in loudspeaker system design and production.

References

[1] Вълчев, Ив. Й., "Електроакустика", Държавно издателство "Техника", София, 1975г., стр. 93÷104

[2] Алдошина И.А., А.Г.Войшивилло "Высококачественные акустические системы и излучатели", Москва, Радио и связь, 1985, стр. 64÷102.

[3] Алдошина И. А. "Електродинамические громкоговорители", Москва, Радио и Связь, 1989. стр.112÷118

[4] A. Rimell and M. Hawksford, "From the cone to the cochlea: Modelling the complete acoustical path", Presented at 100th Convention 1996, May 11-14, Copenhagen, AES, 1996.

[5] Петров П. Н., Акустика. Электроакустические преобразователи: Учеб. пособие, Санкт Петербург, 2003,стр. 48÷55.

[6] Martin J. King, Derivation and Correlation of a Viscous Damping Model Used in the Design of a Transmission Line Loudspeaker System.

[7] Rene Christensen, Active All-Pass Crossover Networks with Equal Resistors and Equal Capacitors, J. Audio Eng. Soc.; Vol. 54, No. ½, 2006 January/February, pp. 45÷53.

[8] Сираков Е.С. Общ модел на електродинамичен високоговорител с директно излъчване за компютърна симулация, *списание Акустика*, год. 7, 2005 г. ISSN 1312-4897

[9] A.A. Angelova, E.S. Sirakov, G.K. Evstatiev. Theoretical analysis of Frequency, Pulse and Transitional characteristics of Loudspeaker (Part I). *"ICEST 2004"*, 16÷19 June 2004, Bitola, Macedonia, pp. 737÷738.

[10] Ekaterinoslav S. Sirakov. Theoretical Analysis as a Function of the Total Q Factor Loudspeaker Characteristics, *ICEST 2006*, 29 June - 01 July, 2006, Sofia, BULGARIA, pp 367-369

Nonlinear Systems Modeling in Broadband Communications

Oleg Borisov Panagiev[1]

Abstract **– In this report modeling of nonlinear systems is investigated. Nonlinear models can be divided to ad hoc type block models, and analytical models, most notably the Volterra series. In broadband communications systems related modeling it is convenient to split nonlinear models to memoryless ones, and to those with memory. In order to obtain optimum results from Volterra methods, it helps to understand a little about the method and the way nonlinear system are modeled.**

Keywords **– CATV/HFC, Volterra series, single-frequency dependent (SFD) model.**

I. INTRODUCTION

Recently, CATV/HFC networks have received much attention in the context of interactive application: telephony, interactive television, home-shopping, video-on demand, high-speed Web browsing, etc. The interactive nature of these services requires however two-way communication on a network that initially was only intended for a one-way broadcasting of television signals. This huge information capacity requires communication system to be very broadband, witch set using of a big number of amplifiers. Each amplifier adds noise and intermodulation distortions to the signal, the number of cascaded trunk amplifiers in the primary lines is thus limited to 40 to 45, in order to ensure a good reliability of the network and a good quality of the signal delivered to the subscribers. In order to provide other services such as digital television, it is necessary to increase the capacity of existing networks [1]. In this case CATV/HFC networks contain many other cascaded nonlinear elements, coaxial and/or optical with coaxial amplifiers too. Each of these nonlinear elements would have deteriorated C/N, respectively BER and CSO/CTB in the transmitted channels. The CSO/CTB frequency spectrum and C/N are adequate characteristics of distortions and noises in channels of CATV network. They can be determinate for the cascaded net elements from the CSO/CTB frequency spectrum and C/N for each of cascaded net elements and coaxial amplifiers too [2]. This leads to make investigation and classification of model types. The next section is devoted to classification and choosing the model of broadband communication systems as CATV/HFC.

[1]Oleg B. Panagiev is whit the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com

II. THE MODEL'S CLASIFICATION

In the process of CATV/HFC nonlinearities research is necessary to choose a good model. The nonlinear system modeling problem can be formulated as follows: Find a nonlinear function f that well describes the input-output mapping obtained by measurements. In general case the function is a mapping from an m -dimensional space to an n – dimensional one, a multiple-input multiple-output system. Usually some noisy measurement results of the input-output mapping are available.

A n^{th} degree polynomial

$$f[x(t), a] = a_0 + a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) + \ldots + a_n x^n(t) \quad (1)$$

is an example of a nonlinear mapping $f : \Re \rightarrow \Re$ with (n+1) parameters. It is assumed that the model is time-invariant. In this report it is assumed that input and output are one-dimensional.

Depending on the amount of prior information known about the modeling task we may talk about white-box, grey-box, or black-box modeling [3]. In white-box modeling the model is perfectly known from prior knowledge. An example of a white-box model is the simple voltage-current law of a diode. In a grey-box modeling something is known a priori, but some model parameters must be determined from measurements. In black-box modeling no physical insight or a priori information is available. The nonlinear black-box modeling problem is among the most difficult in the field of system identification.

The central problem is how to find a proper model class, and model order within the chosen class, of all possible functions. After this it is only a matter of optimizing the parameters of the model. Due to nonlinear nature of the problem, several local optimums usually exist, complicating the optimization phase. Examples of nonlinear model (function) classes include polynomial models, artificial neural networks, and Volterra series. For a grey-box problem one can narrow down the model class based on information known of the modeling problem.

Finding a "good" model class for a black-box problem is a very fundamental problem, for which no single all-around solution seems to exist. Typical methods used to determine model order within a given model class include cross validation and bootstrap methods (statistical methods); Minimum Description Length (MDL) (information theoretical method), Akaike Information Criterion (AIC) (ad hoc); and Structural Risk Minimization (statistical learning theory) [4]. A flowchart of a general modeling task is shown in Fig. 1.

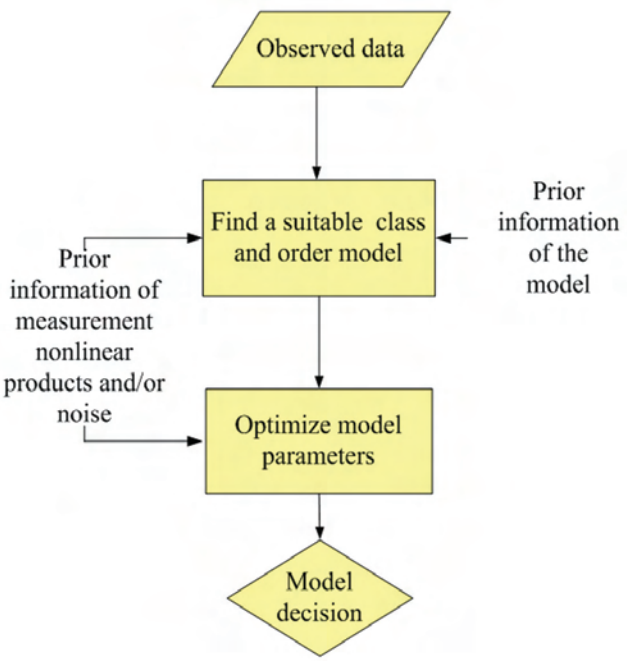


Fig.1. Algorithm for nonlinear system modeling

TABLE 1

Models	Models with memory	Models without memory
Analytical	Volterra series Nonlinear differential equations	Polynomials Simple diode
Block	Saleh model *N* - box model Recurrent neural networks	Two box Table look-up Feedforward neural networks

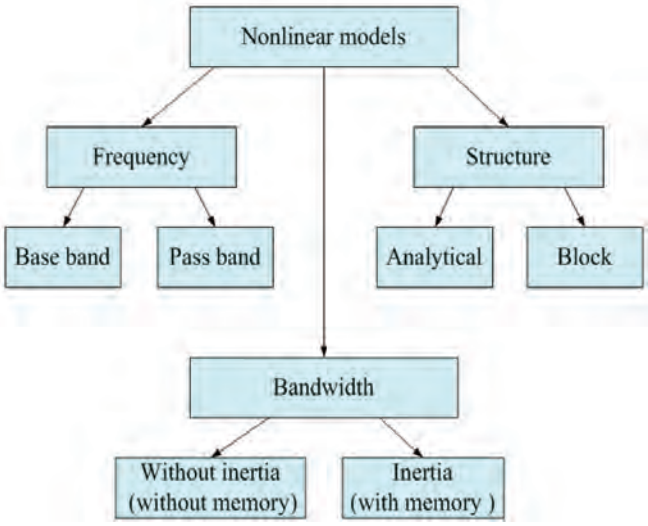


Fig.2. Nonlinear models classification

The classification of mathematical models in dependence of the different symptoms is too multiple. Thus like in a dependence of a level of complexity are isolated: *full models*, forme250 d by immediate union in general system from equations of separate components, building the CATV system; *macromodels*, witches are got by approximation of full models.

From the ways of determinating of the parameters, the models are divided on: *physical*, at whose are examine the physical regularities, inherent for the device (system), as a result of witch the structure of the equations and parameters of the models have definite physical interpreting; *formal*, whose parameters are finding experimentally – by measuring of the reaction of the system of some definite testing signals.

From made above reasoning for system modeling may be formed next classification of models (Fig. 2). Some examples for models are given in Table 1

In the further examined the models will be used the moderating of the system in the pass band, witch for the general case for modern systems for cable television is from 5 MHz to 860 MHz.

III. NONLINEAR SYSTEMS MODELING WITH VOLTERRA SERIES

Well known are many models for nonlinearities with memory, where the nonlinearity system is presented as combination from in series related linear with memory and nonlinear memoryless systems [3], [5], [6]. This model is named summary nonlinear model (SNM) and its decisions may to given with differential equations [6], [7], [8] or power series [9], [10], [11].

Below we offer nonlinear model on advance broadband communication system, which is modeling with Volterra series. This model is named single-frequency dependent (SFD) model (Fig.3). He is consist n numbers of parallel connected branches, each of them contains nonlinear memoryless and linear system with memory from respective order. That helps for the calculating of the spectrum of output signal of the system from each one order ($\leq$ n). The most important priority of the proposed models is the relative simplicity, with witch the kernels (transfer functions of Volterra) can be measured and applied in the analysis of the nonlinear distortions.

Other priorities of the model are:

- Meanwhile compact reading of the inertial and nonlinear characteristics of the tract;
- The block presenting of the transforming of the input influence allow simplifying of the problem about dividing of the nonlinear products from different order. As a result of that studying of the different types nonlinear distortions became possible, instead of studying the process as a whole one.

Although the indicated model may not capture the full aspects of a nonlinear CATV system, its primary usefulness perhaps resides in the associated measurements and computations required for system identification.

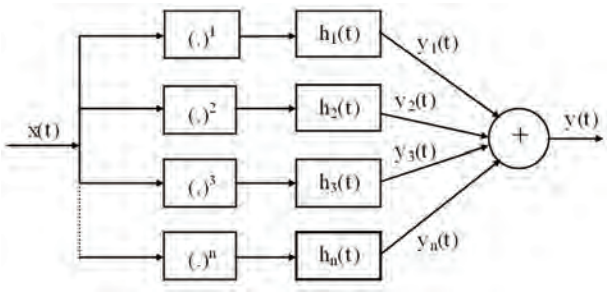


Fig.3. Single-frequency dependent model

Volterra theory is a generalization of the linear convolution integral approach often applied to linear, time-invariant systems. The theory states that any time-invariant, nonlinear system can be modeled as an infinite sum of multidimensional convolution integrals of increasing order. This is represented symbolically by the series of integrals

$$y(t) = h_0 + \int_0^{+\infty} h_1(\tau_1)x(t-\tau_1)d\tau_1$$

$$+ \int_0^{+\infty}\int_0^{+\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2$$

$$+ \int_0^{+\infty}\int_0^{+\infty}\int_0^{+\infty} h_3(\tau_1,\tau_2,\tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3$$

$$+\cdots+$$

$$+ \underbrace{\int_0^{+\infty}\cdots\int_0^{+\infty}}_{n} h_n(\tau_1,\tau_2,...,\tau_n)x(t-\tau_1)x(t-\tau_2)...$$

$$...x(t-\tau_n)d\tau_1 d\tau_2 ...d\tau_n \ , \qquad (2)$$

which are known as the Volterra series. Here, *x(t)* represents the system input while *y(t)* represents the system response. Volterra theory is based on dynamic data, and as such the average values of all input and response data sets are removed. Each of the convolution integrals contains a kernel, either linear (h_1) or nonlinear (h_2,...,h_n), which represents the behavior of the system. Knowledge of these kernels allows the prediction of a system's response to any arbitrary input, and as such is critical to nonlinear Volterra modeling.

The application of the Volterra series requires careful consideration of various factors that can impact on the successful calculation of transfer functions. They are useful tools in observing the behavior of nonlinear systems. Transfer functions convert an input signal into an output signal, and for nonlinear systems can be calculated separately for each order of response using the Volterra series. The Volterra series quantifies the linear and nonlinear (e.g. quadratic, cubic etc.) responses separately for systems with either Gaussian or non-Gaussian inputs, and is particularly useful when calculating transfer functions from experimental data, as the calculations can be performed using a discrete frequency-domain format.

For theoretical examination of the nonlinear distortions we choose (in CATV system spectrum) one channel, who under condition we mark with s and call it test (useful) channel. In each of other channels are transferred unmodulated signals with value of a picture carrier frequency in agreement with European standard CENELEC EN 50083 (B/G, D/K). All picture carrier frequencies are CW. In this case we may to write next equation for output system signal (Fig.3):

$$y(t) = y_1(t) + y_2(t) + y_3(t) + ... + y_s(t) + ... + y_n(t) \qquad (3)$$

According to the (2) in the one-dimensional case, such as the output signal of the system, (3) became:

$$y(t) = \int_{-\infty}^{+\infty} h_1(\tau)x(t-\tau)d\tau + \int_{-\infty}^{+\infty} h_2(\tau)x^2(t-\tau)d\tau$$

$$+ \int_{-\infty}^{+\infty} h_3(\tau)x^3(t-\tau)d\tau + ... + \int_{-\infty}^{+\infty} h_s(\tau)x^s(t-\tau)d\tau + ...$$

$$+ \int_{-\infty}^{+\infty} h_n(\tau)x^n(t-\tau)d\tau. \qquad (4)$$

In correspondence with the accepted upper conditions on the input of the system we apply group signal, witch represents a group of unmodulated signals

$$x(t) = \sum_{i=1}^{N} A_i \cos(\omega_i t + \theta_i); \ \ \omega_i = 2\pi f_i; i \neq s. \qquad (5)$$

Let's replace with (5) in (4) and do the definite mathematical operation: presenting of $cos\omega_i t$ like exponential function, applying the direct Fourier's' transformation, etc. Let group together the obtained by the calculating addends in attitude with the frequencies (summary and differenced). In this case we will accept that the amplitudes of the signals for every picture carrier frequencies are same (just like the request of the CENELEC EN 50083 for every advanced CATV system). The last results from based frequencies, harmonically, two-component and three-component beet are presented in Table 2, (N=3, n=3). The Volterra kernels in the frequency domain from first, second and third order are obtained from following equations:

$$\left|H_1(\omega_i)\right| = \left|\int_{-\infty}^{+\infty} h_1(\tau)\exp(\pm j\omega_i\tau)d\tau\right|; \qquad (6)$$

$$\left|H_2(\omega_b)\right| = \left|\int_{-\infty}^{+\infty} h_2(\tau)\exp(\pm j\omega_b\tau)d\tau\right|; \qquad (7)$$

$$\left|H_3(\omega_b)\right| = \left|\int_{-\infty}^{+\infty} h_3(\tau)\exp(\pm j\omega_b\tau)d\tau\right|, \qquad (8)$$

where:

$\omega_b = \omega_i \pm \omega_j$ in case of second order beat;

$\omega_b = 2\omega_i \pm \omega_j$ in case of two-component beat;

$\omega_b = \omega_i \pm \omega_j \pm \omega_k$ in case of three-component beat

and $i \neq j \neq k = 1, 2, 3;$

$\theta_1 = \theta_1(\pm\omega_i)$; $\theta_2 = \theta_2(\pm\omega_b)$; $\theta_3 = \theta_{23}(\pm\omega_b)$.

TABLE 2

$y(t) = \sum_{i=1}^{3} y_i(t)$	Σ terms
y_1	$A_1\lvert H_1(\omega_1)\rvert\cos[\omega_1 t + \theta_1]$ $A_2\lvert H_1(\omega_2)\rvert\cos[\omega_2 t + \theta_1]$ $A_3\lvert H_1(\omega_3)\rvert\cos[\omega_3 t + \theta_1]$
y_2	$½(A_1^2 + A_2^2 + A_3^2)\lvert H_2(0)\rvert$ $½A_1^2\lvert H_2(2\omega_1)\rvert\cos[(2\omega_1)t + \theta_2]$ $½A_2^2\lvert H_2(2\omega_2)\rvert\cos[(2\omega_2)t + \theta_2]$ $½A_3^2\lvert H_2(2\omega_3)\rvert\cos[(2\omega_3)t + \theta_2]$ $A_1A_2\lvert H_2(\omega_1 \pm \omega_2)\rvert\cos[(\omega_1 \pm \omega_2)t + \theta_2]$ $A_1A_3\lvert H_2(\omega_1 \pm \omega_3)\rvert\cos[(\omega_1 \pm \omega_3)t + \theta_2]$ $A_2A_3\lvert H_2(\omega_2 \pm \omega_3)\rvert\cos[(\omega_2 \pm \omega_3)t + \theta_2]$
y_3	$¾(A_1^3 + 2A_1A_2^2 + 2A_1A_3^2)\lvert H_3(\omega_1)\rvert\cos[2\omega_1 t + \theta_3]$ $¾(A_2^3 + 2A_1^2A_2 + 2A_2A_3^2)\lvert H_3(\omega_2)\rvert\cos[2\omega_2 t + \theta_3]$ $¾(A_3^3 + 2A_1^2A_3 + 2A_2^2A_3)\lvert H_3(\omega_3)\rvert\cos[2\omega_3 t + \theta_3]$ $¼A_1^3\lvert H_3(3\omega_1)\rvert\cos[(3\omega_1)t + \theta_3]$ $¼A_2^3\lvert H_3(3\omega_2)\rvert\cos[(3\omega_2)t + \theta_3]$ $¼A_3^3\lvert H_3(3\omega_3)\rvert\cos[(3\omega_3)t + \theta_3]$ $¾A_1^2A_2\lvert H_3(2\omega_1 \pm \omega_2)\rvert\cos[(2\omega_1 \pm \omega_2)t + \theta_3]$ $¾A_1^2A_3\lvert H_3(2\omega_1 \pm \omega_3)\rvert\cos[(2\omega_1 \pm \omega_3)t + \theta_3]$ $¾A_2^2A_1\lvert H_3(2\omega_2 \pm \omega_1)\rvert\cos[(2\omega_2 \pm \omega_1)t + \theta_3]$ $¾A_2^2A_3\lvert H_3(2\omega_2 \pm \omega_3)\rvert\cos[(2\omega_2 \pm \omega_3)t + \theta_3]$ $¾A_3^2A_1\lvert H_3(2\omega_3 \pm \omega_1)\rvert\cos[(2\omega_3 \pm \omega_1)t + \theta_3]$ $¾A_3^2A_2\lvert H_3(2\omega_3 \pm \omega_2)\rvert\cos[(2\omega_3 \pm \omega_2)t + \theta_3]$ $¾A_1A_2A_3\lvert H_3(\omega_1 \pm \omega_2 \pm \omega_3)\rvert\cos[(\omega_1 \pm \omega_2 \pm \omega_3)t + \theta_3]$

IV. CONCLUSION

In each term in (4) from higher order are contained harmonic and composite nonlinear products from higher order. Thus by examination of the nonlinear distortions from a particular order, it is not necessary to be exanimate the terms from lower order. But the opposite is compulsory, to be much more exact the obtained results.

Even only with three input signals (N=3) you can see, that the undesired vibrating in the output of the system are more then 30 (for n=3), and in one modern system for cable television number of the transferred channels is much bigger (about 100). From the aforesaid, it is not difficult to see that the problem with analyzing of the nonlinear distortions is too complicated and ambiguity solved, even in using of a modern fast speed computers.

REFERENCES

[1] J. Steel, A. Parker, D. Skellern," Characterization of cable amplifiers for broadband network applications. Electronics Department", Macquarie University, NSW 2109, Australia, pp. 39-45, 2001.

[2] V. Germanov,"Calculating the CSO/CTB Spectrum of CATV Amplifiers and optical Receivers", IEEE Trans. on Broadcasting, vol.44 No.3 pp. 363-366, Sep.1998.

[3] J. Sjoberg, Q. Zang, L. Ljung, A. Benveniste and etc. "Nonlinear Black-box modeling in System Identification: a Unified Overview", Automatica, vol.31, pp.1961-1724, 1995.

[4] V. Cherkassky, F. Mulier, *Learning from Data*. John Wiley, 1998.

[5] M. C. Jeruchim, P. Balaban and K. S. Shanmugan,*Simulation of Communications System: Modeling, Methodology, and Techniques*. Second edition, Kluwer Academic/Plenum Publishers, New York, 2000.

[6] G. J. Foshini, A.A.M. Saleh, "Overcoming Optical Amplifier Intermodulation Distortion Using Coding in Multichannel Communications Systems", IEEE Trans.on commun., vol. 38, No.2, pp. 187-191, Feb.1990.

[7] I. W. Sanberg, "On Volterra expansions for time-varying nonlinear systems", IEEE Trans. on Circuits and Systems, vol.CAS-30, No.2, pp. 61-67, Feb. 1983.

[8] Y. Wan, T. Dodd and R. Harrison. "Infinite degree Volterra series estimation". The University of Sheffield, S1 3JD, UK, 2003.

[9] J. C. Pedro, N.B. Garvalho, R.C. Madureira, "Nonlinear distortion model for VCO-PLL FM transmission system", IEEE Trans. on Circuits and Systems, vol.46, No.3, pp. 348-352, Mar. 1999.

[10] M. Weiss, C. Evans, D. Rees, "Identification of nonlinear cascade systems using paired multisine signals". Proc. IEEE Inst. and measur. tech. conf., Canada, pp. 765-770, 1997.

[11] S. A. Maas, "Analysis and optimization of nonlinear microwave circuits by Volterra series", Microwave J., vol.33, No.4, pp. 245, April 1990.

Intermodulation Composite Distortions Theoretically Research in HFC Networks

Oleg Borisov Panagiev[1]

***Abstract* – It is shown that Volterra operator techniques play a key role to study the main structural properties of the CATV/HFC system under consideration. Using the presentation of this operator in the ring of power series allows us to apply some algebraic methods for research.**

***Keywords* – Composite distortions, HFC, trunk amplifier, Volterra series.**

I. INTRODUCTION

During the exploitation of the advanced (HFC – hybrid fiber coaxial) systems for cable television (fig.1) at definite conditions in the active devices, used in the system, spring up nonlinear products. Thus at the subscribers besides the useful signal, proceed and parasitical (disturbing) signals, whose in many cases of big nonlinear prevail over the useful signal. The nonlinearity is being approximated most often with polynomial up to third rank, because of theoretical examinations by mathematical models are too successfully realizable [1], [2].

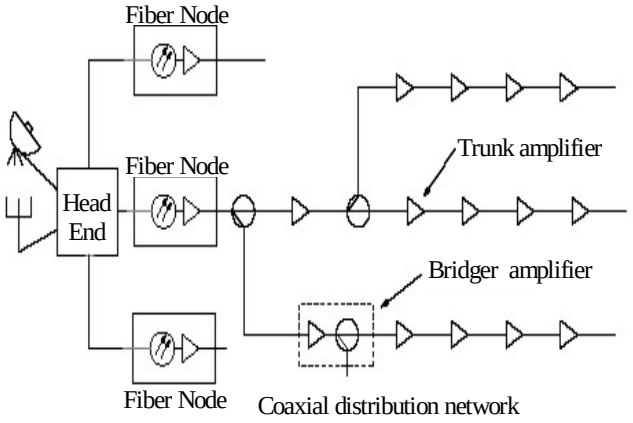


Fig.1. Advanced CATV/HFC system

For getting of optimal results while using the Volterra series is needed profound understanding of the basic methods and way of modeling of the nonlinear systems, witch are presented as a combination of linear and nonlinear subsystems (parts, elements).

The biggest defect of the methods for analyzing by the Volterra series is that the difficulty for finding a resolution at passing of the system from a regime with "weak" nonlinearity to a regime with "strong" nonlinearity [3], [4].

There are some symptoms, which could help us to definite that the initial conditions are out of the admissible borders. For example with increasing of the signal, the intermodulation products do not get into a regime of saturation, but products from transforming of the basic frequencies are weakly saturatable [4].

II. THE RELATIONSHIPS BETWEEN COMPOSITE DISTORTIONS AND SYSTEM PARAMETERS

Below are brought out the relationships that concerns CATV system, which works on a "weakly" nonliner regime, and the polynomial with which the system is definite is from third order. The output signal of the system is from that type:

$$y(t) = \sum_{n=1}^{3} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} h_n(\tau_1,\dots,\tau_n) \prod_{i=1}^{n} x(t-\tau_i) d\tau_i \dots d\tau_n , \qquad (1)$$

where $h_n(\tau_i)$ is the Volterra kernel, n = 1, 2, 3 and $\tau \le t$.
Let's apply, in the accepted model, the approximation of third order when the input signal is

$$x(t) = \sum_{i=1}^{N} A_i \cos(\omega_i t + \theta_i); \omega_i = 2\pi f_i . \qquad (2)$$

Let's replace with (2) in (1) and do the definite mathematical operation: presenting of $\cos\omega_i t$ like exponential function, applying the direct Fourier's' transformation, etc.

In this case in the output of the system we receive

$$y(t) = \sum_{i=1}^{N} A_i \left|H_1(\omega_i)\right| \cos(\omega_i t + \theta_i)$$

$$+ \tfrac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j \left|H_2(\omega_i \pm \omega_j)\right| \cos\left[(\omega_i \pm \omega_j)t + \theta_2\right]$$

$$+ \tfrac{1}{4} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} A_i A_j A_k \left|H_3(\omega_i \pm \omega_j \pm \omega_k)\right| \cos\left[(\omega_i \pm \omega_j \pm \omega_k)t + \theta_3\right] \qquad (3)$$

[1]Oleg B. Panagiev is with the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com

We can see that the Volterra kernels $|H_n(\omega)|$ and the phase corner $\theta_n(\omega)$ are frequency dependant. The valuation of the intermodulation distortions (IMD) in CATV systems is reduced on the base of definite attitudes for the composite distortions from second and third order. For the gain G of advanced amplifiers can be accepted that it is constant in the entire frequency band [5].

A. The estimate of the composite distortions from second order

For estimate of IMD from second order is determinate the ratio of double beat *CSO* (*composite second order*), that is definite as attitude between the level of the basic signal and that, got as a result of totals and differences beating from second order (expressed in dB), i.e.

$$CSO = 20\lg\frac{A|H_1(\omega_b)|}{A^2|H_2(\omega_b)|} \quad (4)$$

where we assume that:

- It is chosen a test channel with ω_r, r = 1, 2, 3... N;
- All levels are even $A_i = A_j = A_k = A$;
- It is accepted a frequency beating $\omega_b = \omega_i \pm \omega_j$, $\omega_b = \omega_i \pm \omega_k$, $\omega_b = \omega_j \pm \omega_k$, while $i \neq j \neq k$ and i,j,k=1,2,3…;
- Volterra kernels $|H_2(\omega_b)|$ are middle value, because of frequency dependence of the device inertia;
- The gain $G = |H_1(\omega_b)|$ is constant for whole spectrum of CATV system.

Then it could be written

$$CSO = 20\lg\frac{G}{|H_2(\omega_b)|} - 20\lg A. \quad (5)$$

When is known the output level of the system (the amplifier) A_{out} and the gain G, as it is concerned that

$$A_{out} = GA, \quad (5a)$$

from expression (5) at a normal Volterra kernel

$$CSO = 40\lg G - 20\lg|H_2(\omega_b)| - 20\lg A_{out}. \quad (6)$$

The first addend in the right part of expression (5) represents reciprocal value of the normalized Volterra kernel from second order or

$$H_{2G} = \frac{|H_2(\omega_b)|}{G}. \quad (7)$$

In association with (5) and (6) for CSO, expressed by the normalized Volterra kernel can be written

$$CSO = -20\lg H_{2G} - 20\lg A. \quad (8)$$

When A_{out} and G for CSO are known, expressed with the normalized kernel, it is obtaining

$$CSO = 20\lg G - 20\lg H_{2G} - 20\lg A_{out}. \quad (9)$$

For system with M numbers of cascade related identical amplifiers

$$CSO_{(M)} = CSO_{(1)} - 10\lg M. \quad (10)$$

B. The estimate of the composite distortions from third order

At similar way it is possible to be definite the ratio of beating from third order CTB (*composite triple beat*), like an attitude between the level of a basic signal and that, got as a result of total and difference beats from third order (expressed in dB), i.e.

$$CTB = 20\lg\frac{A|H_1(\omega_b)|}{\frac{3}{4}A^3|H_3(\omega_b)|}, \quad (11)$$

where we assume that, different from this presented in definition of CSO:

- Beating is from type $\omega_b = 2\omega_i \pm \omega_j$, $\omega_b = 2\omega_i \pm \omega_k$, $\omega_b = 2\omega_j \pm \omega_k$;
- Volterra kernels $|H_3(\omega_b)|$ are middle value, because of frequency dependence of the inertial devices.

Then it could be written

$$CTB = 20\lg\frac{4}{3} + 20\lg\frac{G}{|H_3(\omega_b)|} - 40\lg A. \quad (12)$$

When is known the output level of the system (the amplifier) A_{out} and the coefficient of delivering G, and using (5a), for CSO at two-component beat is validity the following formula

$$CTB = 60\lg G - 20\lg|H_3(\omega_b)| - 40\lg A_{out} + 2{,}5. \quad (13)$$

The second addend in the right part of expression (12) represents reciprocal value of the normalized Volterra kernel from third order or

$$H_{3G} = \frac{|H_3(\omega_b)|}{G}. \quad (14)$$

In association with (12) and (14) for CTB it gets the following appearance:

$$CTB = -20\lg H_{3G} - 40\lg A + 2{,}5. \quad (15)$$

For familiar A_{out} and G for CTB, expressed with the normalized kernel we note down:

$$CTB = 40\lg G - 20\lg H_{3G} - 40\lg A_{out} + 2{,}5\,. \qquad (16)$$

When the beating from third order is from the following type $\omega_b = \omega_i \pm \omega_j \pm \omega_k$, expression (12) takes the following look:

$$CTB = 20\lg\frac{2}{3} + 20\lg G - 20\lg\left|H_3(\omega_b)\right| - 40\lg A\,, \qquad (17)$$

and expression (13) looks like that:

$$CTB = 60\lg G - 20\lg\left|H_3(\omega_b)\right| - 40\lg A_{out} - 3{,}52\,. \qquad (18)$$

The composite distortion from third order for three-component beat, expressed with the normalized kernel, has the following look:

$$CTB = -20\lg H_{3G} - 40\lg A - 3{,}52 \qquad (19)$$

$$CTB = 40\lg G - 20\lg H_{3G} - 40\lg A_{out} - 3{,}52\,. \qquad (20)$$

For system with M numbers of cascade related identical amplifiers

$$CTB_{(M)} = CTB_{(1)} - 20\lg M\,. \qquad (21)$$

III. Calculating of the composite intermodulation distortions

Here we will show the application of the worked out higher up analytic expressions for determination of the composite distortions for a real cable amplifier VX97B/G. It's a product of the German WISI company and it has been used in the trunk lines. His working frequency outflanking is from 47 to 850 MHz, gain G = 35 dB, and maximal output level A_{out} = 121 dBμV [6] (for two programs). Volterra kernels are determinate according to the European standard CENELEC EN 50083 (Appendix C) and the method is proposed in [7]. The results for the Volterra kernels from second and third order (for two- and three-component beat) are presented in Table 1.

From expressions (6), (13), (18) and the technical data for trunk amplifiers (output level and gain) we took out the simplified formulas, with witch we will calculate the composite distortions in dependence from the working frequency outflanking of the cable distribution network (47 to 520 MHz).

1) Composite distortions from second order

$$CSO = -51 - 20\lg\left|H_2(\omega_b)\right| \qquad (22)$$

2) Composite distortions from third order (three-component beat)

$$CTB^{(3)} = -140{,}52 - 20\lg\left|H_3^{(3)}(\omega_b)\right| \qquad (23)$$

TABLE 1

Volterra kernels / f MHz	Second order	Third order	
		$f_i \pm f_j \pm f_k$	$2f_i \pm f_j$
	H_2	$H_3^{(3)}$	$H_3^{(2)}$
49,75	2,04E-06	9,00E-11	1,80E-10
119,25	2,12E-06	8,69E-11	1,74E-10
175,25	2,06E-06	8,10E-11	1,62E-10
191,25	2,00E-06	8,70E-11	1,74E-10
207,25	2,62E-06	8,90E-11	1,78E-10
223,25	2,31E-06	9,06E-11	1,81E-10
231,25	1,48E-06	8,30E-11	1,66E-10
247,25	2,31E-06	8,63E-11	1,73E-10
263,25	2,21E-06	8,75E-11	1,75E-10
287,25	2,04E-06	7,99E-11	1,60E-10
311.25	2,31E-06	8,63E-11	1,73E-10
327.25	2,06E-06	8,81E-11	1,76E-10
343.25	2,02E-06	7,00E-11	1,40E-10
359.25	1,92E-06	7,43E-11	1,49E-10
375.25	2,01E-06	6,71E-11	1,34E-10
391.25	2,31E-06	6,51E-11	1,30E-10
407.25	2,21E-06	7,90E-11	1,58E-10
479.25	1,92E-06	7,43E-11	1,49E-10
495.25	1,68E-06	7,71E-11	1,54E-10
511.25	1,62E-06	7,81E-11	1,56E-10

3) Composite distortions from third order (two-component beat)

$$CTB^{(2)} = -134{,}5 - 20\lg\left|H_3^{(2)}(\omega_b)\right| \qquad (24)$$

Numbered results are given in Table 2, and the graphic presentation is shown on fig.2.

On fig.3 is presented the nomogram for determination of CSO and CATV for f = 175.25 MHz, and A_{out} is changing from 90 dBμV to 120 dBμV. G gets three values: 15 dB, 25 dB and 35 dB. The results are given after replacement with the respective values in expressions (6) and (13), while the simplified formulas (25), (26) are graphical interpreted by MAPLE 6.

$$CSO = 2 * G - A_{out} + 113{,}72\,; \qquad (25)$$

$$CTB = 3 * G - 2 * A_{out} + 198{,}31\,. \qquad (26)$$

IV. Conclusion

The presented theoretical results for the composite intermodulation distortions, successfully can be applied and in the practice. That was demonstrated for one concrete trunk amplifier, but they can be also used for other – subtrunk and house amplifiers.

The presented values in Table 2 and on fig.2 with big accuracy corresponds with the pointed out in [6], witch confirms the obtained results from the theoretically drawing formulas.

TABLE 2

f *MHz*	Intermodulation composite distortions		
	CSO *dB*	*CTB*$^{(3)}$ *dB*	*CTB*$^{(2)}$ *dB*
49,75	62,81	60,40	60,41
119,25	62,47	60,70	60,72
175,25	62,72	61,31	61,30
191,25	62,98	60,69	60,67
207,25	60,63	60,49	60,42
223,25	61,73	60,34	60,30
231,25	65,59	61,10	61,13
247,25	61,73	60,76	60,74
263,25	62,11	60,64	60,60
287,25	62,81	61,43	61,48
311.25	61,73	60,76	60,78
327.25	62,72	60,58	60,60
343.25	62,89	62,58	62,60
359.25	63,33	62,06	62,10
375.25	62,94	62,94	62,91
391.25	61,73	63,21	63,26
407.25	62,11	61,53	61,57
479.25	63,33	62,06	62,10
495.25	64,49	61,74	61,77
511.25	64,81	61,63	61,68

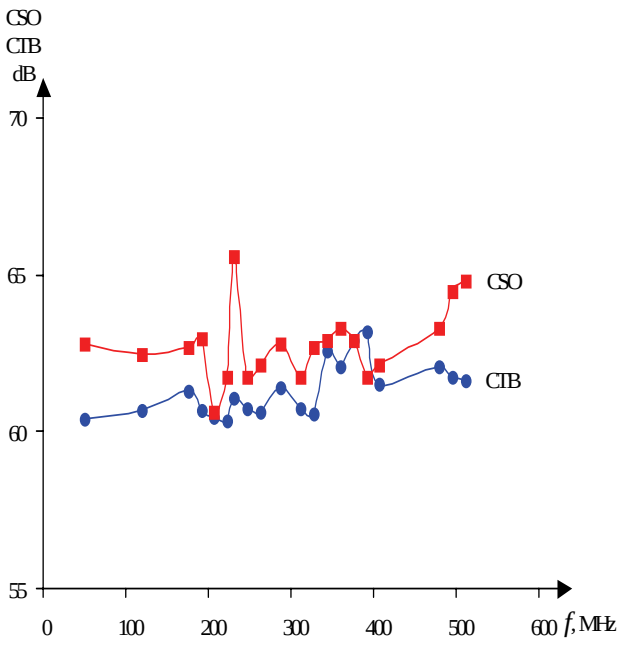


Fig.2. Graphic presentation of the composite distortions from the working frequencies according to Appendix C

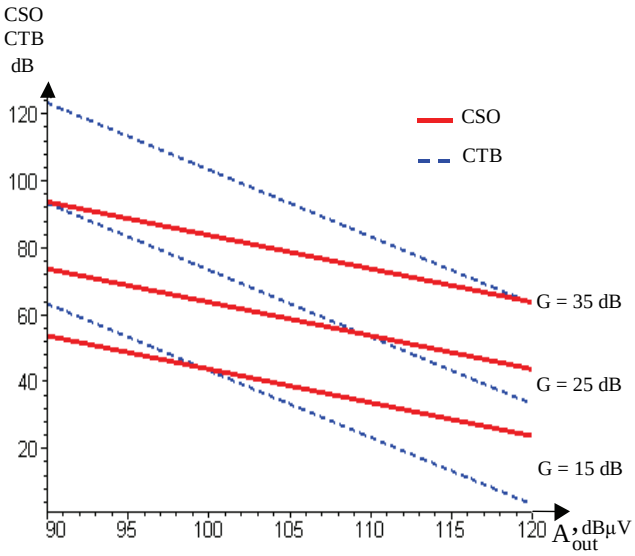


Fig.3. The nomogram for determination of CSO and CATV in dependence from the output/input level and gain on the trunk amplifier VX97B/G

From the nomogram it can be reported values for the composite distortions at different values for the gain and output level, respective for the input level. Let's remind that by exploitation of the CATV systems, the trunk amplifiers work with definite gain. For some reasons, if it gets bigger level on the input, the level in the output also will grow up, while CSO and CTB will change, too. These changes can be reviewed from the nomogram on fig.3 in case when there are three values for G, but they could be multiplicated in dependence of needs of the concrete cable operator.

REFERENCES

[1] J. Steel, A. Parker, D. Skellern, "Characterization of cable amplifiers for broadband network applications", Electronics Department, Macquarie University, NSW 2109, Australia, pp. 39-45, 2001.

[2] W. Ciciora, *Cable Television in the United States - An Overview*. Cable TV Labs. Inc., Louisville, 1999.

[3] D. Gingold, "Integrated digital services for cable networks. Academic Dissertation", MIT, Sept. 1996.

[4] S.A. Maas, "Analysis and optimization of nonlinear microwave circuits by Volterra series". Microwave J., vol.33, No.4, pp. 245, April l990.

[5] Ultra Low Noise and Distortion High-Speed Amplifier for 16-Bit Systems. www.analog.com, pp. 1-12, Dec. 2001.

[6] WISI. "A link to the future". Catalogue'2001.

[7] O. Panagiev, "Analysis and reduction on nonlinear distortions for signals in the broadband cable communication systems". Dissertation, Technical university of Sofia, 2006.

[8] European standard CENELEC "EN 50083", 1998.

Influence of the Refraction in Presence of Various Formed Obstacles

Emil Savov Altimirski[1] and Kliment Nikolaev Angelov[2]

I. INTRODUCTION

Electromagnetic wave propagation in the presence of a wedge-shaped screening obstacle is a well studied phenomenon with corresponding derived mathematical models describing it. It is interesting to examine the field at the receiving point behind an obstacle with non – regular form. It could cause damping or amplification of the field in accordance to the appropriate Fresnel zones and their corresponding degree of closure.

The aim of this examination is to find cases, where obstacle with given form and dimensions is able to significantly influence signal level at receiving point in cases of normal wireless channels and to elaborate a technical method for its rapid determination.

II. THEORETICAL FUNDAMENTALS

A case (fig. 1) is being studied, where in a given cross-section of the wireless channel between transmitter (point A) and receiver (point B) a screening obstacle with a form different from a line lies.

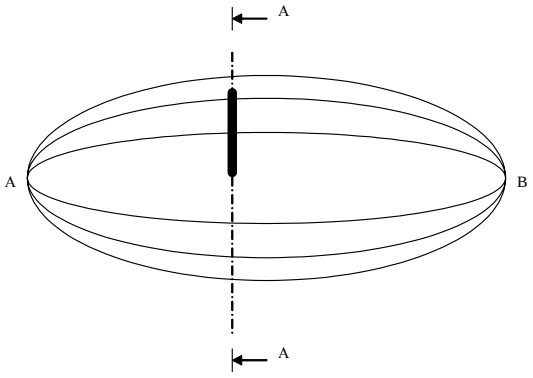


Fig. 1

The surface of the cross–section, depicted in fig. 2 is being divided into a matrix, consisting of points (rectangles with sufficiently small dimensions). The field is being retransmitted by each point of the matrix in accordance with the Huegens-Fresnel principle. Field amplitude and phase created by each of these points in the receiving point is being derived. Subsequently these single fields are being summed taking into account their corresponding amplitudes and phases in order to derive the resultant field. The result is a complex number, whose modulus is important to be defined for engineering praxis purposes.

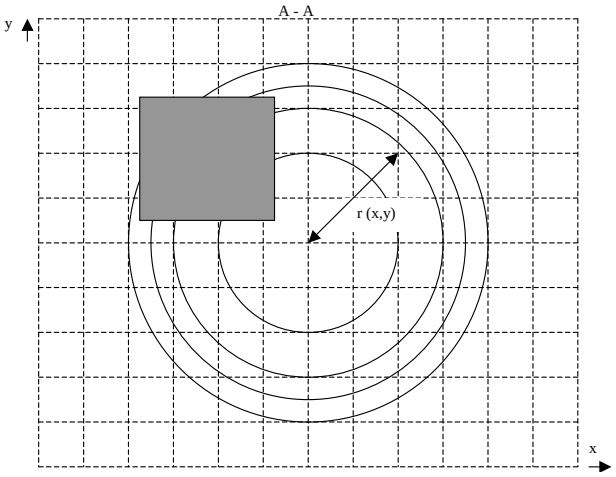


Fig.2

Elementary beams amplitude could be described by means of the following equation (1):

$$|E_r| = \frac{1+\cos[\theta(r)]}{l(r)}, \qquad (1)$$

where ***l(r)*** is the elementary beam way, ***θ(r)*** is the visual angle for the receiving point in accordance with the primary trajectory.

Phases could be derived by means of the following equation (2):

$$\varphi_r = \frac{2.\pi}{\lambda}.\Delta l(r), \qquad (2)$$

where **Δl(r)** is the difference in the ways of the elementary and direct beams.

The following equation is being derived for the resultant field in the receiving point (3):

[1] Ph.D. Emil Savov Altimirski is with Technical University of Sofia, +359 2 965 2230
[2] Kliment Nikolaev Angelov is with Technical University of Sofia, +359 2 965 3447, E-mail: klimbona@gmail.com

$$E_0 = \sum_{r=0}^{r=r\max} |E_r| . e^{j\varphi_r} \qquad (3)$$

where **r** is the current radius of the corresponding matrix point to the centre of the Fresnel zones and is function of **x** and **y** (fig. 2).

The obstacle is represented as a matrix m, consisting only of elements equal to 0 or 1, thus reflecting if a given elementary beam retransmits or not (caused by the obstacle) in receivers point direction. Taking into account, that **r = f(x,y)**, the overall field at the receiver point could be described by means of the following equation (4):

$$E_{0obstacle} = \sum |E_{(x,y)}| . e^{j\varphi(x,y)} m(x,y) \qquad (4)$$

III. RESULTS

In accordance with the above described method, an examination of the elementary beams field amplitude and phase distribution depending on their corresponding trajectory in the case of an obstacle absence has been carried out. The derived results are depicted in fig. 3 and fig. 4. The part in fig. 3, that juts out corresponds to a higher amplitude value, caused by a shorter way of the beams and respectively smaller degree of natural wave energy dispersion. The sectors in fig. 4 that jut out correspond to phases with a value close to 180 degrees, and the hollow ones – to 0.

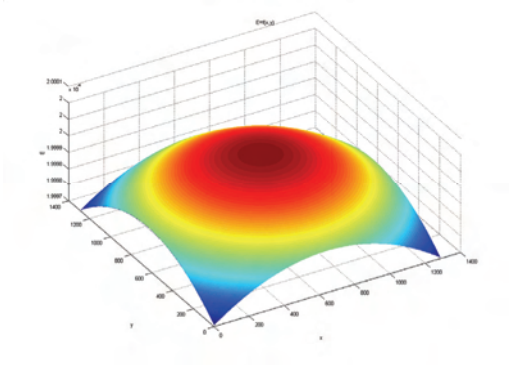


Fig. 3

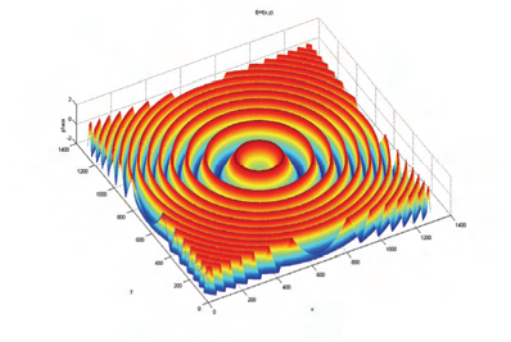


Fig. 4

Fig. 5 represents the influence of a screening obstacle with different forms in the surface of the first ten Fresnel zones. From the figure it is visible that when the center of the obstacle is being moved away from Fresnel zones center at a distance three times greater than the radius of the first Fresnel zone, the influence of the obstacle itself is so week, that it could be further neglected for engineering calculations purposes. In accordance to the above statement and for the reason of a better readability of the figures, only the results, corresponding to distances of a thrice first Fresnel zone dimensions magnitude are being represented further in this work.

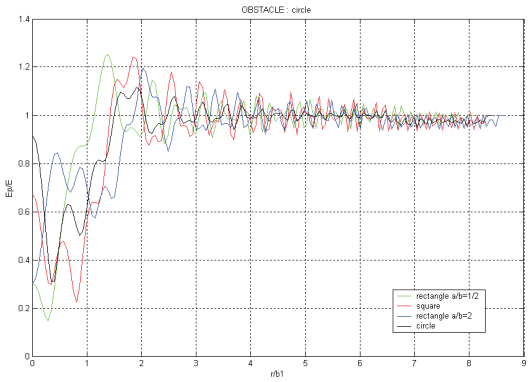


Fig. 5

The influence of an obstacle with a square (fig.6), rectangular 1:2 (fig. 7), as well as rectangular 2:1 (fig.8) forms with different side dimensions has been examined.

It is obvious from the results, that in the case of greater screening object dimensions, damping curve becomes similar to that of a wedge–shaped obstacle. Significant field amplification in the case of a rectangular (2:1) hindrance by a 2,5 – 3 times first Fresnel zone radius dislocation is observed. This phenomenon could be useful from a passive retranslation point of view.

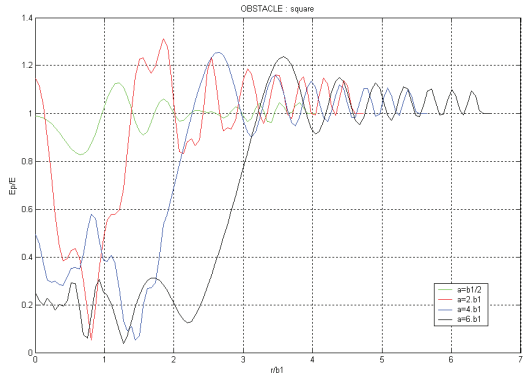


Fig. 6

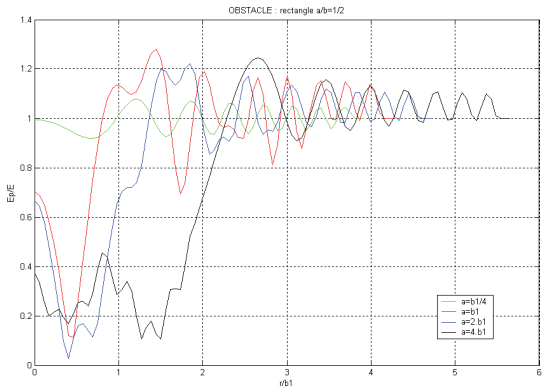


Fig. 7

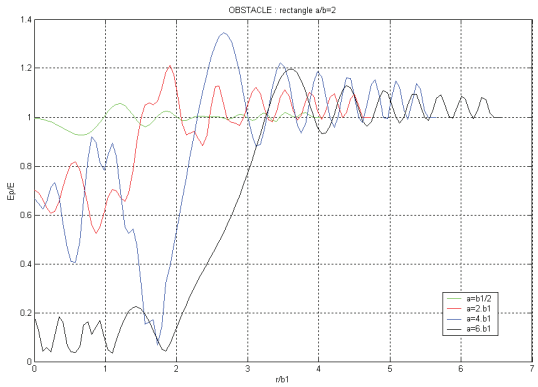


Fig. 8

A comparison of different square dislocations, depending of its side dimensions is being carried out in fig. 9. A greater field peak is being observed in the case of an approximate equality between square side and first Fresnel zone radius and a dislocation two times greater than the radius.

Fig. 10 represents the influence of objects with a different form and the same surface, equal to that of the first Fresnel zone. A similarity between the curves corresponding to square and circle, differing from those, typical for rectangulars is being observed.

An algorithm and a program for MatLab, given in the appendix have been elaborated. They could be used for field determination in the presence of a obstacle with arbitrary form and different distances to receiver and transmitter.

The results from the calculations for determining the influence of an obstacle with a complicated geometric form, depending on the distance between the obstacle and the transceiver, are illustrated on Fig. 11. The form of the complex obstacle is shown on Fig. 12.

The presented method gives the opportunity not only to determine the field in the receiver and to optimize the heights of the transceiving and receiving antenna, but also to search for possibilities for passive retranslation and to calculate the geometric dimensions of the retranslators.

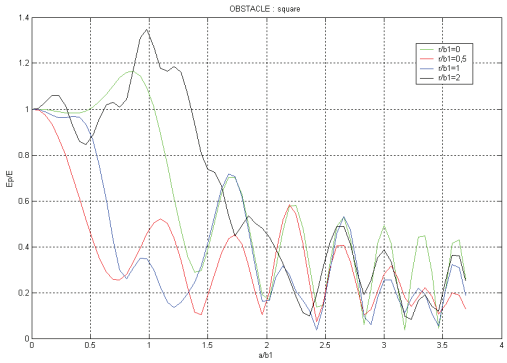


Fig. 9

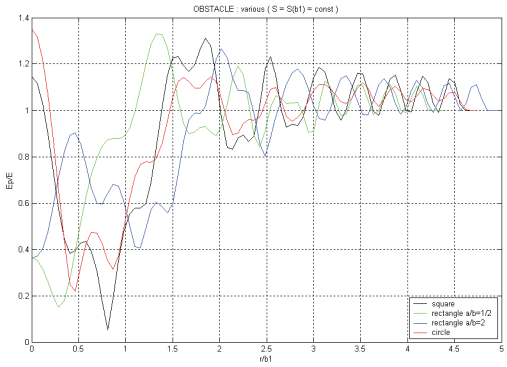


Fig. 10

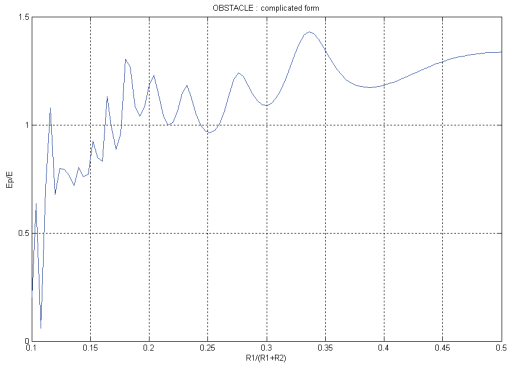


Fig. 11

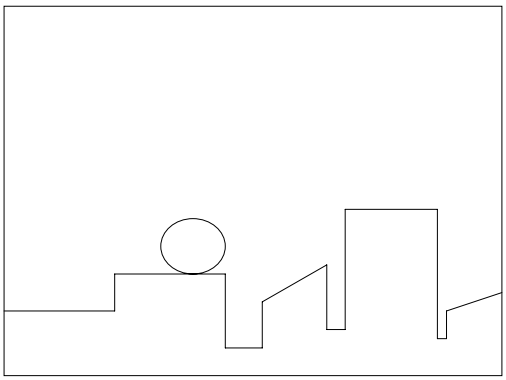


Fig. 12

APPENDIX

```
lm=0.03;
k=32;
s=1;


for p=0:100
rt(p+1)=40*p+1000;
r1=rt(p+1);
r2=10000-r1;

x=-k:s:k;
y=-k:s:k;
[x,y]=meshgrid(x,y);

r=sqrt(x.^2+y.^2);
l1=sqrt(r.^2+r1.^2);
l2=sqrt(r.^2+r2.^2);
dl=l1+l2-r1-r2;
alfa=atan(r./r1);
beta=atan(r./r2);
tao=alfa+beta;

e=(1+cos(tao))./(l1+l2);
f=(2*pi*dl)./lm;
ek=e.*(cos(f)+(i*sin(f)));

e0=0;
for q=1:(2*k./s)+1
for j=1:(2*k./s)+1
e0=e0+ek(q,j);
end
end

e0m=abs(e0);

e0r=0;
for q=1:(2*k./s)+1
for j=1:(2*k./s)+1
e0r=e0r+((ek(q,j))*m(q,j));
end
end


e0mr=abs(e0r);

Lost(p+1)=e0mr./e0m;
end

r1n=rt/10000;
plot(r1n,Lost);
title('OBSTACLE : complicated form');
xlabel('R1/(R1+R2)');
ylabel('Ep/E');
```

GPS/INS Automotive Tracking System

Rossen Miletiev[1] and Rumen Arnaudov[2]

Abstract: **The proposed GPS/INS automotive tracking system includes base station with map software and mobile stations. The combination of the GPS receivers, GPRS service communication, MMC/SD card, micromechanical sensors and map software allows offline and real – time tracking, motion analysis and possible construction of an automotive black box to visualize the mobile station on the map or to reproduce the road situation by a virtual reality at the road accidents. The proposed system is capable to update the navigation data (1Hz) and inertial data (up to 2560Hz) to design an Extended Kalman Filter with very high accuracy due to the high inertial data update rate.**

Keywords: **GPS, micromechanical sensors, MMC/SD card**

I. INTRODUCTION

The realization of the GPS/INS automotive tracking system evolves from the necessity to solve complex transportation tasks for the overloaded transportation traffic in the big cities. The system purpose is connected with the location of the moving objects by using GPS receivers and radio transmission connection like GSM or radio-link network. The combination of the GPS positioning system with the map software enables to find the optimum solutions and allows the quick and adequate decisions of the transportation tasks [1]-[2]. In the same time the inertial system (INS) allows monitoring the automotive drive style, dangerous places and route conditions.

The MEMS accelerometers and gyroscopes are both sensors which can perfectly address active safety systems in the automotive domain. Control of car roll-over, vehicle stability for skidding and antilock braking, parking brake energy, activation of wheel pressure monitoring, suspension adaptation to car and road condition, and other ones are all features are extending more and more any active safety system will be embedded soon in the cars.

Moreover MEMS sensors allows instantaneous detection of any information used by those functions, with an high level of precision and accuracy, with a very low impact in term of space, together with high level of integration with other systems.

Because the Global Positioning System (GPS) and an Inertial Navigation System (INS) complement each other, it is common practice to integrate the two systems in applications calling for continuous high accuracy and reliability [3]-[4].

[1]Rossen Miletiev is with the Technical University of Sofia, 8 Kl. Ohridski Blvd, 1000 Sofia, Bulgaria
E-mail: miletiev@tu-sofia.bg
[2]Rumen Arnaudov is with the Technical University of Sofia, 8 Kl. Ohridski Blvd, 1000 Sofia, Bulgaria
E-mail: RA@tu-sofia.bg

Kalman filtering is the basis for correcting the INS with GPS measurements of satellite range and range-rate. At the same time, the INS provides smooth and accurate short-term measurements of acceleration and velocity that can be used to aid GPS receiver code and carrier tracking [5]. Together the two systems permit improved navigation accuracy especially when GPS is degraded or interrupted because of jamming or interference [6].

The proposed system is capable of measuring acceleration over a bandwidth of 640 Hz for all axes [7] (update rate up to 2560Hz), while the bandwidth of the analogue accelerometer sensors is limited to 100Hz. The higher update rate may significantly decrease the integration errors due to the approximation error reduction. The proposed system also allows full tracking report and analysis and improves the route safety and security.

II. SYSTEM ARCHITECTURE

The block diagram of the GPS/INS tracking system is shown at Figure 1. It consists from the following main modules:

- *Power supply block*. The module transforms the input DC voltage to 3,3V (accelerometer power supply and Vdd_IO lines) and 5,0V (microcontroller and gyroscope power supply). The designed power supply is based on the LDO regulators to produce a low ripple output voltage.
- *GPS receiver*. The module sends the appropriate NMEA 0183 messages to the master board microcontroller;
- *Inertial system*. The module scans the 3 - dimensional linear accelerometer and angular rate sensor (gyroscope) to establish the current object accelerations. The inertial system is upgrade to the basic GPS system and may be installed in the same system case;
- *MMC/SD card*. It stored the GPS RMC messages and inertial sensor data as an archive information (backup function), which allows offline tracking analysis. This module is optional and supports MMC/SD cards up to 2GB (FAT16 file system format);
- *GSM modem*. The module sends the navigation and inertial sensor data via GSM network by GPRS service using the embedded TCP/IP stack to perform a real time tracking function;
- *Serial interface*. The system supports some serial interfaces to enable the communication with external devices, especially with computer based platforms. The supported serial protocols include RS232, RS422 and synchronous serial transfer.

- *USB interface*. The Serial-to-USB module includes FTDI single chip IC, which accomplished the asynchronous serial data transfer to PC according to USB 1.1 and 2.0 protocol specification. The module has UHCI/OHCI/EHCI host controller compatibility.
- *I/O ports*. The optional I/O digital and analog ports allow connecting the additional external sensors such as temperature sensors, fuel level sensors, passenger detectors, etc. The analog inputs include the 10-bit analog microcontroller channels and the both 12-bit analog gyroscope inputs AN0 and AN1. The both microcontrollers support I/O ports to ensure the low and high priority of the external sensors.

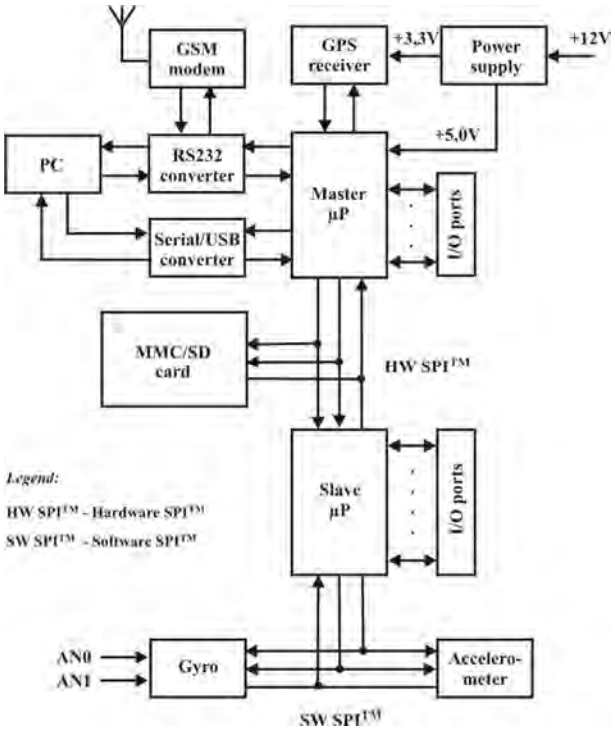


Fig.1. System architecture

The final assembly of the GPS/INS system is shown at Figure 2.

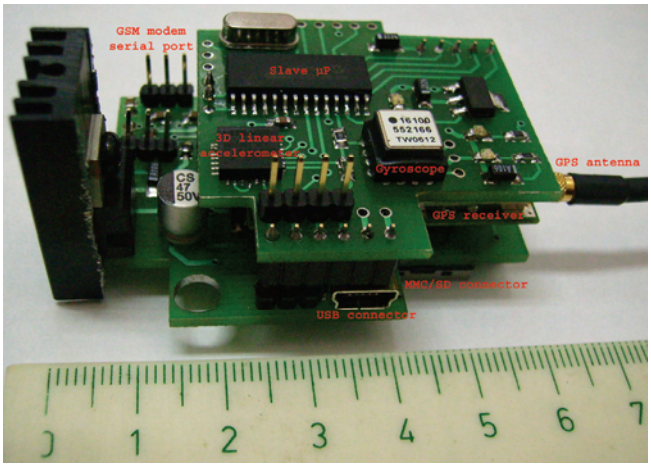


Fig.2. GPS/INS tracking system assembly

III. Accelerometer Design

The accelerometer sensor is based on LIS3LV02DQ – a three axes digital output linear accelerometer, produced by *ST*. This sensor includes a sensingelement and an IC interface able to take the information from the sensing element and to provide the measured acceleration signals to the external world through an I^2C/SPI serial interface. The LIS3LV02DQ has a user selectable full scale of ±2g, ±6g and it is capable of measuring acceleration over a bandwidth of 640 Hz for all axes [7].

The complete measurement chain is composed by a low-noise capacitive amplifier which converts into an analog voltage the capacitive unbalancing of the MEMS sensor and by three ΣΔ analog-to-digital converters, one for each axis, that translate the produced signal into a digital bitstream (Figure 3) [7].

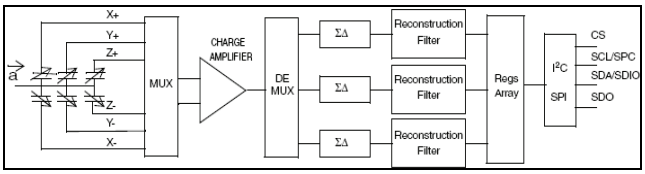


Fig.3. Block diagram of the three axes digital output linear accelerometer [7]

The gyroscope sensor is based on ADIS16100 - a complete angular rate sensor that uses Analog Devices' surface-micromachining process to make a functionally complete angular rate sensor with an integrated serial peripheral interface (SPI). The digital data available at the SPI port is proportional to the angular rate about the axis normal to the top surface of the package. This sensor is able to measure angular rates up to 300°/s while the sensitivity is equal to 0,244°/s/LSB, noise density – 0,10°/s/√Hz rms and bandwidth – up to 40Hz. Also an access to an internal temperature sensor measurement is provided for compensation techniques. Two pins are available to the user to input analog signals for digitization. An additional output pin provides a precision voltage reference [8].

IV. Operation Algorithm Description

The operation algorithm of the GPS/INS tracking system should be capable to:

- Receive all NMEA 0183 messages;
- Read and save the linear accelerometer data according to the selected update rate;
- Read and save the angular rate data, temperature value and scan the both gyroscope 12 – bit analog channels;
- Read and save the I/O port states of the master and slave microcontroller;
- Write all data to MMC/SD card.

Due to the very complex system tasks the operations are shared to the both microcontrollers. The master one reads the GPS navigation data, scans its I/O ports and writes the data to the MMC/SD card while the slave microcontroller controls the inertial block and sends the data on request.

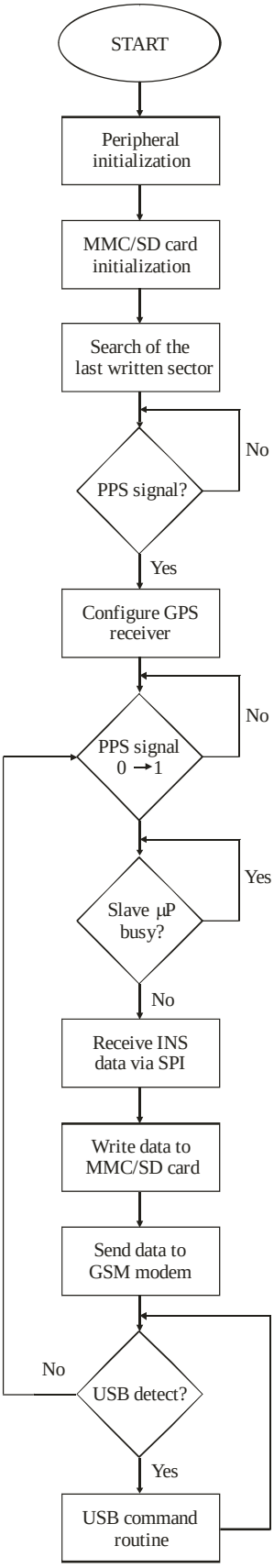


Fig.3. Master μP operation algorithm

The operation algorithm is closely connected with the total number of the navigation and inertial data, analog voltage values and temperature data. The stored data are divided at the following sections:

- Navigation data – up to 80 bytes. The RMC NMEA messages are allocated here;
- Inertial data (320 bytes). This space stores the inertial data (240 bytes) and gyroscope data (80 bytes). The update rate for the inertial block is equal to 40Hz;
- Analog data (100 bytes). The selected space stores the data from the build in 10 – bit ADC. The slave microcontroller may convert up to 5 analog channels while the master one – up to 6. The sampling rate may be independently chosen for each channel;
- Gyroscope peripheral data (6 bytes). The temperature and the analog channel values are updated ones per second.
- Reserved data (6 bytes).

The total amount of data is equal to 512 bytes. This allows reading and writing one MMC/SD card sector (512 bytes) every second, so the navigation data will be saved in the beginning of the sectors which allows a fast data processing.

The operation algorithms for master and slave microcontroller are shown at Figure 3 and Figure 4 respectively.

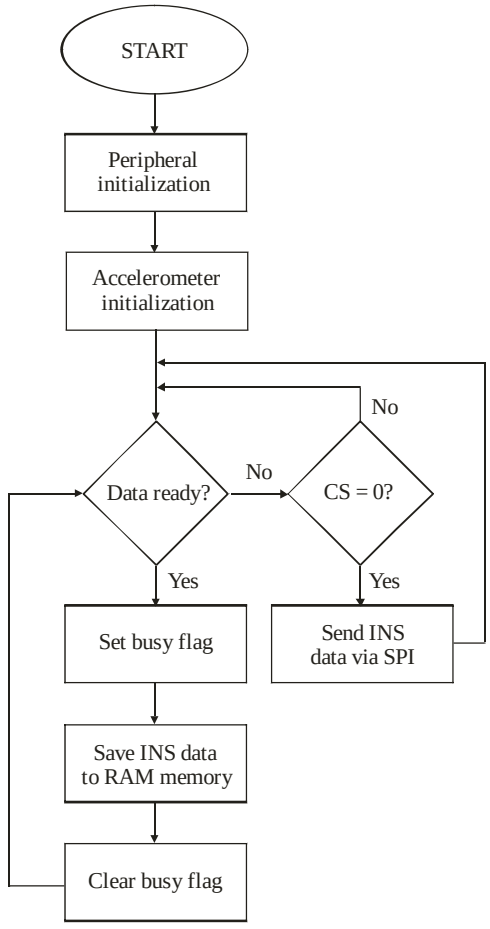


Fig.4. Slave μP operation algorithm

V. Conclusion

The used GPS receivers allow specifying the object location with very high precision – the location error is less than 10m, which is completely sufficient to solve the complex transportation tasks. In the presence of the optional inertial module the system solves the complex analysis of the moving object state – brake system status, vibrations, risky driving, etc. Therefore, the system may be used as a base for automatic obtaining of the actual information, formation of the statistic database, which described the transport traffic, optimal transportation routes calculation and optimal traffic management.

The additional information permit to reinforce the object control according to the following points:

- real – time object tracking;
- speed limitation control;
- driver style management;
- optimization of the transportation traffic.

The main system advantage consists in the integrated inertial sensors, which allows building motor vehicle event data recorder (MVEDR). MVEDRs collect, record, store and export data related to motor vehicle pre-defined events, so MVEDR is turned into the standard automotive system, which storage protocol is described at IEEE P1616.

VI. References

[1] Sivaram C. M. S. L., Kulkarni M. N., GPS-GIS integrated systems for transportation engineering, *IIT*, Mumbai

[2] Jurgen R. K., (1998) Navigation and Intelligent Transportation Systems, Pennsylvania: *Society of Automotive Engineers*, Inc.

[3] Grewal M.S., Weil L.R., Andrews A.P. Global Positioning Systems, Inertial navigation, and Integration, John Willey & Sons, Inc., 2001

[4] J. A. Farrell and M. Barth, The Global Positioning System & Inertial Navigation, McGraw-Hill, 1999

[5] Alonzo Kelly, Modern Inertial and Satellite Navigation Systems, 1994 Carnegie Mellon University, CMU-RI-TR-94-15

[6] Mohinder S. Grewal, Angus P. Andrews, Kalman Filtering: Theory and Practice Using MATLAB, 2nd Edition, Wiley 2001, ISBN: 0-471-39254-5

[7] www.st.com/stonline/products/literature/ds/10175/lis3l02dq.pdf - datasheet of LIS2LV02DQ linear accelerometer

[8] http://www.analog.com/en/prod/0,2877,ADIS16100,00.html – description of ADIS16100 angular rate sensor

New Experiences in Simple Numerical Modelling of Sommerfeld's Integral Kernel

Milica P. Rančić [1] and Predrag D. Rančić [2]

***Abstract* – A number of simple expressions used for modelling the Sommerfeld's integral kernel (SIK), while evaluating the EM filed structure in the surroundings of the vertical Hertz's dipole (VHD) placed above a lossy half-space, are presented in this paper. The results obtained applying the approximate SIK models are compared to exact calculations, confirming a very good accuracy of proposed models.**

***Keywords* – Sommerfeld's integral kernel, Vertical Hertz's dipole, Lossy half-space.**

I. INTRODUCTION

In the process of the EM field characterization in the surroundings of vertical wire structures placed above a lossy half-space, a specific class of integrals, in literature referred to as integrals of Sommerfeld's type, describes the influence of the finite ground conductivity. Since these integrals can not be solved in a closed form, two approaches are possible. One considers a direct numerical integration of the Sommerfeld's integrals, which is a very time-consuming and complex way to deal with this type of integrals. The other one considers modelling of the SIK in a certain way in order to obtain an approximate closed form solution, but of satisfying accuracy. Throughout the last century many authors have dealt with this problem and, as a result, great number of expressions was proposed for modelling the SIK.

Authors of this paper have also dealt with simple SIK modelling ([1]-[9]):

- The first group of the SIK models was obtained developing the spectral reflection coefficient (SRC) into geometry progression series and adopting only the first two or three terms of the obtained series. Model derivation was done under an assumption that the refraction coefficient is $n_{10} >> 1$. Fortunately, in the majority of cases of practical interest, the refraction index is $n_{10} > 3$, which is the range where the model was proven to be satisfyingly accurate ([1]-[3]);
- Developing the SRC into Taylor's series around certain characteristic values of the spectral variable, a number of models was also formed. SIK models were obtained adopting the first few terms, which have extended the application range to the cases of smaller values of the refraction coefficient, i.e. $n_{10} > 2$. The accuracy of this model depends on the number of series' terms that are adopted, the value around which the development is done, as well as the type of the considered wire structure and its location ([4]);
- In the third and forth group of new models, the influence of the lossy ground on the EM field structure is described by one or more fictive images. The position of the images and their weight current coefficients are determined in the process of forming the proposed models ([5]-[8]). The forth group of models is a new form of the so-called two-image approximation of the SIK ([9]).

Adopting good sides of described models, the authors of this paper have proposed a number of SIK models [5-9]. Besides their simple mathematical form, these models are characterized as general since they are not limited by the electrical parameters of the lossy half-space. Also, the application of such approximate expressions gives satisfyingly accurate results for the SIK in the near field, which can be verified comparing them to exact calculations from [10] and [11].

II. THEORETICAL BACKGROUND

A. Description of the model geometry

The VHD is placed in the air ($\sigma_0 = 0$, ε_0, μ_0) at height z'_k above a linear, isotropic and homogenous lossy half-space of electrical parameters σ_1, $\varepsilon_1 = \varepsilon_{r1}\varepsilon_0$, $\mu_1 = \mu_0$ ($\underline{\sigma}_i = \sigma_i + j\omega\varepsilon_i$, $i = 0,1$ - complex conductivity, $\underline{\gamma}_i = \alpha_i + j\beta_i = (j\omega\mu_i\underline{\sigma}_i)^{1/2}$, $i = 0,1$ - complex propagation constant, $\underline{\varepsilon}_{r1} = \underline{n}_{10}^2 = \varepsilon_{r1} - j\varepsilon_{i1}$ - complex relative permittivity, $\underline{n}_{10}$ - complex refraction index, λ_0 - wavelength in air, and $\omega = 2\pi f$ - angular frequency). Schematic illustration of described VHD is given in Fig.1.

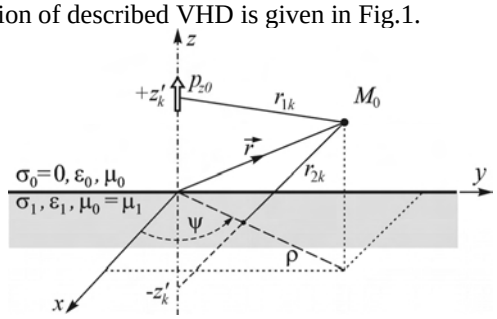


Fig. 1. Schematic illustration of the VHD above a lossy half-space.

[1]Milica P. Rančić is with the Faculty of Electrical Engineering, P.O.Box 73, 18000 Niš, Serbia, E-mail: milica@elfak.ni.ac.yu

[2]Predrag D. Rančić is with the Faculty of Electrical Engineering, P.O.Box 73, 18000 Niš, Serbia, E-mail: prancic@elfak.ni.ac.yu

In the described case of the VHD, the SIK that describes the field scattered from the lossy half-space is defined as

$$S_{00}^{v}(r_{2k}) = \int_{\alpha=0}^{\infty} \tilde{R}_{z10}(\alpha)\tilde{K}_0(\alpha, r_{2k})\,\mathrm{d}\alpha\,, \qquad (1)$$

where: $\tilde{R}_{z10}(\alpha) = \tilde{R}_{z10}(u_0) = (\underline{n}_{10}^2 u_0 - u_1)/(\underline{n}_{10}^2 u_0 + u_1)$ - SRC, $\tilde{K}_0(\alpha, r_{2k})$ - spectral form of the standard potential kernel from the image placed at distance $r_{2k} = \sqrt{\rho^2 + (z + z_k')^2}$, and $u_i = (\alpha^2 + \underline{\gamma}_i^2)^{1/2}$, $i = 0,1$.

B. SIK Approximations

The first model assumes the SRC in a form of a rational function with two unknown constants that are determined by matching the SCR at characteristic points:

$$\tilde{R}_{z10}(u_0) \cong B + A\,\frac{\underline{\gamma}_0}{u_0}\,. \qquad (2)$$

Constants *A* and *B* are determined in the process of approximation as in [5]. Matching the Eq.(2) at points $u_0 = \underline{\gamma}_0$ and $u_0 \to \infty$, constants *A* and *B* are obtained as

$$A = (R_0 - R_\infty) \text{ and } B = R_\infty\,, \qquad (3)$$

where $R_0 = (\underline{n}_{10} - 1)/(\underline{n}_{10} + 1)$, $R_\infty = (\underline{n}_{10}^2 - 1)/(\underline{n}_{10}^2 + 1)$.

Substituting (2) in (1), and adopting Eq. (3), we have:

$$S_{00}^{v}(r_{2k}) \cong BK_0(r_{2k}) + A\underline{\gamma}_0 L(r_{2k})\,, \qquad (4)$$

where $L(r_{2k})$ is a new integral kernel,

$$L(r_{2k}) = \int_{v=z+z_k'}^{\infty} K_0(r_{2kv})\,\mathrm{d}v =$$

$$= -\int_{v=0}^{z+z_k'} K_0(r_{2kv})\,\mathrm{d}v - \frac{\pi}{2}[N_0(\beta_0\rho) + j\,J_0(\beta_0\rho)]\,. \qquad (5)$$

In Eq.(5), $N_0(\beta_0\rho)$ and $J_0(\beta_0\rho)$ are zero-th Neumann and Bessel functions of the first kind, respectively, $\rho = [(x - x_k')^2 + (y - y_k')^2]^{1/2}$ is radial distance, and $r_{2kv} = (\rho^2 + v^2)^{1/2}$.

Comment: Constant *A* can be obtained matching Eq.(2) at a different point ([6]), for example $u_0 = u_B = \underline{\gamma}_0/(\underline{n}_{10}^2 + 1)^{1/2}$. This results in

$$A = -B/(\underline{n}_{10}^2 + 1)^{1/2}\,. \qquad (6)$$

If different values are adopted for constants *A* and *B*, e.g. $A = -2/\underline{n}_{10}$ and $B = 1$, a simple SIK model is also obtained, which actually presents only first two terms of SRC development into a geometry progression series under a condition $n_{10} >> 1$ ([1]-[3]).

The second model is in an exponential form

$$\tilde{R}_{z10}(u_0) \cong B + A\mathrm{e}^{-(u_0 - \underline{\gamma}_0)d}\,. \qquad (7)$$

Again, *A* and *B* are unknown constants, and *d* unknown distance, which are all determined in the process of approximation ([9]). Matching the Eq.(6) at $u_0 = \underline{\gamma}_0$ and $u_0 \to \infty$, and its first derivative at $u_0 = \underline{\gamma}_0$, the following is obtained:

$$A = (R_0 - R_\infty)\,,\ B = R_\infty \text{ and } d = \left|(2/\underline{\gamma}_1)R_0\big/(R_0 - R_\infty)\right|. \qquad (8)$$

Substituting (7) in (1), and adopting Eq. (8), the SIK gains the following two-image approximation form:

$$S_{00}^{v}(r_{2k}) \cong BK_0(r_{2k}) + Ae^{jD}K_0(r_{2kd})\,, \qquad (9)$$

where: $r_{2kd} = \sqrt{\rho^2 + (z + z_k' + d)^2}$ - real distance of the second image in the flat mirror to the observed field point M_0, and $D = \beta_0 d$ - electrical length of the distance *d*.

III. Numerical Results

A number of numerical experiments were performed based on the presented theoretical analysis with an aim to prove the validity of the proposed models. Obtained results are compared to the accurate ones from [10] and [11] and will be graphically illustrated further in the paper. Relative error of accurate results is estimated to be $10^{-7} - 10^{-6}$.

Based on the approximate SIK model given by Eq.(4), modulus of the normalized SIK, S_{00}^{v}/β_0, was determined for different values of the parameters on which it depends, and the results are given in Figs.2a, b and c. Position of the VHD z_k', is taken as a parameter, relative permittivity is $\varepsilon_{r1} = 10$, and Figs. 2a, b and c correspond to the following values of the normalized conductivity $\sigma_1\lambda_0 = 10^{-4}\,S$, $\sigma_1\lambda_0 = 0.175\,S$ and $\sigma_1\lambda_0 = 10\,S$, respectively. In order to investigate accuracy of the proposed model, accurate results from [10] and [11] are also shown (solid circle).

Results obtained applying the proposed model for the SIK are in very good accordance with the ones of accurate calculations, regardless of the electrical parameters of the lossy half-space or the VHD position. In that sense, the proposed model can be, besides simple, characterized as general and accurate.

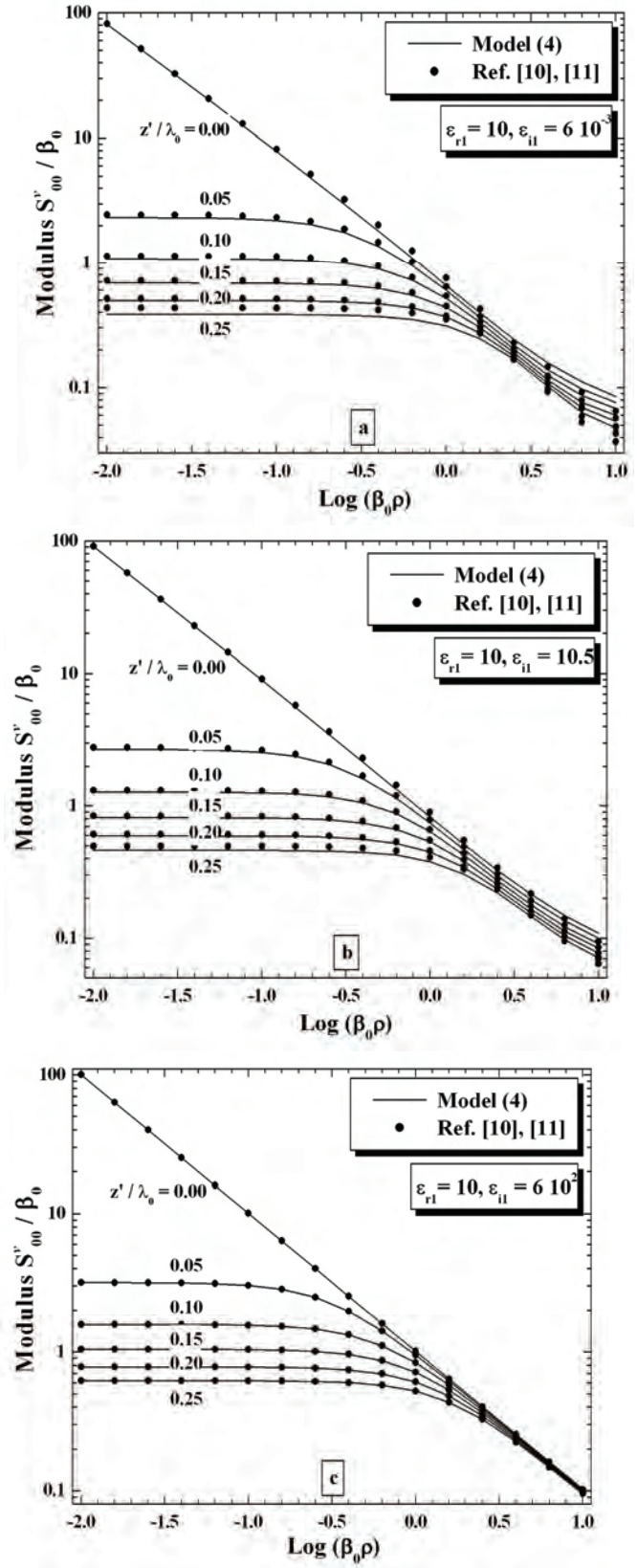


Fig. 2. Modulus of the normalized SIK, at plane $z = 0$, versus normalized radial distance. Relative dielctric constant is $\varepsilon_{r1} = 10$, and VHD position z_k' and imaginary part of dielectric permittivity ε_{i1} are parameters.

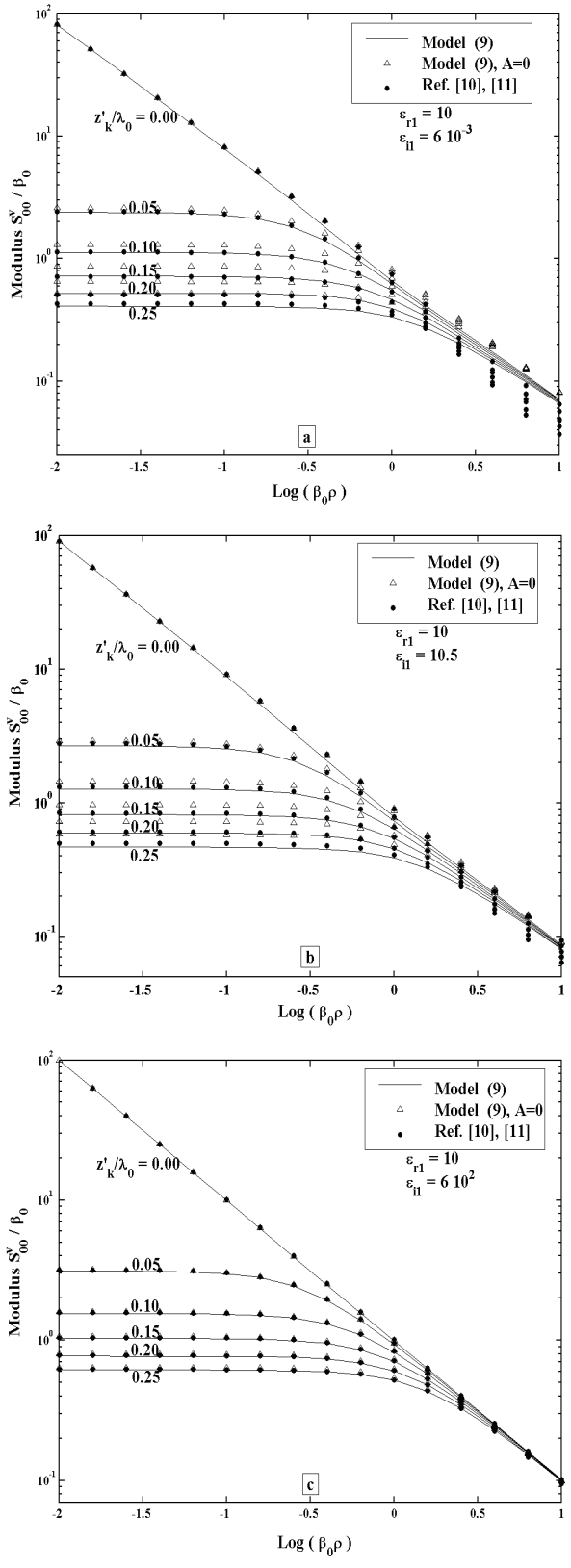


Fig. 3. Modulus of the normalized SIK, at plane $z = 0$, versus normalized radial distance. Relative dielctric constant is $\varepsilon_{r1} = 10$, and VHD position z_k' and imaginary part of dielectric permittivity ε_{i1} are parameters.

Next set of numerical experiments analyses the accuracy of the two-image approximation model given by Eq.(9). Again, the proposed SIK model was tested in the entire range of electrical parameters of the lossy half-space. The results for the modulus of the normalized SIK, S_{00}^{v}/β_0, versus normalized radial distance are presented in Figs.3a, b and c. As previously, parameters are the VHD position z_k' and electrical parameters of the ground. Parallel, the results obtained applying the SIK model given by Eq.(9) when $A=0$, are also given in figures (open up triangle). Comparing was performed using the accurate results from [10] (and [11], solid circle).

IV. CONCLUSION

Two different models for the SIK evaluation in the surroundings of the VHD placed above a lossy half-space are analysed in the this paper. Based on the presented numerical results, one can conclude that both models are, besides of simple form, also satisfyingly accurate. Also, the models can be characterized as general, since no limitations regarding the electrical parameters of the ground were introduced during their derivation. Since the SIK can be easily and very accurately solved using described approximations, the usage of all program tools developed for characterization of vertical wire structures placed above an ideally conducting ground is possible with only minor corrections needing to be done.

This implicates the possibility to completely characterize EM field structure of various vertical wire structures (e.g. vertical dipole antenna, vertical mast antenna, vertical coupled antennas, YAGI structure, etc.) in its near, and using the reciprocity theorem, also approximately in the far field zone. Also, proposed models can be very successfully used for modelling EM field structure of atmospheric discharge ([15]).

REFERENCES

[1] P.D. Rančić, M.I. Kitanović, "A New Model for Analysis of Vertical Asymmetrical Linear Antenna Above a Lossy Half-Space", *Int. J. Electron. Comun. AEÜ*, Vol. 51, No. 3, pp. 155-162, May 1997.

[2] P.D. Rančić, J.V. Surutka, M.I. Kitanović, "The Influence of Finite Ground Conductivity on Characteristics of a Vertical Mast (Monopol) Antenna With Elevated Feeding", Proc. of the II Int. Symp. EMC'96 Roma, L1-2/427-432, Roma, Italy, 1996. (Poster Presentation Best Paper Award)

[3] P.D. Rančić, "H.F. Yagi Antennas Above a Conducting Half-Space", *FACTA Universitatis, Series: Electronics and Energetics* Vol. 14, No. 2, pp. 145-166, 2001. http://factaee.elfak.ni.ac.yu/fu2k12/fu01.html.

[4] P.D. Rančić, "An Efficient Method for Analysis of Horizontal Wire Antennas Placed Above a Ground Plane", Proc. of the IV Int. Symp. EMC'00, Vol.1, pp. 53-58, Brugge 2000.

[5] M.P. Rančić, P.D. Rančić, "Vertical linear antennas in the presence of a lossy half-space: An improved approximate model*" Int. J. Electron. Comun. AEÜ*, Vol. 60, No. 5, pp. 376-386, May 2006.

[6] M.P. Rančić, P.D. Rančić, "An Approximative Model for Analyzing Vertical Antennas Above a Ground Plane", 12th TELFOR'2004, Novembar 23-25, Belgrade, Serbia, 2004. http://www.telfor.org.yu/telfor2004/radovi/PEL-9-28.pdf

[7] P.D. Rančić, M.P. Rančić, D.H. Borisov, "Horizontal Hertz's dipole above a lossy half-space: Review of simple models for the Sommerfeld's integral", Conf. Proc. XLIX ETRAN'05, Vol. 2, , pp. 276-279, June 5-10, Budva, Serbia, 2005. (in Serbian)

[8] M.P. Rančić, P.D. Rančić, "Vertical Hertz's dipole above a lossy half-space: Review of simple models for the Sommerfeld's integral", Conf. Proc. XLIX ETRAN'05, Vol. 2, pp. 272-275, June 5-10, Budva, Serbia, 2005. (Best paper in AP section - in Serbian)

[9] M.P. Rančić, P.D. Rančić, "An approximation of the Sommerfeld's integral kernel in EM field analysis of the VHD", ETRAN, pp.266-269, June 6-8, Belgrade, Serbia, 2006. (in Serbian)

[10] V.V. Petrović, Private communications, ETF Belgrade, 2005.

[11] A.R. Djordjević, M.B. Bažđar, V.V. Petrović, D.I. Olćan, T.K. Sarkar, R.F. Harrington, "AWAS for Windows, Version 2.0, Analysis of Wire Antennas and Scatterers – Software and User's Manual", Artech House, 2002.

[12] P.R. Banister, "Summary of Image Theory Expressions for the Quasi-Static Fields of Antennas at or Above Earth's Surface", *Proc. of the IEEE*, Vol. 67, No. 7, pp. 1001-1008, 1979.

[13] S.F. Mahmoud, A.D. Metwally, "New Image Representation for Dipoles near a Dissipative Earth 1. Discrete Images 2. Discrete Plus Continuous Images", *Radio Science*, Vol. 16, pp. 1271-1283, 1981.

[14] R.M. Shubair, Y.L. Chow, "A Simple and Accurate Complex Image Interpolation of Vertical Antennas Present in Contiguous Dielectric half-Spaces", *IEEE Trans. on AP*, Vol. 41, No. 6, pp. 806-812, 1993.

[15] C. Yang, B. Zhou, "Calculation Methods of Electromagnetic Fields Very Close to Lightning", *IEEE Trans. on EMC*, Vol.46, No.1, pp. 133-141, 2004.

Summing Signals of Push-Pull Broadband Amplifier Modules

Boyan D. Karapenev[1] and Ilya N. Nemigenchev[2]

***Abstract* – This paper presents the qualitative parameters of two identical push-pull broadband amplifier modules implemented with different bipolar transistors and broadband transformers, working in modes: classes B and AB, and five variants for adding their signals by measuring: amplitude characteristics, transmission coefficient, dynamic range and amplitude-frequency characteristics (frequency bandwidth). The received results and parameters of experimental researches have been generalized and presented in graphical and tabular form. Comparative analysis of the experimental results obtained, as well as conclusions and recommendations for application of constructed Variants have been done.**

***Keywords* – summing, signals, push-pull, broadband, amplifier module, broadband transformer.**

I. INTRODUCTION

Up-to-date broadband amplifiers are characterized by their high output power and great efficiency. Specific requirements are set to them with respect to the amplitude-frequency characteristics (AFC), efficiency, gain coefficient in the bandwidth, etc.

Power amplifiers of up to several kW and frequency range up to 100MHz are most frequently constructed on the basis of separate broadband amplifier modules (BAM), made up of bipolar or field-effect transistors and broadband transformers. The output power of such a module reaches 100÷200W. To reach the required level of output power, broadband amplifiers are constructed by summing up the signals of the individual amplifier modules.

II. IMPLEMENTATION AND STUDYING OF BAMS

Broadband amplifier modules are widely used in constructing radio transmitters of wide frequency range. Since the output power of the final stage exceeds considerably the power of an individual amplifier module, it is necessary to implement the principle of summing up the powers of a number of amplifier modules on a common load.

The following ways of summing up signals were used most widely [1], [2], [7]:

- by direct connection of the amplifier modules: parallel and push-pull;
- by bridge connections;
- by means of high-frequency broadband transformers.

A push-pull broadband amplifier module has been designed, operating in modes of B and AB classes. Its schematic diagram is shown in Fig. 1 [6]. The preset basic performance characteristics are: U_{CC}=15V, P_0=40W, $R_i=R_0=R_L$=50Ω, A_U=15, f_B=1MHz and f_H=25MHz. Design calculations have been done and the broadband transmission line transformers (BTLT) used have been constructed [5], [2]. Transformers Tp_1 and Tp_4 provide the required matching at the input and at the output of the amplifier module, and the necessary transformation of the circuit impedances is accomplished by means of Tp_2 and Tp_3 - 4:1 and 1:4, respectively.

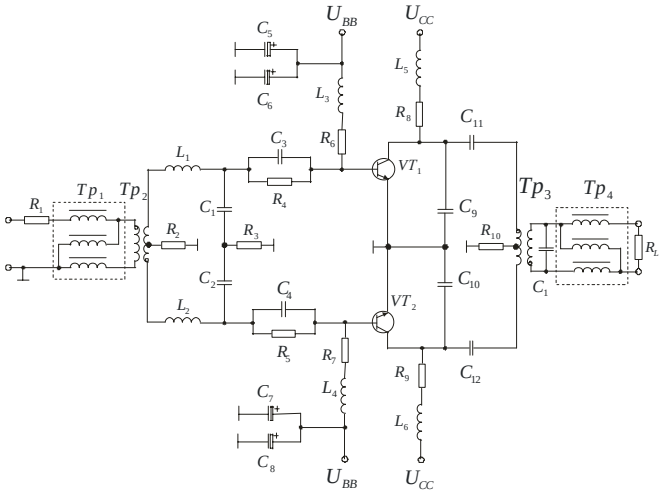


Fig. 1. Schematic diagram of a push-pull BAM

Two identical BAMs (BAM1 and BAM2) have been constructed and studied. They operate in modes of B and AB classes, respectively with transistor couples KT920B and MRF454.

A. Experimental study of BAM1

To find out the effect of the transistors used, experimental study of BAM1 constructed with transistor couples KT920B and MRF454, respectively, has been carried out. The obtained amplitude characteristics (AC) and amplitude-frequency characteristics (AFC) for operation mode of class B are shown in common coordinate systems in Fig. 2 and Fig. 3, respectively.

[1]Boyan D. Karapenev is with the Department of Communication Technics and Technology, Technical University - Gabrovo, Bulgaria, E-mail: bkarapenev@tugab.bg

[2]Ilya N. Nemigenchev is with the Department of Communication Technics and Technology, Technical University - Gabrovo, Bulgaria, E-mail: nemig@tugab.bg

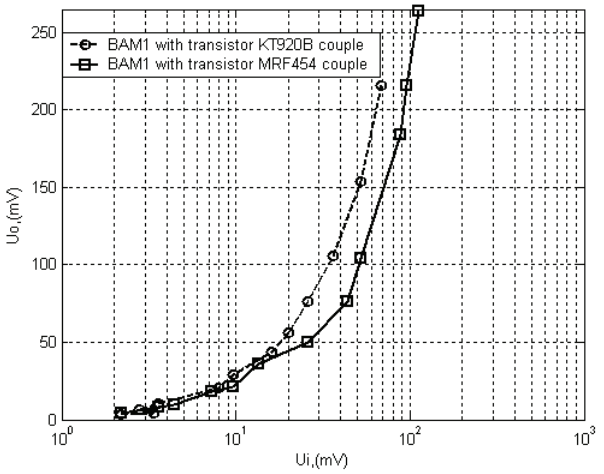


Fig. 2. AC of BAM1 in operating mode of B class, constructed with transistors KT920B and MRF454

It has been found out that when constructing BAM1 with KT920B transistors, larger linearity of characteristics is obtained compared with the construction with MRF454, for a greater dynamic range D_{KT920B}=4,25 with respect to KT920B D_{KT920B}=4,13, and voltage gain Au_{K920B}=3,17 and Au_{MRF454}=2,37, respectively.

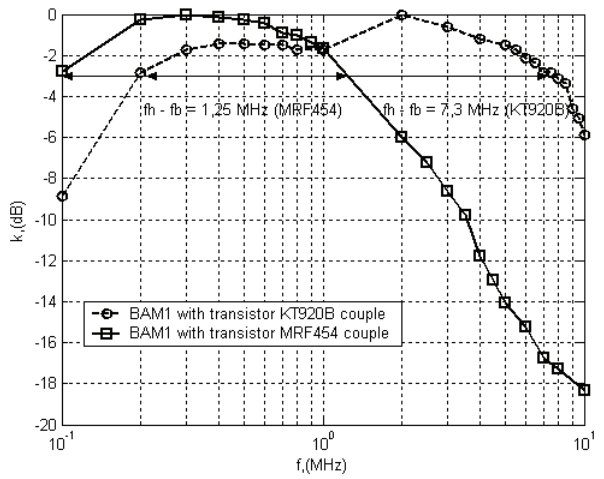


Fig. 3. AFC of BAM1 in B class operating mode, constructed with KT920B and MRF454 transistors

From the characteristics given in Fig. 3 it can be seen that there is a great difference in the frequency bandwidth Δf_{KT920B}=7,3 MHz and Δf_{MRF454}=1,25 MHz, respectively, which is very impressive. It is due to the difference in the catalogue parameters of the transistors used, stray capacitances and inductances.

When constructing input and output BTLT with ferrite toroidal cores of various sizes and respectively various number of windings, AFC of analogous type are the result. Table 1 shows a comparative evaluation of the performance characteristics of each constructed variant.

TABLE I
BASIC PERFORMANCE CHARACTERISTICS

Variant №		1	2
Input transformer		12,7x4,8mm–10 wdg.	12,7x4,8mm –10wdg
Output transformer		**29x7,5mm – 21wdg.**	**21x6,35mm – 17wdg**
transistors **KT920B**	D	Uimin=16mv Uimax=68mV D=4,25	Uimin=18,4,mV Uimax=76mV D=4,13
	Au	3,17	2,78
	Δf	(7,5-0,2)=7,3MHz	(6-0,18)=5,82MHz
Transistors **MRF454**	D	Uimin=44mV Uimax=182mV D=4,13	Uimin=20mV Uimax=70mV D=3,5
	Au	2,37	2,31
	Δf	1,25MHz	0,9MHz

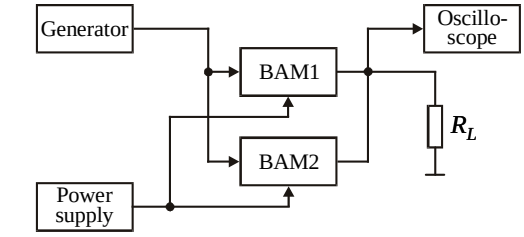
Variant №		3	4
Input transformer		12,7x4,8mm – 10wdg.	12,7x4,8mm -10wdg.
Output transformer		**21x6,35mm – 10wdg.**	**12,7x4,8mm -10wdg**
transistors **KT920B**	D	Uimin=28mv Uimax=144mV D=5,14	Uimin=36,mV Uimax=72mV D=2
	Au	3,31	3,16
	Δf	(5-0,25)=4,75MHz	(7-0,18)=6,82MHz
transistors **MRF454**	D	Uimin=30mV Uimax=112mV D=3,73	Uimin=34mV Uimax=68mV D=2
	Au	2,29	2,32
	Δf	(1,2-0,1)=1,1MHz	(1-0,16)=0,84MHz

Variant №		5	6
Input transformer		**21x6,35 mm – 10wdg**	**21,7x6,35mm-10wdg.**
Output transformer		**21x6,35mm – 17wdg.**	**29x7,5mm -21wdg.**
transistors **KT920B**	D	Uimin=10,4mv Uimax=64mV D=6,15	Uimin=8,8mV Uimax=36mV D=4,09
	Au	4	3,94
	Δf	(0,73-0,02)=0,71MHz	(0,8-0,02)=0,78MHz
transistors **MRF454**	D	Uimin=9,2mV Uimax=72mV D=7,82	Uimin=32mV Uimax=72mV D=2,25
	Au	2,61	2,41
	Δf	(0,75-0,03)=0,72MHz	(0,9-0,1)=0,8MHz

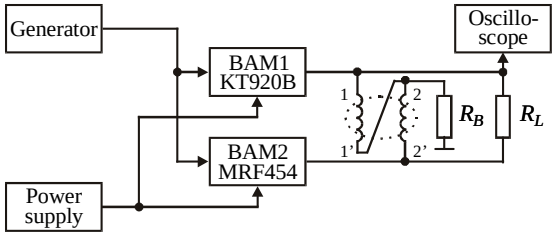
It follows from the above obtained results that the parameters and manufacturing tolerances of the transistors used, as well as the BTLT parameters have an effect on the performance characteristics of BAM by means of the ferrite toroidal core size, the number of windings of the transmission lines used, etc. (in this case a couple of twisted enameled conductors of d=0,62 mm).

B. Experimental studies and outcomes for summing and separating of signals for different connections of BAM1 and BAM2

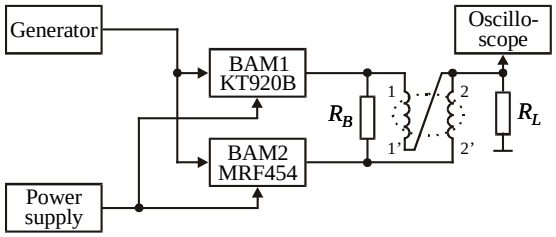
For separating and summing the signals of BAM1 and BAM2 constructed with transistors KT920B, and MRF454 or KT920B, respectively, in order to enhance the output power on a common load, 5 variants of their connection have been made [1], [2], shown in Fig. 4.



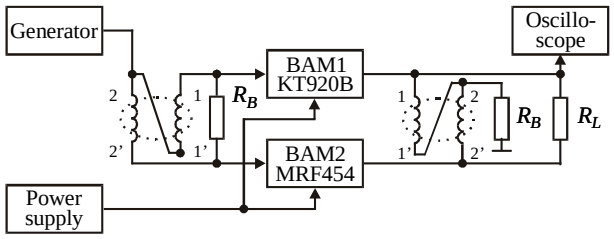
a - Variant 1



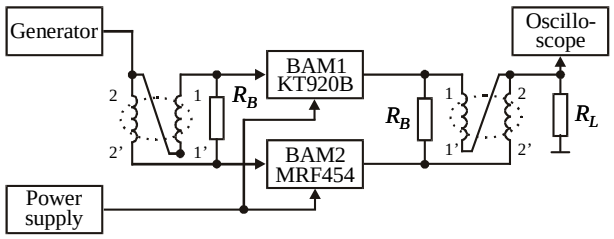
b - Variant 2



c - Variant 3



d - Variant 4



e - Variant 5

Fig. 4. Variants of connecting BAM1 and BAM2 constructed with transistors KT920B, and MRF454 or KT920B

Fig. 5 and Fig. 6 show a comparative evaluation and the graphs of AC and AFC, respectively for Variant 2 of connecting both push-pull BAMs, one of which is constructed with KT920B transistors, and the other - with MRF454 transistors with two BAMs constructed with couples of KT920B transistors each.

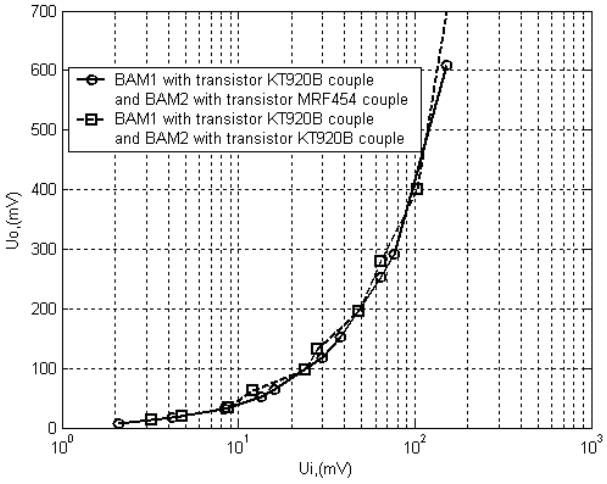


Fig. 5. AC when connecting both push-pull BAMs according to the diagram of Variant 2

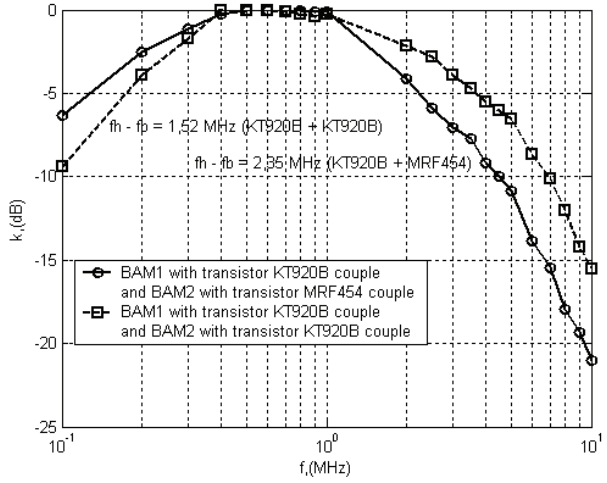


Fig. 6. AFC when connecting both push-pull BAMs according to the diagram of Variant 2

From the characteristics shown in Figs. 5 and 6 it has been found out that BAMs constructed with KT920B and MRF454 transistors and only with KT920B transistors, have identical AC, and the slopes of AFC are displaced to the region of lower frequencies.

The performance characteristics for the five Variants of connecting BAM1 constructed with KT920B transistors and BAM2 constructed with MRF454 are presented in Table 2. Table 3 compares their values for Variants 1, 2 and 3 to those obtained from studying both BAMs constructed only with KT920B transistors.

III. Results Analysis

The use of different couples of bipolar transistors when constructing push-pull BAMs (KT920B and MRF454 in BAM1) results in a considerable difference in the bandwidth 7,3 MHz and 1,25 MHz, respectively.

TABLE II
MAJOR PERFORMANCE CHARACTERISTICS FOR EACH VARIANT OF CONNECTING PUSH-PULL BAMS

Variant №		Variant 1	Variant 2	Variant 3	Variant 4	Variant 5
Input transformers 12,7x4,8mm -10 wdg.; Output transformers 29x7,5mm – 21 wdg.						
MRF454 and T920B	D	Uimin=96mv Uimax=236mV D=2,45	Uimin=38mV Uimax=152mV D=4	Uimin=76mV Uimax=156mV D=2,05	Uimin=52mV Uimax=112mV D=2,15	Uimin=90mV Uimax=212mV D=2,35
	Au	0,28	4	1,38	0,42	46,16
	Δf	(2,5-0,14)= =2,36MHz	(1,7-0,18)= =1,52MHz	(2,2-0,44)= =1,76MHz	(1,2-0,15)= =1,05MHz	(2,5-0,35)= =2,15MHz

TABLE III
COMPARISON OF THE MAJOR PERFORMANCE CHARACTERISTICS FOR VARIANTS 1, 2 AND 3 OF CONNECTING BAM

		Direct parallel	With BTLT	
№		Variant 1	Variant 2	Variant 3
Input transformers 12,7x4,8mm–10 wdg.; Output transformers 29x7,5mm – 21wdg.				
BAM with MRF454 and KT920B	D	Uimin=42mv Uimax=156mV D=3,71	Uimin=38mV Uimax=152mV D=4	Uimin=76mV Uimax=156mV D=2,05
	Au	1,61	4	1,38
	Δf	(2,2-0,4)=1,8MHz	(1,7-0,18)=1,52MHz	(2,2-0,44)=1,76MHz
BAM with KT920B and KT920B	D	Uimin=8,8mV Uimax=53,6mV D=6,09	Uimin=48mV Uimax=152mV D=3,16	Uimin=28mV Uimax=148mV D=5,28
	Au	0,68	4,63	0,59
	Δf	(2,7-0,25)=2,45MHz	(2,6-0,25)=2,35MHz	(1,3-0,8)=0,5MHz

The ferrite toroidal cores used do not have constant parameters in their operating frequency range preset by the manufacturer and their manufacturing tolerances are significant. It is required that their practical manufacture - number of windings, number of twists per unit of length, transmission line used, etc., should be executed with great accuracy and repetition of the key parameters.

The difference in the characteristics and parameters obtained as a result of constructing identical BAMs can reach up to 10% which requires their optimization. It can be fulfilled most easily by a change in the construction of BTLT.

In order to provide a larger frequency bandwidth, the voltage gain should be decreased. A balance at Au_{KT920B}=4,63, Δf_{KT920B}=2,35 MHz and Au_{MRF454}=4, Δf_{MRF454} =1,52 MHz was reached during the experimental studies.

The experimental studies conducted show that a broader frequency band at lower voltage gain can be achieved with a push-pull BAM compared to a single-ended BAM [3].

The BTLT used which help in separating and summing of signals by individual BAMs for summing up their
powers on a common load, also have a strong effect on the performance characteristics obtained. It is of great importance for them to match accurately the input and output impedances which is a prerequisite for making them broadband as required.

A disadvantage of the developed Variants of summing up signals of push-pull BAMs appears to be the great number of BTLT used, which make construction difficult, increase the size, limit the frequency bandwidth and lead to an increase in their manufacturing tolerances.

IV. CONCLUSION

The transistors used, the construction and parameters of the BTLT used and the way of connecting BAMs for summing up signals by criteria of certain performance characteristics: gain coefficient, dynamic scope, and frequency bandwidth, determine the performance characteristics of the constructed amplifier.

The construction of broadband amplifiers by modular principle has the following advantages: identical parameters, interchangeability, easy matching with the help of BTLT, high reliability and low cost.

REFERENCES

[1] Ch. Tihchev, “Radiotransmitter Devices”, Tehnika, Sofia, 1992.
[2] D. Dobrev, L. Jordanova, “Radiocommunications”, part I and II, Sofia, Siela, 2001.
[3] I. Nemigenchev, B. Karapenev, “Study of broadband amplifier modules”, Unitech, Gabrovo, 2005.
[4] I. Nemigenchev, B. Karapenev, “Summing Signals of broad-band amplifier modules”, Unitech, Gabrovo, 2006.
[5] J. Savick, “Transmission line transformers”, New Jersey, NOBLE, 1996.
[6] “RF amplifiers”, Construction documents, Huttinger.
[7] V. Shahgildian, “Design Radiotransmitter Devices”, Sviaz, Moscow, 1976.

SESSION SP I

Signal Processing I

Adaptive Notch Filters for Image Processing

Georgi L. Iliev[1]

***Abstract* – In this paper we develop a very simple 2-D adaptive notch filter design. Second-order all-pass circuits are used to implement the structure of our filter. They guarantee the stability and ensure low sensitivity of the developed structure. Computationally efficient adaptive algorithm is employed for the adjustment of filter coefficients controlling the notch frequency. As a result our filter has lower computational complexity compared to the existing approaches. We study the error surface of our adaptive filter and show that there is only one global minimum which guarantees the convergence. The designed 2-D adaptive notch filter can be successfully applied for image denoising when on-line processing is essential.**

***Keywords* – Adaptive filters, Image processing, Adaptive algorithms, Digital filters, Noise suppression.**

I. INTRODUCTION

Adaptive notch filtering is a useful tool for narrowband interference suppression which has been used intensively for many years in different signal processing applications [1]. Examples can be easily found in the field of telecommunications, radar and sonar signals, adaptive control and sensor systems. During the years a lot of methods for design of one-dimensional adaptive notch filters have been proposed [2-5]. As a rule they employ simplified structures which allow low computational complexity and application of efficient adaptive algorithms [6, 7]. By contrast, two-dimensional (2-D) adaptive notch filter design, which is crucial for some image processing applications, is still an immature area deserving much more research efforts and development. Although several approaches have been proposed [8, 9] their computational complexity is relatively high, which is a result of the too complicated algorithms used for filter parameter adjustment.

In this paper we develop an approach based on very simple second-order all-pass structures. First we show how one-dimensional adaptive notch filters can be designed. Then we construct 2-D adaptive notch filters for image denoising. Finally, some experiments are conducted to evaluate the performance of our method.

II. ALL-PASS BASED ADAPTIVE NOTCH FILTERS

A realization based on second-order Gray-Markel lattice circuit is used [10] – Fig. 1. Using that circuit it becomes possible to implement a second-order adaptive notch filter illustrated in Fig. 2.

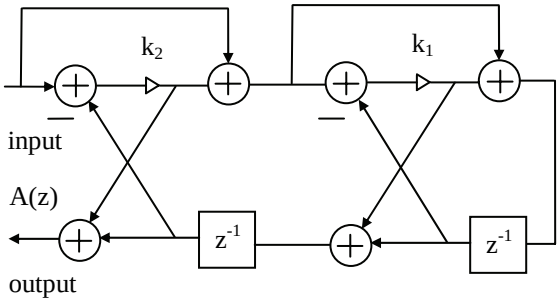


Fig. 1. Second-order lattice Gray-Markel circuit realizing all-pass function A(z)

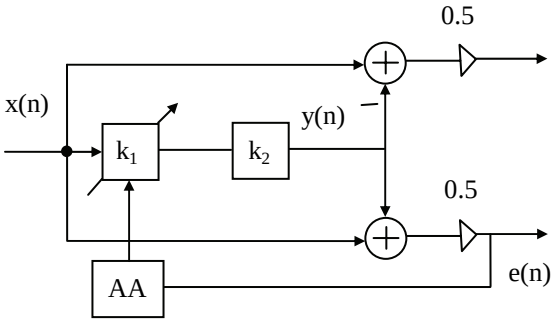


Fig. 2. Second-order adaptive notch filter

This implementation has two very important advantages: first extremely low passband sensitivity that means resistance to quantization effects, second independent control of central frequency and filter bandwidth.

Thus if the all-pass function A(z) is

$$A(z) = \frac{k_2 + k_1(1+k_2)z^{-1} + z^{-2}}{1 + k_1(1+k_2)z^{-1} + k_2 z^{-2}} \qquad (1)$$

then k_1 controls the central frequency ω_0 while k_2 is related to the bandwidth (BW) via

$$k_1 = -\cos\omega_0 \qquad (2)$$

$$k_2 = \frac{1-\tan(BW/2)}{1+\tan(BW/2)}. \qquad (3)$$

BW is directly connected to the distance from the pole to the unity-circle and transforming the structure in Fig.1 into an adaptive filter, it is possible to fix the bandwidth and implement an adaptive IIR filter free of stability problems.

[1]Georgi L. Iliev is with the Department of Telecommunications, Technical University of Sofia, Sofia 1000, 8 Kl. Ohridski St., Bulgaria, E-mail: gli@tu-sofia.bg

Adapting k_1 the central frequency can be shifted around the unity-circle.

For the adjustment of filter coefficient a Least Mean Squares (LMS) algorithm is applied as follows

$$k_1(n+1) = k_1(n) - \mu e(n)[x(n-1) - y(n-1)] \quad (4)$$

where e(n) is the error signal and μ is the step size controlling the convergence speed.

In order to ensure the stability of the adaptive algorithm (AA) the range of the step size μ should be set according to

$$0 < \mu < \frac{K}{L\sigma^2}. \quad (5)$$

In this case L is the filter order, σ^2 is the power of the signal [x(n-1) – y(n-1)] and K is a constant depending on the statistical characteristics of the input signal. In most of the practical situations K is approximately equal to 0.1 [11].

III. Error Surface Analysis

We start the examination of error surface considering the characteristics of the developed fixed notch filter, at this stage neglecting the adaptation. In Fig. 3 magnitude and phase response is shown while Fig. 4 presents pole-zero plot. The figures show the characteristics of a narrowband filter with notch frequency at 0.4 π where the phase response undertakes a 180° shift. That frequency is determined by coefficient k_1 (Eq.2). The second coefficient namely k_2 controls the position of the poles (see Fig.4) and thus the filter bandwidth (Eq.3). The notch frequency is not related to k_2 and on the other hand filter bandwidth has no relation to k_1. That provides the possibility for absolutely independent adjustment of filter parameters which is a very important feature when considering the design of adaptive notch filters.

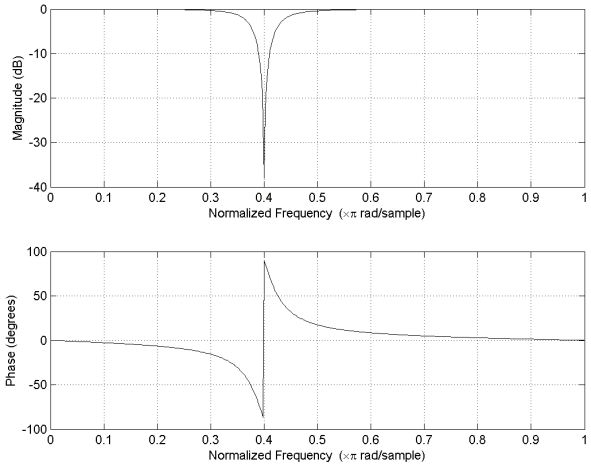


Fig. 3. Magnitude and phase response

Next we examine the characteristics of the cost function. Here we use the mean squared error (MSE)

$$MSE = E\,[e(n)^2] \quad (6)$$

as measure for the performance of our filter. Fig. 5 plots MSE as function of k_1. Several cases are presented for different values of k_2. Examining the error surface we can come to two conclusions. First there is only one global minimum and it is possible to reach it using the simple LMS algorithm. Second making the BW narrower (in Fig. 5 the narrowest BW is for k_2= 0.954) we smooth the error surface and that results in a slower adaptation rate.

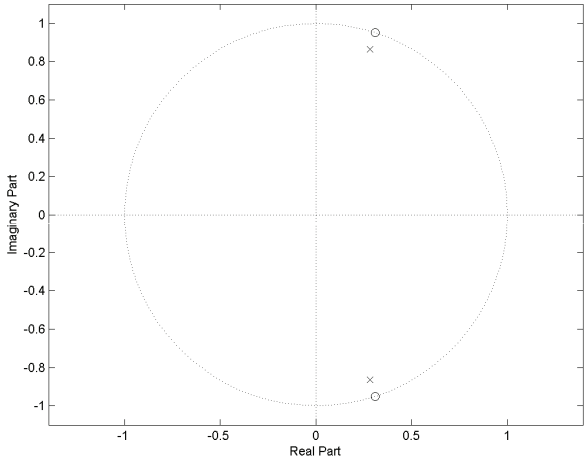


Fig. 4. Pole-zero plot

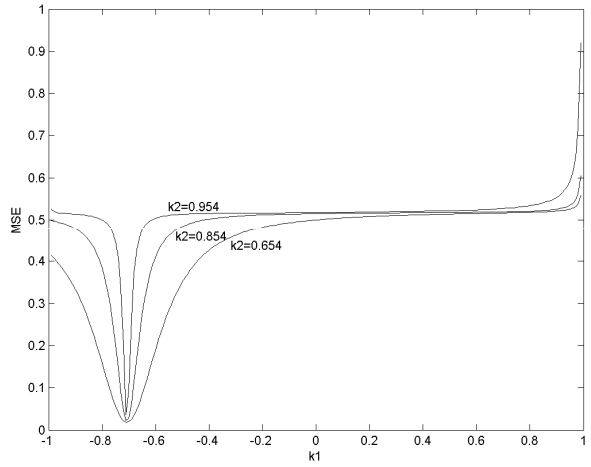


Fig. 5. Cost function

IV. Image Denoising

In Fig. 6 we present the design of our 2-D adaptive notch filter used for image denoising. Two all-pass functions defined via

$$A_1(z_1) = \frac{k_{21} + k_{11}(1+k_{21})z_1^{-1} + z_1^{-2}}{1 + k_{11}(1+k_{21})z_1^{-1} + k_{21}z_1^{-2}} \quad (7)$$

$$A_2(z_2) = \frac{k_{22} + k_{12}(1+k_{22})z_2^{-1} + z_2^{-2}}{1 + k_{12}(1+k_{22})z_2^{-1} + k_{22}z_2^{-2}} \quad (8)$$

are employed. Here z_1 and z_2 represent the delay in the two directions of an image (row-wise and column-wise). Cascading these all-pass structures we form the transfer function of a 2-D notch filter. The adaptation is implemented by using two error signals $e_1(m,n)$ and $e_2(m,n)$. Fig. 7 plots the error surface of MSE for our 2-D adaptive notch filter. Two minima are presented which correspond to the notch frequency determined by the coefficients k_{11} and k_{12}. As with the one-dimensional case BW is controlled by k_{21} and k_{22}. We apply LMS algorithm to find the optimal solution.

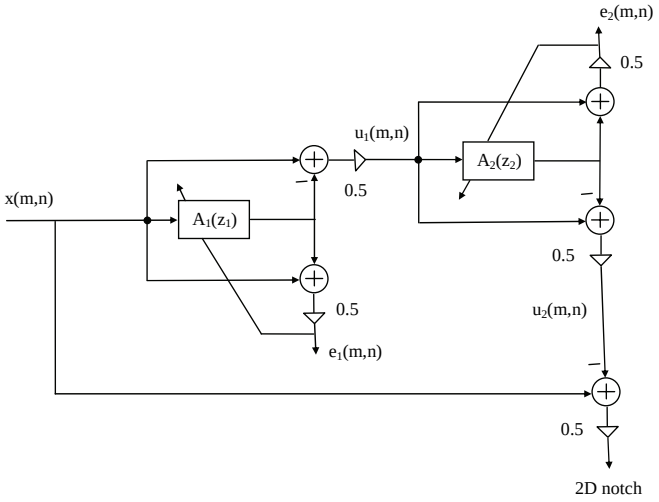


Fig. 6. 2-D adaptive notch filter

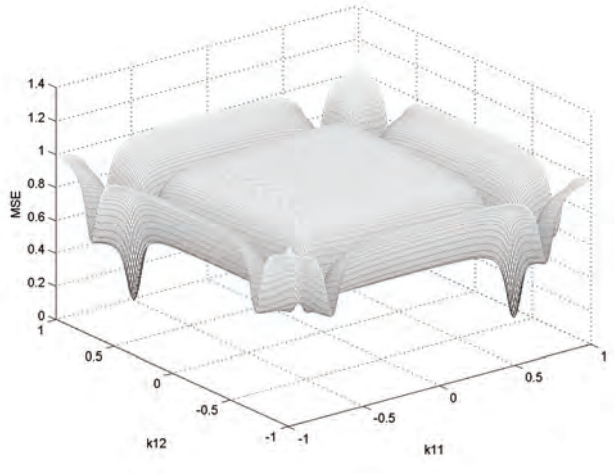


Fig. 7. Error surface

V. EXPERIMENTS

To evaluate the performance of our design we set up an experiment where a 2-D periodical interference in the form of

$$s(m,n) = 0.2\sin(0.3\pi m + 0.4\pi n) \quad (9)$$

is presented. In practice this kind of disturbances are often due to the power supply systems or cross-interference in image transmission channels. Fig. 8 presents the original 256 x 256 image. In Fig. 9 the noisy image is shown. Finally, Fig. 10 displays the image after noise suppression. It can be seen that the interference is removed and the original image is restored. To study the behavior of our adaptive filter we draw the learning curves of coefficients k_{11} and k_{12} (Fig. 11). Obviously the adaptive process converges to the optimal solution. In addition the filter exhibits very high adaptation rate (about 1000 iterations).

Fig. 8. Original image

Fig. 9. Noisy image

Fig. 10. Image after filtration

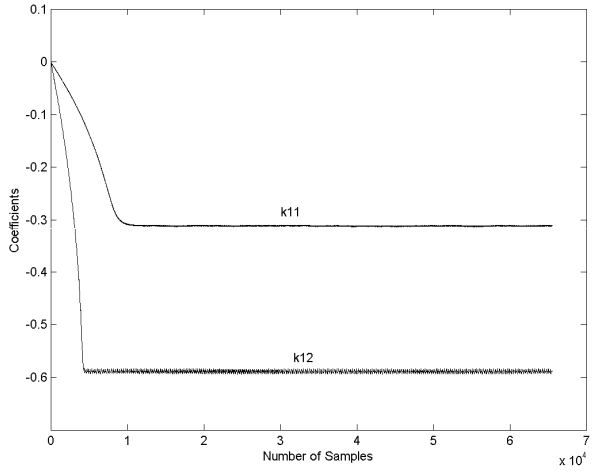


Fig. 11. Learning curves

VI. CONCLUSIONS

In this paper we develop a very simple 2-D adaptive notch filter design. Based on all-pass structures and employing computationally efficient adaptive algorithm the presented method can be successfully applied for image denoising when on-line processing is essential. The main advantages of our approach can be summarized as follows:

- low computational complexity – we adjust only two coefficients and use the simple LMS algorithm;
- stability is easily monitored for the employed second-order all-pass structures;
- error surface has global minimum which guarantees the convergence.

Using the developed method it is possible to construct more complex filter structures in cascade or parallel form for multiple 2-D periodical interference removal which is a challenging topic for further studies in the field of 2-D adaptive notch filter design.

REFERENCES

[1] B. Widrow, J. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, R. Goodlin, "Adaptive Noise Cancelling: Principles and Applications", Proc. of IEEE, vol. 63, no. 12, pp. 1692-1716, Dec. 1975.

[2] P. Regalia, "An improved lattice-based adaptive IIR notch filter", IEEE Trans., Signal Processing, pp. 2124-2128, Sept. 1991.

[3] D. Rao, S. Kung, "Adaptive notch filtering for retrieval sinusoids in noise", IEEE Trans., Acoustics, Speech and Signal Processing, pp. 791-802, Aug. 1984.

[4] T. Kwan, K. Martin, "Adaptive detection and enhancement of multiple sinusoids using a cascade of IIR filters", IEEE Trans., Circuits and Systems, pp. 937-947, July 1989.

[5] K. Martin, M. Sun, "Adaptive filters suitable for real-time spectral analysis", IEEE Trans., Circuits and Systems, pp. 218-229, Feb. 1986.

[6] J. Chicharo, T. Ng, "Gradient-based adaptive IIR notch filtering for frequency estimation", IEEE Trans., Acoustics, Speech and Signal Processing, pp. 769-777, May 1990.

[7] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros", IEEE Trans., Acoustics, Speech and Signal Processing, pp. 983-996, Aug. 1985.

[8] S. Pei, C. Tseng, "Two dimensional IIR digital notch filter design", IEEE Trans., Circuits and Systems, pp. 227-231, Mar. 1997.

[9] S. Pei, W. Lu, C. Tseng, "Analytical two-dimensional IIR notch filter design using outer product expansion", IEEE Trans., Circuits and Systems, pp. 765-768, Sept. 1997.

[10] P. Regalia, S. Mitra, P. Vaidyanathan, "The digital all-pass filter: a versatile signal processing building block", Proc. of IEEE, pp. 19-37, Jan. 1988.

[11] S. Douglas, *Adaptive Filtering*, in Digital Signal Processing Handbook, D. Williams and V. Madisetti, Eds., Boca Raton, CRC Press, 1999, pp. 451-619.

Efficient Compression for Contour Images

Roumen Kountchev[1], Vladimir Todorov[2], Roumiana Kountcheva[3]

Abstract - **In the paper is presented new method for efficient compression of contour images extracted from video files. The work is aimed at processing of video presentations of Sign Language Interpreter, explaining the training lectures for deaf-impaired students. The presented method offers to use the contour image of the interpreter instead of traditional video as an additional tool included in the distance learning procedure. For this the video is processed frame by frame, the contour image of the interpreter is extracted and the obtained contour images from the consecutive TV frames are compressed and prepared for transfer via Internet. The method for lossless compression of the obtained images is based on image histogram analysis and special Run-Length Coding. In result is obtained new video presentation, which retains the understandability of the training information, but is easily accessed via modem. The efficiency of the new method is higher than that of JPEG2000 (lossless version).**

Keywords **– Lossless image compression, Distance learning, Sign language interpretation.**

I INTRODUCTION

The training of hearing-impaired people based on up-to-date means for video communications is a problem, on which significant experience has been accumulated and many investigations have been carried out [1, 2, 3, 4]. One of the most appropriate means is the Internet, which permits distance learning to be done without live sign language interpretation. The most popular systems used to transfer visual information for distance training for hearing-impaired people via Internet, are based on the video compression standards H.261 [5] and MPEG-4 [6]. In these cases, significant part of the transmitted information comprises the training course, together with video sequences, presenting the image of a Sign Language Interpreter, SLI (Fig.1.). The systems used for the transmission of moving images of this kind [7, 8, 9, 10, 11] in accordance with the already mentioned standards as a rule are based on the principle for inter-frame prediction with movement compensation for every block of the predicted frames of type P. The general disadvantage of these systems is that as a result of the standard compression used to store the presented lesson, the quality of the fast moving objects (in this case, the hands of the sign language interpreter) is significantly deteriorated and the understandability of the presented signs is decreased.

[1] Roumen Kountchev is with the FCTT, Radiocomunications Dept., Technical University of Sofia, Bul. Kl. Ohridsky 8, Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg

[2,3] Vladimir Todorov and Roumiana Kountcheva are with T&K Engineering, Mladost 3, P.O.Box 12, Sofia 1712, Bulgaria
E-mails: todorov_vl@yahoo.com. kountcheva_r@yahoo.com

Another important disadvantage is that the lessons are presented as video files, which are large and when transferred via modem, the downloading process is relatively slow. In order to avoid the mentioned dynamic distortions, it was found that it is better to compress the video sequences of moving images with intra-frame compression in correspondence with standards M-JPEG [6] or M-JPEG 2000 [12]. In these cases, the image quality is improved, but the compression ratio is low, which results in even larger files being transmitted and received, making the on-line communications more difficult. Similar image deteriorations are obtained in all cases, when the video processing is performed with MPEG-4 [12].

In order to solve the problems mentioned above, in the paper is presented new approach that aims to promote the accessibility to different scientific courses to students with disabilities. Instead of using the full-color video clips, here is proposed to use the contour image of the SL interpreter, which can be compressed and stored with lossless intra-frame compression for future use (Fig. 2).

The Work Hypothesis is based on the investigations presented in works [3,4,7], in accordance with which, for the successful understanding of the presented information is enough to use the contour images of the sign language interpreter only. The basic requirement in this case is the contours to be extracted from the grayscale images (the consecutive TV frames), compressed and after that - transferred with retained quality to the receiving side. The image quality is extremely important for the contours of the face and hands of the sign language interpreter. For this, the intra-frame compression is better than the inter-frame one. The compression of the contour images performed with MJPEG is with low efficiency and the MPEG compression decreases significantly the image quality. The new method solves these problems to a high degree.

Figure 1. Single TV frame with sign language interpreter

Figure 2. Contour image (single frame), 320 x 240 pixels, 8bpp.

The paper is arranged as follows: in Section II is presented the method for lossless compression of still images, in Section III are given the experimental results and Section IV is the Conclusion.

II ALGORITHM FOR LOSSLESS COMPRESSION

The method for lossless compression of still images containing mainly some kind of graphic information, comprises the basic steps, described below:

Step 1. ***Calculation of the image histogram***

The image histogram is calculated in accordance with the traditional approach.

Step 2. ***Histogram analysis***

In result of the histogram analysis is obtained information about:

- The most frequently met values in the processed image (usually these values are zeros).
- The sequences of not used ("free") brightness values in the processed image. The position (the start value) and the length of the longest sequence of not-used values are stored.
- The single not used ("free") brightness values – they are used for the coding only in case that there were not found sequences of not used brightness values in the processed data.
- Sequences of same values in the processed image data and their lengths.

Step 3. ***Coding***

The coding is performed in accordance with several basic rules:

- Every sequence of same values (zeros), whose length is smaller than that of the longest sequence of "free" values is substituted with one of these values, for example: length of two equal values is substituted by the first "free" value from the sequence; length of three equal values – by the second "free" value from the sequence, etc. In result, every such sequence is coded with one byte only;
- Every sequence of same values (zeros), whose length is longer than that of the longest sequence of "free" values, is coded as "value" and "length". These sequences are coded with 2 Bytes code combinations;
- Every sequence of same values (non-zero) is coded as "value" and "length", i.e. - with 2 Bytes code combination.
- Very long sequences of same values (zero or non-zero) are coded as "value" and "length", but using several bytes for the coding.

Step 4. ***Formatting***

The coded images are stored in new special format (tk). For this in the process of coding is created the format header, containing the information about the length of the sequence of "free" values and their start position in the brightness range, together with the most frequent value in the processed image. In the coded image header are used several flags, indicating the way of coding.

The **Block Diagram** of the method is presented in Fig.3.

III EXPERIMENTAL RESULTS

The algorithm, presented above, is very efficient when contour images or graphics are compressed. In order to obtain high efficiency when the video presentations of Sign Language interpreter have to be processed, the video files should be pre-processed properly. The pre-processing comprises the following two steps:

- The original video sequence (which had not been compressed with H.261 or MPEG-4) is stored in digital form in order to make every TV frame available for processing;
- Each TV frame is processed individually. The processing comprises:
  - Converting to grayscale;
  - Extraction of the main contours.

In result is obtained the contour image of the Sign Language Interpreter. After these procedures the image is ready for the compression.

Some of the obtained experimental results are presented below. In Fig. 4 are shown four contour images, obtained from corresponding consecutive TV frames.

Fig. 4. Test contour images, obtained from consecutive TV frames: 01c, 02c, 03c, 04c.

Image	01c	02c	03c	04c	05c	06c	07c	08c
TKView	38.9	38.9	39.02	38.5	39.02	39.28	39.16	38.7
JPEG	3.34	3.35	3.36	3.36	3.30	3.35	3.37	3.37
JPEG2000	5,82	5,81	5,82	5.80	5.81	5.83	5.77	5.77

Table 1. Compression ratios, obtained with the new method for intra-frame compression and JPEG standards.

The presented results were obtained with the program TKView, developed on the basis of the described algorithm; the results for the JPEG standard were obtained with Microsoft Photo Editor (highest quality, lossy compression), and for JPEG2000 – with Algovision LuraTech (lossless compression).
The mean compression ratio obtained for 90 test images with TKView is 39,08; with JPEG (best quality) it is 3,36 and for JPEG2000 lossless – 5,80. For the tested kind of images the results obtained for the three methods are very consistent.
The comparison results show that the new method is 11,63 times more efficient than JPEG (MS Photo Editor, best quality) and 6,74 times more efficient than JPEG2000 (lossless) for the investigated image class.

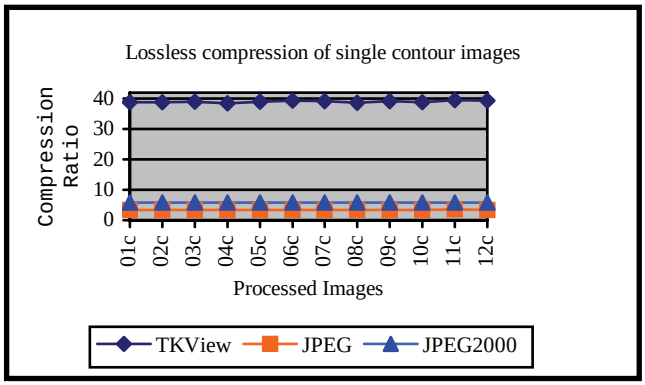


Fig. 5. Graphic presentation of the results obtained with inter-frame compression of contour images

In order to evaluate the method efficiency, the following is assumed: Each TV frame from the original video lesson, representing the greyscale contour image of the SL interpreter is treated as a single still image (320 x 240 pixels) with size 76800 bytes. After compression accordingly the new method (about 40 times) with the special software for lossless compression, the size of the file, obtained in result, is less than 2000 Bytes. The single images obtained from the corresponding consecutive TV frames after that are arranged as video clips and transferred to the receiving side, where they are restored. The high compression ratio permits relatively easy to add such information to distance-learning lessons, accessed via Internet.

IV Conclusion

The presented results show consistent high compression ratios when contour images are processed. These compressions are significantly higher than those obtained with software, based on the JPEG2000 standard (lossless).

In accordance with the presented results, the contour image of the sign language interpreter obtained from a single TV frame with size 320 x 240 pixels after lossless compression is less than 2 KB. This means that the compressed video data could be transferred as additional information, accomplishing the distance-learning lectures for hearing-impaired students. For presentations of 25 TV frames per second the necessary additional information is about 49 KBps. As it is known, the standard MPEG-2 bit-rate is in the range 6-12Mbps. The additional information required for the presentation of the sign language interpretation will require the insertion of these additional 49 KBps.
In case that the size of the used TV frame is smaller (for example, 256x192 pixels) the mean compression ratio is 38 and the additional bit rate is about 30 KBps.
The presented approach is suitable for applications concerning distance learning for students who are deaf or hearing impaired. It could be used for the creation of special handbooks and vocabularies in all cases when some kind of additional information is required and the sign language interpretation will make the explanation easier.

Another possible application is the compressed information to be inserted in some multimedia products like DVDs in order to provide the required information to the hearing impaired people in their most natural way. Same approach could be used in the TV broadcastings instead of the image of the live sign language interpreter. For these applications the contour image could be presented in a small window placed in the lower part of the TV screen.
The comparison results with the JPEG2000 standard for lossless compression of contour images show the advantages of the presented method.
The method is suitable for efficient compression of text and other graphic images, which additionally supports all kinds of distance learning applications [13].

The **further development** of the method is aimed at the following directions:

- Development of special algorithms for efficient extraction of the contours of the interpreter's hands and face in order to optimize the presented information, retaining its understandability.
- Investigation on the influence of the participating brightness levels in the extracted contours on the compression ratio.
- Investigation on the influence of the number of consecutive TV frames, compressed together, on the method efficiency.
- Adaptation of the compressed data header with the requirements of broadcasting standards and DVD formats.
- Development of a special version of the method, aimed at services in mobile communications.

V Acknowledgement

This paper was supported by NSF of Bulgarian Ministry of Education and Science (Contract VU-MI-104).

REFERENCES

[1] Subject-Based Deaf and Hard of Hearing Internet Resources. www.Internet Resouces - Deafness.htm.
[2] Teaching the Learning Disabled: Study Skills and Learning Strategies. www.Teaching the learning disabled.htm.
[3] D. Pearson, J. Robinson. Visual Communications at Very Low Data Rates. IEEE Proc., Vol. 73, 1985, pp. 222-241.
[4] P. Letellier, M. Nadler, J. F. Abramatic. The Telesign Project. Proc. of the IEEE, Vol. 73, 1985, pp. 242-258.
[5] CCITT Recommendation H.261: Video Codec for Audio- visual Services at px64 kbit/s.
[6] M. Ghanbari. Standard Codecs: Image compression to Advanced Video Coding. Institution of Electrical Engineers, 2003.
[7] M. Whybray, E. Hanna. A DSP based videophone for the hearing impaired using valledge processed pictures. Proc ICASSP'89, Vol 3, 1989, pp.1866-1869.
[8] M. Whybray. Moving picture transmission at low bitrates for sign language communication. http://www.zak.co.il/deaf-info/old/videophones.html.
[9] G. Sperling, M. Landy, Y. Cohen, M. Pavel. Intelligible encoding of ASL image sequences at extremely low information rates. Computer Vision, Graphics, and Image Processing 31, 1985, pp. 335-391.
[10] S. Tetzchner (Ed.). COST 219 book: Issues in Telecommunications and Disability. Commission of the European Communities, Catalogue number CD-NA-13845-EB-C, 1991.
[11] A. Heron, N. Mc Donald. Video transmission over a radio link using H.261 and DECT. 4-th International Conference on Image Processing and Applications, IEE, 7-9 April 1992.
[12] D. Salomon. Data compression. Springer-Verlag, New York, Inc. 2004.
[13] Kountchev R., Todorov Vl., Milanova M., Kountcheva R. *Documents Image Compression with IDP and Adaptive RLE*. The 32nd Annual Conference of the IEEE Industrial Electronics Society Conservatoire National des Arts & Métiers – Paris - FRANCE, IECON'06, November 7 – 10, 2006, pp.3261-3266.

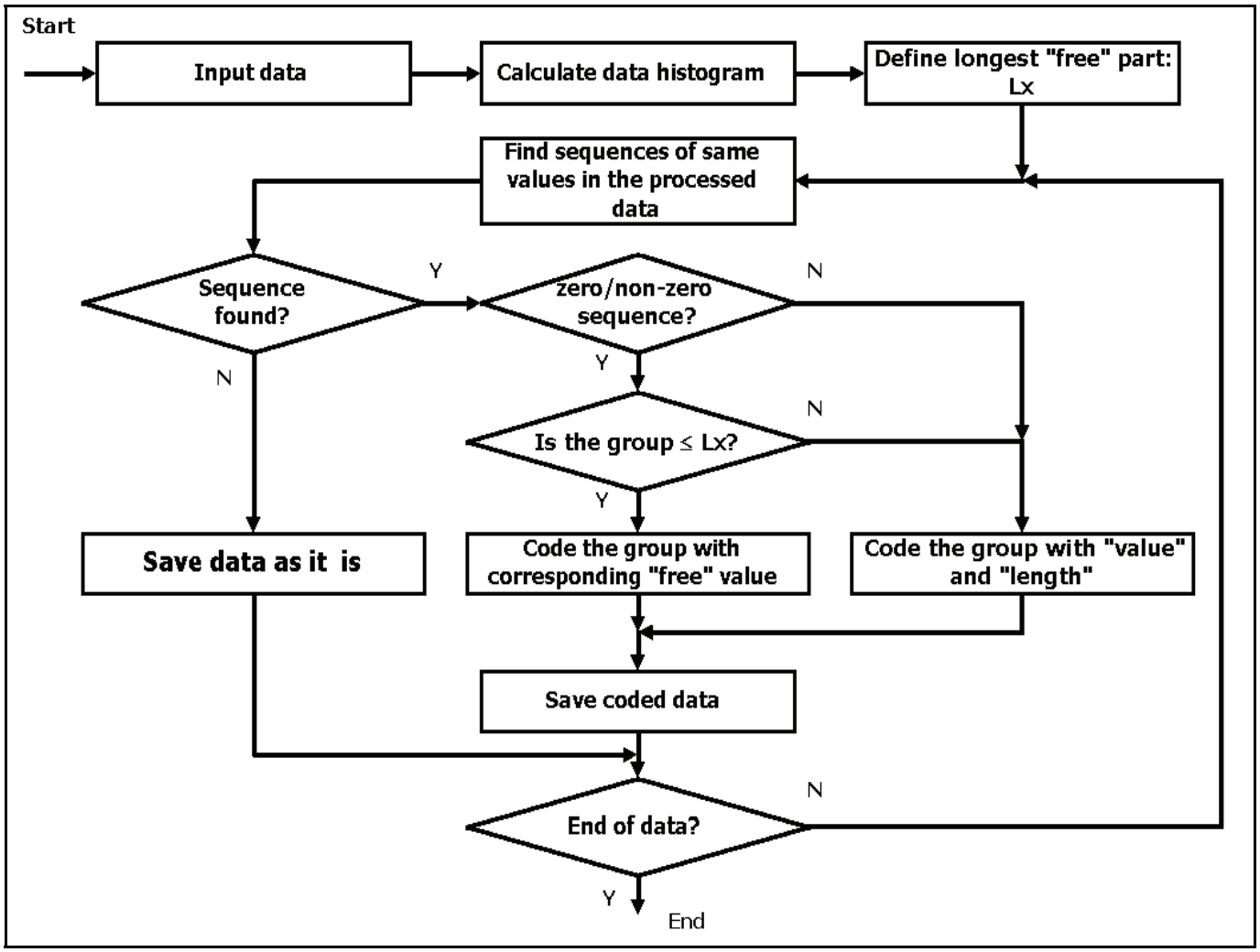


Fig. 3. Block diagram of the method for lossless coding of still images

Using Associative Memories as a Mean for Image Accumulating

Alexander B. Bekiarski[1]

***Abstract* – Image accumulating is a method for collecting the similar pictures form a chosen class. The direct storing of each image is possible but not effectively, because this cause a great amount of occupied memory. The goal of this article is to use the associative memories as a means for image accumulating. The ability of the associative memories for storing images with similar features using association is the main reason to achieving memory space decreasing with the proposed method of image accumulating. The method is examined with a simulation of the associative memory as a neural network.**

***Keywords* – Image Accumulating, Associative Memories, Neural Networks.**

I. INTRODUCTION

The image accumulating is investigated since many years. The reasons for this permanent scientific interest are different. In some of the applications, for example the communications [1], the image accumulating give the advantages in speed of transmission, in medicine [2] - increase the diagnostic precision, in robotic [3] – the recognition etc.

The image accumulating effectiveness is related to the method of image presentation in moment of the image collecting in memory or from the number of calculations for this method.

In this article it is choose to use the associative memories as a mean for image accumulating. This give some important advantages for memory size decreasing and using accumulated images in some basic image applications as features extraction, recognition, image database, fast image searching and retrieving etc.

A common way to represent an artificial model of an associative memory is a neural network [4, 5]. There are some types of neural networks, which can be used as an associative memory presentation. These types can be analyzed and compared for him advantages and effectiveness to image accumulating. A very important condition to achieve this goal is to pose some conditions of this analysis and comparison.

In this article it is used the simulation of the image accumulating in associative memory represented as a neural network. The simulation is made with real images, which are preliminary stored in one of the standard image file formats like BMP, TIFF etc. The results from the simulation are presented, analyzed, compared and followed by the comments and conclusion.

[1]Alexanded B. Bekiarski is with the Faculty of Communications, of Technical University – Sofia, 8 Kliment Ohridski St., 1756 Sofia, Bulgaria, E-mail: aabbv@tu-sofia.bg

II. ASSOCIATIVE MEMORIES REPRESENTATION

The artificial associative memory representation takes the idea from the human ability to perceive, collect and "remembering" images in brain not only from the information of the plain images, but in the cases when eyes see only a representative part of the images or an image with some distortions. These parts or distorted images serve as a "cue" for the plain or original images. This human ability is investigated carefully and it is proved, that it is related of the way the image information is entered, processed and stored in the brain. The way of storing the images in the conventional electronic memories differ mainly from this of the human. This is because the conventional memories work like one, two or multi dimensional linear arrays. The places (or addresses) of the stored images are not related with any features of the images, they are arranged only in size or in time when they are entered and stored. This is a very non effective, but easy method, which is commonly used if there is not so important to manipulate the stored images in the sense of feature searching or retrieving, clustering of the similar images, classification, recognition etc. The implementation of the human model of collecting images in an associative manner is the base of the artificial associative memory representation.

The other fundamental idea of artificial associative memory representation is taken from the physical phenomenon and systems. In some physical systems, for example a bowl in which a ball bearing is allowed to roll freely, it is possible to describe its operations in terms of the network's `energy'. Suppose the ball go from a point somewhere up the side of the bowl with, possibly, a push to one side as well. The ball will roll back and forth and around the bowl until it comes to rest at the bottom.

The physical description of what has happened may be couched in terms of the energy of the system. The ball initially has some potential energy. That is work was done to push it up the side of the bowl to get it there and it now has the potential to gain speed and acquire energy. When the ball is released, the potential energy is released and the ball rolls around the bowl (it gains kinetic energy). Eventually the ball comes to rest where its energy (potential and kinetic) is zero. (The kinetic energy gets converted to heat via friction with the bowl and the air). The main point is that the ball comes to rest in the same place every time and this place is determined by the energy minimum of the system (ball + bowl). The resting

state is said to be stable because the system remains there after it has been reached.

There is another way of thinking of this process which ties in with ideas about memory. Suppose that the ball comes to rest in the same place each time because it `remembers' where the bottom of the bowl is. The location of the bottom of the bowl can be described as a vector $\mathbf{x_0}$, which represents the stored pattern. Thus, the ball position or "state" at any time is given by the three coordinates or the position vector $\mathbf{x}$. By writing the ball's vector as the sum of vector $\mathbf{x_0}$ and a displacement $\Delta\mathbf{x}$ thus,

$$\mathbf{x} = \mathbf{x_0} + \Delta\mathbf{x} \tag{1}$$

It is possible to think of the ball's initial position as representing the partial knowledge or cue for recall, since it approximates to the memory $\mathbf{x_0}$ .If it is now used a corrugated surface instead of a single depression (the bowl) it may store many `memories' (vectors):

$$\{\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_3}, \mathbf{x_4}, \ldots\ldots\ldots\ldots\} \tag{2}$$

If now the ball is started somewhere on this surface, it will eventually come to rest at the local depression which is closest to its initial starting point. That is it evokes the stored pattern which is closest to its initial partial pattern or cue. Once again, this is an energy minimum for the system.

There are therefore two complementary ways of looking at what is happening. One is to say that the system falls into an energy minimum; the other is that it stores a set of patterns and recalls that which is closest to its initial state. If it is build a network which behaves like this, it is completely described by the following key elements:

- a state vector

$$\mathbf{v} = (v_1, v_2, \ldots\ldots\ldots, v_n); \tag{3}$$

- a set of stable states

$$\mathbf{v_1}, \mathbf{v_2}, \ldots\ldots\ldots\mathbf{v_m}, \tag{4}$$

which correspond to the stored patterns and, in, the corrugated surface example, were the bottoms of the depressions in the surface;

- the system evolves in time from any arbitrary starting state $\mathbf{v}$ to one of the stable states, and this may be described as the system decreasing its energy E. This corresponds to the process of memory recall.

The 'energy' principles described here can be used as a base to image accumulating designing an associative memory, which is responding to the mentioned above conditions. There are many possibilities to realize the associative memory working for image accumulating. In this article it is proposed to use a neural network for this purpose. The type of neural network can be chosen from a lot of existing neural network types and each of these types must be analyzed, examined etc.

III. Associative Memories with Neural Networks

A very often representation of an associative memory is a neural network. There are some types of neural networks, which can be used as an associative memory presentation and can be used for image accumulating. After a precision analysis of these types it is decided to use in this work one of the most applicable neural networks type - the Hopfield Neural Network. The architecture, as it is presented in Matlab for simulation of this network is shown in Fig.1.

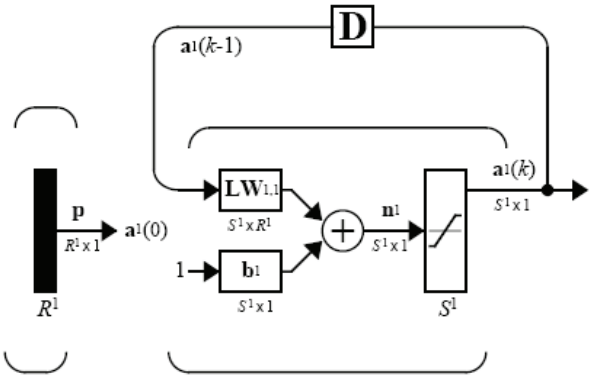


Fig. 1. Figure example

The input vector $\mathbf{p}$ of the Hopfield neural network have R^1x1 dimension. For the image accumulating purpose it is necessary to calculate the dimension for the input vector as:

$$R = N_x + N_y, \tag{5}$$

where

N_x, N_y are image dimensions in horizontal and vertical direction.

The other Hopfield parameters, shown in Fig.1, are described precisely in Matlab Neural Network Toolbox [5], but in this article it is important to discuss the number of neurons in the network - N_{num}. This number can be determined for a concrete Hopfield neural network application as an associative memory for image accumulating, but it must satisfy the following condition:

$$N_{num} \geq N_{icl}, \tag{5}$$

where

N_{icl} is the number of image classes chosen for the desired image accumulating example.

The Hopfield neural network can be consider also as a tree. Their nodes are the states of the network. They can be thinking as a net of images for accumulating. Every node is connected to every other node (but *not* to itself) and the connection strengths or weights are symmetric in that the

weight from node *i* to node *j* is the same as that from node *j* to node *i*.

This means that:

$$w_{ij} = w_{ji}, \text{ and } w_{ii} = 0 \text{ for all } i, j. \quad (6)$$

Notice that the flow of information in this net is not in a single direction as it has been in the nets dealt with so far. It is possible for information to flow from a node back to itself via other nodes. That is, there is feedback in the network and so they are known as feedback or *recurrent* nets as opposed to *feedforward* nets which were the subject of the Backpropagation algorithm. This mean that the transitions for a node forward and backward is a change for an image to another image.

The state of the net at any time is given by the vector of the nodes outputs:

$$(x_1, x_2, x_3, ...). \quad (7)$$

Suppose, that the net now start in some initial state or image and choose a node at random and let it update its output or `fire'. That is, it evaluates its activation in the normal way and outputs a `1' if this is greater than or equal to zero and a `0' otherwise. The net now finds itself either in the same state or image as it started in, or in a new state, that mean new image, which is at Hamming distance one from the old. Then it is chosen a new node or image at random to fire and continue in this way over many steps. For each state, it may evaluate the next state given each of the nodes fires. This means to activate a new image for accumulating. An example of state diagram for using Hopfield neural network as associative memory for image accumulating is given in Table 1.

TABLE I
STATE DIAGRAM FOR IMAGE ACCUMULATING

Current state in time t	New state in time $t+1$
$x_1(t)$	$x_1(t+1)$
$x_2(t)$	$x_2(t+1)$
$x_3(t)$	$x_3(t+1)$
.....	.....
.....	.....

The other way to describe state transition in tree representation of image accumulating is as a *state transition diagram*.

States or images are represented by the circles with their associated state number. Directed arcs represent possible transitions between states or images and the number alongside each arc is the probability that each transition will take place. The states or images have been arranged in such a way that transitions tend to take place down the diagram; this will be shown to reflect the way the system decreases its energy and represent the principle of the association. The important thing to notice at this stage is that, no matter where start in the diagram, the net will eventually find itself in one of the stable states or image. These reenter themselves with probability 1. That is they are stable states or accumulated images - once the net finds itself in one of these it stays there. These state vectors or images are the `memories' stored by the net, it mean stored or accumulated images.

The dynamics of the net accumulating images are described perfectly by the state transition table or diagram. However, greater insight may be derived if we can express this in terms of an energy function and, using this formulation, it is possible to show that stable states or images will always be reached in such a net. Consider two nodes i, j, representing images in the net which are connected by a positive weight and where *j* is currently outputting a `0' while *i* is outputting a `1'.

If for the image *j* were given the chance to update or fire, the contribution to its activation from image *i* is positive and this may well serve to bring *j*'s activation above threshold and make it output a `1'. A similar situation would prevail if the initial output states of the two nodes, representing images, had been reversed since the connection is symmetric. If, on the other hand, both units are `on' they are reinforcing each other's current output. The weight may therefore be thought of as fixing a constraint between *i* and *j* images that tends to make them both take on the value `1'. A negative weight would tend to enforce opposite outputs. One way of viewing these image accumulating networks is therefore as *constraint satisfaction* nets.

This idea may be captured quantitatively in the form of a suitable energy function for the neural network for image accumulating. Define:

$$e_{i,j} = -w_{i,j} x_i x_j \quad (8)$$

If the weight is positive then the last entry is negative and is the lowest value in the table. If is regarded as the `energy' of the image pair *ij* then the lowest energy occurs when both units are on which is consistent with the arguments above. If the weight is negative, the image state is the highest energy state and is not favoured. The energy of the neural network is found by summing over all pairs of nodes or presented images:

$$E = \sum_{pairs} e_{i,j} = -\sum_{pairs} w_{i,j} x_i x_j \quad (9)$$

This may be written:

$$E = -\frac{1}{2} \sum_{i,j} w_{i,j} x_i x_j \quad (10)$$

Since the sum includes each pair twice as $w_{i,j} x_i x_j$ and $w_{j,i} x_j x_i$ and $w_{i,j} = w_{j,i}$, $w_{i,i} = 0$.

It is now instructive to see what the change in energy of the neural network is when a node fires or an image is activated. Suppose node *k* is chosen to be updated with a new image. The energy E of the neural network by singling out the terms involving this node can be written as:

$$E = -\frac{1}{2}\sum_{\substack{i\neq k\\ j\neq k}} w_{i,j}x_i x_j - \frac{1}{2}\sum_i w_{k,i}x_k x_i - \\ -\frac{1}{2}\sum_i w_{i,k}x_i x_k \qquad (11)$$

Now, because $w_{i,k} = w_{k,i}$, the last two sums may be combined

$$E = -\frac{1}{2}\sum_{\substack{i\neq k\\ j\neq k}} w_{i,j}x_i x_j - \sum_i w_{k,i}x_k x_i \qquad (12)$$

For ease of notation, denote the first sum by S and take the x_k outside the sum since it is constant throughout, then the energy E of the neural network is:

$$E = S - x_k\sum_i w_{k,i}x_k x_i, \qquad (13)$$

but the sum here is just the activation of the kth node so that:

$$E = S - x_k a^k \qquad (14)$$

Let the energy E of the neural network after k has updated be $E^{'}$ and the new output be $x_k^{'}$. Then:

$$E^{'} = S - x_k^{'}a^k \qquad (15)$$

Denote the change in energy $E^{'} - E$ by ΔE and the change in output $x_k^{'} - x_k$ by Δx_k , then subtracting (14) from (15) gives:

$$\Delta E = -\Delta x_k a^k \qquad (16)$$

There are now two cases to consider:

- $a^k \geq 0$. Then the output goes from `0' to `1' or stays at `1'. In either case $\Delta x_k \geq 0$. Therefore $\Delta x_k a^k \geq 0$ and so, $\Delta E \leq 0$;
- $a^k \langle 0$. Then the output goes from `1' to `0' or stays at `0'. In either case $\Delta x_k \leq 0$. Therefore $\Delta x_k a^k \geq 0$ and so, $\Delta E \leq 0$;

Thus, for any node or image accumulating being updated we always have $\Delta E \leq 0$ and so the energy of the neural network decreases or stays the same. But the energy of the associative memory is bounded below by a value obtained by putting all the $x_i x_j = 1$ in (10). Thus energy of the neural network E must reach some fixed value and then stay the same. Once this has occurred, it is possible for further changes in the network's state to take place since $\Delta E = 0$ is still allowed. However, for this to be the case $\Delta x_k \neq 0$ and $\Delta E = 0$ it must have $a^k = 0$. This implies the change must be of the form $0 \to 1$. There can be at most N (of course there may be none) of these changes, where N is the number of nodes or images to accumulating in the neural network. After this there can be no more change to the net's state and a stable state has been reached. This means that the neural network and respectively the associative memory is in the stable state – the accumulated image.

IV. Conclusion

The above described equations are used to simulating the work of the proposed associative memory as a mean for image accumulating. The preliminary results are good and show, that the proposed method is applicable in some cases when the goal is to collect images as a database or in some preferred features used as an association for accumulating or retrieving images.

References

[1] R. H. Reeve, P. Tim-Kui Tam, “The Image-Accumulation Technique as a Variable in Multiple-Image Communication”, Annual Meeting of the Association for Educational Communications and Technology, pp. 341-350, Philadelphia, Pa., March 22-26, 1991.

[2] K. Numata, M. Morimoto, T. Isozaki, "Use of accumulation images obtained by arterial-phase contrast-enhanced harmonic grayscale ultrasonography to evaluate tumor vessels in focal nodular hyperplasia and other hepatic tumors", Springer Journal of Medical Ultrasonics, Conference Proceedings, pp. 3-10, Volume 34, Number 1, March, 2007.

[3] M. E. Nilsback, “A Cue-Integration Scheme for Object Recognition Using Discriminative Accumulation”, Thesis, Department of Numerical Analysis and Computer Science, Stockholm, Sweden, 2004.

[4] T. Kohonen, “Self-Organizing Maps”, Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York, 1995, 1997, 2001, Third Extended Edition.

[5] H. Demut, M. Bea, “Neural Network Toolbox”, The MathWork, 2006.

Design of Huffman Decoder FPGA Core

Ivan Mezei[1] and Rastislav Struharik[2]

***Abstract* – In this paper design of an Huffman decoder FPGA core that is part of an motion JPEG system is presented. Core is designed and verified using VHDL hardware description language. It is implemented and tested on an Virtex2 FPGA platform from Xilinx Inc.**

***Keywords* – Huffman decoding, motion JPEG, FPGA.**

I. INTRODUCTION

In this paper the development of a Huffman core that is to be implemented on Field Programmable Gate Array (FPGA) is presented. Core is a part of a project consisting of several cores that are part of an motion JPEG system [1]. The overall system is depicted in Fig. 1. It consists of Discrete Cosine Transformation (DCT), Quantizer and Entropy (Huffman) based encoder and decoder.

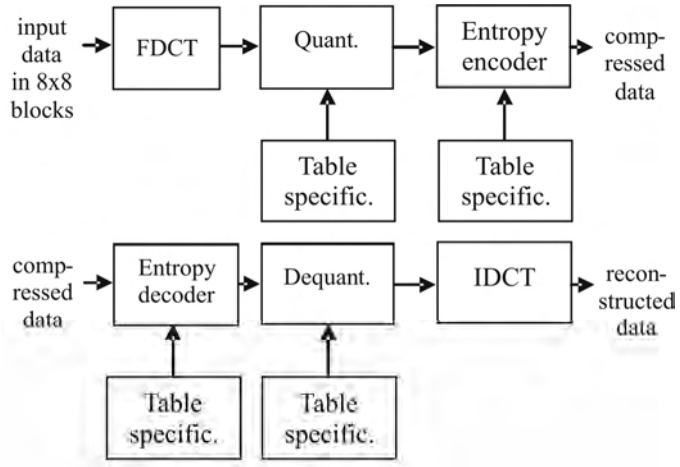


Fig. 1. Motion JPEG system

Motion JPEG is such coding system where all frames of motion pictures are coded as JPEG still pictures and then transmitted. Every picture is decomposed into luminance and chrominance components and after that in macro blocks that are 8x8 pixels sized.

It is well known that DCT is widely used in JPEG encoders [2] since it packs image data into almost optimal number of decorellated coefficients that can be compressed efficiently. Spectral energy of every transformed block is concentrated in top-left corner of macro block, meaning that in that area are coefficients that are significantly different then zero.

After quantization all coefficients that are close to zero become zero and only few coefficients remains non-zero concentrating around top-left corner.

Afterwards Huffman entropy encoding is usually done by run length encoding in zig-zag read order of transformed and quantized image data block thus producing compressed image data.

On the receiver side inverse steps are taken, and after Huffman decoding, run-length decoding, inverse zig-zag reconstruction of every block, dequantization and finally after inverse DCT image data block is decoded. This is not lossless decoding since compression is also not lossless, but this is not the problem in JPEG systems.

II. HUFFMAN DECODING

Theoretical background about Huffman coding is given in [3]. Application in JPEG compression systems can be found in [4-6]. In recent years there are several improvements in Huffman coding [2], but purpose of this paper is not to make improvements of Huffman coding algorithm but to make an core to be implemented on FPGA.

There are several possibilities for Huffman decoder architectures and various optimizations [7]. We decided to make serial sequential implementation with parallel extraction of decompressed data. Algorithm that explains how decoder works is depicted in Fig. 2.

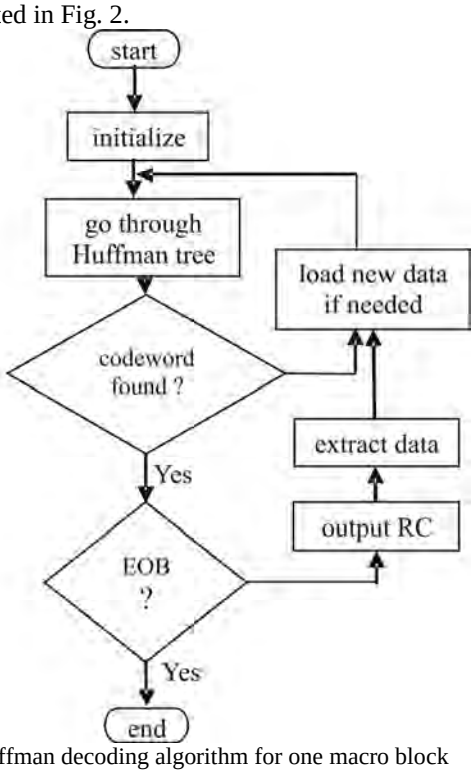


Fig. 2. Huffman decoding algorithm for one macro block

Compressed data bytes, that are input of decoder, are a sequence of Huffman codeword and additional bits in

[1]Ivan Mezei is with the Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia, E-mail: imezei@uns.ns.ac.yu

[2]Rastislav Struharik is with the Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia, E-mail: rasti@EUnet.yu

between. According to algorithm depicted in Fig. 2., decoder operates as follows. After initialization (loading of first compressed byte etc.), decoder goes through Huffman tree. Depending on input control signal, it uses appropriate luminance or chrominance table until it reaches valid codeword. Once codeword is found, RC coefficient can be sent to output of decoder. RC coefficient is concatenated **C**ategory of codeword and zero **R**un-length. According to category of codeword additional bits can be extracted and sent to output as decoded image data.

Using category and zero run-length, image data block can be reconstructed using inverse zig-zag and zero run-length inserting. Dequantizer and inverse DCT transformation complete the decoding process and decompressed image data is obtained as output of decoder.

III. IMPLEMENTATION

Huffman decoder core is to be implemented on Xilinx FPGA circuits. Manufacturer gave some guidelines for implementing decoder in FPGA [8]. There are different FPGA-based decoders to be used with JPEG decoding systems [9-10], and also commercial ones [11]. Our implementation is based on shift registers that serially shift out bit by bit as we go through Huffman tree based on compressed data from current macro block. After codeword have been found, additional bits are sent to output in parallel. Tables with luminance and chrominance RC coefficients are implemented as ROM memories. A small control unit controls how decoder operates.

Simplified block diagram of decoder core is shown in Fig. 3.

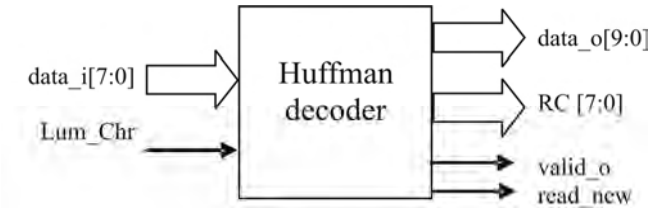


Fig. 3. Simplified decoder block

Huffman decoder core is completely designed and verified using VHDL hardware description language. Core is designed in 3 VHDL files with around 1000 lines of VHDL RTL code (half of which is luminance and chrominance coefficients table implementation). No time or spatial constraints are used.

Core is implemented and tested on 2v1000fg456-6 FPGA from Virtex2 family. Device utilization summary are shown in Table I. Decoder occupies around 5% of FPGA integrated circuit and estimated maximum frequency is around 90MHz. Possibly higher frequency could have been obtained by placing some constraints to critical signals. This was not needed since target clock frequency is 27MHz.

TABLE I
DEVICE UTILIZATION SUMMARY FOR VIRTEX2 FPGA

Device	2v1000fg456-6
Number of Slices	296 out of 5120 (5%)
Number of Slice Flip Flops	76 out of 10240 (0%)
Number of 4 input LUTs	554 out of 10240 (5%)
Number of BRAMs	4 out of 40 (10%)
Maximum Frequency	87.581MHz

Device utilization summary for one of the latest FPGAs from Xilinx (5vlx50ff1153-3 from Virtex 5 family) is shown in Table II. For this FPGA, decoder occupies below 1% of FPGA and estimated maximal frequency is around 176 Mhz.

Table II
DEVICE UTILIZATION SUMMARY FOR VIRTEX5 FPGA

Device	5vlx50ff1153-3
Number of Slice Registers	90 out of 28800 (0%)
Number of Slice LUTs	451 out of 28800 (1%)
Number of BRAMs	2 out of 48 (4 %)
Maximum Frequency	176.947 MHz

IV. CONCLUSION

In this paper design of an Huffman decoder FPGA core is presented. Core is successfully implemented and tested on Virtex2 FPGA family from Xilinx Inc. Design implementation summary for one of the latest families is reported also but testing is to be performed since we don't have that FPGA integrated circuit yet.

It is shown that estimated maximum clock frequency is around 90MHz for this implementation without any constrains applied. Actual frequency with which the core is tested is 27MHz.

V. REFERENCES

[1] P. Symes, *Video Compression Demystified*, McGraw-Hill, 2000.
[2] G. Lakhani, "Optimal Huffman Coding of DCT Blocks", IEEE Trans. on Circuits and Systems for Video Technology, vol. 14, no. 4, pp. 522-527, 2004.
[3] D.A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes", Proc. of the IRE, vol. 40, no. 9, pp. 1098-1101, 1952.
[4] T. Acharya, A.K.Ray, *Image Processing – Principless and Applications*, Wiley, 2005.
[5] A. Cernasov, *Digital Video Electronics*, McGraw-Hill, 2004.
[6] G.K. Wallace, "The JPEG Still Picture Compression Standard", Communication of the ACM 34, pp. 30-44, 1991.
[7] C. Schneider, "A Parallel/Serial Trade-Off Methodology for Look-Up Table Based Decoders", Proc. DAC, pp. 498-503, 1997.
[8] L. Pillai, "Huffman coding", Application note 616, Xilinx Inc.
[9] Z. Yusof, Z. Aspar, I. Suleiman, "Field programmable gate array (FPGA) based baseline JPEG decoder", Proc. TENCON, vol. 3, pp. 218-220, 2000.
[10] M. Elbadri, R. Peterkin, V. Groza, D. Ionescu, and A. Saddik, "Hardware support of JPEG", Proc. Canadian Conference on Electrical and Computer Engineering, pp. 812-815, 2005.
[11] JPEG-D decoder core, www.cast-inc.com

An Algorithm for Accelerated Template Localization in an Image Using Correlation Coefficient

Yulka P. Petkova[1], Vassil J. Smarkov[2] and Milena N. Karova[3]

***Abstract* – An algorithm for accelerated template localization using computation of the correlation coefficient is proposed in the paper. Correlation coefficient is one of the most reliable matching measures used in the tasks of searching and localizing a before chosen template in a larger image. Computations are made in the space area and they are made with a set of before extracted representative points of the template. The computational complexity of the proposed algorithm is kept the same as this of the classic algorithm. The acceleration is due to the reduced number of operations. This is achieved by the "nonpromising" positions elimination. Comparisons with two elimination conditions are used for this purpose. The elimination conditions become more accurate during the calculation process. Besides it is proposed the sets of representative points to be divided into two subsets according to a before sattled feature and the points from one subset only to be involved in the calculations. The points from the second subset take part in the accurate calculation of the corerrelation coefficient in the "promising" positions only. The theoretical evaluations and the experimental results also acknowledge the priority of the proposed algorithm over the classical algorithms. Results are compared on one hand with the classic algorithm for computations over all points of the template and on the other hand it is compared with the algorithm for computations over the set of representative points.**

***Keywords* – template matching, template localization, correlation coefficient**

I. INTRODUCTION

Correlation coefficient overcomes the instability of the normalized cross-correlation to the local intensity changes. This is achieved with the including the mean intensity values of the template and the image in the computations.

Algorithms for template localization using correlation metrics, calculated in the space domain have great computational complexity – O(N.M), where N and M are sizes of the image and the template respectively. When these metrics are computed using a set of representative points, their computational complexity reduces to O(N.P), where P is a number of representative points. Despite all it remains high. This is the reason for searching faster algorithms [1], [4].

In this paper an algorithm for accelerated template localization in a greater image is proposed. It computes the correlation coefficient in the space domain. The acceleration is due to the fast dissimilarity principle and to the computations of the partial correlation coefficient.

II. SUGGESTION FOR ACCELERATED TEMPLATE LOCALIZATION WITH CORRELATION COEFFICIENT COMPUTATION

A. Description of the algorithm

It is known that correlation coefficient which is computed over the chosen set of representative points can be expressed as follows:

$$NCCE(x,y)=\frac{\sum_{q=1}^{P}\left(I_q-\bar{I}\right).\left(T_q-\bar{T}\right)}{\sqrt{\sum_{q=1}^{P}\left(I_q-\bar{I}\right)^2}.\sqrt{\sum_{q=1}^{P}\left(T_q-\bar{T}\right)^2}} \quad (1)$$

where: $\bar{T}=\sum_{q=1}^{P}T_q/p$ is the mean intensity value of the representative points of the template $q = 1, 2, \ldots, P$ and $\bar{I}=\sum_{q=1}^{P}I_q/P$ is the mean intensity value of the corresponding points of the image.

For a set of representative points

$$T=\left\{T_1^L(i_1,j_1),T_2^L(i_2,j_2),\ldots,T_p^L(i_p,j_p),T_{p+1}^H(i_{p+1},j_{p+1}),T_{p+2}^H(i_{p+2},j_{p+2}),\ldots,T_P^H(i_P,j_P)\right\}$$

the correlation coefficient at point *(x,y)* can be represented as:

$$NCCE(x,y)=\frac{\sum_{q=1}^{p}\left(I_q-\bar{I}\right).\left(T_q^L-\bar{T}\right)+\sum_{q=p+1}^{P}\left(I_q-\bar{I}\right).\left(T_q^H-\bar{T}\right)}{\sqrt{\sum_{q=1}^{p}\left(I_q-\bar{I}\right)^2+\sum_{q=p+1}^{P}\left(I_q-\bar{I}\right)^2}.\sqrt{\sum_{q=1}^{p}\left(T_q^L-\bar{T}\right)^2+\sum_{q=p+1}^{P}\left(T_q^H-\bar{T}\right)^2}} \quad (2)$$

If note

$$NCCE^L(x,y)=\frac{\sum_{q=1}^{p}\left(I_q-\bar{I}\right).\left(T_q^L-\bar{T}\right)}{\sqrt{\sum_{q=1}^{p}\left(I_q-\bar{I}\right)^2+\sum_{q=p+1}^{P}\left(I_q-\bar{I}\right)^2}.\sqrt{\sum_{q=1}^{p}\left(T_q^L-\bar{T}\right)^2+\sum_{q=p+1}^{P}\left(T_q^H-\bar{T}\right)^2}} \quad (3)$$

and

$$NCCE^H(x,y)=\frac{\sum_{q=p+1}^{P}\left(I_q-\bar{I}\right).\left(T_q^H-\bar{T}\right)}{\sqrt{\sum_{q=1}^{p}\left(I_q-\bar{I}\right)^2+\sum_{q=p+1}^{P}\left(I_q-\bar{I}\right)^2}.\sqrt{\sum_{q=1}^{p}\left(T_q^L-\bar{T}\right)^2+\sum_{q=p+1}^{P}\left(T_q^H-\bar{T}\right)^2}} \quad (4)$$

than for *NCCE(x,y)* can be written:

$$NCCE(x,y)=NCCE^L(x,y)+NCCE^H(x,y) \quad (5)$$

[1]Yulka P. Petkova is with the Computer Science and Technologies Department, Technical University of Varna, Bulgaria, E-mail: jppet@mbox.digsys.bg

[2]Vassil J. Smarkov is with the Computer Science and Technologies Department, Technical University of Varna, Bulgaria, E-mail: smarkov@ieee.bg

[3]Milena N. Karova is with the Computer Science and Technologies Department, Technical University of Varna, Bulgaria, E-mail: mkarova@ieee.bg

In the ideal case when the compared template and the current window of image coincide, the intensities at every pair of points are equal, i.e. $I_q = T_q^L$ for $q = 1, 2, \dots, p$ and $I_q = T_q^H$ for $q = p+1, p+2, \dots, P$; p and $P \neq 0$. It means that average values are also equal. Than after the substitution in (2), (3) and (4) it can be written:

$$NCCE^{Id} = \frac{\sum_{q=1}^{p}\left(T_q^L - \overline{T}\right)^2}{\sum_{q=1}^{p}\left(T_q^L - \overline{T}\right)^2 + \sum_{q=p+1}^{P}\left(T_q^H - \overline{T}\right)^2} + \frac{\sum_{q=p+1}^{P}\left(T_q^H - \overline{T}\right)^2}{\sum_{q=1}^{p}\left(T_q^L - \overline{T}\right)^2 + \sum_{q=p+1}^{P}\left(T_q^H - \overline{T}\right)^2} \quad (6)$$

It can be written in another way also:

$$NCCE^{Id} = NCCE^{L,Id} + NCCE^{H,Id} \quad (7)$$

The ideal values of partial correlation coefficients $NCCE^{Id}$, $NCCE^{L,Id}$ and $NCCE^{H,Id}$ can be computed in advance and only once.

Let $\overline{I} = \overline{T}$. Than the correlation coefficient at point *(x,y)*, computed over the subset of points T^L can be approximated with:

$$NCCE^{L,Appr}(x,y) = \frac{\sum_{q=1}^{p}\left(I_q - \overline{T}\right).\left(T_q^L - \overline{T}\right)}{\sqrt{\sum_{q=1}^{p}\left(I_q - \overline{T}\right)^2 + \sum_{q=p+1}^{P}\left(T_q^H - \overline{T}\right)^2}.\sqrt{\sum_{q=1}^{p}\left(T_q^L - \overline{T}\right)^2 + \sum_{q=p+1}^{P}\left(T_q^H - \overline{T}\right)^2}} \quad (8)$$

$\overline{T}$, $\left(T_q^L - \overline{T}\right)$, $\sum_{q=1}^{p}\left(T_q^L - \overline{T}\right)^2$ и $\sum_{q=p+1}^{P}\left(T_q^H - \overline{T}\right)^2$ are already computed. Thus the first elimination condition can be formulated:

$$\left|NCCE^{L,Id} - NCCE^{L,Appr}(x,y)\right| > (\varepsilon_{NCCE^L})_{\min_1}, \quad (9)$$

where $(\varepsilon_{NCCE^L})_{\min_1}$ is the current computed minimum value of the deviation of $NCCE^{L,Appr}(x,y)$ from the ideal value $NCCE^{L,Id}$.

If this condition is satisfied, then the current position of the template over the image can be ignored as an "unpromising" without any additional computations and the next position can be checked.

If the elimination condition is not satisfied then the real value of the correlation coefficient $NCCE^L$ at point *(x,y)* must be computed:

$$NCCE^{L}(x,y) = \frac{\sum_{q=1}^{p}\left(I_q - \overline{I}\right).\left(T_q^L - \overline{T}\right)}{\sqrt{\sum_{q=1}^{p}\left(I_q - \overline{I}\right)^2 + \sum_{q=p+1}^{P}\left(I_q - \overline{I}\right)^2}.\sqrt{\sum_{q=1}^{p}\left(T_q^L - \overline{T}\right)^2 + \sum_{q=p+1}^{P}\left(T_q^H - \overline{T}\right)^2}} \quad (10)$$

Thus, the second condition for the elimination of the current position can be formulated:

$$\left|NCCE^{L,Id} - NCCE^{L}(x,y)\right| > (\varepsilon_{NCCE^L})_{\min_2} \quad (11)$$

In this condition $(\varepsilon_{NCCE^L})_{\min_2}$ is the current minimal difference between the ideal and the real correlation coefficient at point *(x,y)*. If this condition is satisfied, this means that the current position is not "promising", i. e. it can not correspond to the correlation maximum and thus the computations are interrupted and this position is skipped.

If the condition is not satisfied, than the real $NCCE^H$ at point *(x,y)* is computed using equation (4). For this purpose only the numerator of $NCCE^H$ has to be computed. The other parts of the equation (the denominator and the average intensity value of the image $\overline{I}$) are already computed.

The maximum value of the correlation coefficient determines the position in the image where the best coinciding between the template and the current window of the image.

The order of computations can be changed and the elimination conditions can be formed on the basis of subset T^H.

B. Sets of representative points

- Edge points

In [3] we propose a new edge definition, based on the interrupted first derivative of the intensity function. In practice at the points of interruption second derivative has local extreme values. The rule for determining if a point belongs to an edge (for one-dimensional case) is:

$$\text{If} \quad \left| \underset{i}{extr}\left(\frac{d^2 I(x)}{dx^2}\right)\right| > \theta, \quad \text{then} \quad i \in E, \quad (12)$$

where *I(x)* is the intensity function; *E* is a set of points *i*, which are edge points; θ is a before settled threshold.

Here we propose to divide such formed set *E* into two subsets, depending on the sign of the local extremum. We propose the following **definition:**

Definition 1: The set *E* consists of two subsets, which are formed according the following criteria:

$$\text{If} \quad \left(\underset{i}{extr}\left(\frac{d^2 I(x)}{dx^2}\right)\right) > 0, \quad \text{then} \quad i \in E_L \quad (13.1)$$

and

$$\text{If} \quad \left(\underset{i}{extr}\left(\frac{d^2 I(x)}{dx^2}\right)\right) < 0, \quad \text{then} \quad i \in E_H \quad (13.2)$$

The set of all edge points *E* is a sum of both subsets

$$E = E_L \cup E_H . \quad (14)$$

- Informative points, selected with the method of equipotential planes

In [2] we propose extracting the representative points to be conformed to the criterion of D-optimality. According to this criterion the most informative points lie on the protruded peripheral wrapping of the object. Thus, we propose the choice to be made by equipotential planes, which are parallel to the plane *xOy* and which cut the three-dimensional image profile (relief) on proper intensity levels. Thus, the extracted points outline the horizontal contours of the local "hollows" and "hills" from the three-dimensional image profile.

Here we propose to divide such formed sets into two subsets, depending on their belonging to local "hollow" or "hill", according to the following **definition:**

Definition 2: Set of points *EP* consists of two subsets, which are formed according to the following criteria:

$$\text{If}\begin{pmatrix}[(PI(i,j)\le Pt_{max})\cap(PI(i+1,j)>Pt_{max})]\cup\\ \cup[(PI(i,j)\ge Pt_{max})\cap(PI(i+1,j)<Pt_{max})]\end{pmatrix},\text{then}(i,j)\in EP_H \quad (15.1)$$

$$\text{If}\begin{pmatrix}[(PI(i,j)\le Pt_{min})\cap(PI(i+1,j)>Pt_{min})]\cup\\ \cup[(PI(i,j)\ge Pt_{min})\cap(PI(i+1,j)<Pt_{min})]\end{pmatrix},\text{then}(i,j)\in EP_L \quad (15.2)$$

where: $PI(i,j)$ is the intensity at the point (i,j); Pt_{min}, Pt_{max} are intensity values, which determine the distance between the *xOy* plane and the planes, cutting low and high parts of the relief ("hollows" and "hills"); EP_L, EP_H are sets of points, outlining "hollows" and "hills".

The set of all informative points *EP* is an union of both subsets:

$$EP = EP_L \cup EP_H \quad (16)$$

C. Advantages of the proposed algorithm

The proposed algorithm is based on the "fast dissimilarity principle". It means that less data are needed to make a conclusion that two object (in our case images) are different than they are equivalent.

When the template and the method for representative point extraction are properly chosen, than the representative points can be situated over the hall image but not only over the bounded part of it.

During the process of computations the elimination conditions are made more precise.

III. EVALUATION OF THE PROPOSED ALGORITHM

A. Analytical evaluation

The computational complexity of the proposed algorithm is equal to the complexity of the classical algorithm for correlation coefficient computation and it is *O(N.P).* The acceleration is due to the reducing the number of operations, because of the "unpromising" positions elimination. The power of acceleration is greatly dependent on the data and has not an absolute evaluation.

According to Nakov P. and P. Dobrikov [5] "… elementary operation is this one, which is executed for a constant time, independently on the size of data. Elementary operations in common case are for example addition, multiplication etc.".

In the algorithms for correlation coefficient computation the most numerous operations are addition and multiplication (of real numbers) and their number is approximately the same. The number of more complicated operations like division and square root is significantly less than the number of the rest operations. This is the reason they are accepted as equivalent with the rest.

Thus the number of elementary operations in the classic algorithm for correlation coefficient computation over the all points of the template is:

$$NEO(NCCE)_{Cl}^{All} = 6.N.M + 4.N \quad (17)$$

where $N = n_1$ x n_2 is the size of the image and $M = m_1$ x m_2 is the size of the template.

The number of elementary operations in the classic algorithm for correlation coefficient computation over the set of representative points is:

$$NEO(NCCE)_{Cl}^{P} = 6.N.P + 4.N \quad (18)$$

where P is the number of representative points of the template.

The number of elementary operations in the proposed accelerated algorithm is:

$$NEO(NCCE)_{Accs} = N.(5p+5) + S_1[4.P + 2.p + 5] + S_2[3.(P-p)+2] \quad (19)$$

where S_1 is the number of points, which remain after the first elimination condition and S_2 is the number of points, which remain after the second elimination condition.

The relative acceleration can be computed using Eq. (20):

$$\frac{NEO(NCCE)_{Cl}}{NEO(NCCE)_{Accs}}. \quad (20)$$

Here $NEO(NCCE)_{Cl}$ is $NEO(NCCE)_{Cl}^{All}$ or $NEO(NCCE)_{Cl}^{P}$.

In Fig. 1 the acceleration (in times) is shown. The size of the image is 128 x 128 pixels; the size of the template is 64 x 64 pixels. Set of representative points consists of 200 points, and it is divided on two subsets, every one of them contains 100 points. It is accepted that 1.25% to 0.25% of all possible positions remain as "promising" after the first elimination condition.

It is seen that the acceleration in comparison with the classic algorithm over all points is significant – over 47 times. The acceleration of the proposed algorithm is not so great in comparison with classic algorithm of consecutive comparison over the set of representative points, but it is over 2.3 times.

B. Evaluation of the experimental results

Series of experiments have been carried out. Parameters of experiments are: size of the image - 128 x 128 pixels, size of the template - 64 x 64 pixels, the set of representative points contains 200 points and the admissible deviation from the right position - $\varepsilon = \pm 1$ pixel. Table 1 contains results of 200 experiments.

The received results are strongly dependent on data. That is the reason that in the table there are shown average, minimal and maximal values for any of parameters.

It can be concluded, that the elimination conditions are very strong – over 90% of all possible position drop out during the first elimination. During the second elimination new about 48% of the rest positions drop out. Thus the full correlation coefficient is computed for maximum 0.56% of all possible

positions. The typical number is about 0.27% of all possible position of the template over the image.

The acceleration of the computations is also significant in comparison with the classical algorithm over all points of the template as well as in comparison with the classical algorithm, which computes correlation coefficient over the set of representative points.

TABLE I
EXPERIMENTAL RESULTS FOR NCCE, COMPUTED OVER DIFFERENT SETS OF REPRESENTATIVE POINTS

Parameter	**Representative points** (Basic subset of 100 points)		
Part of eliminated positions	After the first elimination condition	Minimum number	98,93%
		Maximum number	99,81%
		Average number	99,44%
	After the second elimination condition (from the remaining positions)	Minimum number	0%
		Maximum number	84,85%
		Average number	48,67%
Accele-ration (Reducing the operations) in times	In comparison with the classic algorithm over all points	Minimum	48,09
		Maximum	57,66
		Average	50,65
	In comparison with the classic algorithm over set of points	Minimum	2,36
		Maximum	2,82
		Average	2,48

The reliability of the proposed algorithm (Table 2) is evaluated regarding the possibility of skipping the real position of the searched template. Series of 2500 experiments have been made. Parameters of experiments are: size of the image - 128 x 128 pixels, size of the template - 64 x 64 pixels, the set of representative points contains 200 points and the admissible deviation from the right position - $\varepsilon = \pm 1$ pixel, levels of Gausian noise 10% and 20% and different numbers of representative points.

The received results (which are shown in Table 2) allow concluding that:

- The proposed algorithm for accelerated template localization using correlation coefficient as a measure of similarity does not decrease significantly the reliability of the measure in comparison with the classical algorithms in the conditions of the examined levels of noise and numbers of representative points.
- There are not any unsuccessful localizations in any of series.
- In contrast to other known algorithms using "fast dissimilarity principle", like SSDA algorithm of Castillo [1] and PAIRS algorithm of Lundberg [4], which use comparisons between arbitrarily chosen points, the here proposed algorithm uses points which represent the images in the best way, according to criteria, which are used for their extraction.

TABLE II
EXPERIMENTAL RESULTS – PARAMETER "EXACTNESS OF LOCALIZATION"

Algorithm	**10% noise**				**20% noise**			
	Points	**Successive localizations [%]**	**$\sigma X*10^{-7}$**	**$\sigma Y*10^{-7}$**	**Points**	**Successive localizations [%]**	**$\sigma X*10^{-7}$**	**$\sigma Y*10^{-7}$**
Accele-rated	60	100	0	5,06	100	100	0	8,31
Classic		100	0	3,23		100	0	7,67
Accele-rated	80	100	0	3,92	125	100	0	2,26
Classic		100	0	2,98		100	0	1,84
Accele-rated	100	100	0	1,60	150	100	0	3,57
Classic		100	0	0,61		100	0	0,52

IV. CONCLUSION

It must be pointed out that the evaluations of the relative acceleration are not one to one connected with the evaluations of the run-time. It is due the fact that the different elementary operations have different times for their execution.

To conclude it can be generalized that the proposed algorithm shows good enough results about the examined parameters, to be used in practice for solving different task of the area of computer vision and in particular – for solving the task for template localization.

REFERENCES

[1] Castillo C., C. Chang, An Approach to Vision-Based Person Detection in Robotic Applications, *In 2nd Iberian Conference on Pattern Recognition and Image Analysis*, vol. 2, pp. 209-216, 2005

[2] D. S. Tyanev, Y. Petkova, A New Method for Important Points Extraction, *International Scientific Conference CompSysTech 2004*, Rousse, Bulgaria, 17-18 June 2004, pp. III.11-1 – 11-8.

[3] D. S. Tyanev, Y. Petkova. About the Possibilities of a New Edge Definition in Black and White Images, *ISSE 2005, 28th International Spring Seminar on Electronics Technology*, Wiener Neustadt, Austria, May 19-22, 2005.

[4] Lundberg Frans, Maximum Entropy matching: An Approach to Fast Template Matching, October, 2000

[5] Nakov P., P. Dobrikov, *Programming = ++Algorithms*, Sofia, pages 695

Processing of Sequences of Contour Images of Sign-Language Interpreter

Roumen Kountchev[1], Vladimir Todorov[2] and Roumiana Kountcheva[3]

Abstract - **In the paper is offered a new approach used for the processing of sequences of TV frames, representing the performance of a sign language interpreter. The processing comprises two steps: the extraction of the image contours and their processing with intra-frame lossless compression aimed at retaining the understandability of the presented signs. The efficiency of the lossless compression is increased when the extracted contour images are processed in groups, comprising several consecutive TV frames. The obtained high compression ratio permits the application of the method for distance learning applications in training courses for impaired students and in some real-time applications.**

Keywords – **Lossless image compression, Distance learning tools**

I. INTRODUCTION

The distance learning for hearing impaired people is a problem of high significance. The usual approach is to provide them the training courses as pdf or doc files via Internet, or to use video presentations, containing explanations of sign language interpreter (SLI). The approach based on the participation of SLI is of great importance because the trainees assess the information using their natural language. Insufficiency of the method is that such video presentations offer very bad quality of the fast moving parts such as the interpreter's hands and face, which are of great importance for the signs understandability and the transferred volumes of data are very big. In some cases instead of the color video presentations are used the contours of the SLI image [1,2,6]. For this, the contours are extracted from the video frames and after that - compressed. As it is presented in the special literature [3,4,5], the contour images are understandable and present the information in a quick and acceptable way. The main problem remains the large amount of data, which should be transferred and received [11,12]. The efficient coding of this information will permit its transfer via slow-rate channels [7,8,9,10]. The investigation, presented in this paper is aimed at the efficient compression of the contour images. For this, the consecutive TV frames are processed one by one in order to extract the contours of the SLI image and to prepare them for the compression.

[1] Roumen Kountchev is with the FCTT, Radiocomunications Dept., Technical University of Sofia, Bul. Kl. Ohridsky 8, Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg

[2,3] Vladimir Todorov and Roumiana Kountcheva are with T&K Engineering, Mladost 3, P.O.Box 12, Sofia 1712, Bulgaria
E-mails: todorov_vl@yahoo.com. kountcheva_r@yahoo.com

Usually there are too many extracted contours, which is an obstacle for the efficient compression. In order to preserve the main contours only, retaining the signs understandability, in this work is used presented special pre-processing, based on the image histogram modification.

The paper is arranged as follows: in Section II is presented the algorithm for image pre-processing, in Section III are given the experimental results, obtained for single and multiple frame compression and is investigated the influence of the number of consecutive TV frames processed together; Section IV is the Conclusion.

II. IMAGE PRE-PROCESSING

The method for image pre-processing performs adaptive image contrast enhancement, comprising two consecutive stages: brightness segmentation based on the image histogram analysis, and transformation of the pixels' brightness in accordance with tables, defined by the segments treatment.

In the first stage is performed the image segmentation. For this are used the thresholds k_1 and k_2, which divide the histogram in three segments (A, B, C). The thresholds are set so that to define the second segment (B), which contains the main part of the image objects. In order to make the number of participating brightness levels in the segment B smaller (and correspondingly - to obtain higher compression), this part should be skewed to 1-2 % (instead of 12). The histogram of the SLI image (Fig.1.) is shown in Fig.2 together with the histogram segmentation.

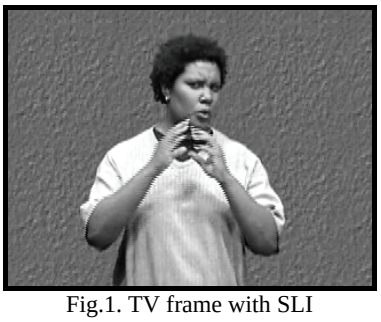


Fig.1. TV frame with SLI

The limits of the central segment (B) are defined performing the following operations:

- The image histogram h(k) is calculated and is defined its maximum:

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

- The value $t = \alpha\, h_{max}$ is defined, for $\alpha < 1$ (for example $\alpha = 0.8$): this value defines the magnitude of the middle segment of the histogram (B).

- The values of the start and end points of the segment B, defined as k_1 and k_2, are calculated in accordance with the relations:

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

$h(k) \ge t$ for $k = k_2+1, k_2+2, .., k_{max}$, for $|k_2 - k_1| \ge \Delta$.

(Δ – a value, set in advance).

In case, that the last condition is not satisfied, the value of the parameter α is decreased, for example to α=0.75, and using the new threshold value $t=\alpha.h_{max}$ are defined the corresponding values for k_1 and k_2. When the requirement $|k_2-k_1| \ge \Delta$ is satisfied, the calculation cycle ends; if not – the process continues. An example is presented in Fig.2.

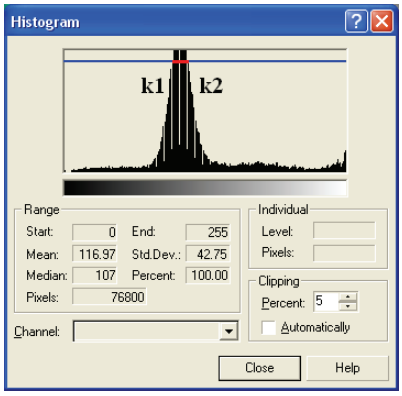


Fig.2.The histogram of the SLI image: the three segments of the image histogram, defined by the points k_1 and k_2

In the second stage of the processing, the brightness level k of the pixels in the three segments of the histogram is transformed in accordance with the relations:

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

The relations $g_A(k)$, $g_B(k)$ and $g_C(k)$ represent the brightness transformation for the pixels in the segments A, B and C correspondingly. In order to perform the required contrast modification, the boundaries k_1, k_2 of the segment (B) are skewed to $(k_1+\delta_1)$ and $(k_2-\delta_2)$, and correspondingly are moved the upper limit value of the segment A and the lower limit value of the segment C. In the presented example δ_1 and δ_2 are parameters, which define the contrast modification of the objects in the segment B and as a consequence – of the segments A and C as well. The brightness transformation tables are defined in accordance with the requirement for histogram equalization of the corresponding segment (A, B or C) with changed (stretched or skewed) brightness 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, for images in which the histogram of the corresponding segment is uniform, i.e. for:

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

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

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

the relations for the brightness transformation are linear and are defined as follows:

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

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

$$g_C(k) = \left(\frac{k_{max} - k_2 + \delta_2}{k_{max} - k_2}\right)(k - k_2 - 1) + (k_2 - \delta_2 + 1).$$

In this case the brightness levels in the range (k_1, k_2) are skewed and correspondingly - the brightness levels in $(0, k_1 + \delta_1)$ and $(k_2 - \delta_2, k_{max})$ are stretched.

III. Experimental Results

A. Single frame compression

In result of the performed histogram modification the number of the participating brightness levels is decreased, which results in higher compression ratio. The histogram modification limits the number of the retained brightness levels to less than 5 in the meaning part of the image. In the remaining part (the image background) the number of the brightness levels is very low too. In result of the performed histogram modification the compression ratio is significantly increased. For the investigation was used more than 120 test images extracted from different video sequences. The results obtained for contour images extracted from different sequences are very close and consistent. Some of the results are presented in Table 1 below. The mean value for the compression ratio after the processing is more than 15 times higher than that, obtained for the original image.

Image	C1	C2	C3	C4
Original	2,54	2,58	2,29	2,49
Treated	36,56	36,78	36,79	37,01

Table 1. Compression ratio obtained with the new method for image processing and lossless compression.

B. Multi-frame compression

In order to obtain higher compression ratio is used multi-frame lossless compression of the contour SLI images extracted from consecutive TV frames. This approach aims to use the high correlation between the consecutive TV frames. The investigation was performed, arranging the processed contour images in one larger image with different size: 2, 3, … up to 12 consecutive frames, arranged horizontally, vertically or as a rectangle. The obtained results are presented below.

Arrangement in horizontal direction

The multi-frame lossless compression is performed arranging the images together and processing the thus obtained compound image as a single one. Here L1h (with size 320 x 240 pixels) is the contour image obtained after contour extraction and histogram modification of the first TV frame in the processed sequence; L2h comprises the contour images obtained from two consecutive TV frames and is with size 640 x 240 pixels; L3h comprises the contour images obtained from three consecutive TV frames and is with size 960 x 240 pixels, etc. The last test image, L9h comprises the contour images obtained from nine consecutive TV frames. Example test images are shown in Fig. 3.

Fig.3. a. Image L1h.

Fig.3. b. Image L3h.

Image	L1h	L2h	L3h	L5h	L6h	L8h	L9h
TKView	40,02	40,39	40,53	40,54	40,62	40,65	40,70
JPEGmax	3,34	3,35	3,36	3,38	3,39	3,39	3,39
JPEG2000	5,82	5,96	5,94	5,97	5,98	5,98	5,99

TABLE 2. Compression ratios for inter-frame compression with horizontal arrangement

The obtained results are presented in the Table 2 and in Fig.4. The last image, L9h comprises the contour images extracted from 9 consecutive TV frames. The comparison was performed with software products, based on the standards JPEG and JPEG2000, and the special lossless compression was performed with TKView. The results, obtained with TKView are much better than the other two. The results for JPEG are not for lossless compression, but for lossy compression with highest possible quality (Microsift Photo Editor, quality factor 100).

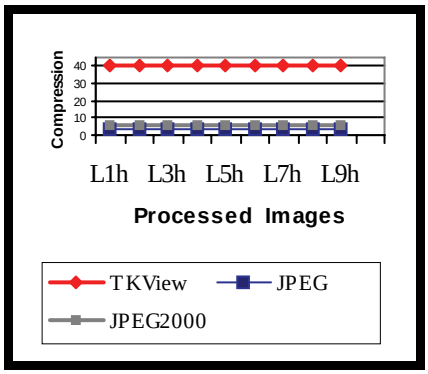


Fig.4. Compression ratio, obtained with horizontal arrangement of the tested images.

Arrangement in vertical direction

Similar investigation was performed with vertical arrangement of the tested images and processing them together. The image L1v is the contour image obtained from the first TV frame in the sequence, with size 320x240 pixels; L2v comprises the contour images obtained from two consecutive frames and is with size 320 x 480 pixels; L3v comprises the contour images obtained from three consecutive TV frames and its size is correspondingly 320 x 720 pixels, etc. The last image, L9v comprises 9 contour images obtained from consecutive TV frames. Some of the obtained results are given in the Table 3.

Image	L1v	L2v	L3v	L5v	L6v	L8v	L9v
TKView	40,02	40,16	40,27	40,24	40,33	40,35	40,39
JPEGmax	3,34	3,37	3,38	3,38	3,39	3,39	3,40
JPEG2000	5,82	5,90	5,91	5,92	5,93	5,94	5,94

TABLE 3. Compression ratios for inter-frame compression with vertical arrangement

The mean compression ratio for TKView is 40,25; for JPEG (lossy compression with highest quality) it is 3,38 and for JPEG2000 – 5,91. The results, obtained for a big number of test images are very consistent. The results show that the compression ratio is increased together with the number of comprising images from 1 to 9; after that the change is negligible. In Fig. 5 the obtained results are presented graphically.

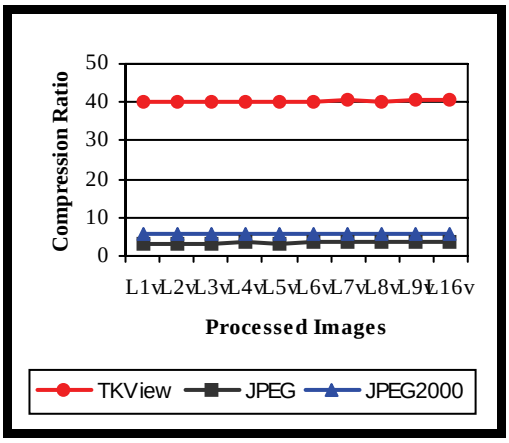


Fig. 5. Graphic presentation of the compression ratio obtained with vertical arrangement

2D Arrangement

The 2D arrangement of the tested images was performed to investigate the influence of the correlation in both directions (vertical and horizontal) on the compression ratio. The processed images were arranged as follows: the image L2x3 comprises 6 consecutive contour images arranged in two rows of 3 images each; the image L2x4 comprises 8 consecutive contour images arranged in two rows of 4 images each, etc. The test image L4x4 is with size 1280 x 960 pixels. The results obtained for the compression ratio are given in Table 4.

Image	L2x3	L2x4	L2x5	L3x4	L4x3	L3x3	L4x4
TKView	40,90	40,98	40,99	41,02	41,06	41,12	41,14
JPEGmax	3,39	3,36	3,30	3,32	3,31	3,32	3,33
JPEG2000	5,92	5,92	5,93	5,93	5,94	5,94	5,94

TABLE 4. Compression ratios obtained with 2D arrangement

The mean value for the compression ratio obtained with 2D arrangement for many test images is as follows: for TKView it is 41.01; for JPEG (lossy compression with highest quality) it is 3.33 and for JPEG2000 (lossless) – 5.93.

The results, obtained with the presented three kinds of arrangement show that highest compression is obtained for 2D arrangement of the tested images. The optimal number of the images processed together is 12. The comparison results with JPEG and JPEG2000 are shown in Fig. 6.

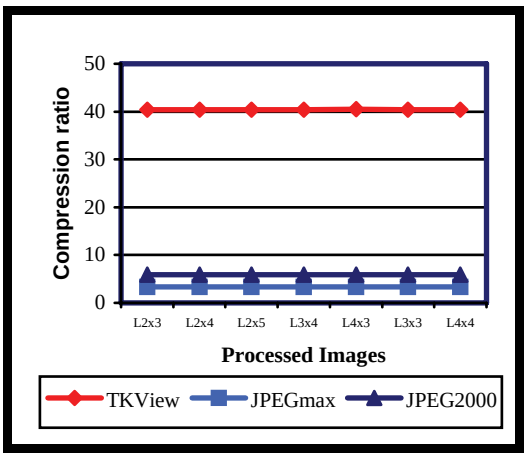


Fig. 6. Graphic presentation of the compression ratio obtained with 2D arrangement

Arrangement	Mean compression Ratio
Horizontal	40.65
Vertical	40.85
2D	41.01

TABLE 5. Mean values of the obtained compression ratio

In Table 5 are presented the mean values for the compression ratio obtained with different arrangement for same images.

IV. CONCLUSION

The presented method for image histogram modification influences the compression ratio of contour images to a high degree. Very important quality of the method is that the understandability of the presented signs is retained because the TV frames are processed individually and after that - compressed. The method permits the compressed data to be transferred as additional information via low-rate channels. The compression results are much better than those, obtained with traditional methods based on the JPEG and JPEG2000 standards. The specific features of the method are:

- The efficiency of the method depends on the retained number of brightness levels in the extracted contour images;
- The consistency of the obtained results permits the parameters, used for the histogram modification to be selected in advance and to perform the pre-processing automatically, which is very important for real-time applications.
- The consecutive TV frames could be arranged after decompression and used as usual video clip.
- In result of the high compression ratio one TV frame, containing contour SLI image is compressed to less than 1900 KB, i.e. the additional information, needed for the video SLI presentation is less than 380 Kbps. Compared with the regular data flow in MPEG-2, which is in the range 6 - 12 Mbps, this amount of data is negligible.
- Additional possibility is to use TV frame with smaller size, for example 256 x 192 pixels. In this case the mean compression ratio is 38 and the additional data flow is about 260 Kbps.
- The method for lossless compression gives very good results when applied for text images and this is an additional tool for distance learning applications [13].

The presented approach is suitable for various applications aimed at training of hearing impaired students.

V ACKNOWLEDGEMENT

This paper was supported by NSF of Bulgarian Ministry of Education and Science (Contract VU-MI-104).

REFERENCES

[1] Subject-Based Deaf and Hard of Hearing Internet Resources. www.Internet Resouces - Deafness.htm.

[2] Study Skills and Learning Strategies: www.Teaching the learning disabled.htm.

[3] D. Pearson, J. Robinson. Visual Communications at Very Low Data Rates. IEEE Proc., Vol. 73, 1985, pp. 222-241.

[4] P. Letellier, M. Nadler, J. F. Abramatic. The Telesign Project. Proc. of the IEEE, Vol. 73, 1985, pp. 242-258.

[5] CCITT Recommendation H.261: Video Codec for Audio- visual Services at px64 kbit/s.

[6] M. Ghanbari. Standard Codecs: Image compression to Advanced Video Coding. IEE, 2003.

[7] M. Whybray, E. Hanna. A DSP based videophone for the hearing impaired using valledge processed pictures. Proc. ICASSP'89, Vol 3, 1989, pp.1866-1869.

[8] M. Whybray. Moving picture transmission at low bitrates for sign language communication. http://www.zak.co.il/deaf-info/old/videophones.html.

[9] G. Sperling, M. Landy, Y. Cohen, M. Pavel. Intelligible encoding of ASL image sequences at extremely low information rates. Computer Vision, Graphics, and Image Processing 31, 1985, pp. 335-391.

[10] S. Tetzchner (Ed.). COST 219 book: Issues in Telecommunications and Disability. Commission of the European Communities, 1991.

[11] A. Heron, N. Mc Donald. Video transmission over a radio link using H.261 and DECT. 4-th International Conference on Image Processing and Applications, IEE, April 1992.

[12] D. Salomon. Data compression. Springer-Verlag, New York, Inc. 2004.

[13] R. Kountchev, Vl. Todorov, M. Milanova, R. Kountcheva. *Documents Image Compression with IDP and Adaptive RLE*. The 32nd Annual Conference of the IEEE Industrial Electronics Society Conservatoire National des Arts & Métiers – Paris - FRANCE, IECON'06, November 7 – 10, 2006, pp. 3261-326.

Ordered Complex Hadamard Transform of Images

Rumen P. Mironov[1], Roumen K. Kountchev[2] and Vladimir S. Mirchev[3]

***Abstract* – A method for new ordered Complex Hadamard Transform of halftone images is presented. The "basis" ordered images for 2D CHT are calculated and distribution of spectral coefficients is obtained and evaluation of coefficients distribution in complex spectrum space for test images is made.**

***Keywords* - digital signal processing, orthogonal transforms, complex Hadamard transform, image processing.**

I. INTRODUCTION

The discrete Walsh Hadamard Transform (WHT) is a fairly simple orthogonal transform [1] and has found applications in data compression, spectral analysis, pattern recognition and watermarking involving image transmission, storage and security. The idea of using complex, rather than integer transforms matrices for spectral processing and analysis has been shown in [2], [3] and [4]. From the Complex Hadamard Transform (CHT), several complex decisions diagrams are derived.

In this paper arranging of spectral coefficients of general Complex Hadamard matrix are investigated, which provide better concentration of energy in low-frequency area of 2D image spectrum. The 'basis' ordered images for 2D CHT are obtained by using a developed MATLAB program simulation. The similar results with well-known Walsh Hadamard Transform are received, what shows that CHT can be used by the same way in more complicated analysis and processing [5], [6], [7].

II. MATHEMATICAL DESCRIPTION

The coefficients of Complex Hadamard Transform matrix $[CH_N]$ with dimension N by N can be represented by the following equations [4]:

$$\begin{vmatrix} c(u,v) = j^{uv} s(u,v) \\ c^*(u,v) = (-j)^{uv} s(u,v) \end{vmatrix}, \quad (1)$$

where: $N = 2^n$, $j = \sqrt{-1}$, u,v= 0,1,…2^n-1 and

$$s(u,v) = \begin{cases} 1 & \text{for } n = 2 \\ (-1)^{\sum_{r=3}^{n} \lfloor u/2^{r-1} \rfloor \lfloor v/2^{r-1} \rfloor} & \text{for } n = 3,4,5..... \end{cases} \quad (2)$$

is the sign function. Here $\lfloor . \rfloor$ is an operator, which represents the integer part of the result, obtained after the division.

The complex Hadamard matrix can be presented in logarithmic form [6], [7]:

$$[CH_N] = e^{j[\Phi_N]}, \quad (3)$$

where $[\Phi_N]$ is a phase matrix of CHT.

From the equation (3) the CHT basis matrix of order 2^n can be calculated for n=2 and u,v=1,2,3,4 by the following way:

$$[CH_4] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} \begin{matrix} \to 0 \\ \to 1 \\ \to 3 \\ \to 2 \end{matrix}, [\Phi_4] = \frac{\pi}{2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 0 & 2 \\ 0 & 3 & 2 & 1 \end{bmatrix} \quad (4)$$

Each number in the right column toward the CHT matrix $[CH_4]$ is obtained as a sum of sign changes between the elements of each row. As a sample for 3rd row of the matrix is implemented:

$$\sum_{i=1}^{3} sign_{(i)} = 1+1+1 = 3$$

The ordered complex Hadamard matrix can be received as rearrange of row numbers in increasing order. In result the ordered CHT matrix of order 4 and the corresponding phase matrices are:

$$[CH_4'] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{matrix} \to 0 \\ \to 1 \\ \to 2 \\ \to 3 \end{matrix}, [\Phi_4'] = \frac{\pi}{2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 3 & 2 & 1 \\ 0 & 2 & 0 & 2 \end{bmatrix}. \quad (5)$$

This approach for rearranging of CHT matrix can be summarized for order N>4.

As a sample, the basis ordered Complex Hadamard matrix of order 8 is:

$$[CH_8'] = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & j & -1 & -j & -1 & -j & 1 & j \\ 1 & j & -1 & -j & 1 & j & -1 & -j \\ 1 & -j & -1 & j & 1 & -j & -1 & j \\ 1 & -j & -1 & j & -1 & j & 1 & -j \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix} \begin{matrix} \to 0 \\ \to 1 \\ \to 2 \\ \to 3 \\ \to 4 \\ \to 5 \\ \to 6 \\ \to 7 \end{matrix}.$$

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

[3]Vladimir S. Mirchev is with the Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky, 8, Sofia 1000, Bulgaria.

In conclusion, the forward and inverse ordered CHT for two-dimensional signals (images) can be generalized in matrix form as follows:

$$\begin{cases} [Y] = [CH'_N][X][CH'_N] \\ [X] = \frac{1}{N^2}[CH'_N][Y][CH'_N] \end{cases}, \quad (6)$$

where the input image is represented by the matrix $[X]$ with the size NxN and the result is a spatial spectrum matrix $[Y]$ with the same size.

III. Experimental Results

The 2D CHT can be expressed in different way by the equation [4],[5]:

$$[X] = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} y_{kl}[T_{kl}] \quad (7)$$

where $[T_{kl}]$ is the matrix of "basis" image with consecutive number (k,l). This expression can be presented as image decomposition $[X]$ in order on N^2 "basis" ordered images with weighted coefficients y_{kl}. Using equations (5) and (7) these images of order 4 are simulated on MATLAB and are shown in Fig.1.

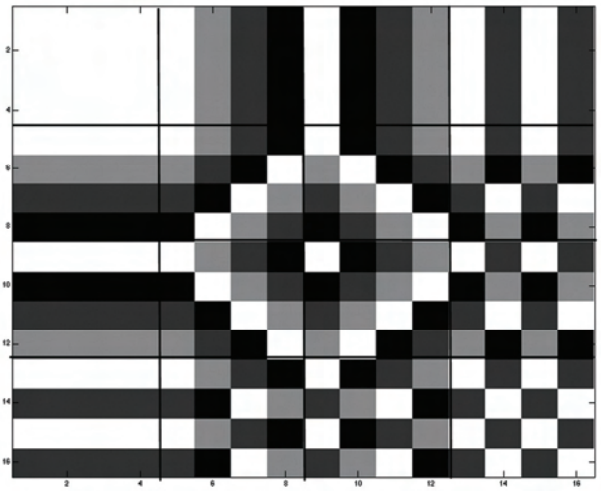


Fig.1. "Basis" function for 2D CHT with ordered matrix of size N=4.

In this figure the values (+1) are colored with white level, the values (-1) – with dark gray level, the values (+j) – with white gray level and the values (–j) – with black level. It is show that the received "basis" images look like the ordered "basis" images of real Hadamard Transform (HT) [5].

On Fig.2 and Fig.3 are shown the "basis" ordered images for 2D CHT of size 8x8 and 16x16 respectively, received in the same way.

For the analyses of spectral distribution between the coefficients of 2D ordered CHT test image "PEPPERS", with size 256x256, 256 gray levels and 64x64, 256 gray levels are used.

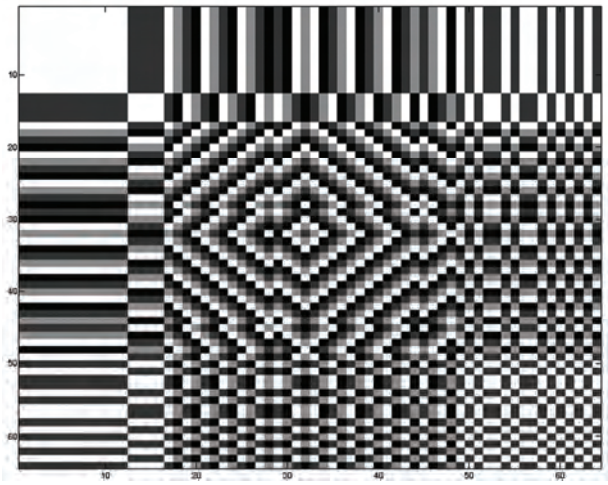


Fig.2. "Basis" function for 2D CHT with ordered matrix of size N=8.

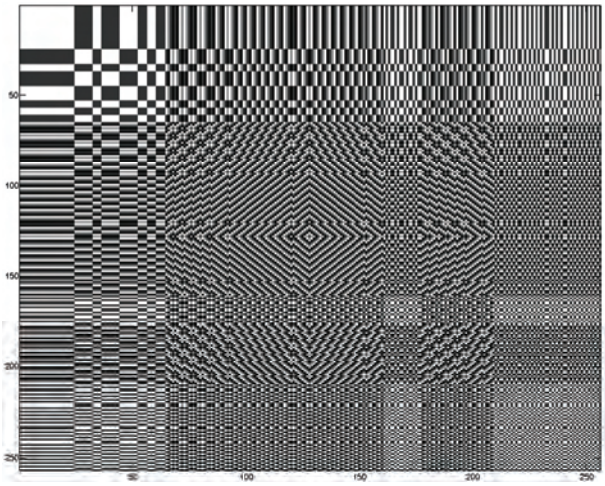


Fig.3. "Basis" function for 2D CHT with ordered matrix of size N=16

On Fig.4a and Fig. 4b 2D amplitude frequency spectrums for each one are present.

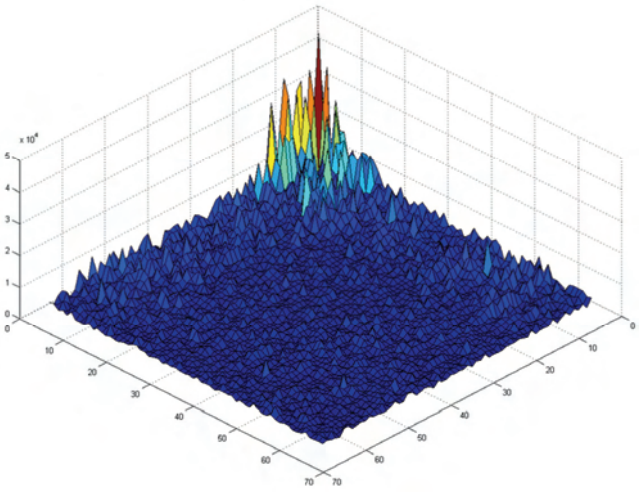


Fig.4a. 2D CHT spectrum for image of size 64x64.

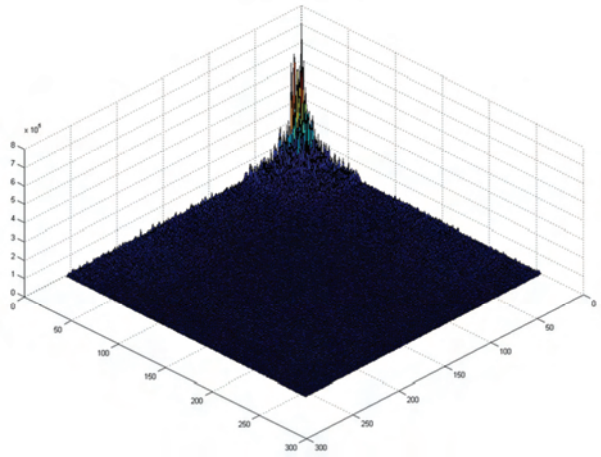


Fig.4b. 2D CHT spectrum for image of size 256x256.

From this figures the high concentration of image energy in the low-frequency coefficients of the spectrum can be seen.

On the Fig.5 the averaged PSNR results from experimental investigation of coding of amplitude and phase spectrums of four test images “LENNA”, “BABOON”, “CAMERAMAN” and “PAPPERS” (Fig.6), received by the 2D ordered CHT and 2D real HT with size 16x16 are shown.

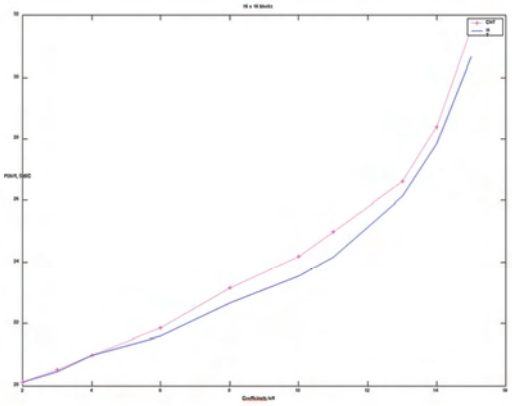


Fig. 5. Dependence of PSNR (dB) from saved coefficients of ordered CHT and HT

The improvement of PSNR for ordered CHT with respect to HT for 16 saved spectral coefficients from each block, with size 16x16, is about 2.5 dB. Therefore, the ordered CHT have some advantages, concerning the real ordered HT by the compression of images, trough the filtration of image transformation.

IV. Conclusion

A class of new ordered Complex Hadamard Transform for images is presented. The general principles of complex matrices construction of high order for 1D and 2D transforms are given. The basic properties of ordered CHT are discussed. The obtained amplitude spectrums for ordered CHT and ordered HT are practically identical and show that both can be used in similar applications.

The presented ordered Complex Hadamard Transform can be used in digital signal processing for spectral analysis, pattern recognition, digital watermarking, coding and transmission of one-dimensional and two-dimensional signals.

Acknowledgement

The authors thank the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science for the financial support by the contract VU-MI -104/2006

References

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

[2] 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.

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

[4] R. Mironov, R. Kountchev. “Analysis of Complex Hadamard Transform Properties”, XLI International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST’2006, Sofia, Bulgaria, July 2006, pp.173-176.

[5] W. K. Pratt. Digital Image Processing, John Wiley&Sons, N.Y., 2001.

[6] J. Wallis. *Complex Hadamard Matrices*. Lin. Mult. Algebra 1, 257, 1973.

[7] W. Tadej, K. Zyczkowski. *A concise Guide to Complex Hadamard Matrices*. arXiv:quant-ph/0512154 v2, 25 May 2006.

Fig.6. Test images “LENNA”, “BABOON”, “CAMERAMAN” and “PAPPERS”
with size 512x512 pixels and 256 gray levels.

Image Sculpture – From 2D Images to 3D Objects

Miglena Dontschewa[1] and Tobias Sturn[2]

Abstract - We introduce a new application for creating 3d objects from 2d images. First we give a motivation for such a program. Then we quickly talk about already known image based modeling programs and their drawbacks. Based on this information we describe our approach and our application. The paper ends with a conclusion and some figures of 3d meshes created with the program.

Keywords **– 3D – Computeranimation, 2D Images, 3D Objects, Programming**

I. INTRODUCTION AND MOTIVATION

Virtual reality is today and will be in the future even more important [5]. With the development of better hardware it is more and more possible to create realistic worlds. But these scenes still look very sterile maybe because the reality consists of many different objects, whereas today's computer scenes contain only a small amount of objects. One reason for this lack of many objects might be that the modeling of objects is a very time consuming and therefore costly process. Traditional modeling with defining ten thousands of polygons, texture coordinates etc might be a frustrating, monotonous work and also a lot of knowledge and creativity is required to create a good quality 3d mesh. The use of hardware like 3d scanners is very costly and they also have their limitations like object size for example. Therefore a flexible program, with which it is possible to quickly and easily create from 2d images 3d objects would be very useful for nearly everyone possesses a digital camera today. Although there are already some applications which create out of 2d images 3d objects they all have their limitations by trying to make the mesh creation automatically. We instead think that the human is unbeatable in the interpretation of 2d images and that therefore the user should do more work by talking as efficient and intuitive with the program as possible not relying on an automatic technique.

II. IMAGE BASED MODELLING PROGRAMS

Programs like iModeller by UZR or ImageModeller by Realviz use automatic techniques to create from 2d images 3d objects. For example with iModeller the objects have to be placed onto a calibration underground so that the program can calculate out of this extra information the camera positions.

Author's are with UCT- Research, FHV, Dornbirn, Austria,
[1]don@fhv.at

Therefore the application can only work with small objects, which can be placed onto such an underground. Another problem is that the images can only be taken from top so that the underground is visible. Working with ImageModeller the user has to set marker points onto the edges of the 3d objects so that the program is able to automatically calculate the camera positions. ImageModeller works very well with buildings and other artificial objects but has its problems with organic objects. So by using automatic techniques the images have high requirements and therefore the usage of the application is limited.

III. OUR APPROACH

Instead of automatically calculating the camera positions, our program uses the fact that the human is in the interpretation of images unbeatable. If we look at only one image we can immediately imagine this object in 3d because we possess information about the object which cannot be found in the images. For computers lack this information and for it is impossible to give the program this information for all objects, objects have to be placed onto calibration undergrounds or markers have to be set onto edges to give the computer this needed information. But because of this automatic object creation there have to be limitations in usage which lead to high image requirements. Therefore to be as flexible as possible and to be able to handle all kinds of objects we have decided to let the user transform the images. For the user immediately recognizes from which position an image is taken, the key point of the application is to communicate as good as possible with the user so that he can tell the program as simple, intuitive and quickly as possible from where the images were taken.

IV. USER INTERFACE

Therefore the user interface represents the core competence of our product /DoBoKe´04/. We have tried hard to create a very intuitive, cool product with which it is as simple and as effective as possible to create out of 2d images 3d objects. In the following we quickly describe the two processes which have to be done by the user.

A. Binary Image Segmentation

First the user has to lay a mask over the object to tell the program what the object on the image is. To place the mask

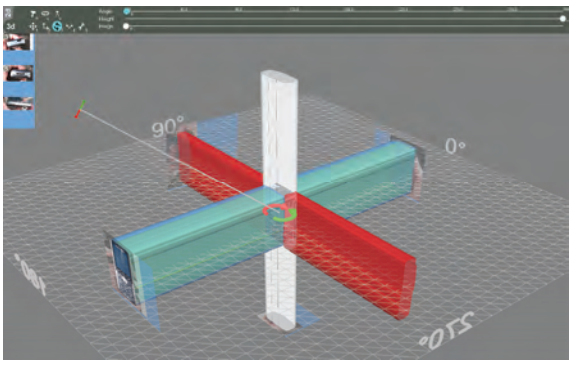
we use many standard tools like brushes, filling, drawing lines and rectangles, shrinking/growing of the mask and threshold. For each operation the user can set specific properties like pen size, fill tolerance, adding or removing the mask etc. It is possible to zoom to the image with the mouse wheel, moving the image around by pressing the right mouse button. The user does not even have to release the brush if he comes to the border which is very frustrating in other applications. The user can use key shortcuts to be even more effective.

B. Image Transformation

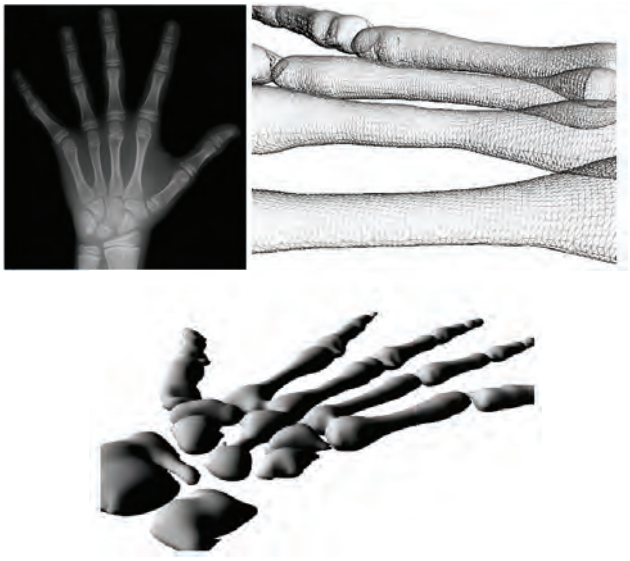
After masking the images the user has to transform the images by rotating, scaling, translating the images and specifying the field of view and how the images should participate in the sculpturing process. Thereby the communication between the program and the user should be as good as possible. ´The user should tell the program as well as possible how the images should be transformed and the program should present the information as good as possible. We have created a new way to specify these values by moving the mouse cursor around the screen center. With this way by moving with the mouse away from the center the user can move the image very accurate without having to input any numbers. An analogy would be a man showing a crane operator with his arms how he should move the crane. This is very intuitive and analogous. For the representation of the images, 3d clippings of each image get created. These can be presented with transparency or in wireframe. By this way the transformation of the images can be seen very well. The user can always sculpture the final mesh to see how the mesh already looks like. 10 seconds are now needed for sculpturing on an Intel Core 2, 2 GHz, 1024 GB Ram but it will be optimized in the future because seeing the final 3d mesh is naturally the best visual help.

V. CONCLUSION

We have presented an application which creates out of 2d images 3d objects. Thereby we don't use an automated approach but let the user transform the images to be as flexible as possible. Therefore the core feature of our product is the good communication between the human and the computer. Bellow you can see the application and how the images get transformed and some 3d objects which were created with the program. Note that the objects are not taken on an underground but are hold in the hand so the image requirements are very low.



The application: Transforming the images



Bones: Only one image as source. Program creates a rotation symmetric mesh.

Handy: Created in ~5 minutes.

Marlboro: Created in ~5 minutes.

REFERENCES

[1] Dimitrov, D., Dontschewa, M. (1995) Computer modelling from magnetic field“, 4.Intern.Konferen „Computer Aided Engineering Education, 13.-15.09.`95, Krakow, Polen

[2] Dimitrov, D., Dontschewa, M., (1999) Medical Image Regocnition, CMK, 05.05.-08.05.`99, Sofia, Bulgaria

[3] Dimitrov, D, Dontschewa, M., (1996) Computermodelieren von 3D-nieder-frequenten Signale, generiert vom NEMEK – Apparat, 31.Wiss. Konferenz „Kommunikative elektronische Computersysteme `96“, 16.05.`96, Sofia, Bulgaria

[4] Dimitrov, D., Dontschewa, M., Nikolova, M. (1997) Computer simulation of 3D-Signals of Nemeks Apparat in Physiotherapy, 7. Multimedia – Fachtagung, 29.-01.10.`97, Dortmund, Deutschland

[5] Dontschewa, M., Bolter, R., Kempter, G., Weidmann, K-H. & Künz, A. (2004) Developments in the area of the virtual reality at the User Centered Technologies Research Institute, 15.-19.09.`04, Ohrid, Mazedonien

[6] Georgieva, V., Dimitrov,D., Neykova, M., (2005) Pulse Modulation Analysis in MATLAB Environment, Proceedings of International Conference “Computer Simulation in Information and Communication Engineering, CSICE’05, October,20-22, `05 , TU-Sofia, Bulgaria, p.79-81

[7] http://pille.iwr.uni-heidelberg.de/~photo01/index.html

[8] http://www.imodeller.com/de/

[9] http://imagemodeler.realviz.com/

[10] von Hellberg, P. (2003) Redesign the Data Dump - Statistical Vector Field, In C. Stephanidis (ed). Human-centred computing: cognitive, social and ergonomic aspects, Vol. 3, pp. 1358-1362, Mahaw: Lawrence Erlbaum.

[11] Raskin, J. (2000), The Human Interface, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA

This page intentionally left blank.

Analysis of the Realism of 3D Computer Animations in the Context of The Human Visual Perception

Miglena Dontschewa[1]

Abstract - A first study is presented in which parameters want to be found and verified which correspond in a virtual scene to the perception of the human so that they feel realistic. It is based on the physiological-psychological knowledge about the visual perception of the human (Helmholz 1866, Kempter, G. & Bente, G., 2004) and from the knowledge about the possibilities of VR Systems (Dontschewa, Kempter et al. 2005) and already known parameters, which influence the realistic impression of a virtual scene (Flemming 1999). The analysis serves as a basis to find parameters which are important for the realistic impact and to empirically prove them. Primarily the paper wants to examine, what the perception of a viewer influences. For this purpose a scene selection, the reference to the human visual perception and its reflection to the possibility of VR Systems are analyzed and adequately considered and included.

Keywords **– 3D - Computeranimation, Perception, Virtual Reality**

I. INTRODUCTION

Normally, if you want to show someone something which does not exist, where the look and the impression is most important, you may model the product with cardboards (Bartenbach – Lichtlabor, Innsbruck) or you use other representation techniques like diagrams, drawings or verbal descriptions. But with such representation techniques you will soon get to a border because it is very difficult to maintain a good impression. For example the modeling with clipboards has big disadvantages if the main objective lies in illuminating the object. You cannot work photogametrically and geometrically correct because for a geometrically scaled object you cannot use the real light sources with their light properties but will have to use smaller lights for example. Today virtual prototypes replace the traditional physical prototypes because they have many advantages. It is not only possible to look at a virtual prototype but it is also possible to interact with it, manipulate it. Also flexible light and material simulations can be done with virtual prototypes which are not possible with physical representations (Dontschewa, Künz et al. 2007).

Also a big advantage is the financial benefit in the development process. Virtual prototypes can also be used in the selling process to give the customer a much better view of the product.

1 UCT- Research Centre, FHV, Dornbirn, Austria, don@fhv.at

II. SELECTION OF A SCENE

Axiomatically, the content of an image plays a big role if the image looks real or not. Therefore is the selection of the scene very important. We describe our thoughts bellow.

- Artificial and non artificial objects

Generally one might distinguish between artificial and non artificial objects. Artificial objects are objects made by human. These can further be classified into simple objects (used in daily life) and complex scenarios (architecture, scenes, complex products etc.). Non artificial objects have a natural root like trees, plants, complex landscapes, animals or humans. Both those object groups get modeled with computer animation systems (CAS) very differently. The objects created by human mostly base on geometrical basic shapes and can therefore be modeled much faster and more accurate. Artificial objects have the advantage that they most often have today a digital source and therefore they look in a 3d scene much more realistic as organic shapes.

- Existing scenes or not existing scenes

An important aspect in the selection of a scene is the necessity of the accessibility. This necessity comes from the try not to compare 2d images (like photos) but visual impressions which come from looking at a real or virtual scene. Therefore we searched for real scenarios to be able to compare the visual impressions. We selected the foyer of a building of the Fachhochschule Vorarlberg, because it is a public location and therefore we have test persons. Also it is possible to take photos of the scene every time or to measure certain objects or to create drafts for the development process. The selected scene contains objects (chairs, tables, bar stool, bar etc) which rarely have potential to look unreliable.

III. REFERENCE TO VISUAL HUMAN CONCEPTIONS AND IT'S REFLECTION TO THE POSSIBILITIES WITH VR SYSTEMS

Some physiologic factors have to be followed that a virtual scene seems realistic. On the one hand the presentation has to be in stereoscopic form to get a depth

perception (Collewijn, H. & Erkelens, C. J. 1990). On the other hand we have to use a "moving camera" because we do not only see single images. Although we only work with one image we respect the fact that the human can move his head. To imitate this "moving camera" an experiment design with sensors for the "tracking" of the head movement was developed. In this study the VR labor (http://www.fhv.at/uct) of the FH Vorarlberg was used.

- Stereo vision
- Field of vision
- Head movements

IV. Tested Reality Parameters

Not much studies exist about „Realism" in the 3d area and even less under the aspect of the human visual conception, therefore we use the ten principles of photorealism in 3d (Fleming, B. 1999) in this study as basis.

➢ *Lightfields*

At the beginning of this analysis we expected that the lighting in the scene plays a big role in the concentration of the viewer of the scene. This assumption was not approved by the eye tracking results.

➢ *Order, Disorder*

A big problem in computer generated images is that the scenes and objects often look very sterile. Each scene, artificial or non artificial, consists of at least a bit of disorder and chaos, figure 1.

Fig. 1: Two figures of the analysis. The aspect of disorder between a ordered and a chaotic scene gets analyzed.

➢ *Detailed information*

The way to tenability of a 3d scene goes over the fluency of the objects in a scene. The more detailed the elements are the more realistic a scene seems.

Fig. 2: All objects have been modeled as detailed as possible to feel as realistic as possible.

Fig. 3: The objects are placed as disordered as possible

V. Verifying Parameter Selection Through the Eye Tracking Analysis

For the analysis an animated illustration of 32 images was created. It consists of synthetic images of the bar and of the foyer, their real archetypes and more synthetic environments. In the self generated images some parameters like order, the detail of the objects and light properties get changed within the same images.

In the eye tracking analysis a camera is set onto the pupils of the eyes to find out the location to which the test person looks at on the 2d monitor. First a calibration is done by telling the camera if the person looks to the left, right, up, down and so on. Then the images are presented onto the screen.

Fig. 4: Eye-Tracking results

VI. DISCUSSION

In this paper a first study on the basis of really existing scenarios was done to find the parameters and criteria for a realistically seeming scene. VR technologies were used to offer stereo viewing, handling the view problematic and the head movement. Now the first analysis was done. More 3d studies have to be done, the scene has to be improved and it is necessary to use comparing studies in real size and with HMD.

ACKNOWLEDGMENTS

I thank all the supporters and cooperators. Prof. (FH) Dr. habil. Guido Kempter and Dipl. Psychologist Pascal Roux for the illuminating inputs. Prof. Dr. habil. Frey for the motivating conversations. Mag (FH) Andreas Künz and Mag. (FH) Philipp von Hellberg for their active assistance with the experiments and Mag. (FH) Philipp Arnold, who created the scene for his diploma thesis. Also all figures in this paper are created by him.

REFERENCES

[1] Arnold, Ph. (2007) Synthetische Visualisierung im Kontext der visuellen Wahrnehmung, Diplomarbeit an der FHV/UCT-Forschungszentrum, Dornbirn, Österreich

[2] Collewijn, H. & Erkelens, C. J. (1990) Eye movements and their role in visual and cognitive processes, chap. Binocular eye movements and the perception of depth. Amsterdam, NL: Elsevier.

[3] Dontschewa, M., Bolter, R., Kempter, G., Weidmann, K-H., Künz, A. (2004) Developments in the area of the virtual reality at the User Centered Technologies Research Institute, 15.-19.09.2004, Ohrid, Mazedonien

[4] Dontschewa, M., Kempter, G., Künz, A. Marinov, M., Weidmann, K.-H. (2004) Using Sensors from a Motion Capturing System for Morphological Movement Examination and Visualization Control. Proc. 13th International Scientific and Applied Science Conference - ELECTRONICS 2004, 22. - 24.09.2004, pp. 88 - 93., Sozopol, BG

[5] Dontschewa, M., Kempter, G., Weidmann, K-H., Marinov, M. (2005) Immersion research in the area of the virtual reality at the User Centered Technologies ResearchInstitute, HCI2005, 12.17.07.`05,LasVegas,USA

[6] Dontschewa, M., Künz, A. & Weber, H. (2007) LowCost-Interaktivität für 3D-Computeranimation mit ComputerSpiel-Engines, uDayV2007, Dornbirn

[7] Fleming, B. (1999) Advanced 3D Photorealism Techniques. New York; John Wiley & Sons Inc

[8] Helmholtz, H. (1866) Handbuch der physiologischen Optik. 2. Auflage. Hamburg; Leipzig: Voss Verlag

[9] Kempter, G. & Bente, G. (2004). Psychophysiologische Wirkungsforschung: Grundlagen und Anwendungen. In Mangold, P., Vorderer, P. & Bente, G. (Hrsg.), LehrbuchderMedien psychologie, 271-295. Göttingen: Hogrefe

This page intentionally left blank.

Analysis and Strategies for Fighting Against Attacks on Watermark Systems

Jovan Spasik[1], Sofija Bogdanova[2] and Dimitar Taskovski[3]

Abstract - **In this paper we analyze the basic watermark characteristics and the strategies for fighting against attacks on watermark systems. This analysis shows that security of the watermark system depends on the watermark application and from information about watermarking system to which attacker can access.**

Keywords – **Analysis and strategies, watermarking system, watermarking applications, watermark characteristics.**

I. INTRODUCTION

Watermarking is a technology that helps the users to protect their intellectual and property rights via embedding invisible information directly into the carrier data. This information, which uniquely identifies true owners, is called watermark. Unfortunately, along with development of process of watermarking, there is an expansion of forgery and illegitimate use of digital data. The attackers represent the biggest problem because they are aiming to destroy the embedded watermark or to make it unreadable for watermark detector. That is the reason why there is a need for some kind of protection of intellectual property materials.

There are many papers in literature of signal processing which discuss about fighting against watermark attacks. The main problem is that these papers offer solutions for fighting only against specific attack or group of attacks. These solutions are not eligible against other attacks, so embedded watermarks remain unprotected.

Until now, nobody presented a complete strategy, which will be adequate for all possible attacks. These facts were incentive for writing this paper in which analyzes of known watermarking systems and watermark characteristic are made. This paper aim to show that robustness of watermarks depends from used watermarking application and its security. For fighting against different attacks according to different watermarking systems, different strategies should be recommended. Also, the paper should signify the negative characteristics and to note what the embedder must do in order to make more robust watermark.

The paper is organized in following way. Basic watermarking applications, watermark characteristic and comparative analysis are presented in the next chapter. In chapter 3 is presented classification of watermarking systems together with strategies for fighting against watermark attacks. This strategies are organized according to information to which attacker can access.

[1] Jovan Spasik, [2] Sofija Bogdanova and [3] Dimitar Taskovski are with the Faculty of Electrical Engineering and information Technologies, University "Ss. Cyril and Methodius", Skopje, R.Macedonia

II. WATERMARKING APPLICATIONS AND WATERMARK CHARACTERISTIC

One of the oldest applications of watermarking, or more precisely data hiding, is "secret communication". Today the watermarking can be used in variety of applications like "broadcast monitoring", "owner identification", "proof of ownership", "authentication", "fingerprinting" and "copy control"[2]. Common watermark characteristics like robustness, tamper resistance and computational cost are application dependent. In practice, it is probably impossible to design a watermarking system that satisfies all of these, so it is necessary to make tradeoffs between them that must be chosen with careful analysis of the application. Further, we present some basic watermark characteristic and strategies for choosing the right watermark in order to achieve right functionality.

"Robustness" is characteristic that demonstrate how resistant the watermark is against common signal processing applied on carrier data. In common, the watermark must be robust only in the period between processes of embedding and detection. On contrary, characteristic that shows how much watermark system is robust against hostile attacks is called "tamper resistance".

Table 1 represents sublimation of most common watermark applications and characteristics together with effects from different types of attacks on individual application.

The goal of **removal attacks** is to remove embedded watermark or to make it undetectable for watermark detector. **Collusion attacks** are important subclass of removal attacks.

The goal of collusion attacker is to create unmarked carrier data, by collecting many copies of the same carrier data marked with different watermarks [8].

Legend: C – Critical, NC- not Critical, R – robust, F - Fragile	**Characteristics**			**Type of attacks**		
Application	Robustness	Speed	Quantity	Removal	Collusion	Forgery
1. Secret communication	F	slow	-	N	N	N
2. Broadcast monitoring	R	fast	+	C	N	N
3. Owner identification	R	slow	-	C	N	N
4. Proof of ownership	R	slow	-	C	N	C
5. Authentication	F	med	-	N	N	C
6. Fingerprinting	R	slow	-	C	C	N
7. Copy control	R	fast	+	C	N	N

Table 1: Analysis of watermarking applications, their characteristic and attacks influence

In contrary, in **forgery attack** group, it is typical for the attacker not to remove the watermark, but to insert another valid watermark in order to fake the watermark detector.
"Computational cost" is characteristic which can be represented as speed of the watermarking system and used devices (embedders and detectors) needed for accomplishing the goal of application.

Conclusion to this chapter is that different set of standard should be used, according to the application for which watermark is created. If users act in accordance with this principle, they can achieve better watermark performances. Along with this, the attacks against watermark will have less success, or the overall performance of the watermarking system will be better.

III. CLASSIFICATION OF WATERMARKING SYSTEMS AND STRATEGIES AGAINST WATERMARK ATTACKS

Analyzing and focusing on the information to which attacker can access are one of the main strategies for designing and classifying security watermarking system [1]. According to this approach, we can assume that for example, attacker does or does not have access to the algorithms used for watermark embedding or that attacker is fair or un-fair player. Fair attacker is attacker that uses only publicly accessible information for his attacks. Unfair attacker can use observations for revealing the secret keys for embedding and detection or any information about functioning of the watermarking system.

In Fig.1 is presented "effort to deal with attacks" against "information to be kept secret" diagram. Here we can see that as "information to be kept secret" increases "the effort to deal with attacks" increases for un-fair attacks and decreases for fair attacks. Keeping in mind this notes, we can make analysis of watermark security with respect to public information ***P*** to which attacker can access, where a is detection algorithm, k_D is detection key and k_E is key used for embedding.

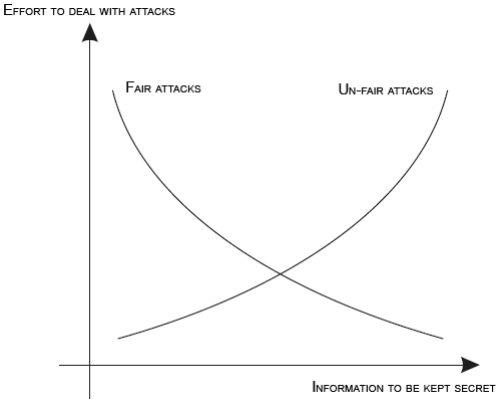


Figure 1: The diagram of "Effort to deal with attacks" versus "Information to be kept secret" in respect to fair and un-fair attacks

According to this information, the following classification can be made:

1. No public information is known: P = 0;
2. Embedding and detection algorithm are public P = {a};
3. Everything except embedding key is public P = {a, k_D};
4. Embedding and detection keys are public P = {a, k_D, k_E}

Scenario:	Public information			Point of view:	
	a	k_E	k_D	**Attacker**	**Owner**
1. Security-by-obscurity	no	no	no	Focused on un-fair attacks	To many secret information
2. Symmetric watermarking	yes	no	no	Balancing between fair and un-fair attacks	It is very hard to keep k_D in secret
3. Asymmetric watermarking	yes	no	yes	It is better to focus on fair attacks	Fair attacks are very dangerous
4. Playing with open cards	yes	yes	yes	Powerful fair attack exist	Nothing but hope

Table 2: Classification of watermarking systems in respect to information publicly accessible by attacker

3.1 Security by obscurity

This is the scenario where no public information about the watermarking system is known. In this scenario, designing of fair attack can be a very hard task. For that reason, the choice of attacker is to concentrate on un-fair attacks, because there could be a possibility some secret information about the watermarking system to outflow, thanks to the observations.

The watermarker's point of view*:* This strategy was common in the beginning of watermarking's history. The main concept was based on the fact that if all information about watermarking system is kept secret, then watermarking would definitely be secure. This was the main reason why research efforts were focused only at robustness requirements and watermarking security was neglected. Unfortunately, it is very hard to keep the watermarking algorithm secret and security-by-obscurity cannot fight against un-fair attacker.

In cryptography, the people are more aware about leaking of information. In 1883, Kerckof presented the basic rules in cryptography [4], where he stated that the cryptographic systems' designers must be aware that the attacker knows their algorithm in details except the secret key, on which the security of cryptographic systems has to be based. The Kerckof's basic rules are some kind of a warning about unsafeness of security-by-obscurity scenario. These rules are foundation of cryptography analysis and also foundation of security analysis of watermarking systems.

The attacker's point of view: Because in this scenario there is no public information available, fair attacker tries to trick the watermark decoder with some transformation of contents of cover data. This is the case of direct confrontation of robustness of watermark and intentional signal processing. Main concern for watermarking systems represents geometrical distortion or geometrical attacks. Possible geometrical counterattacks are for example: embedding of template or extra signal used for synchronization of watermark embedder and detector [13]; embedding of the watermark signal in invariant domains [12]; introduction of redundancy in the watermark signal in order to reduce the space of potential delays; or using of self-registration of the image [10]. From day to day, robustness watermarking techniques become more powerful. If the fair attacker wants to remove the watermark signal, he must distort the cover signal to very low level. We may conclude that in the end the robustness will fight the fair attacker, at least in the context of

given application, so in this scenario the only possible solution for attacker is to use un-fair strategies which will reveal algorithms and keys used for watermarking process. On the other side, the watermark owners are instructed from Kirckof's rules that security of watermarking system can not be protected only by obscurity. According to this, we can conclude that this scenario is not stable state in the table 2.

3.2 Symmetrical watermarking

Under presumption that only embedding and detection keys are secret, in this scenario the fair attacker tries to remove or to make the watermark undetectable based only on a priori information about embedding and detection algorithm. On contrary, un-fair attacker tries to detect the embedded watermark and based on this a posteriori information, to remove or to make watermark undetectable.

The watermarker's point of view: The first concern of the watermarker is choosing a watermark technique that has to be robust against common transformation of watermark content, which attacker can use in current application. Theoretically, it is possible to find watermarking technique, which is robust against common transformation allowed for particular application. On the other side, the attacker can use attacks that are more sophisticated because watermark extraction functions are public and he can act in the embedding domain. More sophisticated attacks are based on noise filtering or more specially Wiener filtering which can be use to separate the watermark from cover signal. Possible countermeasure proposed by Su is satisfaction of Power Spectrum Condition [7] which states that Spectral power density of watermark should be shaped according to Spectral power density of features vector. Another possibility proposed by Le Guelvouit is the owner first to embed watermark and then to attack watermarked signal with Wiener filter [9]. In this way, the efficiency of filters for removing the watermark is diminished. As second, the watermarker must be sure that watermarking systems implemented in public electronic devices can not be hacked. Security followed by rules from this section is possible only if there is a method for securing the keys.

The attacker's point of view: In this scenario attacker assumes that watermarker did a great job and that he has to deal with robust watermark. Because of that, the attacker strategy is to reject the rules of the game. Revealing of the secret keys helps the attacker to forgery watermarked content with very low level of distortion. The keys are used for decoding the secret messages. When the attacker has the keys, the data which has to be embedded and the watermark content, the synthesis of the watermark signal and its subtraction from the watermarked content are considerably easier. It should be noted that this simple attack is not working with models with quantization index, because in that case the embedded signal depends on the cover signal. Similar algorithm for modulation schemes with quantization index is presented in [3]. In this case the carrier signal is more degraded.

3.3. Asymmetric watermarking

In this scenario, both the algorithm **a** and detection key $\mathbf{k_D}$ are public information. The security of embedding key is essential for security of the watermarking system. This concept may sound strange but it is reality in cryptography where, the digital signs are based on asymmetry. The key used for embedding and related verifying key are different, but with verifying key we can check the data which was signed with private key.

If we like to use this concept in watermarking, the processes of embedding and detection should be asymmetrical and they must be based on different keys. This is the reason why the public detection watermarking is usually referred as asymmetrical watermarking.

The attacker's point of view: The asymmetry of embedding and detection keys is not sufficient to offer security in watermarking system. However, asymmetry does not imply robust public key watermarking. So far, all known asymmetrical systems where detection key $\boldsymbol{k}_{\mathrm{D}}$ is public, are hacked. In paper [11] is presented an attack which is valid for almost any known asymmetrical system. As conclusion, we can say that even if asymmetry is helpful, it is far from being adequate to ensure security in a public detection system.

The watermarker's point of view: The knowledge of detection algorithm and key for detection involve knowledge of boundaries of detection region. That is the main reason why asymmetrical systems are not safe methods for detection with public key. In scenarios where attacker knows the boundaries of detection regions, the "closes point" attacks presents deadlock. These closest-point attacks can be prevented only by using sufficiently sophisticated detectors, irrelevant from the methods used for embedding [5] or function that indicates detection region but not reveals boundaries of this region. The function that conveys this principle is fractal function. Unfortunately, until now there is no way to build watermark system that satisfies this function, but this is a good starting point that yields to good characteristic which can be used in public key detection watermarking. The attacker must test every point of the space in order to reveal the boundary of detection region and that process can last extremely long. As addition, the watermarker must verify that there is no other way for revealing the boundary and to be sure, that knowledge of detection keys will not bring any a priori information about embedding key.

3.4 Playing with open cards

The principle presented in previous scenario reduces the amount of information that owner should keep in secret. That yields toward reducing the effort needed for watermarker to resist against un-fair attacks and against stronger fair attacks. At the end, this process will bring the situation in which all information is public. In that situation, the attacker can access to information about detection and embedding keys used by watermarking system.

The watermarker's point of view: It is clear that playing with open cards scenario is most appropriate for un-fair attackers. In this scenario, the watermarker does not care about protection of information because all information is

public. The only question is: Is robustness possible against fair attacker?

In fair version of the "closest point" attack, the attacker knows the boundary of detection region and according to this, watermark can be made unreadable by simply moving the host data in closest point of un-detectable region.

Un-fair version of "closest point" attack differs from its fair version only in determination of detection region. The un-fair attackers estimate boundaries of detection regions through error and trial procedures.

The following chapter presents possibility asymmetric watermarking schemes to give solution for fighting the "closest point" attack.

The knowing of boundaries of detection region is not helpful for the attacker, if detection region is defined in such complicated manner, that moving the point inside and outside from it with respect to visual distortion, would be computational impossible process. Unfortunately, the usage of this complicate detection region makes watermark embedding process to be very complex. Here, asymmetrical watermarking can help the watermarker, by offering him a simple description for detection region. According to this scenario, the needed asymmetry between embedding and detection of watermark (which represent entering and exiting the region) can be accomplish by offering the different description of detection region (different key) to the watermarker and to the attacker as shown in Figure2.

If it is possible to design detection region for which it is very simple to move point inside and very hard to move point outside the region, with acceptable visible distortions, then there is no need to make different keys for embedding and detection and there is no need for keeping them secret too. With other words, the watermarker can play with open cards. It is obvious that designing such detection region is almost impossible, but until we get explicit answer for that, the scenario for playing with open card can not be ignored.

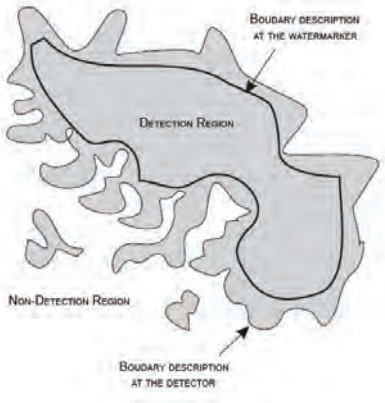


Figure 2: Asymmetrical watermarking; Comparison of boundaries descriptions offered to watermarker (simple) and attacker (complex)

The attacker's point of view: Scenario for playing with open cards is favorable for attacker's point of view because in this case attackers can easily perform lots of different attacks. However, if someone succeeds to design a basic asymmetrical region with features described in previous chapter, the efforts of attacker will become hopeless, because in such situation he will not have possibility to use un-fair attacks.

IV. CONCLUSION:

No matter how robust the watermark can be against one attack, it can be very fragile against another. We demonstrate that for successful fight against attacks on embedded watermarks, it is extremely important to understand the process of watermarking. The lack of understanding the functionality of used watermarking technique is common mistake, which produces weak watermarking systems and un-robust watermarks. This paper clarifies that the watermarker should make good strategy, which includes securing the watermarking system, used techniques and watermarks itself. Security of the watermark system depends on the watermark application in which watermark should be used and from information about watermarking system to which attacker can access. As conclusion, the embedder must have knowledge of all the previous mentioned, in order to design more robust watermarking system.

REFERENCES

[1] M. Barni, F. Bartolini, T. Furon, "A general framework for robust watermarking security", Signal Processing 83, p.2069–2084, 2003.

[2] I. J. Cox, M. L. Miller and J. A. Bloom, "Watermarking applications and their properties", Published in the Int. Conf. on Information Technology'2000, Las Vegas, 2000.

[3] J. Eggers, B. Girod, "Informed Watermarking", Kluwer Academic Publishers, Dordrecht, 2002.

[4] A. Kerckhoffs, "La cryptographie militaire", J. Sci. Militaires 9, p.5–38, 1883.

[5] M. L. Miller, "Is asymmetric watermarking necessary or sufficient?", in: Proceedings of the European Signal Processing Conference - EUSIPCO 2002

[7] J. Su, J. Eggers, B. Girod, "Analysis of digital watermarks subjected to optimum linear filtering and additive noise", Signal Process 81, p.1141–1175, 2001.

[8] I. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia", IEEE Trans. on Image Processing, 6(12):1673–1687, 1997.

[9] S. Pateux, G. Le Guelvouit, C. Guillemot, "Perceptual watermarking of non i.i.d. signals based on wide spread spectrum using side information", Proceedings of the IEEE International Conference on Image Processing, Vol. 3, Rochester, NY, USA, 2002.

[10] P. Bas, B. Macq, "A new video-object watermarking scheme robust to object manipulation", in: Proceedings of the International Conference on Image Processing, IEEE, Thessaloniki, Greece, 2001.

[11] T. Furon, I. Venturini, P. Duhamel, "Unified approach of asymmetric watermarking schemes", in: P.W. Wong, E. Delp (Eds.), Security and Watermarking of Multimedia Contents III, SPIE, San Jose, CA, USA, 2001.

[12] J. O'Ruanaidh, T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking, Special issue on copyright protection and control", Signal Processing 66(3): 303–317, 1998.

[13] S. Pereira, T. Pun, "Fast robust template matching for affine resistant image watermarks", in: A. Pfitzmann (Ed.), Proceedings of the Third International Workshop on Information Hiding, Springer, Dresden, Germany, pp. 199–210, 1999.

SESSION CSIT III

Computer Systems and Internet Technologies III

Interpretational Approach to Service Logic Execution

Ivaylo I. Atanasov[1] and Evelina N. Pencheva[2]

***Abstract* – The paper presents general aspects of implementation of a new mark-up language used for service creation. The language interpreter consists of two parts: lexical analyzer which recognizes the lexical units of the language, and the parser which performs syntax analysis of the input sequence of tokens in order to determine its grammatical structure with respect to language formal grammar. During interpretation, for each language construction correctly recognized by the parser, an appropriate Java construction is activated.**

***Keywords* – Mark-up language for service creation, lexical analysis, parsing, interpreting**

I. INTRODUCTION

Markup languages have found their way into almost every facet of telecommunications from provisioning services through network management systems and even scripting languages for automated voice services.

Because of fast and complicated hardware there is no need to optimize the use and encoding of information in protocols and databases in binary form. The simplicity and ease of understanding make text encoding preferable both in representing information in databases and in encoding protocol messages. This, combined with the view that content means revenue, is giving rise of markup languages for next generation service creation [1], [2].

There exist several markup languages proposed for service creation. Languages like CPL (Call processing Language), VoiceXML (Voice eXtensible Markup Language), CCXML (Call Control eXtensible Markup Language) are mainly oriented towards call control i.e. the creation, manipulation, termination and teardown of communication sessions. None of the existing markup languages supports functions like mobility, user status, terminal capabilities, charging and others. These network functions can be accessed by Parlay/OSA (Open Service Access) interfaces. Parlay/OSA is a service framework which defines application programming interfaces (APIs) for third party service developers. The idea behind API is to hide network specifics and protocol complexity for service developers.

We suggest a new mark-up language called **S**ervice **L**ogic **P**rocessing **L**anguage (SLPL) which is developed to meet challenges of service creation [3]. The language supports the whole palette of network functions exposed by Parlay/OSA interfaces.

One of the factors that determine language usability depends on availability of supporting tools. An SLPL interpreter is developed that makes lexical analysis of service logic description, builds an abstract syntax tree and makes mapping of abstract tree's nodes to appropriate Java constructions. To ease the process of service description a translator from IDL (Interface Description Language) to SLPL is also developed. The IDL is a language used to specify data types and method definitions of Parlay/OSA APIs. Because of the huge amount of data types and methods defined, this tool allows translation of Parlay/OSA definitions into SLPL ones.

In the paper we present some aspects of the implementation of SLPL interpreter and of translator from IDL to SLPL.

II. SLPL INTERPRETATIONAL APPROACH

The interpretation of the logic is separated into four processing phases – front processing, lexical analysis, syntactical analysis, and execution. Fig.1 shows the interpretational approach of execution of service logic described in SLPL.

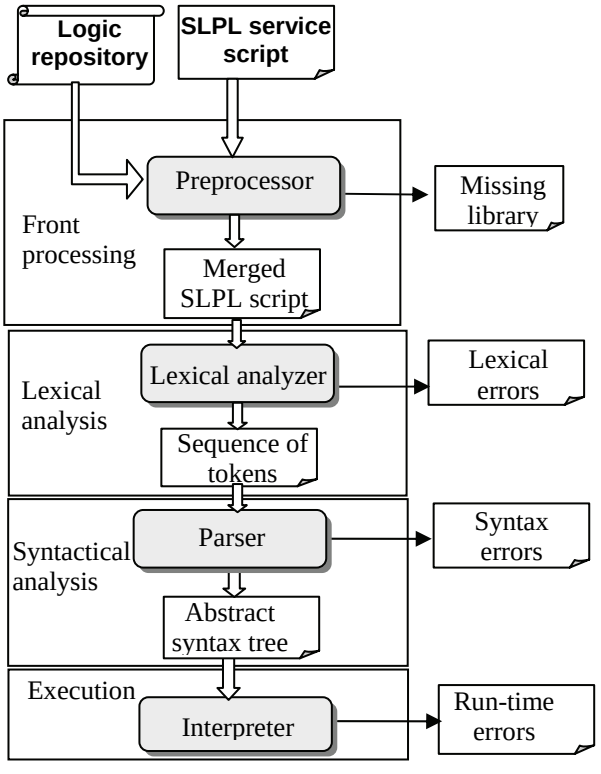


Fig.1 Interpretational approach

The front processing is in charge of loading the service script i.e. making an instance of the logic script. Ready-to-use parts of script containing SLPL data types and methods definitions are imported from the SLPL library. After

[1]Ivaylo I. Atanasov is with the Faculty of Communications and Communication Technologies, Technical University of Sofia, 1000 Sofia, Bulgaria, E-mail: iia@tu-sofia.bg

[2]Evelina N. Pencheva is with the Faculty of Communications and Communication Technologies, Technical University of Sofia, 1000 Sofia, Bulgaria, E-mail: enp@tu-sofia.bg

including definitions in the original instance, the SLPL preprocessor produces a merged, extended instance.

The main task of the SLPL lexical analyzer is to recognize the lexical units of the language like identifiers, key words, literals and so on. The extended instance of service script is lexically converted into a sequence of tokens and then the sequence is passed as input to the parser.

The SLPL parser performs syntax analysis of the input sequence of tokens in order to determine its grammatical structure with respect to SLPL formal grammar. The SLPL parser is to decide whether the sequence is acceptable in the terms of the syntactic rules of the language. If it is to be rejected, then error log is open. Parsing transforms input sequence of tokens into a data structure (an abstract tree), which is suitable for later processing and which captures the implied hierarchy of the input.

During real processing for each language construction recognized correctly by the SLPL parser, an appropriate Java construction is activated.

III. SLPL INTERPRETER

The main task of SLPL lexical analyzer is to break SLPL service logic description into tokens. The lexical analyzer processes the input sequence of characters to recognize the lexemes and to evaluate their type.

The input sequence scanning is based on finite state machine. It has encoded within it information on the possible sequences of characters that can be contained within any of the tokens it handles.

The finite state machine (FSM) of the lexical analyzer is table-based. The state-table defines all details of the behavior of FSM. It consists of three columns: in the first column state names are used, in the second the virtual conditions built out of input names using the positive logic algebra are placed and in the third column the output names appear.

Fig.2 shows an example of FSM that recognizes the key words *val*, *valref* and *value*.

The lexeme types in SLPL are the following:

- literals (integer, long, float, double, octet, logical and character)
- operators (for example equation)
- separators (for example space, slash, angular brackets, double quotes)
- identifiers
- key words.

In order to construct a token, the lexical analyzer needs to evaluate the characters of the lexeme that is to produce a value. The lexeme's type combined with its value is what properly constitutes a token, which can be given to a parser.

For example, let us consider the SLPL language construction for synchronization. For the input

<wait timeout="20"/>

the SLPL lexer ignoring spaces recognizes the type and value of the tokens listed in Table 1.

The sequence of tokens is passed as input to the SLPL parser.

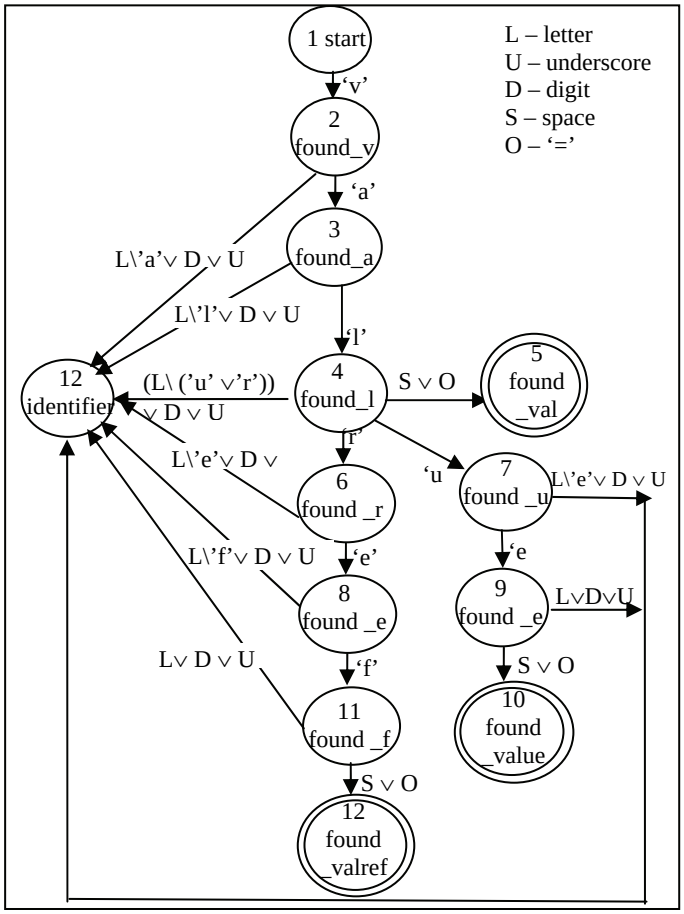


Fig.2 FSM recognizing SLPL lexemes

TABLE 1
LEXEMES AND THEIR TYPE

<	separator
wait	key word
timeout	key word
=	operator
"	separator
20	integer literal
"	separator
/	separator
>	separator

The SLPL parser analyzes the sequence of tokens to determine its grammatical structure with respect to the SLPL formal grammar. Its task is essentially to determine if and how the input can be derived from the start symbol within the rules of the SLPL formal grammar. This is done in a top-down parsing manner. The SLPL starts with the start symbol and tries to transform it to the input. The parser starts from the largest elements and breaks them down into incrementally smaller parts. It is LL parser i.e. it parses the input from left to right, and constructs a leftmost derivation of the sentence.

The parser builds abstract syntax tree, which is a finite, labeled, directed tree, where the internal nodes are labeled by grammar rules, and the leaf nodes represent the terminal or nonterminal symbols.

The SLPL grammar is formally described in Augment Backus Naur Form (ABNF). For example, the language construction for synchronization “wait-statement” used to wait for result of network provided service is formally defined with the grammar rules shown in Fig. 3.

```
1.  wait_def        = LB "wait" [timeouted](slash GB/ GB
                      event_list); wait for an even or timeout
2.  timeouted       = "timeout” is DQUOTE integer-val DQUOTE
3.  event_list      = 1*event_list_item LB slash "wait" GB
4.  event_list_item = LB“event” a_name slash GB
5.  a_name          = "name" is DQUOTE simple_name DQUOTE
6.  integer_literal = 1*DIGIT
7.  simple_name     = ALPHA*(ALPHA/ DIGIT/ underscore)
8.  slash           = %d47; ‘/’
9.  is              = %d61; ‘=’
10. underscore      = %d95; ‘_’
11. LB              = %d60; ‘<’
12. GB              = %d62; ‘>’
13. ALPHA           = %d65-90 / %d97-122; A-Z a-z
14. DIGIT           = %d48-57: 0-9
```

Fig.3 Grammar rules for definition of wait-statement

The SLPL parser has an input buffer, a stack on which it keeps (terminal and nonterminal) symbols from the grammar, a parsing table which tells what grammar rule to use given the symbols on top of its stack and its input type.

In order to construct the parsing table, that is to establish what grammar rule the parser should choose if it has a nonterminal *A* on the top of its stack and a symbol *a* on its input stream, we have used the algorithm described in [4].

For example, Table 2 shows the parsing table for the grammar rules defining the “wait-statement”.

TABLE 2
PARSING TABLE FOR WAIT-STATEMENT

	$	wait	timeout	event	name
wait_def					
wait_def_next			3		
wait_def_ext					
timeouted			6		
event_list					
event_list_ext					
event_list_item				10	
a_name					11

TABLE 2
PARSING TABLE FOR WAIT-STATEMENT
(CONTINUE)

	slash	GB	is	DQUOTE	integer_ literal	simple _name
wait_def						
wait_def_next	2	2				
wait_def_ext	4	5				
timeouted						
event_list						
event_list_ext	9					
event_list_item						
a_name						

The terminal ‘$’ indicates the bottom of the stack. The resulting abstract tree from the input sequence of tokens for

<wait timeout="20"/>

is shown in Fig. 4

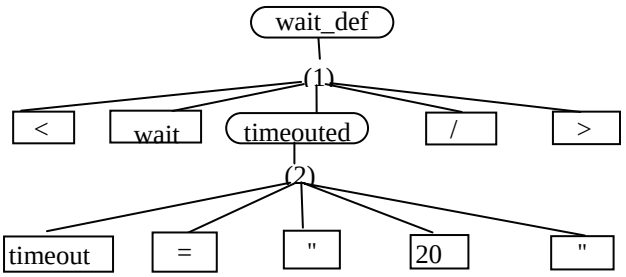


Fig. 4 Abstract syntax tree built for ‘wait’-statement

The overview of the abstract syntax tree generated from an SLPL service logic description is shown in Fig. 5.

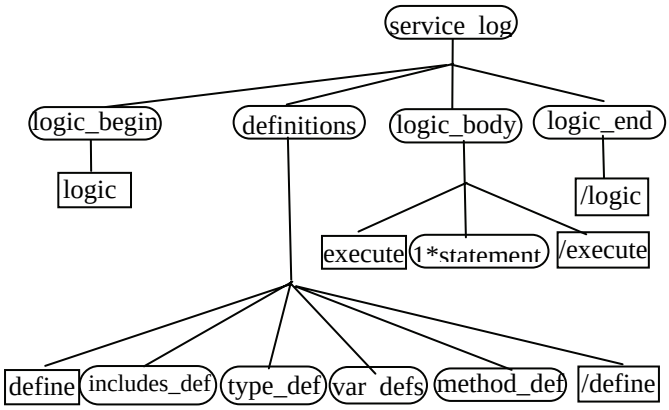


Fig.5 Overview of syntax tree of SLPL grammar description

During processing correct language structures are compared with templates of semantic descriptions and converted into Java calls.

For example, the SLPL language construction for method invocation with capturing exceptions is shown in Fig. 6 and the corresponding Java code activated is shown in Fig. 7.

```
<try>
   <invoke>
      <method name = "locationReportReq">
         <arguments>
            <argument name = "application" valref = "theAppl"/>
            <argument name = "users" valref = "theUsers"/>
         </arguments>
      </method>
   </invoke>
   <catch>
      <exception name = "P_APPLICATION_NOT_ACTIVATED"/>
      <exit/>
   </catch>
</try>
```

Fig. 6 SLPL construction for method invocation

```
try {locationReportReq( theAppl, theUsers );
    } catch ( ApplicationNotActivatedException e )
   { exit(); }
```

Fig. 7 An example of Java code corresponding to method invocation

IV. TRANSLATION FROM IDL TO SLPL

The Parlay/OSA interfaces are specified in IDL which is programming language independent. A huge amount of data types on which methods operate are included in IDL specification. To reduce efforts needed for data types and method definition in SLPL an "include-statement" is provided. This construction allows including SLPL descriptions of methods and data types in the definition part of the service logic script. These ready-to-use parts of script are located in the script repository, accessed in shared library manner.

To create SLPL script repository a translator form IDL to SLPL is developed. As it is shown in Fig. 8, the translator is supplied with IDL description as input and the result is SLPL description as output.

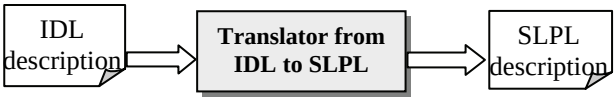


Fig. 8 Translator from IDL to SLPL

The formal IDL definition of an interface includes data types and methods. Each method has a type (the result type), a list of arguments of certain type, and a list of exceptions that it can raise.

The translator makes lexical analysis of an IDL description to determine tokens and then performs syntactical analysis to build an abstract syntax tree corresponding to IDL description. During code generation the translator converts the syntactically correct IDL descriptions to SLPL ones.

Fig.9 shows a part of SLPL grammar rules that formally define an interface with data types and methods.

V. CONCLUSION

The suggested markup language SLPL possesses greater expressive power in comparison with existing markup languages for service creation. The language supports all network function exposed by Parlay/OSA interface. Provided language construction for flow control brings SLPL near to programming languages.

The language usability is provided by development of language supporting tools such as SLPL interpreter and translator from IDL to SLPL. These tools are written in Java to achieve portability. Another reason for Java choice as implementation language is that we use Ericsson Network Resource Gateway SDK (version R5A02) that simulates Parlay/OSA interfaces to verify the SLPL functional capabilities. The interface method calls of Parlay/OSA interface simulator are in form of Java code.

The approach enables rapid prototyping, rapid application development, and easy end-user customization. These facts free developers to focus on the creation of new types of applications that may bring new revenues to service providers.

1. interface_def = LB "interface" a_name GB interface_body LB slash "interface" GB
2. interface_body = inheritance type_defs method_defs
3. inheritance = LB “inherits” GB inherit_instances LB slash “inherits” GB
4. inherit_instance= 1*(LB “interface” a_name slash GB)
5. type_defs = (LB "types" 1*type_def slash "types" GB) / (LB "types" slash GB)
6. type_def = simple_type / complex_type / alias_type
7. simple_type = "integer" / "long" / "float" / "double" / "boolean" / "string" / "char" / "octet" / "reference" / "Timer" / "Time" / "Date" / "DateAndTime" / "Duration“
8. complex_type = enumerated_type / structured_type / sequence_type / union_type
9. method_defs = (LB "methods" slash GB) / methods_def
10. methods_def = LB "methods" GB 1*method_def LB slash "methods" GB
11. method_def = LB “method” a_name method_args [body_def] return_type [exceptions_raised] LB slash "method" GB
12. method_args = LB "arguments" GB 1*method_arg LB slash "arguments" GB
13. method_arg = LB "argument" a_name a_type slash "argument" GB
14. return_type = (LB "returns" slash GB) / return_def
15. return_def = LB "returns" a_type slash GB
16. exceptions_raised= LB "raises" 1*exception_def slash "raises" GB
17. exception_def = LB "exception" (exception_by_type / exception_by_name) slash "exception" GB
18. exception_by_type= a_type
19. exception_by_name= a_name
20. body_def = LB "body" GB 1*statements LB slash “body“ GB
21. statements = assign_statement / invoke_statement / wait_statement / if_statement / case_statement / while_statement / goto_statement / try_statement / arithmetic_operations / new_statement / exit_statement

Fig.9 Formal grammar tool for interface definition in SLPL

REFERENCES

[1] Bakker, J. Tweedie, D. and M. Umnehopa, “Evolving service Creation; New developments in Network Intelligence”, http://www.argreenhouse.com/papers/jlbakker/Bakker-elenor.pdf

[2] Bakker J-L, R. Jain, “Next Generation Service Creation Using XML Scripting Languages”, http://www.argreenhouse.com/papers/jlbakker/bakker-icc2002.pdf

[3] Atanasov I., E. Pencheva, “A Mark-up Approach to Add Value”, IJIT Enformatika, vol.3, Number 4, pp 267-276

[4] Aho, Sethi, Ullman, *Compilers: Principles, Techniques, and Tools*, Addison-Wesley, 1986.

[5] Common Object Request Broker Architecture (CORBA), v3.0, *OMG IDL Syntax and Semantics*, http://www.omg.org/docs/formal/00-10-07.pdf

[6] Ericsson Network Resource Gateway SDK (version R5A02) http://www.ericsson.com/mobilityworld/sub/open/technologies/parlay/tools/parlay_sdk

Software for Estimation of xDSL Service Quality

Mladen Ranđelović, Aleksandar Atanasković, Nebojša Dončov and Bratislav Milovanović

***Abstract* – In this paper, estimation of xDSL service performances is presented. Since the transmission medium quality is one of the key requirements for performance capabilities of these services, an analytical method for modelling of primary per-unit length parameters of twisted pair is described. This model is implemented in the existing software extended to allow for estimation of data rates on telephone subscriber loops for xDSL service. For the example of particular twisted pair telephone cable, an experimental measurement of its parameters is done, which is followed by their modelling using described software. At the end, comparison of experimental and modelled results of capacity for specific cable is presented.**

***Keywords* – xDSL, twisted pair, modelling, estimation, characterization.**

I. INTRODUCTION

The demand for high-speed data communications has been on the increase over the past several years due to drivers such as the internet, including the World Wide Web (WWW), video service, consumer services, and entertainment to name a few. High-speed voice band modems have reached a maximum transfer rate of 56 Kbps and alternative ISDN technology (*Integrated Services Digital Network*) access speeds reaching rates of 144 Kbps. Due to bandwidth limitations, new technologies are desired to exceed this limitation. There are different broadband access technologies on the market that are capable of transferring voice, data and multimedia content with high speed. These technologies can be classified in three groups due to transmission medium used: twisted pair, cable (fiber and coaxial) and wireless.

DSL (*Digital Subscriber Line*) technologies, called xDSL, which are equivalent to broadband access technology on twisted pair, are developed by telecommunication industry in order to use millions of kilometers of existing twisted pair infrastructure. One of the main advantages of DSL technology is equal usage and sometimes even sharing of telephone lines (local loops) with plain old telephone service (POTS) [1]. However, due to narrowband nature of POTS (0-4 kHz for voice signals), local loops are not suitable for broadband xDSL service.

The main idea of xDSL technology development is using existing infrastructure of telephone subscriber loops. Performance analysis of DSL technologies is of vital importance to DSL service providers. This type of analysis requires accurate prediction of the local loop structure and precise identification of the cable parameters.

Different models of cables have been derived in order to accurate describe the behaviour of primary parameters using transmission line model [2]. Although many of them are rather empirical [3], there is also an analytical model based on physical and electromagnetic characteristics of twisted pair [4]. Both empirical and analytical models are implemented in developed software used for characterization of twisted pairs on physical layer [5-6]. In addition, by using these models, it is possible to estimate achievable capacity of DSL systems, as illustrated in this paper. Data rates for xDSL service are calculated using power spectral densities (PSD) given in ITU-T 996.1 Recommendation [7]. Number of bits per subchannel and data rates for twisted pair are the characteristics that are incorporated in modified software for estimation of xDSL performances. The software is developed in MATLAB programming environment [8]. Based on available input parameters, software calculates the twisted pair characteristics and compares them with available measured characteristics. Verification of software is given on an example of particular twisted pair in frequency band of interest for xDSL service.

II. TWISTED PAIR MODELLING AND DATA RATE CALCULATION

The electrical characteristics of the twisted pair cables are defined using the classical transmission line model. This model incorporates a set of four parameters per unit length, including a series inductance and resistance and a shunt capacitance and conductance, also known as the RLCG parameters of the cable. The series inductance represents the total self-inductance of the two conductors, and the shunt capacitance is due to the close proximity of the two conductors. The series resistance is due to the finite conductivity of the conductors, and the shunt conductance is due to the dielectric loss in the material between the two conductors. Graphical representation of a segment of twisted pair cable per unit is given on a figure 1.

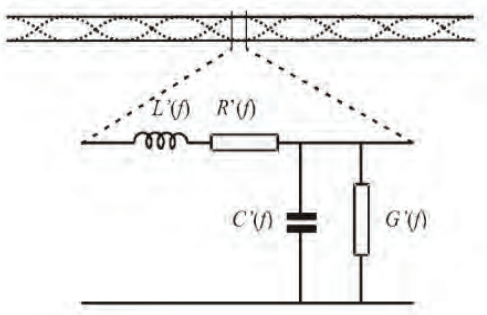


Fig.1 Segment of twisted pair modelled with transmission line model

[1]Authors are with the Faculty of Electronic Engineering, Aleksandra Mevdvedeva 14, 18000 Niš, Serbia, E-mails: {mladen, beli, doncov, bata}@elfak.ni.ac.yu.

Although physical principles have inspired the models, many are rather empirical. The empirical models are developed by fitting of measured primary and secondary parameters. These kind of models cannot be used for some frequency characteristics of the twisted pair, like crosstalk and transfer function. Their main deficiency is that they are not based on physical characteristics of twisted pairs. In reference [4], an example of an analytical model is described.

This model, which includes the skin effect and the effect of permittivity of isolation material, can precisely define primary parameters of twisted pair on frequencies of interest for xDSL service. Figure 2 shows physical characteristics of a cross section the twisted pair cable, which consists of two conductors with the diameter d and distance D between their centers. Conductivity of the metallic conductor is σ while permittivity of the insulating material is $\varepsilon_r' - j\varepsilon_r''$.

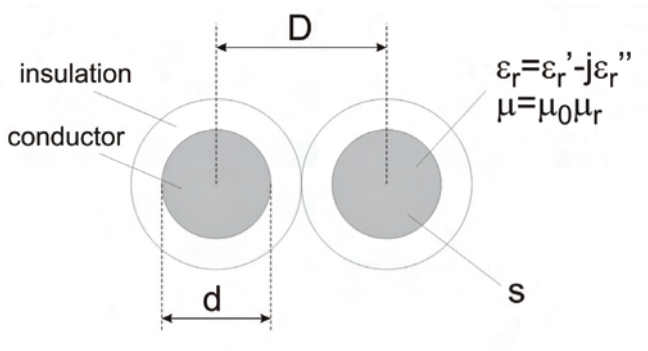


Fig.2 Physical characteristics of a cross section of twisted pair cable

Primary parameters per unit length according to analytical model can be calculated as (1-4):

$$C'(f) = \frac{\pi\varepsilon_0\varepsilon_r'}{\cosh^{-1}(D/d)}, \quad (1)$$

$$L'(f) = \frac{\mu_0}{\pi}\cosh^{-1}(D/d) + \frac{R_s(f)}{2\pi^2 fd}\frac{D/d}{\sqrt{(D/d)^2-1}}, \quad (2)$$

$$R'(f) = \frac{R_s(f)}{\pi d}\frac{D/d}{\sqrt{(D/d)^2-1}}, \quad (3)$$

$$G'(f) = \frac{2\pi^2 f\varepsilon_0\varepsilon_r''}{\cosh^{-1}(D/d)}, \quad (4)$$

Parameter R_s in equations (2) and (3) represents resistance which is frequency dependable as:

$$R_s(f) = \sqrt{\pi f \mu/\sigma}\ , \quad (5)$$

Using primary parameters, secondary parameters of twisted pair, characteristic impedance Z_c and propagation constance $\gamma=\alpha+j\beta$, can be calculated.

$$Z_c(f) = \sqrt{[R'(f)+j2\pi f L'(f)]/[G'(f)+j2\pi f C'(f)]} \quad (6)$$

$$\gamma(f) = \sqrt{[R'(f)+j2\pi f L'(f)][G'(f)+j2\pi f C(f)']} \quad (7)$$

Uniform portion of line of length l can be modelled as a two-port network, where voltage V_1 and current I_1 of the left-hand port are linked to the voltage V_2 and current I_2 of the right-hand side port through so-called ABCD model:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} V_2 \\ I_2 \end{bmatrix}, \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cosh(\gamma l) & Z_0\sinh(\gamma l) \\ \frac{1}{Z_0}\sinh(\gamma l) & \cosh(\gamma l) \end{bmatrix} \quad (8)$$

The values A, B, C and D are dependent only on the secondary parameters; they are independent of the termination impedance. Insertion loss can be computed according to:

$$H_{iloss}(f) = \frac{Z_S(f)+Z_L(f)}{Z_s(f)(CZ_L(f)+D)+AZ_L(f)+B} \quad (9)$$

where $Z_s(f)$ and $Z_L(f)$ are the source and termination impedances at the ends the loop. Power spectral density (PSD) of the received signal is then given by:

$$S_p(f) = S_x(f)\left|H_{iloss}(f)\right|^2 \quad (10)$$

,where $S_x(f)$ is the PSD of the transmitted signal.

Crosstalk is the primary noise source in DSL systems. For the purpose of estimating the performance of a DSL system, the crosstalk PSD, $S_n(f)$, needs to be computed, in most cases using the models presented in [7]. The crosstalk noise is modelled as additive Gaussian noise.

For ADSL and second-generation VDSL (VDSL2), DMT (*Discrete Multi Tone*) technology of multi-carrier modulation is standardized. In DMT systems, a discrete Fourier transform (DFT) is used to combine many narrowband signals to form one broadband signal. For each narrowband channel (each subchannel or tone), the signal and noise are almost white within the band. Therefore, data rate for each subchannel can be approximated by:

$$C_k = \frac{1}{T}\log_2(1+\frac{SNR_k}{\Gamma}) \quad (11)$$

where 1/T is the data symbol rate of each subchannel, SNR_k is the average signal-to-noise ratio in the subchannel and Γ is the overall gap from Shannon capacity at the desired error rate. Based on the assumption of almost flat signal and noise spectrum within each channel, the SNR of each channel can be approximated by:

$$SNR_k = \frac{S_p(f_k)}{S_n(f_k)} \quad (12)$$

where f_k is the sub-carrier (center) frequency for each subchannel. If DMT system operates with a noise margin or if coding is used, then the margin and coding gain are incorporated into Γ factor according to:

$$\Gamma = \frac{\Gamma(P_e)\cdot\gamma_m}{\gamma_c} \quad (13)$$

where $\Gamma(P_e)$ is a constant reflecting the Shannon gap (for a bit error rate of 10^{-7} this constant is approximately $\Gamma(P_e)=9.89$, γ_m is noise margin (typically 5-6 dB) and γ_c denotes to coding gain (due to trellis coding it is typically in the range from 3 dB to 5 dB).

The aggregate data rate using all the narrowband subchannels is the sum of the data rates of the individual subchannels:

$$C = \sum_{k=K_0}^{K_1} C_k = \frac{1}{T}\sum_{k=K_0}^{K_1}\log_2(1+\frac{SNR_k}{\Gamma}) \quad (14)$$

III. SOFTWARE FOR DATA RATE CALCULATIONS

In order to calculate capacity of twisted pairs for xDSL service, the existing software for modelling of primary and secondary parameters [5,6] is modified to allow transmission rate calculation. For software design, the option of creating Graphical User Interface (GUI) in MATLAB programming environment is used.

An analytical model based on physical and electromagnetic characteristics of twisted pair is implemented in software. Besides physical characteristics, other input parameters are length of line, system margin, error probability and coding gain. As output characteristics, user can choose among primary parameters (RLCG parameters), secondary parameters, insertion loss, transfer function, signal and noise power spectral densities, number of bits per channel and data capacity for twisted pair. If capacity of transfer is chosen, user needs to know which systems may interfere with the line of interest. In addition, software can give comparison of the modelled and measured results for characteristics of interest.

IV. MEASUREMENTS OF TWISTED PAIR PARAMETERS

In order to experimentally obtain primary parameters per unit length, measuring of input impedance of tested twisted pair in frequency band of interest is done. Impedance is measured when line has short and open end. Devices and instruments used for calibration and measuring are shown on figure 3.

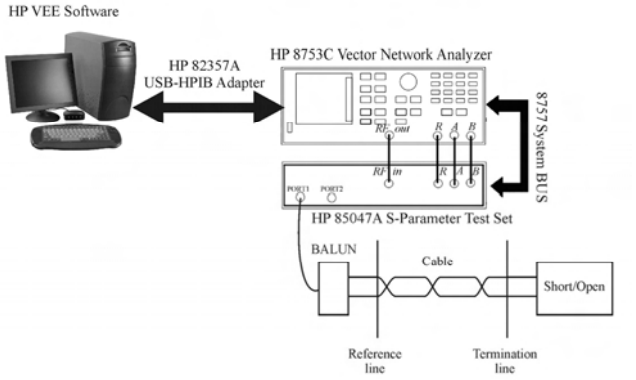


Fig 3. Measuring system

Through these measurements it is possible to calculate secondary parameters in frequency band of interest: Zc(f) i γ(f). As input impedance of twisted pair of length l in the case of open end is: Zoc(f)=Zc/tanh(γl)), and in the case of short end is: Zsc(f)=Zctanh(γl); characteristic impedance and the constance of propagation can be calculated as:

$$Z_c(f) = \sqrt{Z_{oc}(f) Z_{sc}(f)}, \tag{15}$$

$$\gamma(f) = \frac{1}{l} \tanh^{-1} \sqrt{\frac{Z_{sc}(f)}{Z_{oc}(f)}}, \tag{16}$$

Based on these results, primary parameters can then be calculated as:

$$R'(f) = \mathrm{Re}\{\gamma(f) Z_c(f)\}, \tag{17}$$

$$L'(f) = \frac{1}{2\pi f} \mathrm{Im}\{\gamma(f) Z_c(f)\}, \tag{18}$$

$$C'(f) = \frac{1}{2\pi f} \mathrm{Im}\left\{\frac{\gamma(f)}{Z_c(f)}\right\}, \tag{19}$$

$$G'(f) = \mathrm{Re}\left\{\frac{\gamma(f)}{Z_c(f)}\right\}, \tag{20}$$

Other characteristics of interest can then be computed by using equations (8)-(14).

V. NUMERICAL ANALYSIS

For the verification of analytical model implemented in software described on Section 3, twisted pair of the following physical characteristics is analyzed: d=0.65 mm and D=1.35 mm. Conductor of twisted pair is copper with specific conductance σ=5.65·107 S/m. Real part of insulating material permittivity is $\varepsilon_r' = 2.4$. Primary and secondary parameters of the described twisted pair are computed using equations (1)-(4) and equations (6)-(7), respectively. To verify modelled parameters, experimental analysis of considered twisted pair in frequency band up to 10 MHz is done too. Secondary parameters are extracted from measurements results as described in Section 4, and then primary parameters are calculated using equations (17)-(20). Either using modelled or measured parameters, other twisted pair characteristics can be easily obtained through equations (9-14). Comparison of the modelled and measured results for some of the software output characteristics is shown on figures 4-7.

Number of bits per channel for FDD (Frequency-Division Duplexing) and EC (Echo Cancelling) techniques, obtained using modelled and measured parameters, is shown on figs. 4 and 5, respectively.

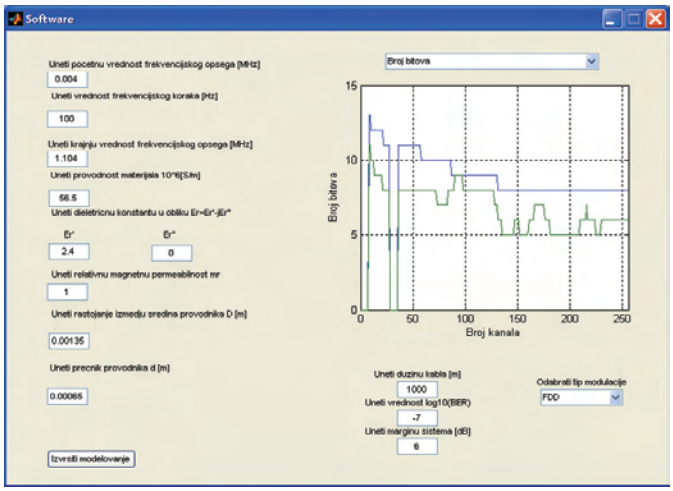


Fig 4. Number of bits per channel for FDD technique (modelled parameters – blue; measured parameters - green)

FDD channels, that are used for xDSL service, are from 8 to 27 for upload and from 36 to 256 for download transfer. Channel between 28 and 35 are not used for this service. Channels used for EC are between 8 and 31 for upload and between 8 and 256 for download transfer. For number of used channels, estimated data rate for EC is greater than for FDD.

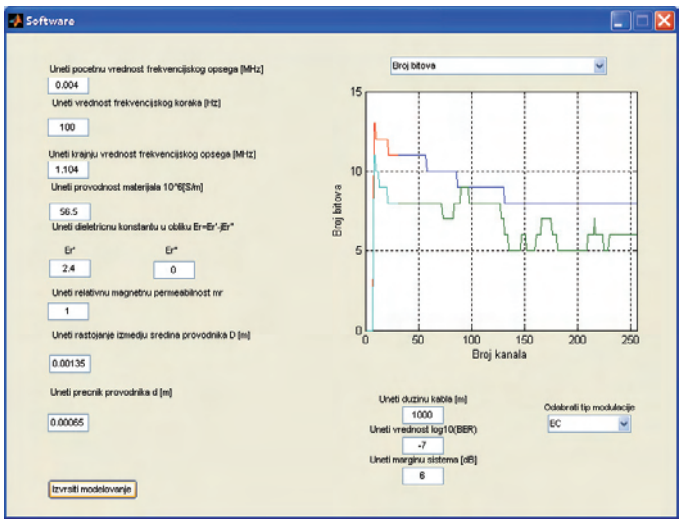


Fig 5. Number of bits per channel for EC technique (modelled parameters: red - upstream, blue – downstream; measured parameters: cyan – upstream, green - downstream)

Power Spectral Densities (PSD) of signal and noise for FDD and EC techniques are given on figs. 6 and 7, respectively, where modelled parameters are marked with red line for signal and green line for noise and measured parameters are marked with blue line for signal and cyan line for noise. Data rates for upstream and downstream, obtained from modelled and measured parameters, are given for a system with Bit Error Probability $BER=10^{-7}$ and system margin of 6 dB. Length of the local loop is l=1000 m.

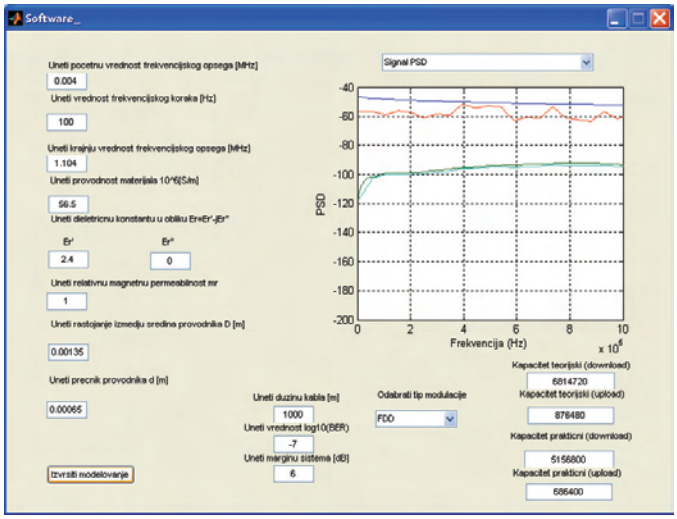


Fig 6. PSD and data rates for FDD technique

For FDD, data rates calculated from modelled parameters are 6.499 Mbit/s for download and 856 kbit/s for upload transfer. Data rates computed from measured parameters for specific cable are 4.918 Mbit/s for download and 670 kbit/s for upload. As mentioned before, capacities are greater for EC for the same cable and the same system. For EC, estimated data rates from modelled results are 7.63 Mbit/s and 1007 kbit/s for download and upload direction, respectively. Data rates calculated form measured results are 5.80 Mbit/s and 780 kbit/s for download and upload transfers, respectively.

VI. Conclusion

In this paper estimation of performances for telephone local loop, as the key component of xDSL service is presented. For the purpose of modelling of twisted pair, theory of transmission lines and analytical model based on physical and electromagnetic characteristics are used. Calculations of power spectral densities and data rates are based on ITU-T Recommendations. These characteristics, as well as primary and secondary parameters, transfer function and insertion loss, are implemented in developed software for evaluation of xDSL service quality. Through this software operaters of xDSL service are able to get an overview of the estimated performances of xDSL access technology on existing local loops.

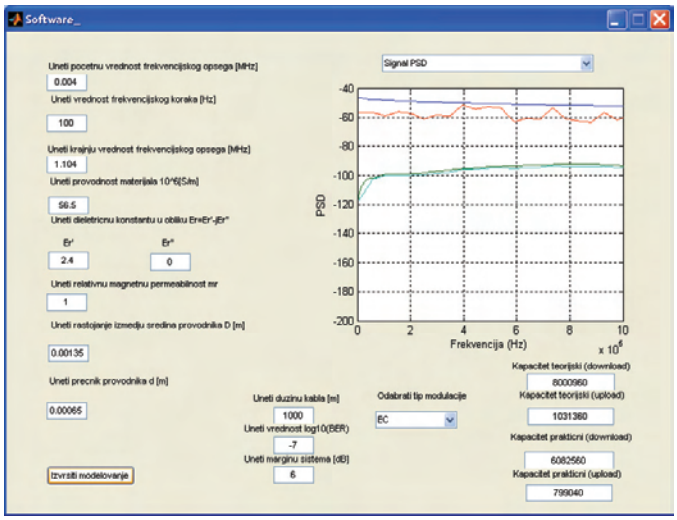


Fig 7. PSD and data rates for EC technique

VII. References

[1] L.Cohen, Heaviside's Electrical Circuit Theory, London: McGraw Hill Book London, 1928.

[2] P.F. Combes, Microwave Transmission for Telecommunications, John Wiley&Sons, 1991.

[3] P. Golden, H. Dedieu and K. Jacobsen, Fundamentals of DSL Technology, Auerbach Publications, Taylor & Francis Group, 2006.

[4] J.J. Yoho, Physically-Based Realizable Modeling and Network Synthesis of Subscriber Loops Utilized in DSL Technology, PhD Thesis, Faculty of the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, USA, 2001.

[5] Mila Stojanović, Aleksandar Atanasković, Mladen Ranđelović, Nebojša Dončov, Bratislav Milovanović, "Karakterizacija telefonskih parica za potrebe xDSL servisa", Proc. of L Conference ETRAN 2006, Vol. 2, Belgrade, Serbia, pp. 325-328, 6-8. June, 2006.

[6] Bratislav Milovanović, Mladen Ranđelović, Aleksandar Atanasković, Nebojša Dončov," Softverska realizacija i eksperimentalna verifikacija modela xDSL kablova ", POSTEL 2006, Belgrade, Serbia, pp. 277-287, 12-13. December 2006.

[7] ITU-T G.996.1 (06/1999), Test procedures for digital subscriber line (DSL) transceivers, ITU-T Recommendation, 1999.

[8] S.E. Lyshevski, Engineering and Scientific Computations Using MATLAB, John Wiley&Sons, 2003.

Efficiency of Parallel Graph Processing on Multicomputer Platform

Plamenka Borovska[1] and Milena Lazarova[2]

***Abstract* – The paper is aimed at investigation of parallel algorithms for finding minimum spanning tree of a graph. Prim's algorithm and genetic Prim's algorithms are regarded. The impact of cluster size and workload on the parallel performance are explored. Performance profiling and analysis are presented.**

***Keywords* – minimum spanning tree, genetic algorithm, parallel computational model, multicomputer.**

I. INTRODUCTION

The Minimum Spanning Tree (MST) problem is aimed at finding a least cost spanning tree in an edge weighted graph. The problem has been well studied and many efficient polynomial-time algorithms have been developed by Dijkstra, Kruskal, Prim [1, 2]. Finding of the MST is a task with a very wide application in the field of communication network design and communication network design problems.

Genetic algorithms (GA) are part of evolutionary computation techniques, which is a rapidly growing area of artificial intelligence. GAs have been successfully applied in many research fields as biogenetics, computer science, engineering, economics, chemistry, manufacturing, mathematics, physics.

The objective of this paper is to investigate the efficiency of parallel algorithms for finding minimum spanning tree in a graph. Prim's algorithm and genetic Prim's algorithm are regarded. Parallel computational models based on "manager-worker" parallel programming paradigm are presented. The algorithms are implemented using parallel MPI program. The target parallel platform is a multicomputer. The impact of multicomputer size and computational workload on the parallel performance parameters are explored and performance profiling and analysis are presented.

II. MINIMUM SPANNING TREE OF A GRAPH

A. Prim's algorithm

Let's consider undirected graph $G = (V, E)$ where:

- $V = \{v_1, v_2, \ldots, v_n\}$ is a finite set of vertices;
- $E = \{e_1, e_2, \ldots, e_m\}$ is a finite set of undirected edges.

Each element $e_k \in E$ ($k = 1, 2, \ldots, m$) is an ordered pair (v_i, v_j), $v_i, v_j \in V$, $1 \le i,\ j \le n$.

A graph is weighted when a function $f(i, j)$ relates an integer value (cost) to each edge $(i, j) \in E$ with $f(i, j) = f(j, i)$.

Minimum spanning tree (MST) is a tree $T(H, D)$ of the graph G with minimum sum of the weights of all edges included in the tree, where H is the set of vertices in the tree and D is the set of edges, connecting the vertices of the MST. Prim's algorithm is iterative algorithm for finding MST. It runs (n-1) iterations for a graph with n vertices. A vertex connected with the vertices already added to the tree by an edge with a minimum weight is selected at each iteration. The procedure of building MST according to Prim's algorithm is as follows:

Start from an arbitrary vertex s: H = {s}, D = Ø
Repeat (n–1) iterations
Choose edge (i, j) ∈ E such that
i ∈ H, j ∈ V/H and f(i, j) is minimum
Include the vertex j in H and edge (i, j) in D

B. Genetic algorithm of Prim

Genetic algorithms (GA) are metaheuristic method for global search and optimization that mimics the evolution of life organisms. The tree main principles of the natural evolution are applied to the GA: reproduction with variation, natural selection based on fitness and repetition with diversification of the individuals expressed by differences between a population and the next one. GA utilizes a population of data structure, called chromosome. Chromosomes represent possible solution of the regarded problem. Each chromosome has a fitness which is a numerical value indicating the quality of the solution the chromosome represents. The GA selects chromosomes to survive or reproduce so that those with better fitness are more likely to be selected. Crossover, also called recombination, combines genetic information from two parent chromosomes. Mutation randomly modifies one parent chromosome. When enough offspring is generated the parents are replaced and the process continues so that chromosomes that represent better solutions evolve.

GA algorithms are successfully applied to solve several graph and network problems including building of minimum spanning tree [3, 4, 5].

In order to apply genetic algorithm to the problem of finding MST the following issues are considered:

- the GA will generate only MST of a given graph;
- the correspondence between chromosomes and spanning tree will be 1:1;

[1]Plamenka Borovska is with the Faculty of Computer Systems and Control, Department of Computer Systems, 8 Kliment Ochridsky str., 1756 Sofia, Bulgaria, E-mail: pborovska@tu-sofia.bg

[2]Milena Lazarova is with the Faculty of Computer Systems and Control, Department of Computer Systems, 8 Kliment Ochridsky str., 1756 Sofia, Bulgaria, E-mail: milaz@tu-sofia.bg

- chromosomes' coding, crossover and mutation operators will not depend on the problem size.

Several spanning tree representations and chromosome encodings of the possible solution of MST of given graph for GA exists [6, 7, 8]. The coding of a spanning tree of a graph for GA is based on a representation of the spanning trees as strings of numerical weights associated with the graph's vertices.

The operations used for GA for finding of MST of given graph are as follows:

- Initialization – in order to guarantee that the solution found by GA will be correct, a random population of trees is generated at the initialization stage of the GA algorithm. The generated trees have to be spanning trees of the given graph and in order to satisfy that requirement vertices and edges are added to the tree according to Prim's algorithm but without observing the requirement that the edge added is minimal. The MST will be searched in the generated population of spanning trees.
- Selection – the selection is performed based on the roulette wheel selection rule which states that parents are selected according to their fitness. The weight of each generated tree is calculated and the total sum of all weights of population is found. A random number is generated between 0 and the total sum of the weights in the population.
- Crossover – for the crossover operator the following idea is applied: the solution after crossover has to inherit as many of the parents' edges as possible. At first step the new spanning tree takes the edges that are common for both two parents. At the next step edges belonging to only one of the parents are checked if they can be included in the new solution without making cycles with the edges that have already been added. At this step edges are again randomly added to the new spanning tree as was during the initialization stage. The addition of edges stops when a full tree is built, i.e. the number of added edges is one less than the number of added vertices.
- Mutation – mutation is based on addition of randomly selected edge and deletion of an edge that lies in the cycle between the added edge and the other tree edges, going out of the vertices when the edge was added. In brief the mutation stage is as follows:
  - a random edge connecting two vertices is selected to be added to the tree;
  - a set of edges that lie between the two vertices is constructed; the edges in this set together with the new added edge make a cycle;
  - an edge from the cycle is selected to be deleted;
  - the new edge is selected while the old one is deleted.

III. Parallel algorithms for minimum spanning tree

A. Parallel computational model of Prim's algorithm

There are several possibilities for parallel calculations in the sequential Prim's algorithm for finding MST of a graph on a multicomputer platform:

- At each iteration of the algorithm finding the minimum weighted edge connecting a vertex not included in the spanning tree so far and a vertex that is part of the spanning tree – since the problem is solved utilizing several processors each processor can perform a local search for minimum weight edge only between part of the set of graph vertices and then by global reduction operation the global minimum is obtained. The result is the minimum weight of an edge to be added to the spanning tree at the current iteration. This edge connects two vertices – the first one is already part of the spanning tree built so far and the other one is the new added vertex. The processor that has found the minimum weight edge broadcasts the number of new added vertex to all other processors.
- Preparation for the next iteration – update of the edges of minimum weight connecting each vertex not included in the spanning tree and the vertices in the spanning tree. Each processor works on a strip of the graph adjacency matrix and calculates the updated minimum weight edges for part of the vertices out of the spanning tree.

Data decomposition is utilized for parallelization of the Prim's algorithm for finding MST. The set of vertices, from which a vertex to be added to the current spanning tree is selected, is distributed among the parallel working processes. Utilization of data parallelism has the advantage of load balancing between the processors. Each of *p* parallel working processes operates on either $\lceil n/p \rceil$ or $\lfloor n/p \rfloor$ vertices of total n vertices in the graph. The upper and lower indices of the strip of each process are easily calculated.

Parallel computational model of Prim's algorithm for multicomputer platform is based on the parallel algorithmic paradigm "manager-worker" (fig.1).

The manager process is responsible for the following activities:

- Sends the graph adjacency matrix to all worker processes (function MPI_Bcast);
- Sends the vertex number of the last vertex added to the spanning tree to all worker processes (function MPI_Bcast);
- Gathers values of the minimum weight edges to a vertex from the spanning tree locally calculated by the worker processes (function MPI_Gather);

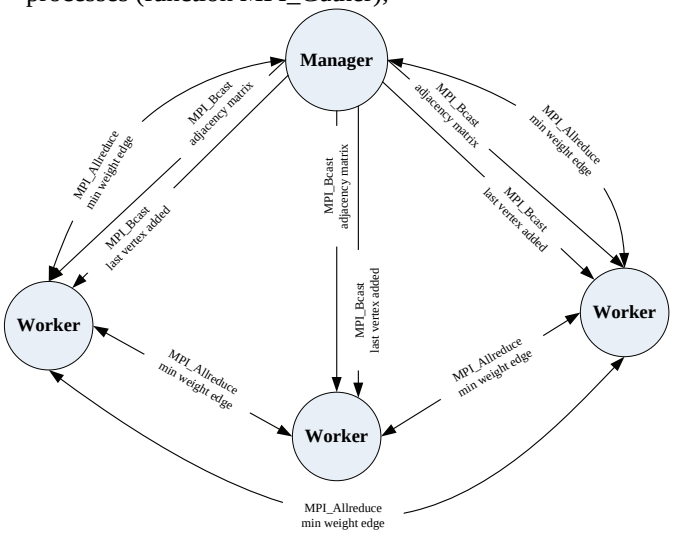


Fig.1. Parallel computational model of Prim's algorithm

- Selects a vertex to be added to the current spanning tree on the base of the results received by the worker processes.

Each worker process performs the following tasks:

- Receives the graph adjacency matrix by the manager process (function MPI_Bcast);
- Receives a vertex number of the last vertex added to the spanning tree (function MPI_Bcast);
- Working on the local set of graph's vertices, for each vertex determines an edge of a minimum weight connecting the vertex with a vertex in the spanning tree;
- Sends the calculated values to the manager process (MPI_Scatter).

B. Parallel computational model of genetic Prim's algorithm

Finding the solution of a specific problem by GA is inspired by the natural evolution and is a native parallel process that can be easily parallelized. Three main approaches of parallelizing GA exists:

- Common population, parallel calculation of the cost function of each individual by a worker process and genetic operators performed by the manager process. This approach is useful when the calculation of cost function is time consuming.
- Separation of the population into subpopulations and calculations for each subpopulation by a different process. This approach provides independent population evolution and thus higher diversification of the movements in the search space.
- Distribution of the genetic operations to different processes. This approach is an extension to the above one since genetic operators are applied only to consecutive populations and thus high diversification of the individuals is kept.

The purpose of parallelization of the GA algorithms is to reduce the time of finding the solution.

The presented genetic Prim's algorithm uses the third approach for parallelization. The crossover and mutation operators are carried out at the worker processes while initial population generation and selection are executed by the manager process.

The parallel computational model of genetic Prim's algorithm is based on parallel programming paradigm "manager-worker" (fig.2).

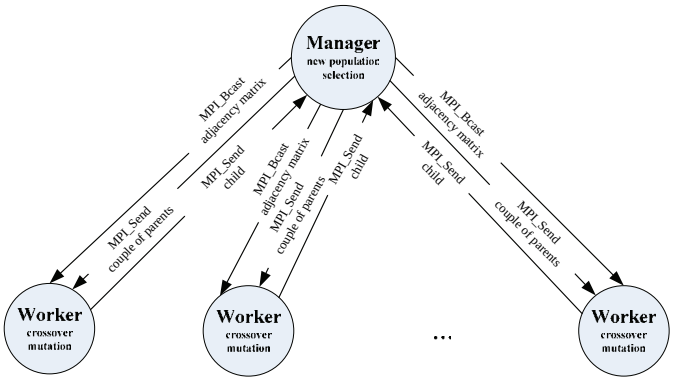


Fig.2. Parallel computational model of genetic Prim's algorithm

The manager process performs the following activities:

- Sends the graph adjacency matrix to all worker processes (function MPI_Bcast);
- Generates new population;
- Performs selection operation to select a couple of parents for a crossover operator;
- Sends the selected couple of parents to each worker process (function MPI_Send);
- Waits for a result of the crossover by any worker process (MPI_Probe);
- Receives the result from a worker process and sends to it new parents for crossover;
- Gathers all results from all worker processes for a given population
- Starts next population and repeats the above activities;
- Sends termination signals to all worker processes when the above activities are completed for all populations;

The tasks of each worker process are as follows:

- Receives the graph adjacency matrix from the manager process (function MPI_Bcast);
- Waits for a message from the manager process;
- When a message from a manager is received checks if it contains termination signal and if so then terminates;
- If the received message is not a termination signal interprets it as signal to start the crossover and mutation operators;
- Receives a couple of parents by the manager process;
- Performs the crossover of the two parents;
- Sends the result to the manager process;
- Waits for a new message;
- Performs mutation;
- Sends the result to the manager process.

B. Experimental framework

The experimental study of the parallel algorithms for finding MST described in the previous section is performed on multicomputer platform comprising 5 workstations (Intel Pentium IV 1.5GHz, 256MB RAM) connected by Fast Ethernet 100 Mbps switch. For the implementation of the algorithms Microsoft Visual Studio 2005 and MPICH-2 are used.

The speedup and efficiency of Prim's algorithm applied for finding MST of a graph with 1000 nodes are shown in fig.3 and fig.4 respectively. Gantt's diagram and communication transactions profiles of the Prim's algorithm are shown in fig.5 and fig.6 respectively.

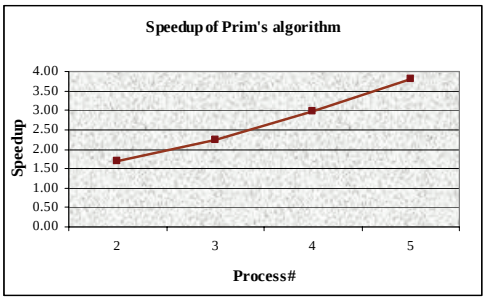


Fig.3. Speedup of parallel Prim's algorithm for finding MST on multicomputer platform

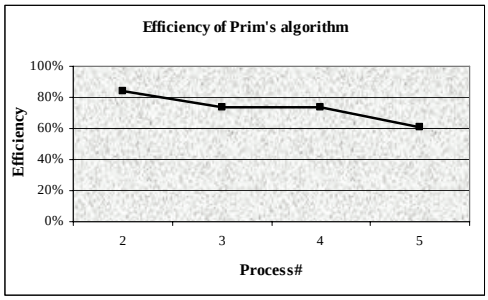


Fig.4. Efficiency of parallel Prim's algorithm

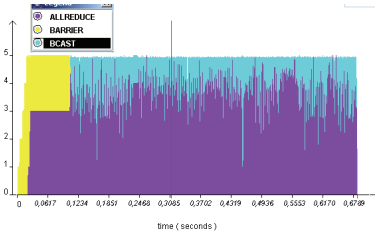


Fig.5. Communication transactions profile of parallel Prim's algorithm

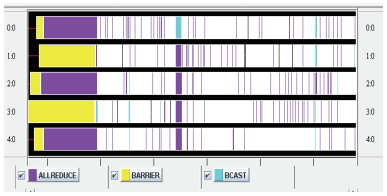


Fig.6. Gant's diagram of parallel Prim's algorithm

The speedup and efficiency of genetic Prim's algorithm applied for finding MST of a graph with 1000 nodes are shown in fig.7 and fig.8 respectively. Gantt's diagram and communication transactions profiles of the genetic Prim's algorithm are shown in fig.9 and fig.10 respectively.

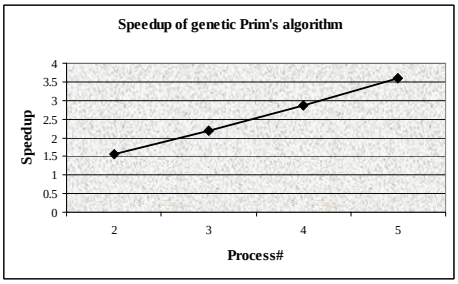


Fig.7. Speedup of parallel genetic Prim's algorithm

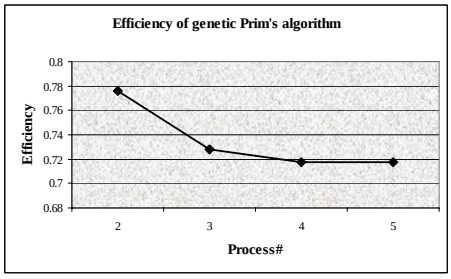


Fig.8. Efficiency of parallel genetic Prim's algorithm

The experimental results show almost linear speedup and high efficiency of the parallel calculations of MST using both Prim's and genetic Prim's algorithm.

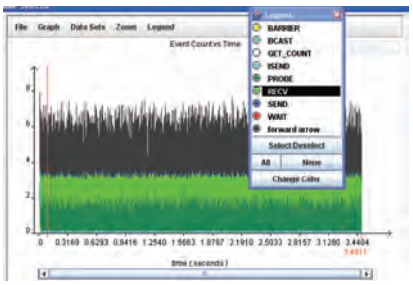


Fig.9. Communication transactions of parallel genetic Prim's algorithm

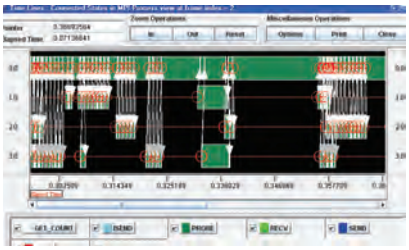


Fig.10. Gant's diagram of parallel genetic Prim's algorithm

IV. Conclusion

The paper investigates the efficiency of parallel Prim's and parallel genetic Prim's algorithms on multicomputer platform. Parallel computational models based on "manager-worker" parallel programming paradigm are presented. The algorithms are implemented using parallel MPI program. Experimental results show that both algorithm scale well with respect to the multicomputer size.

References

[1] Prim R., Shortest Connection Networks and Some Generalizations, Bell System Technical Journal, Vol.36, pp.1389÷1401, 1957.

[2] Kruskal J., On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem, Proc. of the American Mathematics Society, Vol.7, No.1, pp.48÷50, 1956.

[3] Zhou G., M. Gen, Genetic Algorithm Approach on Multi-Criteria Minimum Spanning Tree Problem, European Journal of Operational Research, Vol.114, No.1, 1999.

[4] Hanr L., Y. Wang, A Novel Genetic Algorithm for Degree-Constrained Minimum Spanning Tree Problem, IJCSNS International Journal of Computer Science and Network Security, Vol.6, No.7A, pp.50÷57, 2006.

[5] Lin L., M. Gen, Node-Based Genetic Algorithm for Communication Spanning Tree Problem, IEICE Transactions on Communications, E89-B, pp.1091÷1098, 2006.

[6] Palmer C., A. Kershenbaum, Representing Trees in Genetic Algorithms, Proc of the First IEEE Conference on Evolutionary Computation, pp.379÷384, Orlando, FL, 1994.

[7] Raidl G., B. Julstrom, A Weighted Coding in a Genetic Algorithm for the Degree-Constrained Minimum Spanning Tree Problem, 2000 ACM Symposium on Applied Computing, Como, Italy, 2000.

[8] Julstrom B., Codings and Operators in Two Genetic Algorithms for the Leaf-Constrained Minimum Spanning Tree Problem, International Journal on Applied Mathematics and Computer Science, Vol.14, No.3, pp.385÷396, 2004.

Parallel FFT Implementation on FPGA

Plamenka I. Borovska and Petyr G. Manoilov

Abstract - **The paper examines the parallel execution of the Fast Fourier Transformation (FFT) computations in a multiprocessor system with shared data memory. The multiprocessor system is designed and implemented on a FPGA – chip. Some experimental results – the speed-up of parallel FFT execution and the amount of occupied FPGA-resources for different number of processors in the system are exposed.**

Keywords **– Fast Fourier Transformation, Butterfly, Multiprocessor System, Field Programmable Gate Array.**

I. INTRODUCTION. THE FFT COMPUTATION.

The Fast Fourier Transform (FFT) [1] is essentially a computationally efficient algorithm for extracting spectral information from signal waveforms, which may be in real time or recorded form (i.e. a transformation from the time domain to the frequency domain). It is often used to dramatic effect in a growing range of applications including radar and sonar processing, speech recognition and image processing.

According to the FFT transform, the underlying computation for the spectral representation of a signal with N samples(points) can be expressed as:

$$\dot{R}_k = \dot{X}_k + \dot{Y}_k e^{-j\frac{2\pi k}{N}},\ 0 \le k \le \frac{N}{2} - 1 \qquad (1)$$

and

$$\dot{R}_k = \dot{X}_{k-N/2} - \dot{Y}_{k-N/2} e^{-j\frac{2\pi}{N}(k-N/2)},\quad \frac{N}{2} \le k \le N-1. \qquad (2)$$

Equations (1) and (2) represent the basic computational requirements for the FFT algorithm and are commonly known as the "butterfly" macro-operation due to the sample of the flow graph often used to represent the this operation, as illustrated in Figure 1. The arguments and the results in these equations are complex numbers. The circle in the butterfly graph notes a complex adding operation, when the result line is upwards and notes a complex subtracting operation, when the result line is downwards.

Plamenka I. Borovska, PhD, is with the Technical University of Sofia, Faculty of Computer Science and Control, 8 boul. "Kl.Ohridski", Sofia, 1756, Bulgaria, E-mail: pborovska@tu-sofia.bg

Petyr G. Manoilov, PhD, is with the Technical University of Sofia, Faculty of Computer Science and Control, 8 boul. "Kl.Ohridski", Sofia, 1756, Bulgaria, E-mail: p.manoilov@mail.bg.

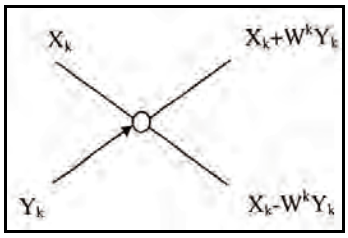


Fig.1. Flow graph, representing a "butterfly" FFT macro-operation

The arrow line notes a complex multiplying operation with multiplier (factor) $W^k = e^{-j\frac{2\pi}{N}k}$ .A single butterfly macro-operation is a FFT for a signal with 2 samples (points). We find, that it requires one complex multiplication, one complex addition and one complex subtraction. Since the multiplications and additions of the FFT algorithm are complex, the algorithm is ideally suited to handle complex input sequences. In transforming real data sequences an additional computational saving can be realized by combining two real N-point sequences into a single N-point complex sequence.

The operations on real and imaginary parts of numbers according to the butterfly of Figure 1 are shown in flow graph of Figure 2. In Figure 2 the symbols R_1, R_2, X_1, X_2, Y_1, Y_2, W_1, W_2 mean the real and imaginary parts of complex numbers $\dot{R}, \dot{X}, \dot{Y}, \dot{W}$. Using the butterflies with the appropriate factors W^k, any N- point FFT can be designed.

Figure 3 shows the flow graph of the 16- point FFT. This computational task requires execution of 32 butterfly macro-operations, arranged in four passes and each butterfly includes the operations, presented in flow graph of Figure 2 – four multiplications, three additions and three subtractions.

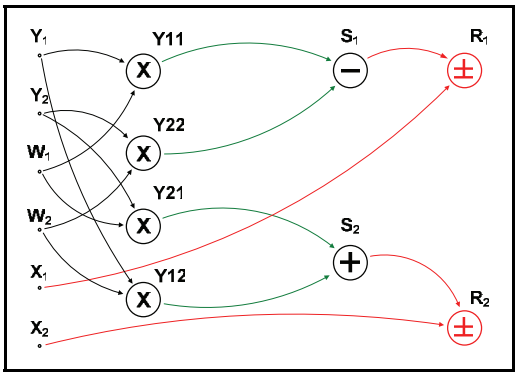


Fig.2. Flow graph, representing the "butterfly" FFT operations on real and imaginary parts of numbers

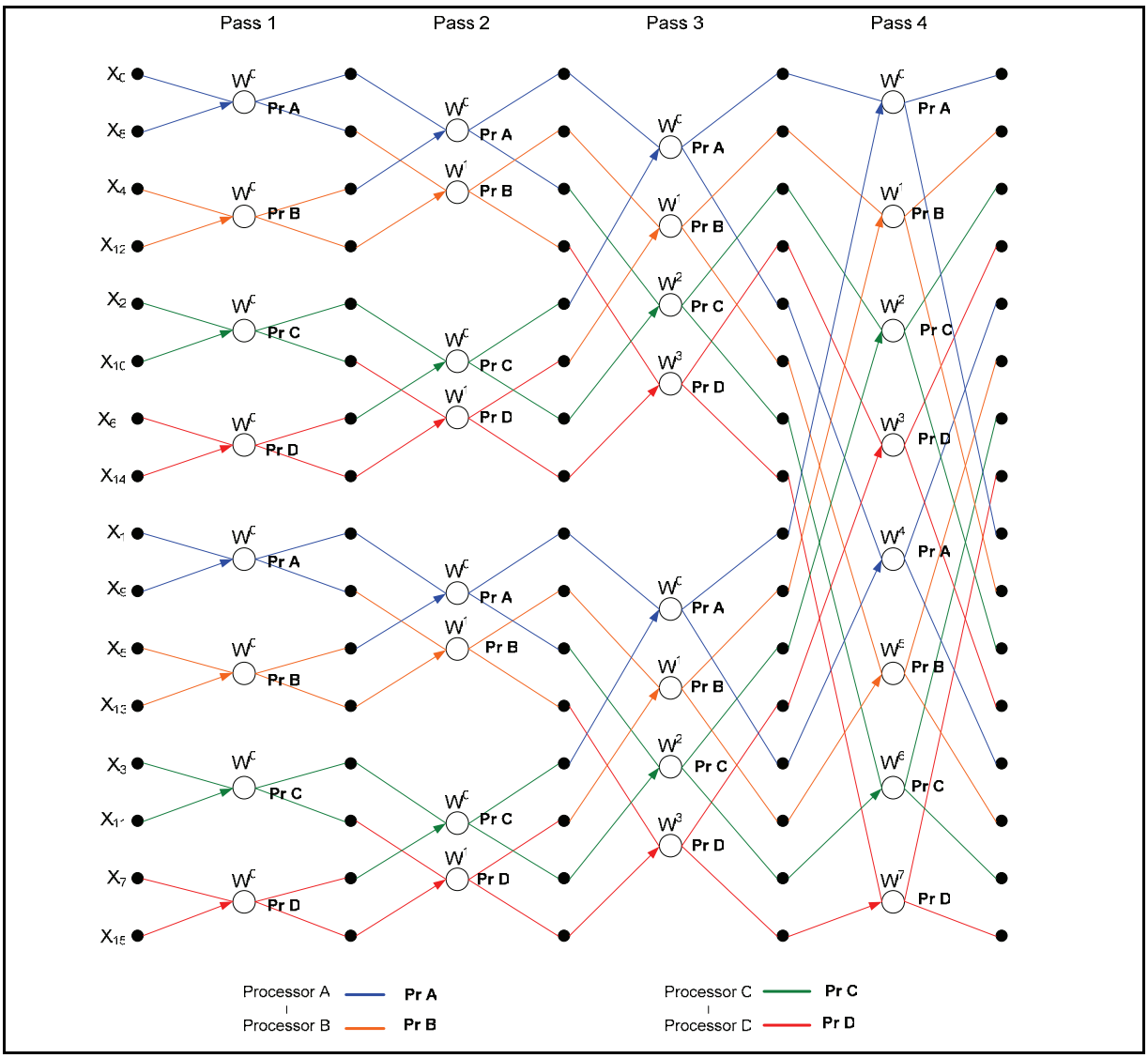


Fig.3. Flow graph of the 16-point FFT

II. Multiprocessor System, Used for Parallel FFT Implementation

The high – level block diagram of a multiprocessor system, designed and implemented on a Field Programmable Gate Array (FPGA) chip is shown on the Figure 4. The system consists of several processor modules (controller KCPSM3 + FFT accelerator) and common shared data memory, used for read and write processor operations by means of memory arbiter.

In accordance with Flinn's taxonomy this parallel computer architecture is classified in MIMD category - each processor module executes its own instructions and operates on its own data [2]. The system presents a symmetric multiprocessor (SMP), because all the data memory is in one block and has the same access time from every processor module.

The memory arbiter block contains a control logic that allows access to shared memory and solves collisions in case of simultaneous access. The arbiter accepts requests from each processor module and arbitrates which one is granted access to the shared memory at any time. Each processor module initiates a memory request signal by its I/O logic device when it wants access to the shared memory and deactivates it when finished. If more than one processor requests the memory at the same time, access is granted by using "round robin" priority access algorithm.

The block diagram of a system processor module is shown on the Figure 5. We use the KCPSM3 controller and the PicoBlaze processor (processor core) in it, designed by Xilinx. The processor core is delivered as synthesizable VHDL source code. PicoBlaze executes code (program for FFT implementation in this instance) from it's own program memory.

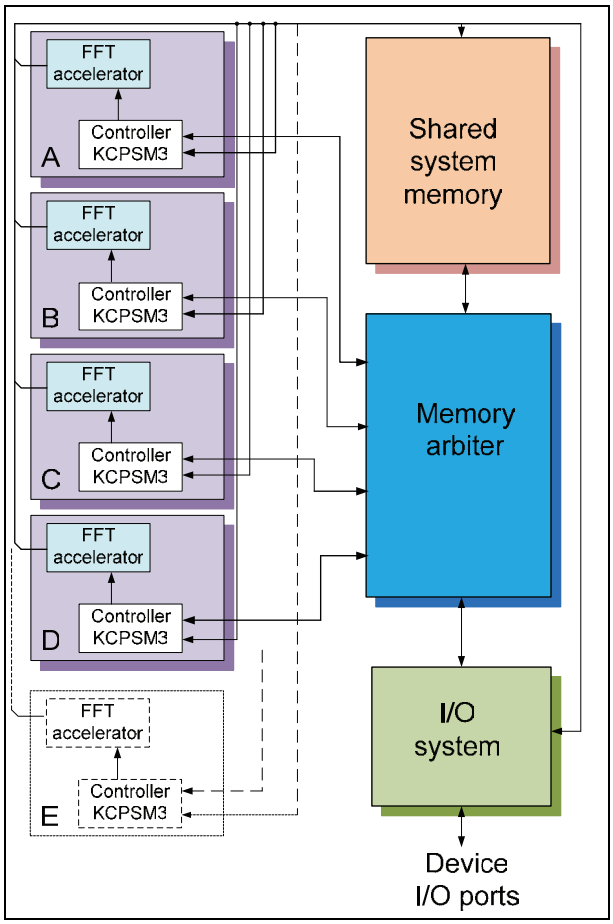


Fig.4. Block diagram of the multiprocessor system

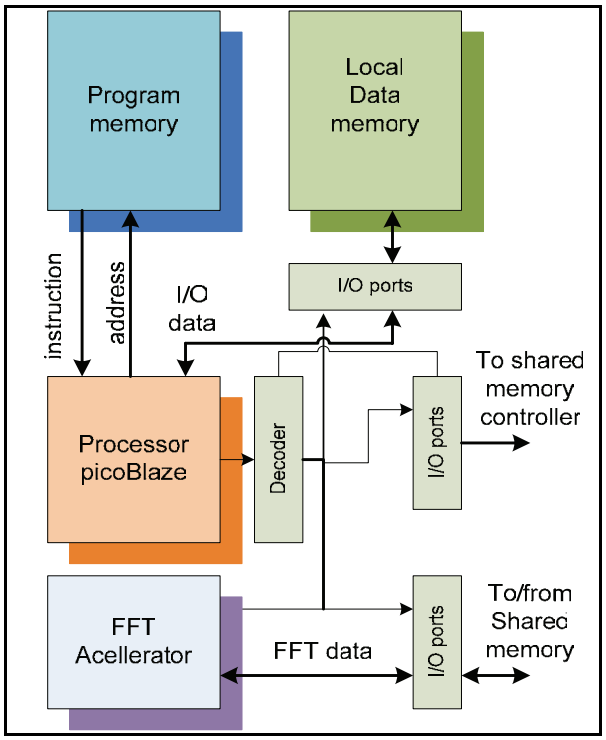


Fig.5.Block diagram of the processor module

The FFT accelerator block is a specialized coprocessor for FFT butterfly implementation. It executes the butterfly operations (shown on Figure 2) in pipeline mode with three pipeline stages [3].

Under instruction of PicoBlaze processor, the accelerator reads the butterfly operands and writes the butterfly results in the shared memory, avoiding the processor. Thus, the FFT computation speed grows. The PicoBlaze processors generate the operand and result addresses in the shared memory, only.

The organization of the parallel FFT execution is shown by the flow graph on the Figure 3. The processor modules (marked by **Pr#** on the graph) execute in parallel the butterflies of every task pass. The **barrier synchronization** is used by the multiprocessor system for this parallel execution [4]. That sort of synchronization guarantees, that no butterfly execution will proceed beyond any pass end, called the barrier, until every other butterfly execution in this pass has reached the barrier.

The multiprocessor system is designed, simulated and implemented on Xilinx FPGA – chips of Spartan3E family by means of WebPACK design environment, simulator ModelSym and VHDL source code. Figure 6 shows the chip area utilization (FPGA-chip xc3s500e) for a four-processor system.

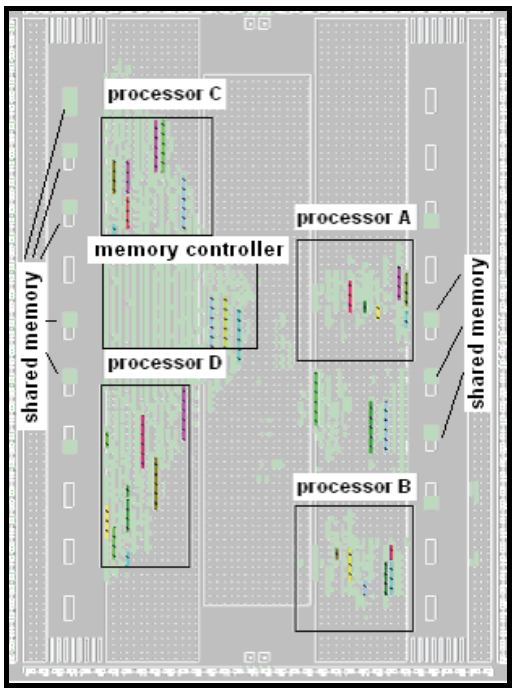


Fig.6. Four-processor system on FPGA-chip

III. Experimental results.

We have examined the parallel FFT implementation in multiprocessor systems on FPGA-chips with different number of processors. It is well-known, that the common shared memory is a “bottleneck” in shared memory systems. For that

reason, we have examined systems with different number of ports (one, two, four). Some of the important results of our investigations are shown on Figures 7-9.

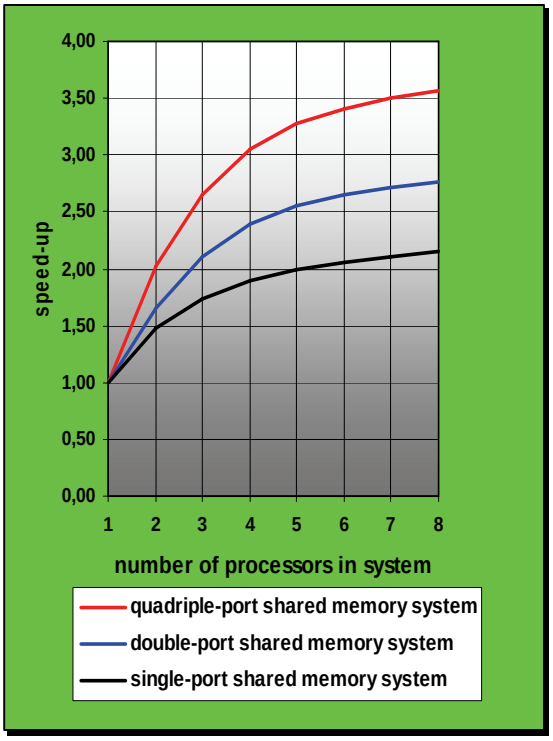


Fig.7. Multiprocessor speed-up as function of processor number and memory port number

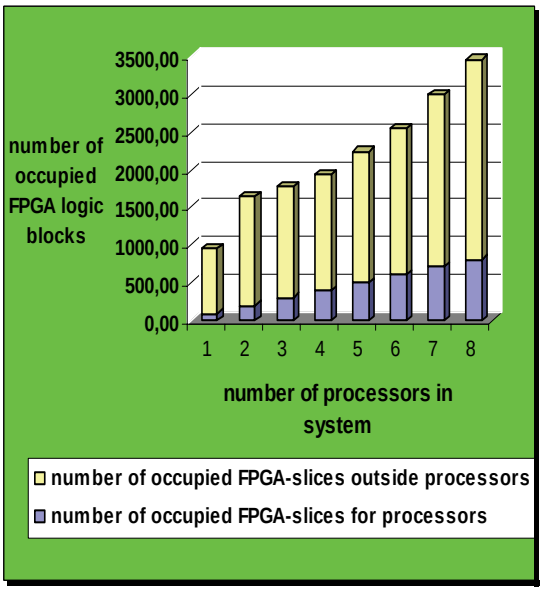


Fig.8. Number of occupied FPGA logic blocks as function of processor number

The results of our investigations are obtained by program and hardware simulation, parameter calculations, graphical and text documents of design environment working.

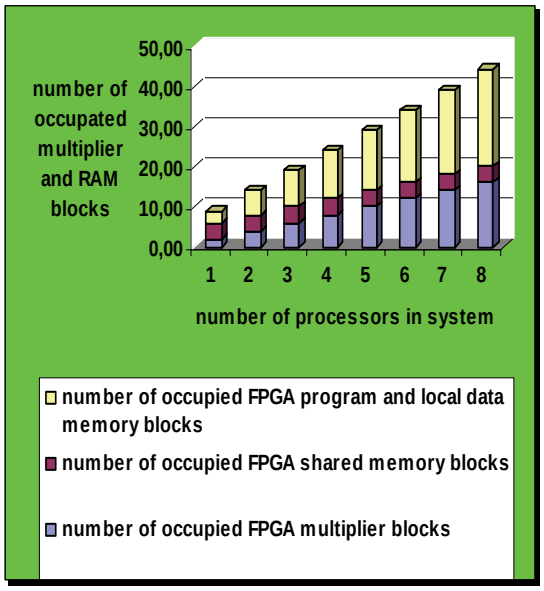


Fig.9. Number of occupied FPGA memory and multiplier blocks as function of processor number

IV. Analysis of results. Conclusion

Speed-up (the ratio between sequential FFT execution time of single processor and parallel execution time of multiprocessor) is probably the most important parameter of the multiprocessor system effectiveness. The results of Figure 7 suggest, that the processor number growth is low effective for more than 4-6 processors in the system (saturation threshold). Figure 8 and Figure 9 show, that the amount of occupied FPGA resources continue to grows, though.

In conclusion, we consider, that the results, exposed in this paper, will be useful to designers of Systems-On-Chip and Digital Signal Processing.

References

[1] Bowen B., W. Brown, “VLSI Systems Design for Digital Signal Processing”, Prentice-Hall, 1982.

[2] El-Rewini H., M. Abd-El-Barr, “Advanced Computer Architecture and Parallel Processing” John Wiley & Sons, 2005.

[3] Fleury M., A. Downton , “Pipelined Processor Farms – Structured design for embedded parallel systems”, John Wiley & Sons,2001.

[4] Kamburov G., P.Manoilov, P. Zaykov, P. Borovska, “Parallel Architecture Implemented in FPGA Based on Shared Memory System”, Third International Scientific Conference COMPUTER SCIENCE, Conference Proceedings, pp.96-101, Instanbul, Turkey, 2006.

Architecture of HASIS-3D System Designed for Hail Suppression Purposes

Vladan Mihajlović[1], Slobodanka Đorđević-Kajan[1], Dejan Rančić[1], Bratislav Predić[1], Igor Antolović[1], Predrag Eferica[1] and Zoran Babić[2]

***Abstract* – The main task of hail suppression system is efficient seeding of potentially hail containing clouds. The estimation whether the cloud is potentially hail containing or not is made by tracking the development of cloud systems. Three-dimensional Hail Suppression Information System (HASIS-3D), presented in the paper, provides tracking and measurement of specific parameters of cloud systems. Cloud tracking is enabled by efficient visualization of radar data. The data being used by the system is acquired using meteorological radars and represents cloud reflectivity for electromagnetic waves. Visualization of the data in 2D view is performed using data in polar coordinate system and representing each reflectivity level with appropriate color. Availability of different 2D view modes in HASIS-3D system allows precise analysis of clouds and more adequate estimation of hail probability. Representing clouds with iso-surfaces in 3D view allows early detection of potentially hail containing clouds by inspection of the cloud configuration and size in space. The system, also, provides tools for precise measurement of hail cells that need to be seeded with chemical reagents. The system automatically determines the parameters for launching rockets to efficiently seed the cloud with hail potential. The calculation of parameters uses the 3D iso-surfaces of the cloud that is enclosed by the points with specified reflectivity.**

***Keywords* – Hail suppression, cloud visualization, volume visualization.**

I. INTRODUCTION

Many different information systems are involved in meteorology domain, from simple ones that manage meteorological measurements to complex systems that monitor weather conditions [1]. The complex systems use Doppler's radars for data acquisition. The data represents different parameters about the atmosphere and provides various types of analyses.

In this paper, the system for hail suppression - HASIS-3D (Three-dimensional Hail Suppression Information System) is presented, as a specific information system in weather modification domain. It is developed to support hail suppression activities in Hydrometeorological Service of Republic of Serbia (RHMSS). The process of hail suppression is based on seeding the clouds with high probability with chemical reagents. It operates on the infrastructure provided by the RHMSS. HASIS-3D is complete system which uses stationary Gematronik radars for data acquisition. Gematronik radars are delivered with Rainbow 4.0 software for radar data processing [2]. This software controls the radar and provides data visualization using projections on 2D planes. Drawback of the system is lack of support for extraction of cloud area that will be seeded with chemical reagents and calculation of elements for launching rockets on cloud with high probability. Rainbow software neither provides 3D representation of the cloud. The well-known composite hail suppression system TITAN [3], also, has the mentioned shortcomings.

The system HASIS-3D is beeing developed to resolve the emphasized drawbacks and to support the methodology developed for HASIS system [4]. Criteria used in decision-making are combination of measured parameters of clouds and visual impressions that include heuristics similar to criteria presented in [5,6]. HASIS-3D offers two modes for clouds monitoring: two-dimensional (2D) and three-dimensional (3D) view. In 2D view, radar data is represented as colors from chosen color spectra. Analyzing created picture and using zooming option, the operator can detect hail cells. The system provides the possibility to create iso-contour lines in order to improve support for hail cells detection. In 3D view, operator can analyze configuration and dimensions of the clouds represented with several iso-surfaces corresponding to different reflectivity intensity. In 3D view it is possible to detect cloud with high probability in early phase of development and make more accurate decision for seeding the cloud. Using extracted iso-surfaces, the system calculates parameters for launching the rockets to optimally seed the hail cells. Hence, HASIS-3D is a complete system that supports all phases in cloud seeding process, which differentiate it from the other similar solutions in this field.

The paper is organized in four sections. The section 2 describes methodology for cloud seeding that is supported by HASIS-3D system. The architecture of the system is described in section 3. The ending section summarizes the advantages of the proposed system and specifies the interests for further research.

II. HASIS-3D METHODOLOGY

The cloud seeding methodology has been defined along with request for HASIS-3D system. HASIS-3D system

[1] Vladan Mihajlović, Slobodanka Đorđević-Kajan, Dejan Rančić, Bratislav Predić, Igor Antolović, Predrag Eferica is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: {wlada,sdjordjevic,ranca,bpredic,antol,efa}@elfak.ni.ac.yu

[2] Zoran Babić, is with the Republic Hydrometeorological Service of Serbia, Kneza Višeslava 66, 11000 Beograd, Serbia, E-mail: zbabic2003@yahoo.com

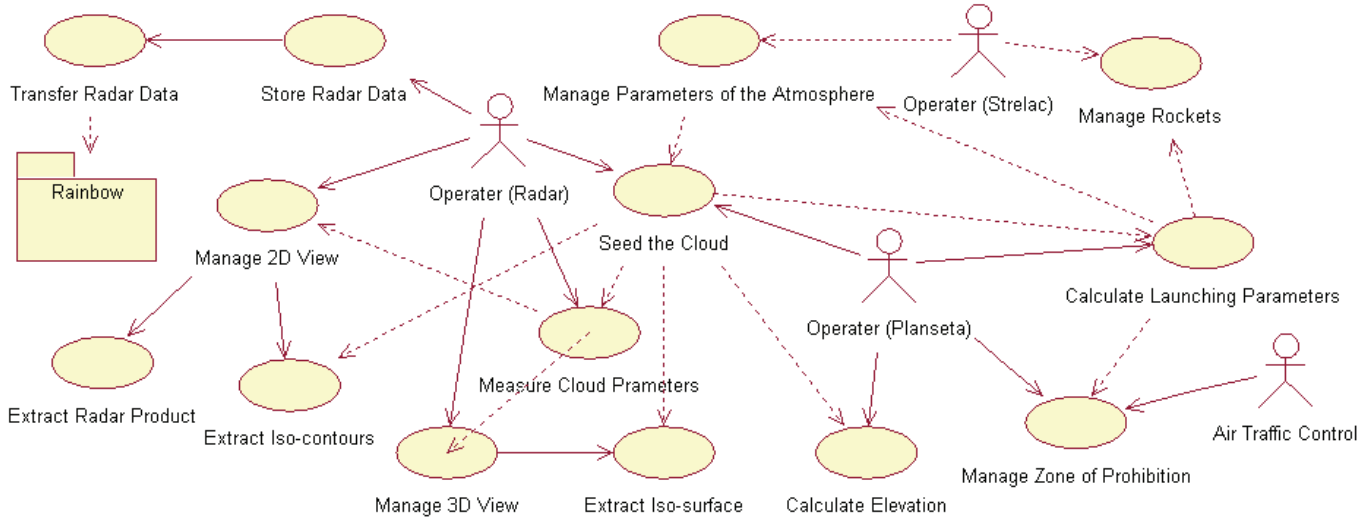


Fig. 1. Global use case diagram of HASIS-3D system

consists of the three subsystems: Main workstation, GIS workstation and Database workstation. The configuration is defined to support hail suppression methodology defined by meteorologist from RHMSS and approved in previously conducted researches [4]. Figure 1 depicts functionalities provided by HASIS-3D system.

System HASIS-3D is developed to function at the radar centers equipped with Gematronik 400S and 500S radars. The radar control is provided by the Rainbow system. It gathers the radar data and creates files with different products. The first task of HASIS-3D Main station is to acquire the product with cloud reflectivity as it is created. The next step is to reference data in coordinate system of the corresponding geographic area. After that, radar data is prepared for visualization and analysis. The analyses can be performed in 2D and 3D view. In 2D view, the cloud reflectivity is represented using the color spectra projected on the horizontal and vertical planes. To improve the cloud analysis, operator can zoom and the created radar data picture is stretched to cover corresponding geographical area. In this way, operator can more precisely detect hail cells and determine the parameters of the cells. If the cloud is estimated as potentially dangerous, operator measures the parameters either in 2D or 3D view. The system can automatically extract the iso-contour from 2D view to help operator measure cloud parameters more precisely. Even better monitoring perspective is provided in 3D view. In 3D view, operator can analyze configuration and dimensions of the clouds represented with several iso-surfaces corresponding to different reflectivity intensity (using rotation and flying through the scene). This view can help the operator to estimate the hail cells dimensions and to make decision whether to seed the cloud or no. The extraction of iso-surfaces is done automatically by the system and iso-surfaces are transferred to the GIS workstation.

The GIS workstation accepts the forwarded iso-contour or iso-surface, georeferences it and visualizes it on the geographical map in the 2D view. This workstation also provides 3D view of extracted cloud iso-surface with the terrain that is underneath the cloud. The 2D and 3D view provide information about position of rocket launching stations. The main task of this workstation is to calculate ballistic parameters for launching the rockets with chemical reagent, to optimally seed extracted iso-contour or iso-surface. The calculation is performed automatically, according to geographic location of the launching station, ballistic elements of the rockets and regions of space that are allowed for seeding. The user can adjust parameters for calculation to improve the efficiency. The only task that is not done automatically is requesting and obtaining the permission from Air-traffic control for the regions to seed. The ballistic parameters are forwarded to Database station.

The Database station stores data about meteorological situation in the atmosphere at that day and the results of seeding process. These include data about ballistic parameters of rockets, location and availability of the launching station and the rockets that they possess. Operator on this system issue the commands to each launching station using voice radio according to the ballistic parameters calculated earlier on GIS workstation.

III. HASIS-3D Architecture

This section describes architecture of the HASIS-3D system that implements the methodology described in previous section. The subsystem RADAR and PLANSETA are designed and implemented as one integral part, and during installation of the system user can chose witch of two set of functionalities to activate. The subsystem consists of six components: Observer, Processor, Extractor, Viewer, Seeder and Manager (Fig. 2).

Rainbow software that controls Gematronik radar works on *Solaris* operating system which is installed on *Sun* workstation. This workstation is delivered as bundle with the Gematronik 400S radar. Rainbow software provides visualization of basic radar canvas created from row radar data. HASIS-3D system uses only product that contains cloud reflectivity gathered by radar. HASIS-3D system works on MS Windows operating system. The optimal solution to

transfer data from Solaris to Windows workstation is to use file transfer protocol. For that purpose FTP server is set up on *Sun* workstation and Rainbow software is configured to store files with reflectivity of the atmosphere and be accessible through FTP. This solution provides independence of two systems. During scanning process radar files with products are periodically created and stored on FTP server. Product creation period depend on scanning parameters that can be defined using Rainbow software. Aside HASIS-3D system, component **Observer** is designed according to *Observer* design pattern. The basic task of the component is to detect appearance of new product on FTP server and transfer the file to RADAR workstation. Component **Observer** is developed to memorize snapshot of the contents of the FTP server after starting the system and begin transfer when new reflectivity radar product file is detected on the server.

To prevent the system to instantaneously replace the radar product that operator analyzes at the moment the component **Observer** stores up to 10 last products in circular bafer. To keep operator informed that new data are available system will visually notify him. Operator can carry on with ongoing analysis or begin analysis of new data. Except providing contiguity in analysis process, circular buffer offer operator possibility to examine the history of performed scanning. Last 10 products cover the time from 30 to 100 minute. All products are stored locally on RADAR workstation and can be

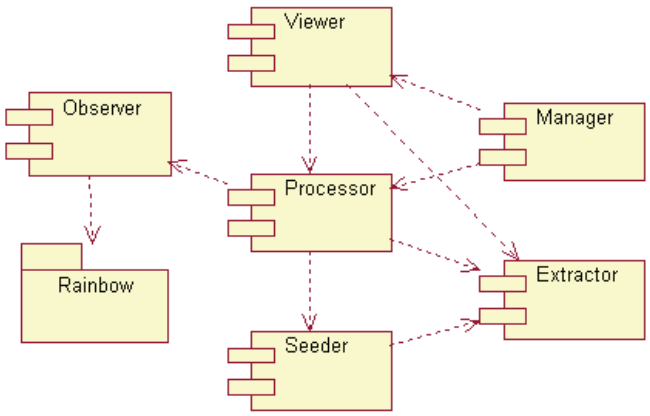


Fig. 2. Component model of RADAR and PLANSETA subsystem

loaded in the circular bafer for further analysis.

The processing of converted radar data is performed by component **Processor**. This component is specific for RADAR subsystem. The first task of this part of the system is to create products of radar canvas for visualization in 2D view. Basic product that can be created are CAPPI (*Constant Altitude Plan Position Indicator*), RHI (*Range Height Indicator*), MAX and AB. Every radar canvas product represent the cross section of atmosphere reflectivity in the planes parallel to horizontal and vertical. This component processes data in polar coordinate system to extract product of radar canvas [7]. The extraction process uses algorithms from computer graphics to achieve requested efficiency and precission. To provide better precision in visualization three-linear interpolation is used in creating products. The next task of this component is to determine parameters of the cloud. It provides, also, storage of radar product canvas, after its loading and analysis. **Processor** is the heart of the system and

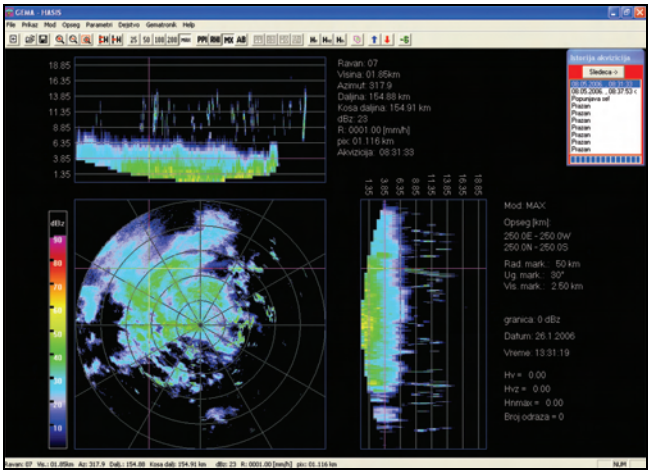


Fig. 3. 2D visualization of radar canvas in HASIS-3D system

it distributes data to all other parts of subsystem RADAR.

The **Extractor** component has two mayor tasks in automatic analysis of clouds, extraction of iso-contours and iso-surfaces. Based on the extracted iso-contours and iso-surfaces it is possible to calculate cloud's parameters that help make more precise estimation of potential hail risk. **Extractor** extracts iso-contours for the specified reflectivity value in either horizontal or vertical 2D cross section using flood fill algorithm, first, and then vectorizes it [4]. The **Extractor** uses marching cubes algorithm for iso-surface extraction [8,9]. Basic idea of this algorithm is to analyze reflectivity in 3D matrix vertices. The algorithm uses data in Descarte coordinate system, therefore system must convert row radar data from polar coordinate system. Transformation is performed in two steps. The first step is conversion from polar to cylindrical, and then from cylindrical to rectangular. The conversion uses interpolation by virtue of two nearest points. Efficiency conversion is critical condition for this component of the system because it is the most time consuming operation in the system. The major advantage of marching cubes volume visualization algorithm is high level of optimization. Therefore, algorithm use less time than other algorithms and quality is suitable for hail suppression purposes.

Component **Viewer** provides realistic representation of radar canvas. It works in three modes, Radar 2D view, Planseta 2D view and 3D view. Radar 2D view consists of three rectangular cross sections according to the selected radar canvas product (Fig. 3). The cross section data are created by **Processor**. It uses color palette defined by operator to visualize reflectivity of point in the product. The reflectivity of the atmosphere is raster layer in 2D view. Besides raster data, 2D view visualizes vector data too. The vector layers are extracted iso-contours, regions of responsibility of radar center, lakes, rivers, cities etc. Planseta 2D view visualizes two raster and three vector layers. Raster layers are georeferenced map of terrain and horizontal cross section. Vector layers are iso-contours, locations of rocket launching stations and prohibited zones. Prohibited zones are square

regions that represent units of space defined by Air-traffic control for witch permition from the Air Traffic Controll must be obtained before launching rockets for seeding. 3D view visualizes part of 3D space selected in 2D view using iso-surfaces created by **Extractor** (Fig. 4). 3D view provides also visualization of horizontal and vertical cross section selected by operator. This view includes also terrain representation extracted from real data elevation model.

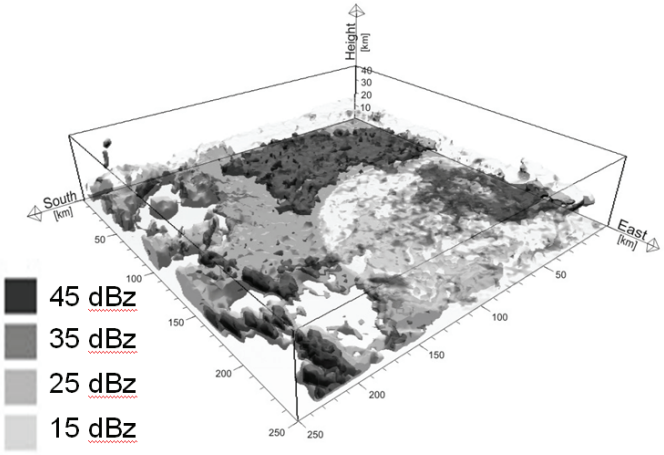


Fig. 4. 3D visualization of clouds in HASIS-3D system using iso-surfaces specified reflectivity values (15dBz, 25dBz, 35dBz, 45dBz)

Component **Seeder** is specific component for subsystem PLANSETA. It performs automatically two major tasks in cloud seeding process. The first is calculation of span of elevations for launching rockets which is effective for current parameters measured in the atmosphere. The second is to automatically determine distribution of rocket trajectories to achieve near optimal seeding of the cloud with hail potential. The result of this phase is sequence of azimuths and elevations for rocket launching stations that can seed the cloud and it is stored on subsystem STRELAC.

Component **Manger** is connection between system and operator. This component manages and synchronies work of all other components. **Manager** forwards the extracted iso-contours and iso-surfaces in subsystem RADAR and receives it in subsystem PLANSETA. The communication with database is done by this part of the system.

IV. Conclusion

HASIS-3D is a complete information system supporting tracking of cloud systems and estimating probability of hail. HASIS methodology was successfully enhanced and adapted and at the same time retained all techniques from the legacy system. HASIS-3D successfully fulfills following tasks:

1. Viewing and tracking cloud systems. System fulfills this task by providing precise and efficient visualization of radar data in two types of displays, 2D and 3D. Data in 2D display is available by creation of different types of cross-sections representing reflectivity levels of cloud system in real space. Iso-surfaces created from preset levels of reflectivity provide adequate representation of the cloud in 3D view.

2. Early detection of clouds posing hail threat. Quick conversion of data acquired from the radar and their efficient visualization shortens the time required for decision-making wheather to engage hail-threatening cloud or not. Viewing reflectivity of cloud systems at various 2D cross-sections and, more efficiently, their 3D representation based on their size, shape and cloud transformation enables recognition of clouds with high hail probability in early phases of their forming.

3. Automatization of the seeding process. Impact of operator error on the seeding process is minimized. Using iso-surfaces to calculate the parameters for launching rocket improves precision of identifying zones within cloud that need to be seeded and improve the effect of seeding the cloud.

The next step in further HASIS-3D development will be realization of virtual environment that will provide animation of three-dimensional clouds motion and rockets trajectories visualization, as well as parallelization of used algorithms to improve efficiency of the seeding process.

Acknowledgement

Research presented in this paper is partially funded by Serbian Ministry of Science and Environment Protection and Republic Hydrometeorological Service of Serbia as part of the project “Development of virtual environment for three-dimensional cloud visualization and three-dimensional cloud seeding”, ref. no. TR-6213A.

References

[1] L. A. Treinish, Visual data fusion for applications of high-resolution numerical weather prediction, *IEEE Proceedings of the conference on Visualization '00*, 2000, pp. 477 - 480.

[2] Rainbow 5.0 application software, http://www.gematronik.com/website/selex/pdf/datasheets/rainbow.pdf

[3] TITAN (Thunderstorm Identification. Tracking, Analysis, and Nowcasting) radar processing software, http://www.rap.ucar.edu/staff/vidal/titan/

[4] D. Rančić, M. Smiljanić, S. Đorđević-Kajan, A. Kostić, P. Eferica, P. Vuković, Z. Vučinić, Radar data processing for cloud seeding in hail suppression information system, *RADME 98 - Theoretical, Experimental and Operational Aspects of Radarmeteorology*, Rome, Italy, 1997, pp. 137-149.

[5] J. Johnson, *et al.*, The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm, *Weather and Forecasting 13*, 1998, pp. 263-276.

[6] T. Johnson, C. Barrere, M. D. Eilts, N. Kuhnert, M. Mathis, D. Axisa, Weather modification operations with NEXRAD level-II data and products, *Proceedings of 16th Conf. Inadvertent and Planned Weather Modification*, San Diego, USA, 2005

[7] V. Mihajlović, S. Đorđević-Kajan, D. Rančić, B. Predić, Z. Babić, 2D Cloud Visualization in Hail Suppression Information System, *YUINFO’07*, 2007, Publication on CD.

[8] D. Rančić, V. Mihajlović, B. Predić, I. Antolović, P. Eferica, A. Milosavljević, Clouds Visualization for Hail Suppression Purposes, *Proceedings of the Sixth IASTED International Conference on Visualization, Imaging and Image Processing*, Palma de Mallorca, Spain, 2006, pp.116-121.

[9] W. E. Lorensen, H. E. Cline, Marching cubes: a high resolution 3D surface construction algorithm, *Computer Graphics (SIGGRAPH '87 Proceedings)*, 1987, vol. 21(4), pp. 163-169.

Low Cost-Interactivity for 3D-Computeranimation with Computer Game-Engines

Andreas Künz[1], Miglena Dontschewa[2] and Heidi Weber[3]

Abstract **- This paper describes a project that examines possibilities to use a 3D game engine as a basis for a building-signpost in the form of a Point of Interest. The user shall virtually experience a real existing location with the help of multimedia content. Thereby the aspect of entertainment as far as the used technical possibilities of a 3D game has an important role. It is shown that the given ways of interactivity facilitate completely new approaches of using a Point of Interest. Furthermore the work shows that the technological basis of a 3D game engine also can be used for a lot of other purposes than just for playing for fun. This work was developed at the University of Applied Sciences Vorarlberg in cooperation with the Virtual Reality Laboratory of the User Centred Technologies Research Centre.**

Keywords **– 3D-Computeranimation, Game Engine, POI, Interactivity**

I. INTRODUCTION

The User Centered Technologies Research Center at the University of Applied Sciences Vorarlberg in Austria focuses especially on the topics Usability and Virtual Reality (VR). The in this article introduced project was acquired within a diploma thesis by Mag. (FH) Daniel Peherstorfer und Mag. (FH) Stefan Neururer and coached by the authors of this article. The work deals with a new interaction platform referring to results of previous works (Dontschewa& Schade, 1994; Dontschewa, Stamatov et al. 1995). It should communicate information of specific buildings in a playful way. For the prototype of the Discovery Terminal we decided to take the new building at our University as a basis. Therefore all the interior and exterior regions were created virtually and integrated in a game engine. So it's possible to explore the architecture with the use of a simple gamepad and gain information about specific rooms in a play full way.

II. METHOD

A. *The Challenge*

Within the framework of the work at hand we are examining if a 3D game engine is appropriate to make an existing, real location interactively accessible.

[1] User Centered Technologies Research Institute at the University of Applied Sciences Vorarlberg Hochschulstrasse 1, A-6850 Dornbirn ank@fhv.at
[2] User Centered Technologies Research Institute at the University of Applied Sciences Vorarlberg Hochschulstrasse 1, A-6850 Dornbirn don@fhv.at
[3] University of Applied Sciences Vorarlberg Hochschulstrasse 1, A-6850 Dornbirn hwe@fhv.at

The resulting product named „Interactive Discovery Terminal" (IDT) consists of a Windows-PC and a TFT screen that are both integrated in a robust case as seen in Figure 1.

Fig. 1: Terminal

The terminal is controlled by a customary gamepad, acoustic information is mediated by earphones. It is positioned at a frequented spot within the visualized building and allows the user to experience the University of Applied Sciences in a playful way.

B. The Game Engine

The term Game Engine is composed by the two words game and engine. The game itself consists of levels, textures, models and other game-specific elements, but is also to classify as an information system (Dimitrov&Georgieva, 2001). The engine portrays the working basis for the development of the game; it is the technology for the graphical representation in realtime, the sound environments, physical calculations and other gaming experiences. Modern game engines for titles like Half Life 2 by Valve Software or Max Payne by Remedy Entertainment are programmed in a way that is is possible to make own modifications and levels. Therefore own editors are shipped with the game that allow the user to create own models, texture them and implement them within own levels. Because the game editor for Max Payne 2 is especially powerful and makes it possible to not only model own levels but also program several different interactions and character animations this engine by Remedy Entertainment was chosen as a basis for the application.

III. WORK STEPS

A. Modeling

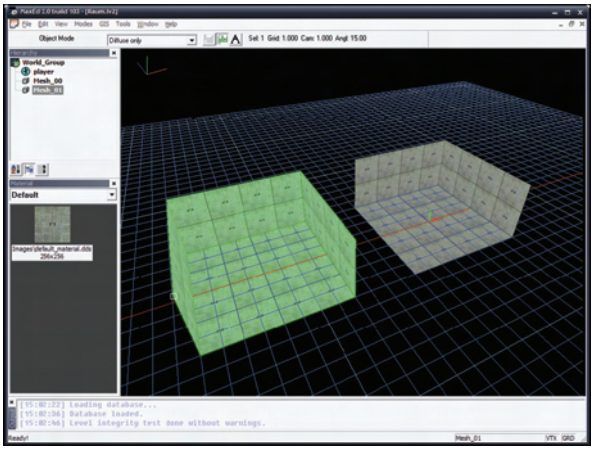


Fig. 2: Room modeling

There are several different ways to implement geometry. A special plugin allows importing complex objects from other 3D-animation software, but MaxEd – the game editor of Max Payne 2 – offers the possibility to create own simple objects directly. For an architecture project like in the work at hand only simple geometric shapes are needed. That lends itself to construct rooms, corridors, elevators and similar objects within the editor. Figure 2 shows two rooms modeled in MaxEd that are further connected by a corridor and integrated in the complete building model.

B. Texturing

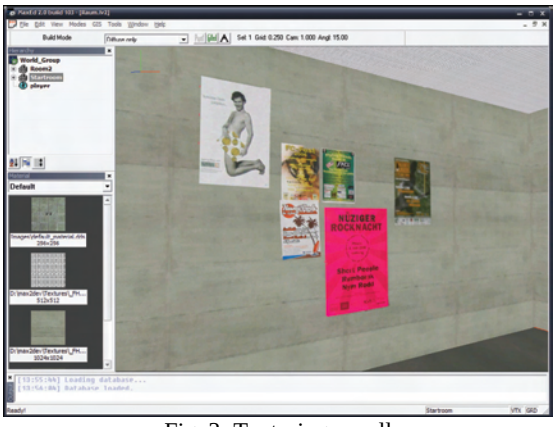


Fig. 3: Texturing a wall

The creation of textures starts with taking pictures of real existing surfaces in a most possible distortionfree way. These pictures are optimized in a photoediting software as far as color and perspective are concerned. Finally they are exported as DDS-Files including Alpha-Maps and different MipMap-channels. In the editor the prepared textures are applied to the appropriate objects. Figure 3 shows how posters are mapped onto the virtual wall. In this way all rooms and objects are provided with pictures and material properties.

C. Interaction

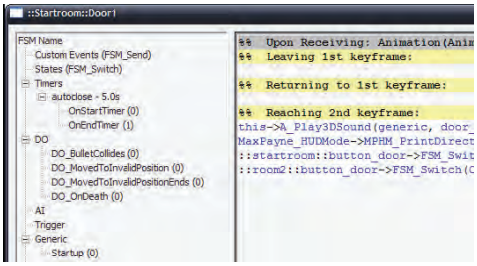


Fig. 4: Interaction by scripting

Virtual scenes can be equiped with interaction by Scripting, keyframeanimation and dynamic simulations. In that way commands can be applied to switches or objects that communiate with other elements through messages and recall predefined animations on demand (Künz, Dontschewa et al. 2005). The Havoc Physics Engine is used for collision detection and simulation of physical behaviour of the objects in the scene.

D. Lighting

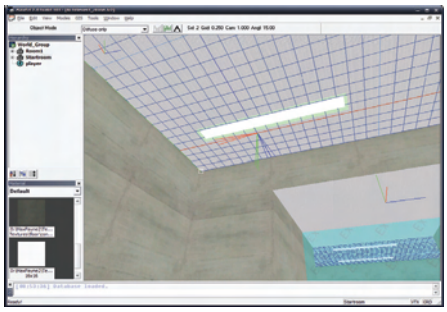


Fig. 5: Placing lights

The engine used in Max Payne 2 offers a variety of methods to visualize virtual worlds as realistic as possible. Environment Mappings make realistic reflections on surfaces possible – for example mirroring on car paint. To garantuee a good performance (Gerullis, Dontschewa et al. 1997) the lighting situation is already pre-calculated. Radiosity lights are based on the law of conversation of energy: the light that a surface receives but not absorbs is reflected. The result is saved into so called lightmaps that are multiplied by the object textures. That leads to a quite realistic presentation.

E. Characters

Fig. 6: Enrichen places with virtual humans

One of the biggest advantages of the Max Payne 2 editor is the quite simple possibility to implement human characters from an extensive library including animations. The virtual persons can walk on pre-defined paths, talk, react to the user and do specific actions. Therefore so called Artificial Intelligence Nodes are created that are combined to a network to exchange information within a level. Separate character scripts can be started either automatically or by a trigger. They send a row of commands to the characters that accomplish them in a specific order.

IV. Application

The product developed during this work is predestinated for all kinds of presentations as well as for interactive signposts or applications in museums. In all three areas architecture plays a big role. The terminal can be seen as a multimedia platform that allows visitors to explore bigger complexes of buildings and to gather information about that location without any stress and effort. Additionally to an application in big public buildings the terminal could be used for architecture projects that do not exist yet or even anymore. Especially in museums or historical diggings such a platform would offer visitors a completely new experience. They could directly walk through formerly existing buildings and explore old cities that are already destroyed. Such an interactive visualization in 3D lets the visitors understand and experience old churches, arenas, streets, etc. much better than simple paintings.

V. Discussion

Generally you may ask why the IDT was developed with the specific game engine described in this article and not with probably more appropriate software. In this connection it is necessary to elicit the alternatives for creating such interactive 3D scenarios:

- Open Source Game Engines: Most of the engines like OGRE, openSceneGraph or Blender are completely free and fully adaptable. But also quite often there is a lack of support and usability. Besides it is difficult to find a product that really implements all relevant tasks such as phsyics, shaders, scripting, characters, etc. Additionally a lot of these engine demand solid programming knowledge.
- Commercial Game Engines: software like Torque or 3DgameStudio mostly provide all the features for a high price. They also afford quite a long time to learn using all the features.
- VR-Tools: Software like Quest3D and WTK are explicitly capable of creating interactive virtual worlds. Anyhow they often have constricted graphical qualities and also demand a higher budget.
- Other Game Editors: Of course there are a lot of different games out there that ship complex editors that offer a lot of editing. Max Payne 2 was the final choice because it is cheap, flexible and easy to handle. The most important advantage is the offer of human-like characters that lead to a believable and vivid staging.
- Quicktime VR: This technology is too restricted as it only offers pre-defined views and transitions.

References

[1] Dimitrov, D., Georgieva, D. (2001) Information system for telemedicine pp.343-347 Proceedings of 12th European EAEEIE conference, 14.-15.05., Nancy, France

[2] Dontschewa, M., Schade, H.-P. (1994) Leistungsfähiges lowcost multimediales Präsentationssystem. Bestandteile, Voraussetzungen, 39.IWK, 27.-30.09.1994, Band 1, S. 384, Ilmenau, Deutschland

[3] Dontschewa, M., Stamatov, D., Schade, H.-P. (1995) Verschmelzung der Audio-, Video- und Computertechnik. Konzept und erste Ansätze multimediaunterstützten Wissenaustausches, 40.IWK, 18.-21.09.1995, Band 2, S. 90, Ilmenau, Deutschland

[4] Gerullis, J., Dontschewa, M., Schade, H.-P. (1997) Renderer für Wavefront Explore auf derBasis von SIPP, 7. Multimedia – Fachtagung, 29.-01.10.1997, Dortmund, Deutschland

[5] Künz, A., Dontschewa, M., Claudia Heitz, Kempter, G. (2005) Virtual Prototyping Computer Simulation in Information and Communication Engineering - CSICE05, Sofia, 20-22.10.2005

[6] Mullis, K. (2004) Game Development Using an Open Source Engine, Charles River Media

[7] Peherstorfer, D., Neururer, S. (2006) Interactive Discovery Terminal, Diplomarbeit

[8] Pharr, M., Randima F. (2005) GPU Gems 2, Addison Wesley

[9] Rotzoll, Fr., Dontschewa, M., Stamatov, D. (1997) „Computergrafik multimedial“, 15.-16.10.1997, Wismar, Deutschland

This page intentionally left blank.

B-tree Index Structures for Multimedia Data

Vladimir T. Dimitrov[1]

***Abstract* – There are specific requirements to Data Base Management Systems (DBMS) for multimedia data support. Conventional index structures are designed for storage and retrieval of conventional data. They have demonstrated some limitations in the case of multimedia data. B-tree index structures are more appropriate for storage and retrieval of multidimensional data.**

***Keywords* – Database Systems, Multimedia Data, Tree-Like Index, Multidimensional Data.**

I. INTRODUCTION

Conventional index structures are one dimensional, in sense that they assume a single search key, and they retrieve records that match a given search key value. There are applications that view data as existing in a 2-dimensional space, or even in higher dimensions. This kind of applications are hardly supported by conventional DBMS, instead specialized systems are designed for multidimensional applications. One important way in which these specialized systems distinguish themselves is by using data structures that support certain kind of queries that are not common in SQL applications.

Geographic Information Systems (GIS) are typical example of multidimensional applications. They store its objects (points or shapes) in two-dimensional space. These databases are maps, where the stored objects represents houses, roads, bridges, pipelines, and other physical objects.

The queries used in GIS are not typical of SQL queries, although many of them can be expressed in SQL. These queries can be classified as follow:

- **Partial match queries**. Values are specified for one or more dimensions and DBMS has to search for all points matching those values in those dimensions.
- **Range queries**. Ranges are given for one or more of the directions, and DBMS search for the set of points within those ranges. Shapes can be searched partially or wholly within the range.
- **Nearest-neighbor queries**. DBMS search for the closest point to a given point.
- **Where am I queries?** DBMS search for shapes in which a given point is located.

There are four tree-like structures useful for range queries or nearest neighbor queries on multidimensional data:

1. Multi-key indexes.
2. kd-trees.
3. Quad trees.
4. R-trees.

1-3 are intended for sets of points. The last one is commonly used to represent sets of regions, but is also useful for points.

II. MULTIPLE-KEY INDEXES

In this case, several attributes are representing the data points. A simple tree-like scheme for accessing these points is an index of indexes, or more generally a tree in which the nodes at each level are indexes for one attribute.

In Fig. 1 the idea is illustrated for the case of two attributes. The "root of the tree" is an index for the first of the two attributes. This index could be any type of conventional index, such as a B-tree or a hash table. The index associates with each of its search-key values a pointer to another index. If V is a value of the first attribute, then the index which is returned by following key V and its pointer is an index into the set of points that have V for their value in the first attribute and any value for the second attribute.

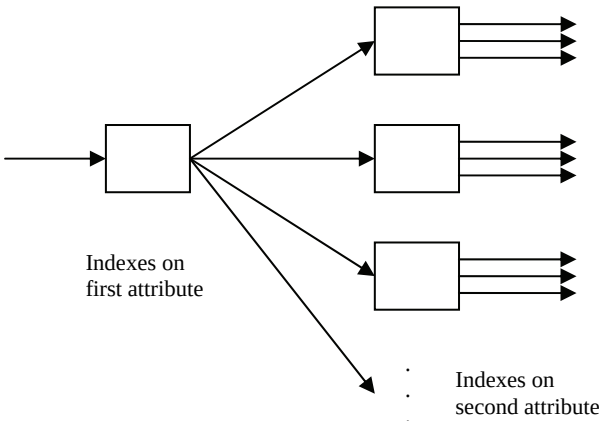


Fig. 1. Nested indexes on different keys

In a multiple-key index, some of the second or higher rank indexes may be very small. Thus, it may be appropriate to implement these indexes as simple tables that are packed several to a block.

Partial-match queries. If the first attribute is specified, then the access is quite efficient, all what have to be done is to find the one subindex that leads to the desired points. If the root is a B-tree index, then two or three disk I/O's have to be done to get the proper subindex and then to use whatever I/O's are needed to access all of that index and the points of

[1]Vladimir T. Dimitrov is associate professor at the Faculty of Technical Sciences, University of Sofia, POB 1829, 1000 Sofia, Bulgaria, e-mail: cht@fmi.uni-sofia.bg

the data file itself. If the first attribute does not have a specified value, then every subindex have to be searched, a potentially time-consuming process.

Range queries. The multiple-key index works quite well for a range query provided the individual indexes themselves support range queries on their attribute. To answer a range query, root index has to be used and the range of the first attribute to find all of the subindexes that might contain answer points. Then each of these subindexes has to be searched using the range specified for the second attribute.

Nearest-neighbor queries. To find the nearest neighbor of point (x_0, y_0), a distance d has to be find first, such that several points are expected to be within distance d of point (x_0, y_0). After that the range query can be applied: $x_0 - d \le x \le x_0 + d$ and $y_0 - d \le y \le y_0 + d$. If there are no points in this range, or if there is a point, but distance from (x_0, y_0) of the closest point is greater than d, then the range has to be increased and search repeated. Search can be ordered so that the closest places are searched first.

III. KD-TREES

kd-tree is generalization of the binary search tree to multidimensional data. It is a main-memory data structure. A kd-tree is a binary tree in which interior nodes have an associated attribute a and a value V that splits that data points into two parts: those with a-value less than V and those with a-value equal or greater than V. The attributes at different levels of the tree are different, with levels rotating among the attributes of all dimensions. In the classical kd-tree, the data points are placed at the nodes, but for the sake of block model of storage two modifications are done:

1. Interior nodes have only an attribute, a dividing value for that attribute, and pointers to left and right children.
2. Leaves are blocks, with space for as many records as a block can hold.

An example of kd-tree is presented in Fig. 2. There are two dimensions (Speed and Age), which are alternatively splitting data points at each level. Leaves contain data points and they can be placed at each level of the kd-tree.

For the lookup values for all dimensions could be given. A lookup of a tuple is performed as in a binary search tree. At every interior node the search is redirected to a subtree which is possible to contain a leaf with the tuple.

Insertion starts as lookup to find the leaf. If the block of the leaf has enough room – the new data point is inserted. If there is no room, the block is divided into two new blocks. Content of the block is distributed into the new two blocks using the attribute corresponding to the leaf level. New interior node is created whose children are the two new blocks. In the new interior node the splitting value is put.

Partial-match queries. If values are given for some of the attributes, then at every level belonging to attribute whose value is known search direction is clear. If there is no value for the attribute at a node, then both its children have to be explored.

Range queries. Sometimes, a range will allow search direction to be directed only to one child of a node, but if the range straddles the splitting value at the node, then both children have to be explored.

Nearest-neighbor queries. They are executed in the same way as in the case of multiple-key indexes.

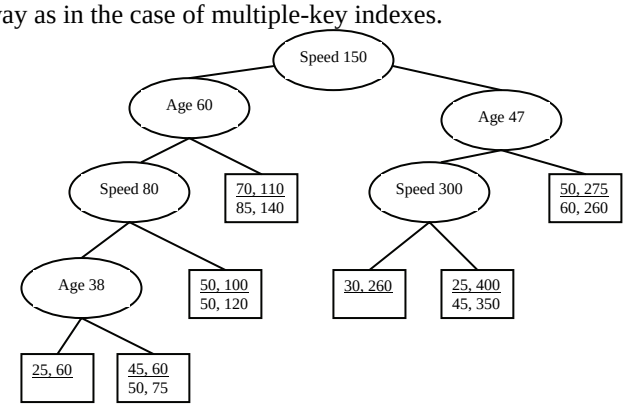


Fig. 2. A kd-tree with two dimensions

Adapting kd-trees to secondary storage. Let kd-tree with n leaves is stored in a file. Then the average length of a path from the root to a leaf will be about log2 n, as for any binary tree. If each node is stored in a block, then to traverse a path one disk I/O per node must be done, which in summary is more than for the typical B-tree. In addition, since interior nodes of kd-tree have relatively little information, most of the block would be wasted space. The twin problems of long paths and unused space cannot be solved completely, but there are two approaches that will make some improvements in performance:

- **Multiway branches at interior nodes**. Interior nodes of a kd-tree could look more like B-tree nodes with many key-pointers pairs. If a node contains n keys, then values of an attribute could be split into n + 1 ranges. If there are n + 1 pointers, then could be followed appropriate one to a subtree that contain only points with attribute in that range. Problems are when reorganization of the nodes has to be done, in order to keep distribution and balance.
- **Group interior nodes into blocks**. In this approach the tree nodes have only two children, but many interior nodes are packed into single block. In order to minimize the number of block that have to be read from disk while traveling down one path, the best is to include in one block a node and all its descendants for some number of levels. That way, once the block with this node is retrieved, it is possible to use some additional nodes on the same block, saving disk I/O's.

IV. QUAD TREES

In quad tree, each interior node corresponds to a square region in two dimensions, or to a k-dimensional cube in k dimensions. If the number of points is no larger than what will fit in a block, then this square is a leaf of the tree, and it is represented by the block that hold its points. If there are too many points to fit in one block, then the square is an interior

node, with children corresponding to its four quadrants. See Fig. 3 for an example of quad tree.

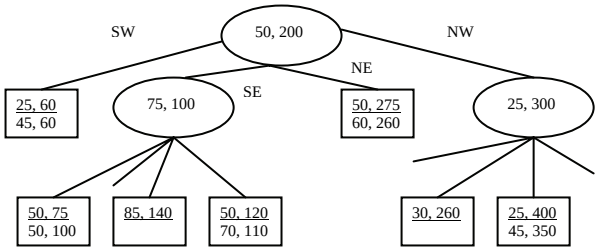


Fig.3. A quad tree

Since interior nodes of a quad tree in k dimensions have 2k children, there is a range of k where nodes fit conveniently into blocks. However, for the 2-dimensional case, the situation is not much better than for the kd-tree; an interior node has four children. In quad tree splitting point for a node has to be the centre of a quad-tree region, which may or may not divide the point in that region evenly. When the number of dimensions is large many null pointers in the interior node could be found. In this case, only non-null pointers can be represented.

Standard operations on quad tree resemble those for kd-tree.

V. R-TREES

An R-tree (region tree) is a data structure that captures some of the spirit of a B-tree for multidimensional data. B-tree node has a set of keys that divide a line into segments. Points along that line belong to only one segment and it is easy to determine a unique child of that node where the point could be found.

An R-tree represents data that consists of 2-dimensional or higher-dimensional regions, which are called data regions. An **interior node** of an R-tree corresponds to some interior region, which not normally a data region. The region can be of any shape, but in practice it is usually rectangle or other simple shape. The R-tree node has, in place of keys, subregions that represent the contents of its children. The subregions are not needed to cover entire region, which satisfactory as long as all the data regions that lie within the region are wholly contained within one of the subregions. The subregions are allowed to overlap, although it is desirable to keep the overlap small. An R-tree for a map is presented in Fig. 4.

R-tree is useful for “where-am-I” queries, which specify a point P and asks for the data regions in which the point lies. Search starts from the root, with which the entire region is associated. The subregions at the root are examined to determine which children of the root correspond to interior regions that contain point P.

If there are zero regions, then P is not any data region. If there is at least one interior region that contains P, then P must be recursively searched at the child corresponding to each such region. When one or more leaves are reached, then actual data regions should be found or a pointer to that record.

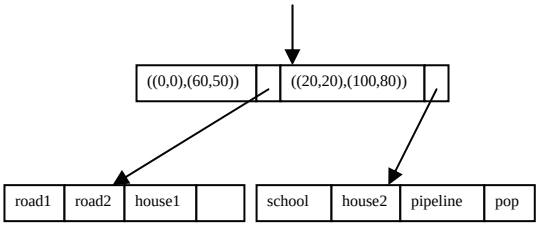


Fig. 4. An R-tree for a map

When new region R into R-tree should be inserted, procedure starts from the root and looks subregion into which R fits. If there is more than one such region, then one of them is picked and process is repeated there. If there is no subregion that contains R, then one of the subregions has to be expanded. Which one to pick may be difficult decision. Idea is to expand regions as little as possible, children’s subregions have to be asked to increase their area as little as possible, change the boundary of that region to include R, and recursively insert R at the corresponding child.

Eventually, a leaf is reached where region R is inserted. If there is no room for R at that leaf, then the leaf must be divided. The new two subregions have to be as small as possible, but they have to cover all the data regions of the original leaf. Having split the leaf, the region is replaced and pointer for the original leaf at the node above is replaced too by a pair of regions and pointers corresponding to the two new leaves. If there is a room in the parent, process is finished. Otherwise, recursively nodes are divided going up the tree.

VI. BITMAP INDEXES

In this index structure file records have permanent numbers and there is some data structure that is used easy to find record by their numbers. Such a structure can be a conventional index on the data file, but in the data file there is no need to support key field (the record number).

A **bitmap index** for a field F is a collection of bit-vectors of length n, one for each possible value that may appear in the field F. The vector for value v has 1 in position i if the record with that number has v in field F, and has 0 there if not.

Bitmap indexes require too much space, when there are many different values for a field, since the total number of bits is the product of the number of records and the number of values. That is why, compression has to be used.

Bitmap indexes have management problems, but they answer partial-match queries very efficiently in many situations. They can also help answer range queries. These kind of queries are supported by using AND/OR operations with bitmap-vectors.

Let bitmap index on field F of a file has n records, and there are m different values for field F that appear in the file. Then the number of bits in all the bit-vectors for this index is mn. This number can be small compared to the size of the file itself, but the larger m is, the more space the bitmap index takes.

But if m is large, then 1’s in a bit-vector will be rare - the probability that any bit is one is 1/m. If 1’s are rare, then bit-

vectors can be encoded to take much fewer than n bits on the average. **Run-length encoding** approach is encoding sequence of i 0's followed by a 1 (**a run**), by some suitable encoding of the integer i. The codes for each run are concatenated together, and that sequence of bits is the encoding of the entire bit-vector.

Representing the integer i as a binary number do not work, because there is no way to divide the concatenated encoded bit-vector on runs. So, the encoding of run length must be more complex. There are many encoding schemas, but they work well only when typical runs are very long. Such a schema uses the number of bits j needed to represent in binary the number i. This number j is approximately $\log_2 i$, and is represented by j - 1 1's and a single 0.

Concatenated sequence of above mentioned codes easy can be divided into j codes and then original vector can be recovered. For that purpose are used 0's as separators – every j-code is a sequence of 1's ending with 0.

Every bit-vector so decoded will end in 1, and any trailing 0's will not be recovered, but the number of records in the file is known, so additional 0's can be added. Since 0 in a bit-vector indicates the corresponding record is not in the described set, then trailing 0's can be ignored.

Let m = n, i.e., each value for the filed on which bitmap index is constructed, has a unique value. The code for a run of a length i have about $2\log_2 i$ bits. If each bit-vector has a single 1, then it has a single run, and the length of that run cannot be longer than n. Thus, $2\log_2 n$ bits are an upper bound on the length of a bit-vector's code in this case. Since there are n bit-vectors in the index (m = n), the total number of bits to represent the index is at most $2\log_2 n$. Without the encoding n^2 bits would be required.

Because only one run can be decoded at a time, operations (AND/OR) on bitmap-vectors can be interleaved with decoding at runtime.

Bit-vectors can be indexed with secondary indexes on values of the vector field, but instead value – a pointer to the bit-vector can be used. Bit-vectors can be packed in blocks of index data file, but if they are very long they can cross block boundaries in a chain of blocks.

There are two aspects to the problem reflecting data-file modifications in a bitmap index:

1. Record numbers must remain fixed once assigned.
2. Changes to the data file require the bit map index to change as well.

When a record is deleted its number is not used again and in the data file a "tombstone" is put. This is a consequence of point 1. The bitmap index must also be changed – for record position in all bit-vectors 0 has to be put.

When a new record is inserted, the next number is assigned to it. This means that next available number has to be supported persistently. In all bit-vectors at the end 1 has to be added where this is appropriate. Technically, 0 must be added too, but this operation can be postponed till next insertion of new record with a value in the corresponding field.

Modification a record will influence bit-vectors – value of 1 has to be changed to 0 and vice versa where is appropriate. Bit-vectors for new values have to be created, as in the case of insertion of new records.

VII. Conclusion

Above presented data structure are goods tools for effective storage and retrieve of persistent multimedia data. Some of them are used in currently available DBMS, but others are used only in specialized systems. Our intension is to extent an open source DBMS like MySQL with some of these indexed data structure for the support effectively and efficiently search and retrieval of multimedia data.

These data structures are at the physical level of an implementation of multimedia DBMS, but more research has to be done at higher levels: query language, query compilation and execution.

Acknowledgement

Presented overview is funded under contract VU-N-202/2006 "Multimodal biometric analyses: Methods and algorithms" with Bulgarian National Fund for Scientific Research.

References

[1] N. Beckmann, H.-P. Kriegel, R. Schneider, B. Seeger, The R*-tree: an efficient and robust access method for points and rectangles. Proc. ACM SIGMOD Intl. Conf. on Management of Data, 1990.

[2] J. L. Bentley, Multidimensional binary search trees used for associative searching, Comm. ACM 18:9, 1975.

[3] J. L. Bentley., Multidimensional binary search trees in database applications, IEEE Trans. On Software Engineering SE-5:4, 1979.

[4] J. L. Bentley, J. H. Friedman, Data structures for range searching, Computing Surveys 13:3, 1979.

[5] W. A. Burkhard, Hashing and tree algorithms for partial match retrieval, ACM Trans. on Database Systems 1:2, 1976.

[6] R.A. Finkel, J. L. Bentley, Quad trees, a data structure for retrieval on composite keys, Acta Informatica 4:1, 1974.

[7] A. Guttman, R-trees: a dynamic index structure for special searching, Proc. ACM SIGMOD Intl. Conf. on Management of Data, 1974.

[8] J. T. Robinson, The K-D-B-tree: a search structure for large multidimensional dynamic indexes, Proc. of ACM SIGMOD Intl. Conf. on Management of Data, 1981.

[9] T. K. Sellis, N. Roussopoulos, C. Faloutsos, The R+-tree: a dynamic index for multidimensional objects, Proc. Intl. Conf on Very Large Databases, 1987.

[10] J. B. Rothnie Jr., T. Lozano, Attribute based file organization in a paged memory environment, Comm. ACM 17:2, 1974.

SESSION TST IV

Telecommunications Systems and Technology IV

Balancing the Reverse Path in CATV Systems

Dobri M. Dobrev[1] and Lidia T. Jordanova[2]

***Abstract* – The paper deals with some problems in balancing the reverse path of HFC CATV systems and the methods to solve them. Criteria are given to determine the parameters of the reference signal fed to the input of each optical node and amplifier in the reverse path. The sequence of balancing operations is described. A special attention is drawn to the design of the subscribers' distribution network and to the equalization of the signal levels in the subscribers' cable modems at the drop amplifier input in particular. Two methods to set and maintain the balance of the reverse path are described that make it possible for some of the signal characteristics to be monitored.**

***Keywords* – HFC CATV, Reverse Path, Unity Gain, BER, laser clipping, drop equalizer.**

I. INTRODUCTION

CATV systems make it possible for unlimited number of subscribers to access different services that cab be divided into two general groups: commonly provided and additional. Cable television and broadcasting are associated with the first one, while the second one includes high-speed Internet, on-demand multimedia, voice and phone communications, domestic and office alarm systems etc. Modern CATV systems are of the hybrid type: they implement fiber optics for the trunk rings and sub-trunk lines and coaxial cables for the peripheral branches.

One of the main causes to worsen communication quality over the reverse path of CATV systems is the funnel effect due to the tree-and-branch topology of the cable distribution networks. As a result, noise and inter-modulation products from all the network branches interfere with the signal in the subscribers' cable modems. Thus, the value of the bit error ratio (BER) at the receiver output of the cable modem terminal system (CMTS) is increased and communications over the reverse path get worse or cut-off. The influence of the funnel effect can be reduced significantly through optimization of the cable distribution network's topology and of the RF signals' dynamic range at the modulation input of the reverse path lasers or through appropriate filtering [1]. Another way to estimate the operability of the reverse path is to check the level of the digitally modulated RF carrier at the receiver input of the CMTS. As referred in CMTS specifications, a typical value for the cable modems is 0 dBmV. Reverse path balancing can be applied to provide that level, thus avoiding any undesirable worsening of BER in the optical transmitters of the reverse path.

When balancing the reverse path some difficulties can be met because of the different level of the signals that come from the subscribers' cable modems and are passed to the input of the reverse path amplifiers. This is due to the unequal paths passed by the signals, i.e. to the different attenuation along the coaxial cable and the passive devices included (splitters and taps). Though the level of the radio pulse transmitted by the cable modem is automatically controlled in the CMTS (within the limits of 92 dBμV to 112 dBμV) no alignment of the signals' levels is possible, so variations within 6 dB are quite admissible. Hence, when designing the reverse path it is of great importance to achieve a level equalization of the subscribers' signals at the input of the drop amplifier.

II. THE CONCEPT OF UNITY GAIN

The idea is based on the requirement that the loss between each amplifier section equals the gain in that section, i.e. a gain of 0 dB. Not meeting the requirement means that the signals of some sections when passed to the optical transmitter input of the reverse path will differ in level from the necessary one. This will cause an inadmissible drop in the carrier-to-signal noise ratio (CNR) and in the ratios of the carrier-to-aggregate distortion signals such as composite second order (CSO) and composite triple bit (CTB) at the receiver input of the CMTS which will result in an increase of BER.

Balancing the forward path causes no problems because all the signals are of the same level as aligned in the headend. The aim is to obtain one and the same signal level at the cable amplifiers' inputs of the coaxial line. To this aim an attenuator and an equalizer are used in the amplifiers' input to compensate for temperature and frequency effect on the coaxial cable attenuation. The optimal signal level in the forward path amplifiers can be calculated when given the necessary values of the following parameters: CNR, CSO and CTB at the subscriber's outlet, number of RF channels transferred, parameters of the amplifier chosen, number of amplifiers implied in each coaxial line [2].

The input of each amplifier and optical node in the upstream direction is used as a check point to measure signals' levels of the reverse path. Alignment begins at the return path transmitter/amplifier closest to the headend – typically a fiber optic node – and aims at obtaining a reference signal level that is preliminary set. To this aim either the transmission coefficient of the preceding optical node or the attenuation of the attenuator and the slope of the equalizer characteristics is adjusted at the output of the preceding amplifier.

[1]Dobri M. Dobrev is with the Faculty of Communications and Communications Technologies, 1756 Sofia, Bulgaria, e-mail: dobrev@tu-sofia.bg

[2]Lidia T. Jordanova is with the Faculty of Communications and Communications Technologies, 1756 Sofia, Bulgaria, e-mail: jordanova@tu-sofia.bg

When choosing the reference input level a consideration must be made for optical and RF noise floors as well as clipping due to overdrive. A typical manufacturer's specification for optimum video input level to the return laser is 20 dBmV. Modern sweep systems are designed to operate 10 dB (or more) below the optimum carrier level, i.e. at a reference level of 10 dBmV. CATV systems are designed to operate at a reference level of 15 dBmV (75 dBμV), that must be set both in the input of each optical node and cable amplifier of the reverse path and in the headend [3]. Level variations within ± 2 dB (i.e. variations from 73 dBμV to 77 dBμV) are admissible.

The test signal frequency must correspond to the operation frequency of the cable modems, a frequency of 23 MHz to 25 MHz being the ideal one. In systems that employ a very wide reverse path (5… 65 MHz) it may be better to utilize two test frequencies, one at 25 MHz and the other at 50 MHz to properly correct for even the reverse path slope.

III. Level equalization of the upstream signals at the drop amplifier input

High quality of the received CATV signals requires a level of 70 ± 5 dBμV at the subscriber's tap. Appropriate values of the tap attenuation in each subscriber's tap must be chosen as shown on fig. 1. As far as the reverse path is concerned losses in both the coaxial cable and passive devices of the subscribers' distribution network are much less than those of the forward path. On the diagram one can see two digits, the first one referring to signals in the reverse path and the second one referring to signals in the forward path. It is evident that with tap attenuation thus distributed the signals of the subscribers' modems will be of different level at the subscriber's amplifier input the highest value corresponding to the most-distant homes. This would make balancing the reverse path quite difficult.

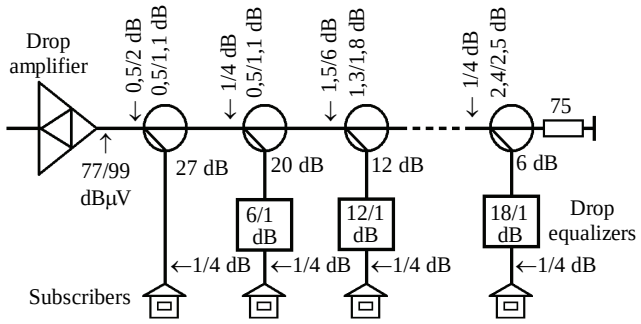


Fig. 1.

To compensate for signal level differences in the upstream direction drop equalizers of chosen characteristics must be implemented in the subscribers' taps. When choosing the value of attenuation some considerations must be made referring to the subscribers' distribution network, attenuation in the coaxial cable and in passive devices, the level of signals transmitted to home terminals etc. The aim is to make all signal levels coming to the subscriber's amplifier from all the subscribers' cable modems equal to the reference value (75 ± 2 dBμV). In the case a reference level of 77 dBμV is applied and drop equalizers of attenuation of 6, 12 and 18 dB are used.

Drop equalizers provide selectivity of attenuation at a chosen step within the frequency band of the reverse paths thus making possible the inter-active optimization of the system design. They help suppressing the ingress within the reverse path. Another advantage is that they enable the subscribers' cable modems to transmit signals of higher levels, hence to improve the CNR. Drop equalizers compress the dynamic range of the signals passed to the drop amplifier for the reverse path. As a result signal limiting is avoided and isolation between subscribers is improved.

IV. Algorithm for balancing the reverse path

The algorithm for balancing the reverse path can be explained by means of fig. 2 where a section of the cable distribution network of a CATV system is shown. Its optical part has a tree-and-branch topology, where signals at some points of the optical trunk line branch out towards optical nodes to feed the local area coaxial distribution networks.

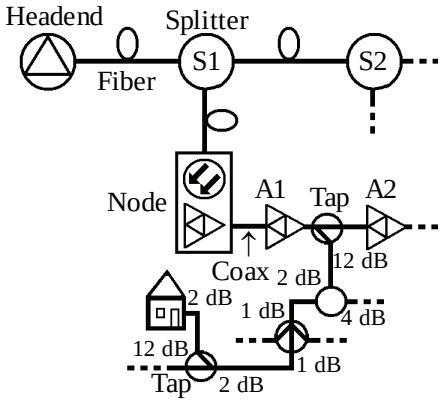


Fig. 2.

The procedure of balancing starts from the optical node that is closest to the headend. A test signal (75 dBμV and frequency about 25 MHz) is fed to the modulation input of the return laser module. Its gain is adjusted in a way to make the level of the RF signal at the headend input equal to the reference level. A spectrum analyzer or a sweep receiver in the range of 5 MHz to 65 MHz can be used to monitor the test signal in the headend. After balancing the reverse path section of the first optical node the test generator is injected into the next optical node input port and the procedure is repeated.

Balancing the coaxial sections of the reverse path can not start before gain of all optical nodes is set to the desirable value. Initially the test signal is fed to the input of amplifier A1 in the first local coaxial network that is closest to the optical node. The test signal level at the headend input is made equal to the reference one (75 dBμV), the attenuator and the equalizer connected in its output being properly adjusted. After balancing the section of the first amplifier the test generator is injected into the next amplifier input port A2 and the procedure is repeated, etc.

V. METHODS TO SET AND SUPPORT BALANCE IN THE REVERSE PATH OF A CATV SYSTEM

The above algorithm for balancing the reverse path requires co-operation of two operators: one to adjust the corresponding optical node or amplifier and the other to monitor the signal level in the headend and to communicate with the first operator (over radio or wire channel). This makes it necessary for more effective methods of balancing to be applied during the whole lifetime of the system. In the paper two methods are considered as illustrated in fig. 3 and fig. 4.

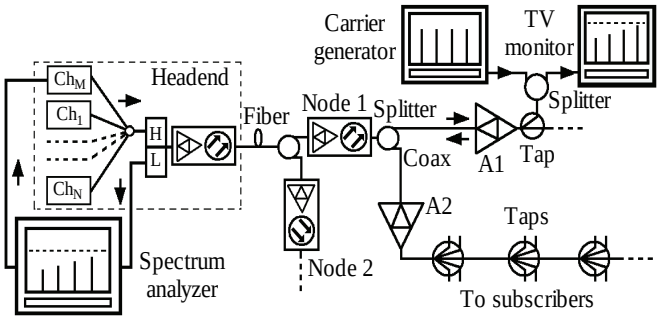


Fig. 3.

With the first method a spectrum analyzer, a TV monitor and a 2- or 4-carrier generator (from 5 to 65 MHz) are needed. The spectrum analyzer is connected at a test point of the headend to control setting the necessary reference level of the reverse path signal. A video modulator and an up converter are used to transfer the produced video signal to a given (unused) TV channel and to add it to the other signals in the forward path of the cable system. Then all optical nodes and amplifiers in the reverse path are adjusted one after the other, 2 or 4 carriers from the signal generator being fed to their inputs. The TV monitor is connected at the same points after being adjusted for the forward path by means of an image from the spectrum analyzer in the headend. The signal generator is connected at the input of the next-to-come optical node or amplifier in the reverse path and its gain is adjusted by means of the TV monitor with the aim to equalize the head up signals' level with the reference level.

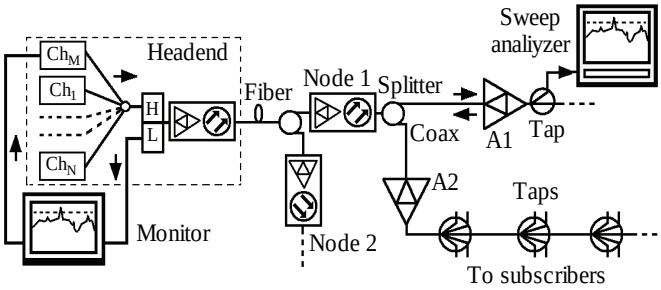


Fig. 4.

The second method makes use of a sweep analyzer and a return monitor. The sweep analyzer is connected at the input of the reverse path amplifier to be adjusted and the return monitor is in the headend, its image being transmitted over a separate forward TV channel. Thus a direct connection is established between the 2 sets of instruments. This method is more effective because it provides possibilities to monitor the frequency-response curve of the whole reverse path and to evaluate ingress, reflection and frequency-dependent attenuation in all the units.

VI. CONCLUSION

Balancing the reverse path of HFC CATV can cause troubles to the operators due to random factors that can not be considered at the design phase. Thus the balancing procedure requests all requirements here considered to be strictly observed. The proposed methods for balancing the reverse path are rather easy to implement. It makes it possible for the received signals' level in the headend to be monitored on the very spot where each amplifier or optical node is connected. They make operators' activities easier, besides they provide facilities to build up a fully automated system for maintaining the desired quality of communications over the reverse path of CATV systems.

REFERENCES

[1] L. Jordanova and D. Dobrev. Improvement of the HFC system reverse path performance. Proceedings of the 14th International Scientific and Applied Science Conference "Electronics ET'2005", Sozopol, book 1, pp. 156-161, 2005.

[2] D. Dobrev and L. Jordanova. Noise and Distortions Limitations in Designing of HFC CATV Systems. Elsevier: Optical Fiber Technology, Vol. 12, Issue 2, April, pp. 196-204, 2006.

[3] Hewlett-Packard. Cable System Preparation Guide for Two-Way Data, Draft V2.0, 2003.

[4] C. A. Eldering, N. Himayat, and F. C. Gardner. CATV Return Path Characterization for Reliable Communications. IEEE Communications Magazine, No 8, pp. 62-69, 1995.

[5] S. Bette, V. Moeyaert, V. Tamgnoue, M. Blondel, P. Mégret. Comparative Analysis of the Impact of CATV Return Path Ingress and Impulse Noises in a 16-QAM Transmission System. Network & Optical Communications (NOC), pp. 229-236, Eindhoven (NL), 29/06-01/07, 2004.

This page intentionally left blank.

Boundary Conditions when using Quadrature Modulations in Coaxial Cable Television Networks

Kiril Koitchev[1], Krasen Angelov[2] and Stanimir Sadinov[3]

***Abstract* – Estimation of potential noise stability is made concerning downstream and upstream channels when using quadrature modulations of the type QPSK and QAM along coaxial CATV networks. It is assumed that, within the channel, there is white Gaussean noise of zero mathematical expectation and one-side power spectrum density of power. Error probability is calculated during transmission of individual symbols. Recommendations are made concerning the optimal choice in using a certain type of modulation with regard to cable network quality and the selected method of coding.**

***Keywords* – CATV, QAM, M-PSK, BER, DOCSIS**

I. INTRODUCTION

In digital TV broadcasting along coaxial lines by the standard DVB (Digital Video Broadcasting) the error probability *BER* (Bit Error Rate) at the receiver input should not exceed 10^{-11} [1, 8]. A larger value of *BER* will decrease the quality of picture since the signal is continuously being transmitted and, therefore, there will be no opportunity for correction of errors.

Should the CATV network be used for transmission of digital traffic (LAN, Internet and others) the situation then is quite different. In this particular case, if data are transmitted with error, this error could be detected through the incorrect control sum total of the digital stream. If IP protocol is used, then in case of error generation on channel level it can be corrected at transmission level or at the level of the package transfer route. Such mechanism allows successful operation at values for *BER* of $10^{-8} – 10^{-6}$.

Frequencies distribution according to DOCSIS standard is shown on Fig. 1. The downstream channel occupies the frequency band from *87* to *862 MHz*. Data transmission is done along a channel of width *6MHz* (or *8MHz* for Europe – EuroDOCSIS). The used type of modulation can be 64-QAM or 256-QAM [1, 2]. Symbols transmission rate depends on the type of cable and on amplitude-frequency response of the filter in the downstream channel and amounts to *30,342Mbit/s* and *42,88Mbit/s*, for 64-QAM and 256-QAM respectively.

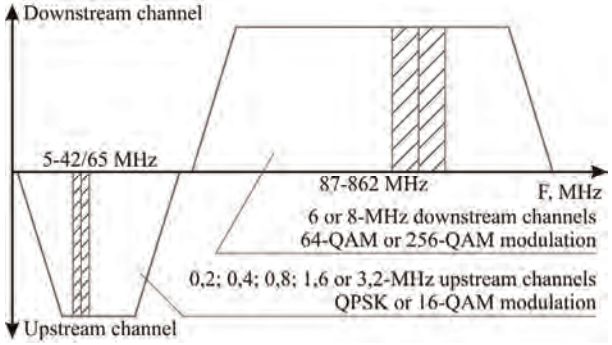


Fig. 1. Distribution of frequencies of downstream and upstream channel in accordance with DOCSIS

The upstream channel operates within the frequency band from *5* to *42 MHz* (from *5* to *62 MHz* for Europe). This range can be divided into channels of widths *0,2*; *0,4*; *0,8*; *1,6* and *3,2 MHz*. In the upstream channel it is possible to use either QPSK or 16-QAM modulation. In case we use a modulation of the type QPSK, the maximum possible rates will vary from *0,32* to *5,12 Mbit/s*, and with 16-QAM – from *0,64* to *12,24 Mbit/s*.

II. ESTIMATION OF NOISE STABILITY IN THE DOWNSTREAM AND UPSTREAM CHANNEL

Potential noise stability within the downstream and upstream channel will be calculated first. It is assumed that in it there is a white Gaussian noise of zero mathematical expectation and one-side power spectrum density of the power N_0. At first the error probability of an individual case will be calculated followed by an estimation of the *BER* magnitude.

In the common case the error probability P_a of an individual symbol is [3, 4]

$$P_a = P\left[\bigcup (A_j \geq A_i)\right],$$

where

$$A_i = \int_0^T \left|N(t) + S_i(t)\right| S_i(t)dt - 0{,}5E_j \ . \qquad (1)$$

In the above formula $N(t)$ stands for fluctuation disturbance; $S_i(t)$ – the symbol being transmitted; E_j – the energy of the symbol.

The estimation of this probability is difficult since in the use of quadrature modulation of the type QPSK and QAM, the amplitude of symbols is not identical. The task can be simplified if we assume that there will be an error in the

[1]Kiril R. Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

[2]Krasen K. Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@mail.bg

[3]Stanimir M. Sadinov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: murry@tugab.bg

reception of the symbol only when noise power exceeds the half of the difference between the symbol being transmitted and the symbol which is closest to it. This is the so called equivalent power of the two symbols P_E [2, 6].

By employing this approach in equally probable symbol transmission through Gray code and 4-PSK modulation we get

$$P_{E\,4PSK} = \frac{1}{2}\left[P_E(a_0) + P_E(a_1)\right] = \frac{1}{2}(1+1)P_a, \quad (2)$$

where a stands for the basis coordinates with regard to axis "a", P_a – the error probability related to the solution based in relation to the same axis.

By means of analogy it is possible to write equations for 8-PSK and 16-PSK, respectively:

$$P_{E\,8PSK} = \frac{1}{3}\left[P_E(a_0) + P_E(a_1) + P_E(a_2)\right] \approx \frac{1}{3}(1+1+2)P_a, \quad (3)$$

$$P_{E\,16PSK} = \frac{1}{4}\left[P_E(a_0) + P_E(a_1) + P_E(a_2) + P_E(a_3)\right] \approx \approx \frac{1}{4}(1+1+2+4)P_a \quad . \quad (4)$$

Summarizing for M-multiple PSK modulation we get

$$P_{E\,MPSK} = \frac{1}{\log_2 M}\left(1+1+2+4+\ldots+2^{\log_2 M-2}\right)P_a \approx \approx \frac{M}{2\log_2 M}P_a \quad . \quad (5)$$

Error probability P_a can be expressed as [6, 7]

$$P_a = \frac{4}{M}\sum_{i=1}^{M/4}\left[\sqrt{\frac{2E_b\log_2 M}{N_0}}\sin\frac{(2i-1)\pi}{M}\right], \quad (6)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{t^2}{2}}dt$ is Gaussian function, N_0 – the spectrum density of noise.

Taking Eq. (5) into account, we get for Eq. (6)

$$P_{E\,MPSK} \approx \frac{2}{\log_2 M}\sum_{i=1}^{M/4}Q\left[\sqrt{\frac{2E_b\log_2 M}{N_0}}\sin\frac{(2i-1)\pi}{M}\right], \quad (7)$$

In case of $M \geq 4$, then the ratio $E_b/N_0 >> 1$, and the first member of the sum will be dominating thus $i = 1$. Following that Eq. (7) gets simplified to the kind

$$P_{E\,MPSK} \approx \frac{2}{\log_2 M}Q\left[\sqrt{\frac{2E_b\log_2 M}{N_0}}\sin\frac{\pi}{M}\right]. \quad (8)$$

Eq. (8) can be regarded as approximation characteristics for MPSK modulated signal, *BER* being taken into account.

Fig. 2 shows the analysis results in using Eq. (8) and simulation by the Monte Carlo method.

In case of equal probability symbols *BER* for 16-QAM can be rendered by the expression by way of analogy with Eq. (2)

$$P_{E16QAM} \approx \frac{1}{2}(1+2)P_Q, \quad (9)$$

where P_Q stands for error probability related to the solution with regard to the Q axis. Thus, for the general case of M-QAM modulation

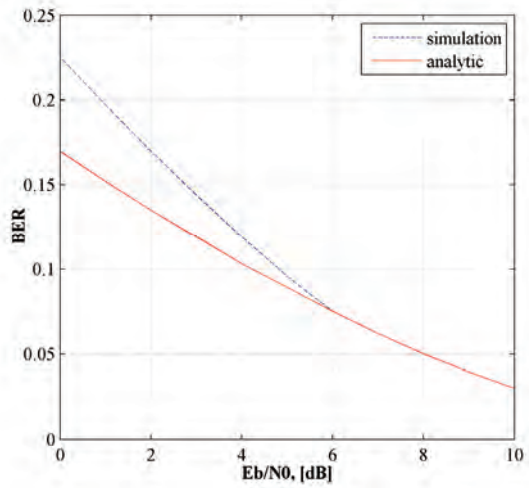


Fig. 2. Monte Carlo simulated and estimated from Eq. (8) *BER* as a function of E_b/N_0 for M-PSK modulated signal

$$P_{E\,MQAM} \approx \frac{1}{\log_2 M}\left(1+2+4+\ldots+2^{\log_2\sqrt{M}-1}\right)P_Q = = \frac{2}{\log_2 M}\left(\sqrt{M}-1\right)P_Q \quad . \quad (10)$$

Taking account of the constellation diagram for M-QAM in employing Gray code, it follows that

$$P_Q = \frac{2}{\sqrt{M}}\sum_{i=1}^{\frac{\sqrt{M}}{2}}Q\left[\frac{(2i-1)d}{\sqrt{\frac{N_0}{2}}}\right], \quad (11)$$

and d can be calculated from

$$E_b\log_2\sqrt{M} = \frac{M-1}{3}d^2. \quad (12)$$

Then Eq. (12) is reduced to

$$P_{E\,MQAM} \approx \frac{4}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right)\sum_{i=1}^{\frac{\sqrt{M}}{2}}Q\left[(2i-1)\sqrt{\frac{3E_b\log_2 M}{(M-1)N_0}}\right]. \quad (13)$$

In this way Eq. (13) appears to be an expression of approximation for M-QAM and as is in Eq. (8) there can be drafted the characteristics shown on Fig. 2.

Since the signals transmitted along the line appear to be simple ones then it follows that it is possible to determine the necessary relation E_b/N_0 for every type of modulation, provided the error probability is adequately assigned. Assuming that Gray code is used, (in this case each symbol will be distinguished from the next closest one by *1bit* only in

order to ensure minimum error), it is possible to calculate the magnitude of *BER* as a result of dividing the symbol error by *8* for 256-QAM; by *6* for 64-QAM; by *4* for 16-QAM and, finally by 2 for QPSK. For these cases the graphs of *BER* dependence from E_b/N_0 ratio are shown on Fig. 3.

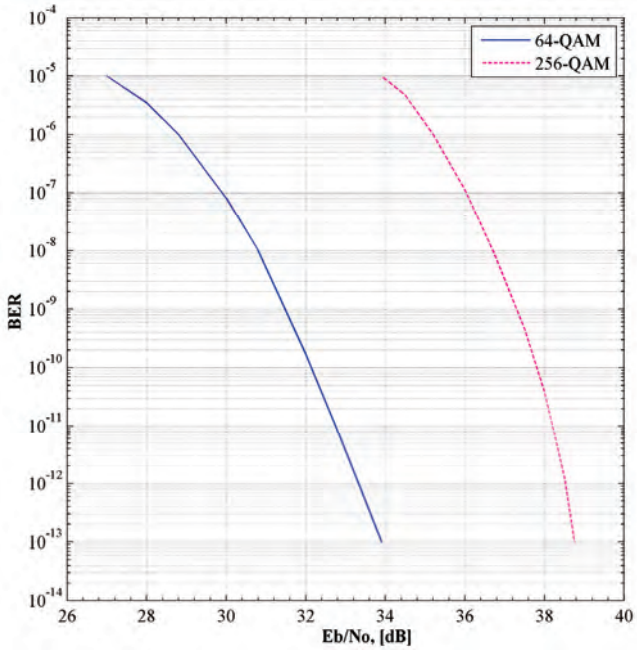


Fig. 3. *BER* = $f(E_b/N_0)$ for 64-QAM and 256-QAM with Gray code using

Table I contains the required values for E_b/N_0 with error probability 10^{-11}, 10^{-8} and 10^{-6} during data transmission along the downstream channel of CATV.

TABLE I
REQUIRED VALUES OF E_B/N_0 WITH ASSIGNED *BER* FOR THE DOWNSTREAM CHANNEL FOR 64-QAM AND 256-QAM MODULATION

BER	64-QAM, [dB]	256-QAM, [dB]
10^{-6}	29	35
10^{-8}	30,7	36,7
10^{-11}	32,4	38,4

When transmitting data along computer networks the sufficient value of *BER* is approximately 10^{-8}. In case of using coaxial networks it is necessary to select E_b/N_0 ratio of *36,7dB* and more for 256-QAM modulation since it ensures maximum transmission rate of *52 – 55 Mbit/s* with *8MHz* (according to EuroDOCSIS). For the sake of comparison it is worth noting that for transmission of digital TV signals using the Read-Solomon code according DVB-C standard it is required that the ratio E_b/N_0 is *32dB* with 256-QAM.

In case of using the upstream channel it should be borne in mind that it is a case of data transmission from numerous subscribers. Thus, the channel quality is defined in connection with *PER* (Package Error Rate) [4]. Table II shows the values of signal-to-noise ratio for *PER* = 10^{-6}, 10^{-8} and 10^{-11} with various modulations for the upstream channel.

TABLE II
REQUIRED VALUES OF E_b/N_0 WITH ASSIGNED *PER* FOR UPSTREAM CHANNEL FOR QPSK, M-PSK AND 16-QAM MODULATION

PER	QPSK, [dB]	36, 16, 10 QPSK, [dB]	16-QAM, [dB]
10^{-6}	16,5	7,2	23,5
10^{-8}	18,0	8,4	25,0
10^{-11}	19,5	9,5	26,5

By knowing *BER* (or *PER*) and in relation to E_b/N_0 ratio it is possible to determine the rates of the digital traffic along the transmission channel. Fig. 4 shows the relationship between the rate of the digital traffic *V*, according to DOCSIS 1.1/2.0, depending on E_b/N_0 for channel width of *3,2MHz* and error probability (*BER*) of 10^{-8}.

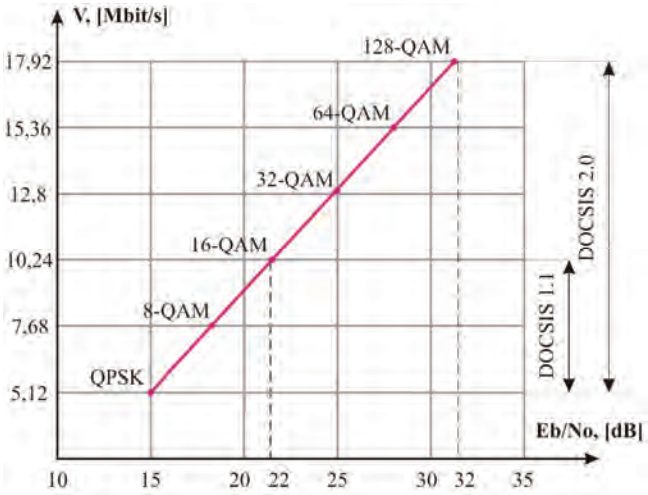


Fig. 4. *V* = $f(E_b/N_0)$ for upstream channel in accordance with DOCSIS 1.1/2.0 when using *3,2MHz* channel bandwidth and *BER* = 10^{-8}

III. CONCLUSION

The optimum modulation format used should be selected on the basis of the quality parameters of the CATV and the selected way of coding. Signal power can be increased to help improve signal-to-noise ratio. This in turn may cause some other parasitic effects such as the penetration of the signal at the input of the TV set, disturbances due to the presence of second and third harmonic in the signal along the upstream channel data transmission, etc.

Signal-to-noise ratio can be improved at the design and construction stage of networks set-up by using higher quality components such as coaxial cables with triple screening, passive dividers and deflectors of maximum screening (not less than *120dB*), or through the network topology, for example, by employing HFPC technology – a hybrid network

featuring passive coaxial part also referred to as “fiber to the home”. This particular type of technology appears to be the most attractive one both in terms of signal quality performance and network transmission capacity.

ACKNOWLEDGEMENT

This paper has been sponsored under the auspices of the “Planning of Multimedia Telecommunications Networks with Traffic and QoS Management” project – a part of the “SCIENTIFIC INVESTIGATIONS” Fund, contract ВУ-ТН-105/2005.

REFERENCES

[1] European Standart CENELEC EN 50083. Cable distribution systems for television, sound and interactive multimedia signals, December 2002.

[2] Jianhua Lu, K. B. Zetaief, J. Chuang, M. L. Liou. M-PSK and M-QAM BER Computation Using Signal-Space Concept. IEEE Transaction of Communication, Vol. 47, pp. 181-184, February, 1999.

[3] Irshid M. I., S. Salons. Bit Error Probability for coherent M-ary PSK systems. IEEE Transaction of Communication, Vol. 39, pp. 349-352, March, 1991.

[4] Ciciora W., J. Farmer, D. Large. Modern Cable Television Technology. San Francisco, MK Publishers Inc., 1999.

[5] Ovadia S., Broadband Cable TV Access Networks: From Technologies to Applications, Prentice Hall, Inc., 2001.

[6] Pavan K., S. Alonini. BER Computation of 4/M-QAM Hierarchical Constellations. IEEE Transaction of Broadcasting, Vol. 47, pp. 228-239, September, 2001.

[7] Yoon D., K. Cho, J. Lee. Bit error probability of M-ary Quadrature Amplitude Modulation. IEEE VTC’2000-Fall, Boston, Massachuessets, pp. 2422-2427, September, 2000.

[8] O. B. Panagiev, CATV systems – parameter’s optimization, ICEST 2005, Proc. of Papers, vol.2, Nish, pp.497-499, 2005.

Planning of VoIP System for Hybrid Fiber-Coaxial Television Networks

Nataliya Varbanova[1], Kiril Koitchev[2] and Krasen Angelov[3]

***Abstract* – Introducing VoIP into cable television network is an efficient and inexpensive alternative of the traditional switched telephone network. It is assumed that the cable television network corresponds with the EuroDOCSIS standards. When a VoIP system is planned, the cable network capacity and architecture, required number of simultaneously performed calls, and number of allowed call attempts per second are taken into consideration. VoIP system capacity is estimated depending on network loading and used codec for transmission channel parameters defined in the DOCSIS specifications.**

***Keywords* – Voice over IP, DOCSIS, HFC Network, CATV**

I. INTRODUCTION

Hybrid fiber-coaxial (HFC) cable television network is a wide used urban access network that offers large variety of interactive services [1, 2]. For voice services delivering in this type of networks voice over IP (VoIP) technology is the most used.

VoIP service operates by converting voice signals to data packets, sending these data packets through the IP network, converting these packets back into voice signals, and managing the overall call setup (dialing), connection, and termination (hang-up) [3].

VoIP introduces new traffic loads and traffic that is synchronous in nature, versus asynchronous like typical residential high speed data. VoIP also requires support for low delay and low packet loss. This is because VoIP is a streaming service where retransmission is not feasible. In addition VoIP requires high service availability. The level of availability depends on whether the VoIP service is intended to be a primary line service, available after a power outage or a secondary line service where service availability is not as critical.

For just the CMTS and HFC portions of a cable operator's network, VoIP design choices span:

- Multimedia Terminal Adaptor (MTA) configuration parameters;
- HFC node and DOCSIS parameters;
- CMTS configuration parameters.

[1]Nataliya Varbanova is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: nataliavarbanova@abv.bg

[2]Kiril Koitchev is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: koitchev@tugab.bg

[3]Krasen Angelov is with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail: kkangelov@mail.bg

II. ARCHITECTURE OF HFC CABLE TELEVISION NETWORK WITH VOIP SERVICE DELIVERING

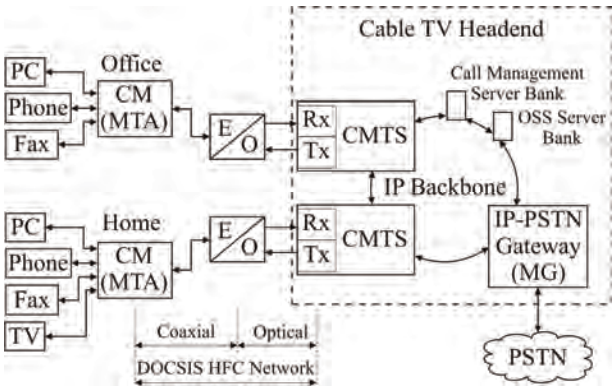


Fig. 1. Architecture of HFC cable TV network with VoIP service delivering

IP telephony equipment can be easily included into existing cable architecture (Fig. 1) [4, 5]. Cable modem termination system (CMTS) is an essential network component that enables voice packet transmission over HFC network. CMTS consists of upstream RF receivers and a downstream RF transmitter. It is connected by optical links to HFC access network, and to three large groups of servers. The first group is responsible for the operation system support (OSS). The second group is a call management server, and the third group contains IP-PSTN Gateways. The OSS server group is responsible for installation, configuration, operation, and management of the entire system. The call management server group tends on the signaling information condition and the connection control. For connectivity with the public switched telephone network PSTN-gateways are needed [6].

Described components correspond the PacketCable specifications for voice over cable television network [7].

III. PLANNING OF VOIP SYSTEM FOR CABLE TV NETWORK

A. Multimedia Terminal Adaptor VoIP configuration parameters

There are many MTA configuration parameters but those with the largest impact on VoIP service include VoIP codec, packetization period, voice activity detection (VAD) state, and jitter-buffer size. The VoIP codec defines the method by which voice is encoded and is often referenced in terms of its output bit rate. The VoIP packetization period is the period

over which encoded voice bits are collected for encapsulation in packets. The VAD state refers to whether VoIP packets are sent all the time or only during talk periods, with VoIP packets not being transmitted during quite periods. The VoIP jitter-buffer size defines the maximum delay and nominal play-out delay of a jitter buffer.

The choice of the MTA configuration parameters impacts capital cost from a CMTS utilization perspective and operational cost from a VoIP QoS perspective. Basic MTA configuration parameters are shown in Table I.

TABLE I
MTA CONFIGURATION PARAMETERS

Configuration parameters	Typical configuration today
Codec	G.711, but G.728 and G.729E are optional recommendations iLBC and BV16 are mandated in PC1.1
Packetization period	10ms or 20ms (20ms and 30ms only options for iLBC)
VAD state	Avoided in cable applications
Jitter buffer size	Variable, but not uncommon to see 15ms play out and 30ms maximum

A lower rate codec will reduce VoIP payloads allowing higher channel utilization, but will also results in lower voice quality, and may degrade FAX and modem performance and prevent the passing of inband DTMF tones. A lower rate codec also imposes a higher processing burden on the MTA as well as the multimedia gateway (MG). This may result in MTA call-mixing limitations and lower MG capacity. Most cable VoIP trials and early deployments today use G.711 encoding with its *64kbps* pulse code modulated output. But PacketCable, the cable VoIP standards forum, has several low-rate codec recommendations and requirements that are categorized as providing a "toll-grade" voice performance with much lower output rates [7].

Table II provides an example of the reduced-rate benefits of low rate encoding.

TABLE II
EXAMPLE OF CODEC IMPACT ON CALL RATE

Codec	Codec output rate	QPSK upstream call flow rate	Downstream call flow rate
G.711	64 kbps	115.2 kbps	109.6 kbps
G.728	16 kbps	57.6 kbps	61.6 kbps
G.729E	12 kbps	57.6 kbps	57.6 kbps

It shows raw output rates and DOCSIS service flow rates (including overhead from RTP, UDP, IP, Ethernet, and DOCSIS protocols) for G.711, G.728, and G.729E encoding. The example assumes *10ms* packetization periods and use of DOCSIS PHS, BPI+, and FEC (PHS assumptions are for *41* byte savings in upstream and *12* byte savings in downstream. Upstream FEC assumptions are for long data grants with RS *K=220, T=8,* short data grants with RS *K=78, T=5,* and use of shortened last code words. In the example G.711 packets are assigned to long data grants and G.728 and G.729E packets are assigned to short data grants.).

Referring back to Table I, we next see that a higher packetization period produces less packet overhead thereby allowing higher channel utilization. However a higher packetization period also causes higher VoIP path delay and can increase packet error rate (*PER*). Packetization delay can be codec dependent and nominally contributes around *1.5* times the packetization period to VoIP path delay. The path delay contribution is the result of the packetization delay itself plus DOCSIS unsolicited grant service (UGS) grant uncertainly. UGS grant uncertainty is the time a newly generated voice packet must wait at the MTA's CM before receiving a UGS grant transmission opportunity. Nominally this delay is half the packetization period. Because of these concerns, most cable VoIP deployments today use *10ms* or *20ms* packetization periods.

Table III provides an example of the trade-off between packetization-period delay and rate reduction. It shows nominal path delay and DOCSIS service flow rates for *10, 20,* and *30ms* packetization periods. The example assumes G.711 encoding and use of DOCSIS PHS, BPI+, and FEC.

TABLE III
EXAMPLE OF PACKETIZATION PERIOD TRADE-OFF

Packetization period	Nominal path delay contribution	QPSK upstream call flow rate	Downstream call flow rate
10 ms	15 ms	115.2 kbps	109.6 kbps
20 ms	30 ms	89.6 kbps	86.8 kbps
30 ms	45 ms	85.3 kbps	79.2 kbps

Referring back to Table I, we next see that enabling VAD will reduce VoIP load since voice will only be sent when talking occurs. This allows more calls and/or data to exist concurrently, resulting in higher channel utilization. However enabling VAD will also degrade voice quality as the result of a voice clipping at the beginning of each talk period. All cable VoIP trials and early deployments that we know of today avoid VAD for this reason. All cable VoIP trials and early deployments that we know of today avoid VAD for this reason.

Finally from Table I, we see that MTA jitter-buffer size has no impact on CMTS utilization and should be chosen based on VoIP QoS considerations. Jitter buffer size should provide an appropriate balance between jitter-induced packet drops and its contribution to VoIP path delay. Choosing a large jitter-buffer reduces packet dropping from jitter but increases VoIP path delay. In particular, its selection should be coordinated with VoIP *PER* considerations for a total packet loss rate target, and with the choice of VoIP packetization-period to meet a VoIP path delay target.

To determine an appropriate VoIP path delay target, first consider an appropriate maximum round-trip VoIP delay. *300ms* is often used as a target for maximum round trip delay between an MTA and PSTN phone over a long distance connection. From this delay, *150ms* is often targeted for end-to-end delay, with around *100ms* assigned to PSTN long-

distance propagation delay. The result in this case is a remainder of *50ms* for local VoIP path delay.

Jitter-buffer size, packetization period, and packetization-related grant uncertainty combine with other delays to form a VoIP path delay.

B. HFC node and DOCSIS parameters

HFC node parameters that impact VoIP service include average node size and the maximum number of nodes per CMTS receiver group. Average node size defines the average number of homes served by a fiber node. Larger node sizes require fewer fiber terminating nodes and fibers. Smaller nodes generally result in less ingress noise in an upstream channel.

The maximum number of nodes per CMTS receiver group defines the maximum number of nodes that can be connected to one or more CMTS receivers. Higher node allowance per receiver group permits more nodes to be supported per receiver when service take rate is low. But larger allowance will also funnel more ingress noise in an upstream channel.

The choice of these parameters impacts capital cost from a node and CMTS utilization perspective and operational cost from a VoIP QoS perspective.

TABLE IV
DOCSIS CHANNEL PARAMETERS

DOCSIS channel parameter	**Typical configuration today**
Upstream bandwidth	1.6 or 3.2 MHz
Upstream modulation	QPSK or 16QAM (only choices)
Downstream modulation	64QAM or 256QAM (only choices)
PHS	~40 bytes upstream, not clear for downstream
Upstream FEC	Dependent on upstream conditions

DOCSIS parameters that impact the VoIP service include upstream channel bandwidth, upstream and downstream modulation, upstream and downstream packet header suppression (PHS), and upstream forward error correction (FEC) (Table IV). EuroDOCSIS allows for a variety of upstream bandwidth and upstream and downstream modulation choices [8, 9]. Choosing *3.2MHz* upstream bandwidth provides double the capacity of *1.6MHz* channel but requires support for a contiguous chunk of *3.2MHz* bandwidth within upstream band. It also requires a *3dB* higher signal to noise ratio (*SNR*). A 16-QAM upstream modulation format provides double the channel capacity of QPSK. But 16-QAM also has approximately *7dB* higher *SNR* requirements than QPSK at 10^{-6} *BER*, and greater amplitude and phase noise sensitivity. A 256-QAM downstream modulation format provides over *40%* higher channel capacity than 64-QAM. But 256-QAM also has approximately *6dB* higher SNR requirement than 64-QAM at 10^{-6} *BER*, and significantly greater amplitude and phase noise sensitivity.

PHS defines the amount of packet overhead savings in the upstream and downstream channels accompanying VoIP packets using UGS. The PHS reduces VoIP overhead allowing higher channel utilization, and has no impact on VoIP QoS.

FEC defines Reed Solomon coding parameters used in upstream channels. Limiting the amount of FEC will help limit VoIP overhead and allow higher channel utilization. But increasing the amount of FEC improves noise and interference robustness, possibly permitting support for higher bandwidth and modulation choices.

The choice of these parameters impacts capital cost from a CMTS utilization perspective and operational cost from a VoIP QoS perspective.

C. CMTS configuration parameters

CMTS configuration parameters that impact VoIP service include the maximum bandwidth to allocate to VoIP service and the ratio of node-to-CMTS receiver assignment. The maximum bandwidth allocation for VoIP defines the maximum upstream and downstream bandwidths allowed for VoIP service at peak VoIP loading. This allocation is essential when data is sharing the same DOCSIS channel resources because it prevents VoIP calls from monopolizing all available bandwidth. The ratio of node-to-CMTS receiver assignment defines the number of nodes assigned to a CMTS receiver group and the size of the receiver group in terms of number of receivers. For example the ratio may be *1* node per receiver group of *2* receivers.

The choice of CMTS parameters impacts capital cost from a node and CMTS utilization perspective and operational cost from a CMTS and node reconfiguration and VoIP QoS perspective (Table V).

TABLE V
CMTS CONFIGURATION PARAMETERS

Configuration parameters	**Typical configuration today**
Maximum VoIP bandwidth allocation	Dependent on data traffic load, but typically between 40% to 60%
Ratio of node-to-CMTS Rx assignment	Unclear, but Rx group size >1 is likely to increase as VoIP take rate grows

The use of current versus padded traffic margin when choosing a maximum VoIP bandwidth allocation provides a trade off between CMTS utilization and frequent reconfiguration. In addition, a node-to-CMTS receiver assignment that allows more than 1 receiver per receiver group can avoid the need for node splitting at high service take rates.

The optimal choice of maximum VoIP bandwidth allocation and the ratio of node-to-CMTS receiver assignment requires an iterative computation to determine the maximum take rate for a given node size (or node size maximized for a given take rate). The computation requires voice and data traffic load assumptions along with traffic margin (typically added to take rate but could involve setting DOCSIS parameters, as well as CMTS hardware design and DOCSIS

parameters, as well as CMTS hardware design and performance constraints.

TABLE VI
EXAMPLE OF NUMBER OF LINES PER CMTS ESTIMATION

Used codec	Packetization Period, ms	QPSK upstream call flow rate, kbps	Number of simultaneously possible calls	Total traffic based on Erlang B model, Erl	Lines per upstream receiver	Total number of lines per upstream receiver	Lines per CMTS	Total number of lines per CMTS
G.711	10	115,2	40	29,01	414	591	1656	2364
	20	89,6	52	39,70	567	810	2268	3240
	30	85,3	55	42,41	605	864	2420	3456
G.728	10	57,6	81	66,29	947	1352	3788	5408

D. Example of VoIP service over HFC CATV network estimation

After the short description of the main parameters of voice telephony service, a scenario for VoIP service estimation will be considered. An example for such estimation is to evaluate the number of subscribers that are possible to be served at the same time. An area that is served by a CMTS consists of one downstream RF transmitter and four upstream RF receivers (such as Motorola BSR 1000) is assumed. The estimations are made for the following assumptions:

- Upstream modulation format: QPSK;
- Upstream channel bandwidth: *3,2MHz* (nominal data rate *4,6Mbps*);
- G.711 codec (Tables II and III);
- G.728 codec (Table 2);
- Blocking probability 1%;
- Traffic per residential user *0,07Erl* (*4,2min* call in average) ;
- *70%* of the CATV users are subscribed to VoIP service.

The estimations based on above assumed parameters are shown in Table VI.

IV. CONCLUSION

The numbers in Table VI are comparable to the traditional switched telephone networks. If the additional PHS options are applied, and low rate codecs or a 16-QAM modulation is used for the upstream channel (*3,2MHz@10,21Mbps*), than the VoIP network capacity will be greater than the switched telephone network capacity.

Lower rate encoding would realize a larger rate reduction for increased packetization periods. For example G.728 realizes nearly a *50%* flow rate reduction between *10ms* and *30ms* packetization compared to around *25%* for G.711.

IP-based telephony has a substantial disadvantage. Like all the IP-based services the telephony is very sensitive to packet loss and delays when the transmission rate is too low.

It is recommended to coordinate the choice of node and DOCSIS parameters so as to reach a target maximum PER for VoIP service while minimizing aggregate node and CMTS costs (as driven by utilization considerations). The maximum VoIP PER target should be chosen from a total VoIP packet loss rate target that also ac-counts for jitter-induced packet drops. The total VoIP packet loss rate target should be chosen according to voice service needs.

PER may increase significantly with packetization period if error conditions tend to be random in nature, such as from thermal noise or short impulse noise.

If or when VoIP and data traffic loading grows beyond the capacity of a current CMTS configuration, CMTS reconfiguration will be required. In this case optimal selection of maximum VoIP bandwidth allocation and the ratio of node-to-CMTS receiver assignment should be recomputed using new traffic load assumptions and parameter choices.

ACKNOWLEDGEMENT

This paper has been sponsored under the auspices of the “Planning of Multimedia Telecommunications Networks with Traffic and QoS Management” project – a part of the “SCIENTIFIC INVESTIGATIONS” Fund, contract BY-TH-105/2005.

REFERENCES

[1] Н. Долгополова, И. Рытвинский, Н. Быстрицкий, Эффективное внедрение услуг Triple Play операторами связи, Телемультимедиа №4, 2005.

[2] И. Колпаков, О. Васькин, С. Смирнов, Универсальная мультисервисная транспортная среда на базе сетей кабельного телевидения (Часть 2), Теле-Спутник №2, 2003.

[3] L. Harte, D. Bowler, R. Flood, IP Telephony Basics, ALTHOS Publishing, 2004.

[4] K. Koitchev, S. Sadinov, St. Nemtsov, IP Telephony Over Coaxial Cable Television Network, TELSIKS 2003, Proc. of Papers, pp.250-253, Nish Serbia and Montenegro, 2003.

[5] Voice over IP – Grundlagen, Regulierung und erste Erfahrungen, Rundfunk & Telekom Regulierungs GmbH, Band 1, 2006.

[6] В. Спирин, Оборудование для организации телефонии и передачи данных в сетях кабельного ТВ, Теле-Спутник №2, 2001.

[7] PacketCable™ 2.0, Codec and Media Specification, PKT-SP-CODEC-MEDIA-I02-061013, CableLabs, 2006.

[8] eDOCSIS™, eDOCSIS Specification, CM-SP-eDOCSIS-I11-070223, CableLabs, 2007.

[9] DOCSIS 2.0, Radio Frequency Interface Specification, CM-SP-RFI2.0-I11-060602, CableLabs, 2006.

Estimation of the Effectiveness of Some of the Access Algorithms in Terms of Servicing Subscribers of the CATV Systems

Lidia T. Jordanova[1] and Jordan I. Nenkov[2]

Abstract **- In this article is made comparative analysis of some of the access algorithms in terms of servicing subscribers of the CATV systems. There is shown a method allowing us to solve the collision problem and there is made comparison between some MAC protocols, used in CATV networks, using the following parameters: access delay, throughput, loss ratio and fairness.**

Keywords **– HFC, MAC protocols, 802.14, Access algorithms, Collision resolution.**

I. INTRODUCTION

The modern CATV networks are two-way and they are based on tree and branch topology. Each branch serves 125 to 500 subscribers. One of the limitations here is that the station can not listen directly to the upstream transmissions from other stations; hence, they are incapable of detecting collisions and coordinating their transmissions all by themselves. A multiple access technology other than carrier sensing is required so that all subscribers within a branch can share the available reverse bandwidth.

The MAC protocol for CATV systems must answer the following important requirements: dynamic bandwidth allocation to CBR (Constant Bit rate), VBR (Variable Bit Rate) and ABR (Available Bit Rate) traffic type; high channel throughput; low access delay; Support for a large number of stations and metropolitan area coverage.

II. THE MAIN COLLISION RESOLUTION ALGORITHMS (CRA)

A. p-persistence

This algorithm is based on the well-known ALOHA protocol. Each station transmits its request in the available contention slot, with probability p. Unlike in traditional ALOHA, this rule applies to new requests as well as to retransmissions. The maximum achievable throughput of ALOHA is 36.7 % ($1/e$). The probability of successful transmission P_{SUCC} for p- prersistence is given by

$$P_{SUCC} = n.p.(1-p)^{(n-1)}, \quad (1)$$

in which n is the number of contenders at the beginning of the slot. The system is stabilized when $p = 1/n$, under the Poisson traffic assumption. In order to estimate n, which is generally unknown, the pseudo- Bayesian algorithm is used at the head-end (HE).

B. N-ary Splitting Tree

In this group of algorithms, all stations involved in a collision are split into n subgroups. Then, each station that was involved in the collision randomly selects one of these subgroups. The first of the n subgroups retransmits in the subsequent available contention slot. All other stations enter waiting mode, until the resolution of the previous subgroup. The collision resolution process can be represented as a tree, in which each collision produces n new nodes. the best throughput is achieved with $n = 3$ (ternary tree). The following figure shows an example of the collision resolution process with ternary tree, in which the initial collision multiplicity equals 5. The numbers inside the tree nodes represent the stations that transmitted, and the numbers next to each node show the resolution order.

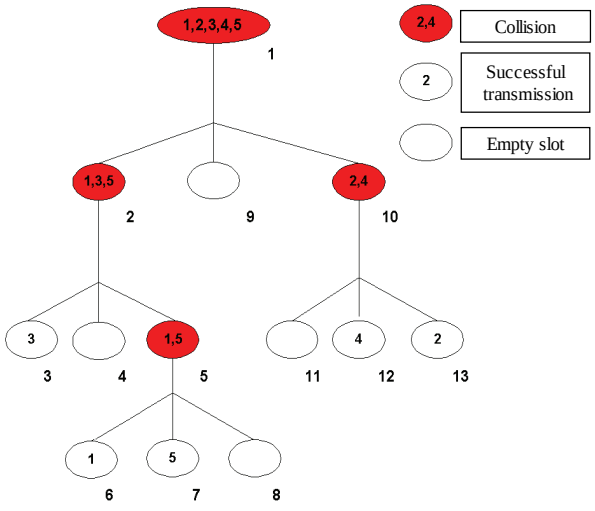


Fig. 1

The newcomers' transmission policy has a strong influence on the performance of the tree-algorithms. *Blocked access* retains new requests until the current contention is resolved, i.e. newcomers are not allowed in the contention slots used for retransmission. *Free access* allows newcomers

[1]Lidia T. Jordanova is with the Faculty of Communications and Communications Technologies, 1756 Sofia, Bulgaria, e-mail: jordanova@tu-sofia.bg

[2]Jordan I. Nenkov is with the Faculty of Communications and Communications Technologies, 1756 Sofia, Bulgaria, e-mail: jordan_n2002@yahoo.com

to send immediately, in any contention slot. The order of collision resolution in the tree can be LIFO, in which new collisions are resolved first, or FIFO.

III. MAC LAYER ALGORITHMS

Here are presented some examples of MAC protocols and their collision resolution algorithms used in CATV networks. MAC protocols can be classified into two categories: distributed protocols and centralized protocols. There is no central controller in distributed protocols, like CSMA/CD and R-ALOHA protocols. The centralized protocols provide better timing mechanisms in avoiding collisions.

The Reservation slotted ALOHA scheme (*R-ALOHA*) [1], a modification of Slotted ALOHA is originally proposed to improve the throughput of a satellite channel beyond that of slotted ALOHA. R-Aloha is a distributed protocol. Each time slot here matches one cell. In case of successful transmission in one slot, the corresponding slot in the subsequent frame is reserved for the station. Stations that have new data check the current frame. Any idle slot will be available in the subsequent frame.

Extended Distributed Queuing Random Access Protocol (*XDQRAP*) [2] is a distributed algorithm in which each station maintains queues for transmission of both data and requests. The contention resolution algorithm is tree-based, and short one-cell messages can preempt long data messages. In the upstream channel, a data slot is followed by two (or three) contention slots. All stations must monitor for the feedback from the request transmission and update their data and request queues accordingly. In this way, the "source" station knows when to commence transmission, and the "destination" station knows when to commence reading the message. The HE remains passive throughout this scheme. The distributedschemes, however, do not use the inherent central control point of the network – the HE. It is more difficult to meet QoS demands with distributed implementations and they are more susceptible to errors.

The MAC Level Access Protocol (*MLAP*) [3] divides the upstream into frames of variable lengths, which are called blocks. Each block contains a number of contention slots and a number of data slots, and each data slot encapsulates an ATM cell. MLAP assumes that the HE scheduler can prioritize transmissions, as the stations can have a number of queues for different data sources, based on priorities, and can send priority information with requests. Stations can also use "piggybacking". The algorithm used to resolve collisions in MLAP is START-*n* (*n*-ary Stack Resolution). START-*n* actually runs a free-access LIFO n-ary tree for each collided slot.

The Adaptive Digital Access Protocol (*ADAPt+*) [4] also relies on centralized control by the HE. The protocol defines frames of fixed sizes, in which the HE allocates the first regions for isochronous traffic (i.e. telephony) and the rest for available bit rate traffic. In the latter part, bandwidth is available for both request and data transmissions (contention mode) and the rest of the bandwidth is left for the reservation mode. The protocol supports data carriage in ATM cells. No original CRA is proposed in *ADAPt+*, and the authors suggest using any well-known algorithm.

The Centralized Priority Reservation (*CPR*) [5] uses the HE to manage the request and data channels in the upstream, and the grant and data channels in the downstream. This is achieved with knowledge of the exact delay for each station, and by "sandwiching" a number of contention slots between each data slot in the upstream. The *p*-persistence algorithm is used to resolve collisions.

The *Continuous mode with p-persistence* represents better kind of CPR but with the difference that it is required a frame structure of the upstream channel. The mechanism is self-regulating – at low loads there are plenty of contention slots, and when the load is high and there are not enough contention slots, the requests cannot be sent and therefore more contention slots will be allocated. The CRA that is used is once again *p*-persistence, and "piggybacking" is permitted. When the distance to the HE is sufficiently long, many requests are accumulated during the round trip delay (RTD) time, which leads to a long burst of data slots allocated by the HE. The proposed solution to this problem is to periodically insert a number of contention slots "by force". These slots are called FMS (forced mini-slots). In general, if "piggybacking" is used and the average request size is *k* cells, the formula used to calculate the number of N_{FMS} is

$$N_{FMS} = \left(1 - v_p\right) e / k \,, \tag{2}$$

where v_p is the "piggybacking" arrival rate. the use of FMS reduces the proposed self-regulating continuous mode to a kind of clustered mode scheme with small frames.

Unilink protocol [6] is also a centralized protocol. The central node, called pacer, is not necessary to be located at the HE. The Unilink frame length is kept constant, but supports the transmission of variable-length messages by using concatenation, that is, a station is allowed to transmit in several consecutive slots, thus saving PHY and MAC overheads. The frame is divided into three regions: a periodic dedicated region (for synchronous traffic), a reservation region and a contention region. The boundaries between these regions are changed by the system controller according to load. A station wishing to communicate tries to seize one or more slots in the contention region using CSMA/CD. After successfully transmitting, the station may move transmissions to the reservation region.

Pipelined Cyclic Upstream Protocol (*PCUP*) operates in two modes: cyclic transmission mode and negotiation mode. In negotiation mode, the HE runs a membership control algorithm in order to permit the off-line stations to join in. Every 0.5 seconds or more, the HE sends a special *invitation* frame to all the inactive stations. Stations that were inactive since the last membership control become off-line and do not receive the current invitation. Then the HE performs *positioning*. In positioning, a transmission start time is assigned to each station, in a way that neutralizes propagation offsets. Data from different stations arrives in sequence to the HE and further away stations can start transmission before closer stations complete theirs. During positioning, the HE performs ranging and classifies the stations by an ascending order of distance.

The HE allocates transmission quota, t_i, for station i, as follows:

$$if \sum (b_i + W_i) < CBT,\ t_i = \max(b_i, 1) + W_i \quad 3)$$

$$if \sum (b_i + W_i) > CBT,\ t_i = \max(\frac{CBT.\beta_i}{\sum \beta_i}, 1) + W_i \quad 4)$$

$$if \sum (G_i + b_i + W_i) < CBT,\ t_i = \max(G_i + b_i, 1) + W_i \quad 5)$$

$$if \sum (g_i + b_i + W_i) < CBT < \sum (G_i + b_i + W_i),$$

$$t_i = \max(g_i + b_i, 1) + W_i + (CBT - \sum (g_i + b_i)).\frac{G_i - g_i}{\sum (G_i - g_i)} \quad 6)$$

$$if (CBT - \sum b_i) < \sum (g_i + W_i) < CBT,$$

$$t_i = \max((CBT - \sum (W_i + g_i)).\frac{\beta_i}{\sum \beta_i}, 1) + W_i + g_i \quad 7)$$

$$if \sum (g_i + W_i) > CBT,$$

$$t_i = \max((CBT - \sum W_i).\frac{\alpha_i}{\sum \alpha_i}, 1) + W_i \quad 8)$$

where CBT is cycle time; W_i is guard-band time; α_i and β_i are traffic urgency parameters; g_i is the minimum number of guaranteed cells at station i; G_i is the requested number of guaranteed cells at station i.

After that, the HE then computes the transmission starting time S_i for station i, precisely as follows:

$$S_i = \sum_{j=1}^{i-1} t_j - \tau_i , \quad (9)$$

where t_j is the allowed transmission duration for station j and τ_i represent propagation delay. Prioritized bandwidth scheduling (i.e. t_j computation) is performed at the end of the cycle, for two traffic classes: guaranteed and best-effort. A special frame sent to the station specifies the new transmission frequency. In the last slot, the station sends its buffer status to the HE in order to facilitate the subsequent cycle scheduling.

IV. Simulation Study

On the following figures are shown the results from the comparison of access delay, throughput, loss ratio and fairness of three MAC protocols - PCUP, R-Aloha and Unilink. The simulation researches are made in the following assumptions:

– A 40 MHz band can be divided into multiple upstream channels. Each of them can be 1 MHz to 6 MHz wide and 1.6 Mbps to 10 Mbps in capacity. We assume the upstream frequency range to be 8-26.5 MHz and 1.544, 2.048, 6 and 10 Mbps transmission rate per channel. By this assumption, 17, 14, 5 and 3 upstream channels can be used simultaneously. Each branch serves 125 to 500 homes. That means an upstream channel is shared by about 30, 35, 100 and 167 subscribers respectively. The scale of network is assumed to be 80 km.

– The length of an upstream cell is assumed to be 424 bits, which is equal to 275 μS transmission time with the 1.544 Mbps transmission rate. Moreover, stations have a limited buffer size of 500 cells. Each station‘s hardware addresses is 48 bits.

Three types of traffic models, recommended by IEEE 802.14, are applied in the simulations. Traffic model A contains only ABR service. In traffic model B, CBR applications contribute 50% traffic load, the rest 50% remains to be ABR. And in traffic model C, the VBR applications present 30% traffic, CBRs pump 30% and the rest is for ABR service.

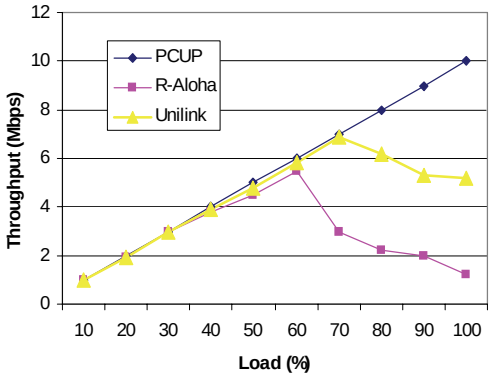


Fig. 2

On the Fig. 2 is shown the ratio between the PCUP, R-Aloha and Unilink throughput and Load under the traffic model B. The channel capacity is 10 Mbps. The Load (%) is defined as ratio between arrival cells and link capacity. Obviously, the throughput of R-Aloha and Unilink are not ideal under heavy traffic. Because of PCUP's centralized bandwidth scheduling, it achieves excellent channel throughput under various traffic types.

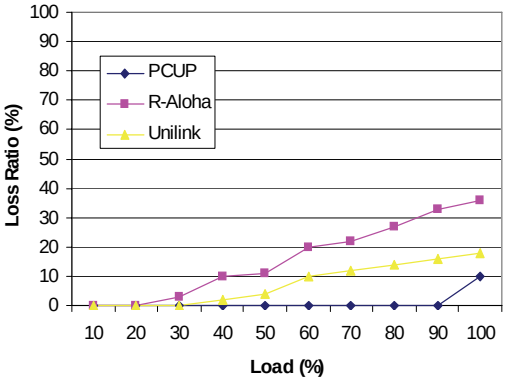


Fig. 3

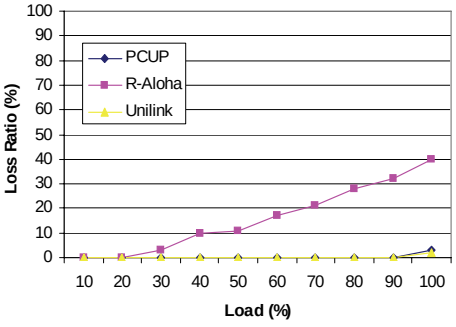


Fig. 4

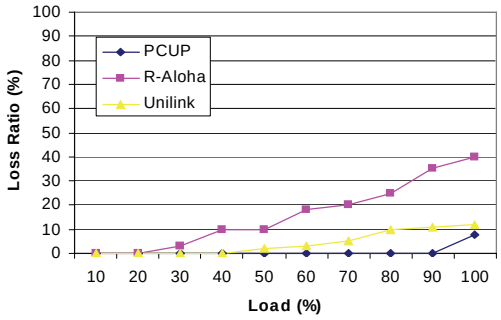


Fig. 5

On the fig. 3, 4 and 5 are shown the ratios between loss ratio and load equals to ABR, CBR and VBR traffic for the three protocols under traffic model C. In Fig. 3, we find that UniLINK has better performance than PCUP. UniLINK has only 3.1% loss ratio (lost cells/ arrival cells [%]) when the system is fully loaded, while PCUP has 5.1% loss ratio. Comparing Fig. 5 with Fig. 3, the curves are similar, but UniLINK achieves lower loss ratio of VBR traffic. This scenario is due to applying reservation scheme over VBR traffic. In R-ALOHA protocol, every cell is treated in the same way, i.e. no priority. So it behaves in a similar way under various traffic types.

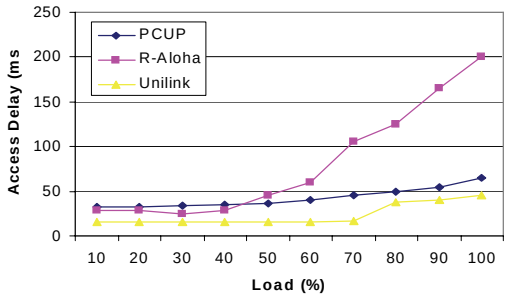


Fig. 6

Fig. 6 presents the dependences between the Access Delay and the Load for the three protocols. R-ALOHA has larger access delay when traffic load is high. Because of higher collision ratio under heavy traffic, collided cells may be retransmitted many times. Thus access delay grows up accordingly. There is the shortest access delay in Unilink protocol, which is a little better than the access delay in PCUP and in Unilink tne access delay is relatively constant.

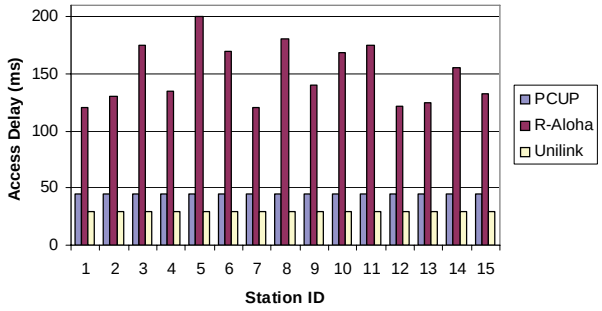


Fig. 7

This constance in the access delay in these two protocols is obvious on Fig. 7, where is shown the histogram of the distribution of access delay. For its obtaining are used statistic data for randomly select 15 stations from 167 stations sharing the 10 Mbps channel. There can be made more conclutions from that histogram for another important parameter of the CATV systems – fairness.

V. Conclusion

The conducted analisyses and simulation reserches shows that in PCUP and Unilink can be guaranteed constant low rate of access delay while in R-Aloha it is not. This shows that R-Aloha is not suitable in CATV networks. Unilink is more suitable when is needed short access delay and PCUP is prefered when it is required high throughput of the system. That is confirmed by the statistic information, taken from the research of the upstream parameters.

References

[1] Simon S. Lam, “Packet Broadcast Networks-A Performance Analysis of the R-ALOHA Protocol”, November 1979
[2] C. Wu, G. Campbell, “The Extended DQRAP(XDQRAP) algorithm”, DQRAP Research Group Report 94-3, Dec. 9, 1994.
[3] http://www.comsoc.org/ci/private/1996/mar/bisdikian.html
[4] http://www.comsoc.org/ci/private/1996/mar/sriram.html
[5] Dolors Sala "MAC Protocols for Multimedia Data over HFC Architecture"Georgia Tech Technical Report GIT-CC-95-48, October 27, 1995.
[6] Chris Grobichki, J. M. Ulm, “UniLINK as a media access protocol for community cable

Defining the Channel Parameters and Their Influence on SER for OFDM Transmission

Kiril Koichev [1], Stanimir Sadinov[2] and Georgi Traikovski [3]

***Abstract* – In this report we examine the fundamental characteristics of OFDM based systems which feature transmission of very high rates. The latter allow the systems to be used for providing progressive services and applications. Based on the requirements defined in the standards these systems should satisfy some criteria such as BER and SER which are the main parameters under examination in here.**

***Key words* - OFDM, FFT(IFFT), BER, SER, SNR.**

I. INTRODUCTION

Generally, we regard as well known the methods for orthogonal frequency division and multiplexing of signals (OFDM) which are connected with (*IFFT-Inverse Fast Fourier Transform)* and addition of cyclic expansion by using the window method [1].

Fig.1 shows an exemplary block diagram of an OFDM model where in its upper part is the transmitter (Tx) and the lower part corresponds to the receiver (Rx). In the centre is seen the block – *IFFT* which shapes input data into a certain number of sub-carriers. Within the receiver the subcarriers are demodulated by Fast Fourier Transform *FFT*, in which the functions from *IFFT* are reversed. Another interesting peculiarity of the *FFI/IFFT* conversion is the fact that they are almost identical; that is, *IFFT* can be converted into *FFT* by means of a complex conjugation of inputs and outputs of *FFT* and further on through their division by the output length of *FFT*. This allows the use of identical hardware on both transmitter and receiver sides provided there is no transmission and reception at one and the same point of time..

So far all activities involved prior to *IFFT* have not been the subject matter of a serious discussion Their investigation entails binary data to be encoded by corrective code in straight direction. Further this data is offset and after that converted to a quadratic amplitude modulated (*QAM)* value.

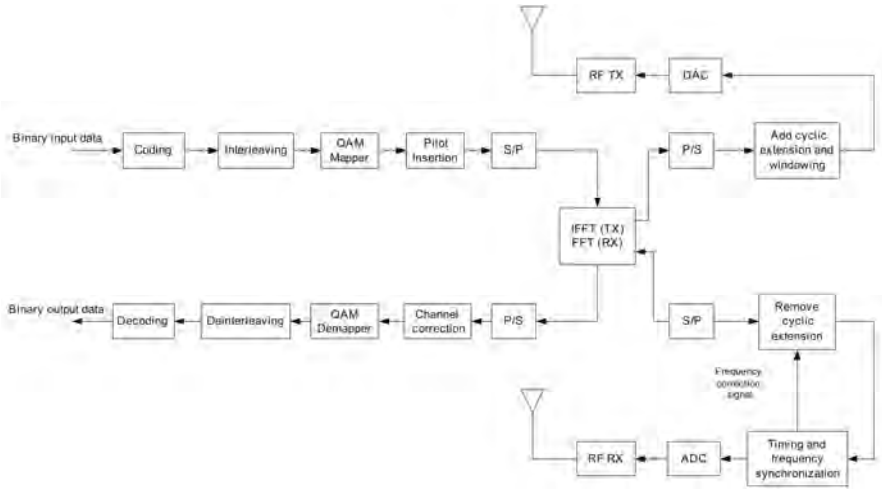


Fig. 1. Block diagram of OFDM transmitter

II. MULTI-PATH DISSIPATION

Within a certain radio connection the transmitter signal can be reflected by a number of objects such as blocks of flats, cars, mountains etc. This is the reason for the multi-path radio connection as well as for the reception of reflected signals at the receiver side. Fig.2 illustrates the possible ways of occurrence of multi-path transmission.

[1, 2] Kiril R. Koitchev and Stanimir M. Sadinov are with the Faculty of Electrical Engineering and Electronics, Technical University – Gabrovo, 4 H. Dimitar St., 5300 Gabrovo, Bulgaria, E-mail's: koitchev@tugab.bg; murry@tugab.bg

[3] Georgi Traikovski,MSc is with the Communications Technology and Equipment Department of Technical University- Gabrovo, 5300Gabrovo,Bulgaria, No4 "H.Dimitar" street e-mail:gorgi_to@yahoo.com

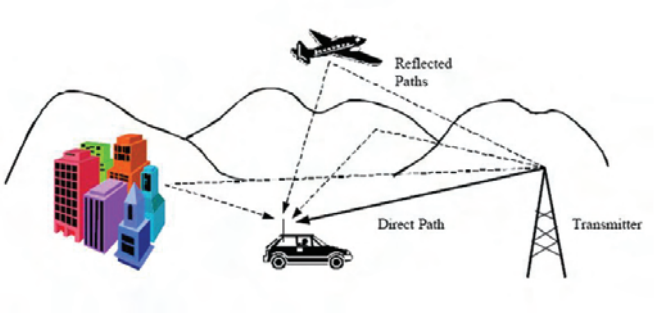


Fig. 2. Multi-path dissipation

This effect may cause serious problems which most often are regarded as undesired interference at the receiver side. Fig. 3 shows the levels of the reflected signal which may appear as resulting from the fading effect [2] [8].

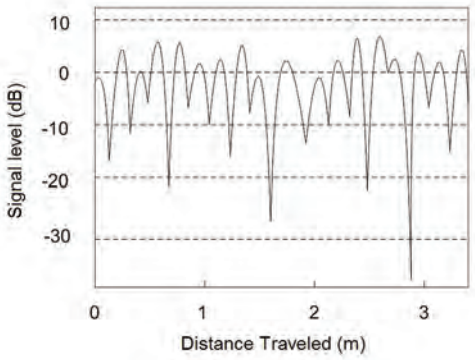


Fig. 3. Typical fading of Rayleigh when the mobile station is on the go (for at 900 MHz)

It is known that the distribution of Rayleigh is usually used for statistical description of time variation in the character of the received signal. Table 1 shows the variation in probabilities for the levels of a certain signal by using Rayleigh's distributions.

TABLE 1. CUMULATIVE DISTRIBUTION FOR RAYLEIGH DISTRIBUTION

Level of signal [dB]	Probable variation in signal level [%]
10	99
0	50
-10	5
-20	0.5
-30	0.05

III. DOPPLER EFEKT

Another well known fact is the difference in frequency of the source and reception signals when the distance between the source and the receiver varies due to their motion.When both move away from each other ,the frequency of the received signal is greater than that of the source; alternatively, as they come closer to each other, the frequency grows smaller [3].

Frequency variation with regard to Doppler effect depends on the relative movement between the data source and the receiver as well as on the rate of dissipation of the wave.Doppler displacement then can be expressed as

$$\Delta f \approx \pm f_0 \frac{\upsilon}{c},$$

where Δf is the frequency variation of the source, which is related to the received signal, f_0 , ν is the speed difference between the receiver and transmitter and c is the speed of light.

For example, if $f_0 = 1GHz$ and $\nu = 16{,}67 km/h$ than Doppler displacement will be

$$\Delta f = 10^9 \frac{16{,}67}{3.10^8} = 55{,}5Hz .$$

Basically, this displacement of $55Hz$ will not substantially mar the transmission itself. Yet, on the other hand, it may occur as a real problem if the transmission technology employed is very sensitive to frequency displacements in the carrier signals as is the case with OFDM.

IV. INVESTIGATION OF NOISE EFECT IN OFDM SYSTEMS

By means of the Simulink module of the software package Matlab it was possible to implement the simulation model from Fig. 1.Based on the simulation results it was found out that the signal-to-noise ratio (*SNR*) within OFDM systems is similar to a standard digital transmission with two carriers. Fig. 4 shows the results which bring about to correlation depending on the type of modulation used. It is clearly seen that in the case of employing quadrature phase manipulation (QPSK) the permissible value of signal-to-noise ratio *SNR* is from 10 to 12 dB. The value of *BER* becomes non-permissible when the signal-to-noise ratio drops below 6 dB. However, by using binary phase manipulation (BPSK) it is possible to eliminate this fault at the expense of the capacity of the transferred data.

By using BPSK with available OFDM the permissible value of *SNR* is from 6 to 8 dB.

In a channel of particularly low noise it is possible to use 16PSK which will boost system's capacity.

If *SNR* is 25 dB and more, 16PSK is used thus increasing the capacity two times as compared with QPSK.

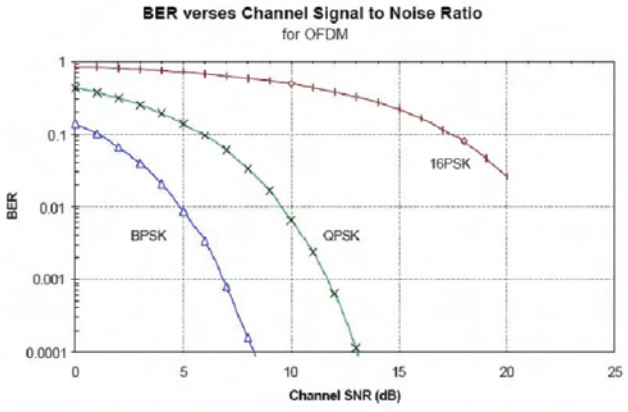


Fig. 4. *BER* compared to *SNR* for OFDM in using BPSK, QPSK и 16PSK

V. Simulation Results

The simulation results in these models are based on the change of the basic parameters which these systems feature such as the signal-to-noise ratio, Doppler effect and the number of symbol errors plus SER coefficient. The results are shown on a graph. Fig. 5 shows the graphic dependence of SER on the signal energy/the spectrum density of noise (E_s / N_o) with various Doppler frequencies which are in full compliance with the standards. It should be noted that these results are obtained before the RS coder which implies that external coding is not taken into account.

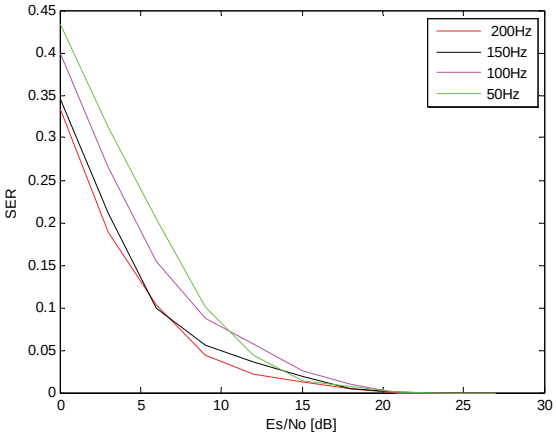


Fig. 5. Graphic presentation of *SER* with regard to E_s / N_o and $f_{dop.}$ before RS coder

As is evident from the graphs the coefficient *SER* decreases substantially as the signal-to-noise ratio (E_s/N_0) increases and is retained within the permissible boundaries.SER equals naught with large values in this ratio. Similarly, it is seen that better results are obtained with Doppler frequency values being high; the most optimal ones are obtained with frequency of *175 Hz.*

Fig.6 shows the variation in the number of errors *SER*, which appear under the same conditions but after the RS coder.

It is evident that these results are inter-related and the number of errors varies by the same law, i.e with greater signal-to-noise ratio they drop low and are of lesser value with higher Doppler frequencies.

The results of the simulation investigation are shown on a graph, however, in this case *Reed-Solomon* coding has been taken into account, therefore ,the bulk of transmitted data will be greater.

These simulation results are produced based on the change in the same parameters. The graph on fig.6 shows the dependence of *SER* from E_s / N_o with various Doppler frequencies.

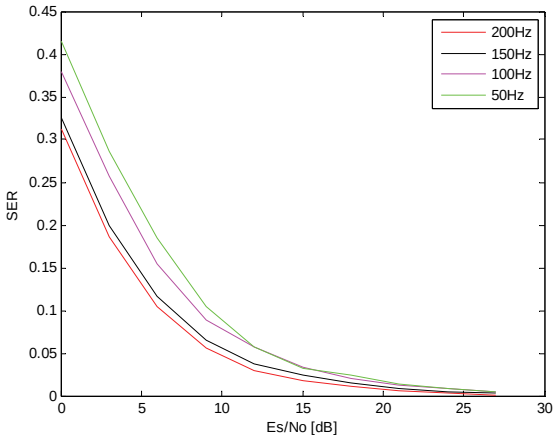


Fig. 6. Graphics determination of *SER* depending on E_s / N_o and $f_{dop.}$ after RS coder

Here, as is in the graphs of fig.5, it is evident that the *SER* error is negligible at higher values for the E_s / N_o ratio.

VI. Conclusion

The use of OFDM technology is appropriate as a type of modulation technique which can be successfully employed in high-tech telecommunications wireless networks as is the case with Wireless Local Area Network (WLAN) or the European Standard HIPERLAN/2. Here have been done just a number of tests of the OFDM technology followed by presentation of major parameters and characteristic features. Simulation results have been obtained on the basis of *SER*, *SNR*, E_s / N_o, $f_{dop.}$, *BER* parameters as well as some other parameters which are important for the delay in signal distribution. There are other parameters which have not been taken into account during these simulations such as the efficiency of frequency stability and the efficiency related to noises generated by the use of rectangular pulses, etc.

OFDM technology yields very good results when it is jointly used with CDMA (Code Division Multiple Access) no matter whether CDMA is used as a system of single carrier frequency or as a multiple-path system.

The difference in consumer capacity between OFDM and CDMA is reduced to whether there is sectorization of the cells and some available information concerning customers activity within the systems.

It may well be claimed that CDMA technology ensures good results only with multi-cellular systems where a single frequency is used for all cells. This enlarges the capacity for joint operation with the other systems and reduces intercellular interference.

A possible further investigation will be connected with the problems which arise in the employment of OFDM with numerous subscribers. One such potential problem could be the requirement for the transmitter to cover a much larger area in order to cope with the high level of signal instability between subscribers. At present this is accomplished by OFDM sys-

tems through the use of *forward error correction* which helps to improve the system's effectiveness and productivity.

This particular area is still not well investigated and more should be done to improve the patterns of forward error correction as is the case in telephony and data transmission systems [6] [7]. Several modulation techniques are used with OFDM such as BPSK, QPSK, 16-QAM and 256-PSK, but the real system configurations have the capacity to dynamically select the technique of modulation with regard to the data being transferred.

ACKNOWLEDGEMENT

This paper has been sponsored under the auspices of the "Planning of Multimedia Telecommunications Networks with Traffic and QoS Management" project – a part of the "SCIENTIFIC INVESTIGATIONS" Fund, contract BY-TH-105/2005.

VII. REFERENCES

[1] Keller, Thomas, and Lajos Hanzo. "Adaptive Multicarrier Modulation: A Convenient Framework for Time-Frequency Processing in Wireless Communications." *IEEE Proceedings of the IEEE* 88 (May, 2000): 609-640

[2] Wang, Zhengdao, and Georgios B. Giannakis. "Wireless Multicarrier Communications." *IEEE Signal Processing Magazine* (May, 2000): 29-48

[3] Bingham, John A. C. "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come." *IEEE Communications Magazine* (May, 1990): 5-14

[4] Van Nee, Richard, and Ramjee Prasad. *OFDM for Wireless Multimedia Communications*. Boston: Artech House, 2000.

[5] Naguib, Ayman F., Nambi Seshadri, and A. R. Calderbank. "Increasing Data Rate over Wireless Channels." *IEEE Signal Processing Magazine* (May, 2000): 76-92

[6] Mitra, Sanjit K. *Digital Signal Processing: A Computer-Based Approach*. New York: McGraw-Hill, 2001.

[7] O'Leary, Seamus. *Understanding Digital Terrestrial Broadcasting*. Massachusetts: Artech House, 2000.

[8] Penchev, P. Introduction to Telecommunications Network Measurements, *iCEST, XLI International Scientific Conference on Information, Communication and Energy Systems and Technologies*, 29th June – 01st July, 2006, Sofia, Bulgaria.

Cross-Phase Modulation Induced Nonlinear Phase Noise in Spectral Efficient WDM DPSK-OOK systems

Georgi K. Georgiev[1] and Nikolay M. Stoyanov[2]

***Abstract* – This paper provides a simple and efficient model for involving both SPM and XPM - induced nonlinear phase noise in WDM systems with mixed DPSK-OOK channels. An analytically based new model is derived from the two-channel pump-probe model and later extended to multi-spanWDM systems. The proposed model is developed to determine the effects of XPM induced nonlinear phase noise on DPSK channels from neighbouring OOK channels.**

***Keywords* – Cross-phase modulation (XPM), differential phase-shift keying (DPSK), nonlinear phase noise, WDM systems.**

I. INTRODUCTION

Recently, 40Gb/s advanced modulation formats such as DPSK and DQPSK, have been investigated to upgrade transmission capacity [2-9]. The tolerances to linear and nonlinear transmission impairments are studied in single channel transmission, where an implementation of novel NRZ-advanced modulation format, provides a better dispersion tolerance, but suffers from strong nonlinear impairments. The impairments caused by the cross phase modulation in high bit rate DWDM systems will affect the performance of these systems to a greater extent. To evaluate the optical system performance under the influence of cross phase modulation XPM, an analytical model for mixed DPSK and OOK multi channel system is developed in this paper, which is faster, simple and as accurate as possible. Previous analytical studies involved analyzing the intensity modulation generated phase noise in single channel or pure DPSK systems. This paper provides a simple and efficient model for involving both self phase modulation (SPM, single channel effect) and XPM (multi-channel effect) induced nonlinear phase noise [1] in mixed WDM optical systems. The roots of this model lie in the derivation of the analytical formula for sensitivity of 40Gb/s DPSK or 40Gb/s DQPSK channels from 10Gb/s OOK channels. These modulation formats can be an effective way to increase spectral efficiency and overcome transmission impairments. DQPSK is a very attractive format because it is just a little more complex than binary DPSK and suitable for upgrade of the existing DWDM systems, specially applied for high capacity transmission. It seem to have the potential for reducing cost by increasing dispersion tolerance and allowing transmissions to 40 Gb/s per channel or higher, using cheaper, established 10Gb/s components.

[1]Georgi K. Georgiev is with Ministry of the Interior of Bulgaria, E-mail: ggeorgiev@ncc.mvr.bg

[2]Nikolay M. Stoyanov is with Ministry of the Interior of Bulgaria, E-mail: nstoyanov@ncc.mvr.bg

II. THE ERROR PROBABILITY OF DPSK SIGNALS AS A FUNCTION OF SNR

The error probability is derived analytically for differential phase-shift keying (DPSK) signals contaminated by both self- and cross-phase modulation (SPM and XPM) induced nonlinear phase noise. XPM-induced nonlinear phase noise is modeled as Gaussian distributed phase noise. When fiber dispersion is compensated perfectly in each fiber span, XPM-induced nonlinear phase is summed coherently span after span and is the dominant nonlinear phase noise for typical wavelength-division-multiplexed (WDM) DPSK systems. For systems without or with XPM-suppressed dispersion compensation, SPM-induced nonlinear phase noise is usually the dominant nonlinear phase noise. For calculating the error rate of DPSK transmission we employ the model given in [2]. It uses a matched optical filter in front of the receiver for optimization of the receiver performance and assumes single polarization. Combined with the Gaussian approximation for the distribution of the XPM-induced nonlinear phase noise (Nicholson Model [3]), form [2] and [4] we get analytical expression for the error rate of DPSK:

$$p_e \approx \frac{1}{2} - e^{-\frac{\rho_s}{2}} \sqrt{\frac{\rho_s}{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \left(I_n\left(\frac{\rho_s}{2}\right) + I_{n+1}\left(\frac{\rho_s}{2}\right) \right) \times \times \left| \Psi_{\Phi NL}\left(\frac{(2n+1) < \Phi_{NL} >}{\rho_s + 1/2}\right) \right|^2 e^{\left(-\frac{2n+1}{2}\sigma^2_{XPM}\right)} \quad (1)$$

where the influence of neighboring channels is determined by phase variance σ^2_{XPM}, n is a summation index, $I_n(.)$ denotes modified Bessel function of first kind and n^{th} order, $<\Phi_{NL}>$ is the mean nonlinear phase shift, ρ_s is the SNR and the characteristic function of the nonlinear phase noise [4] is

$$\left|\Psi_\Phi(j\nu)\right| = \sec\sqrt{j\nu}\exp(\rho_s\sqrt{j\nu}\tan\sqrt{j\nu}). \quad (2)$$

The trigonometric function with complex argument can be calculated by the following relationship

$$(\sec\sqrt{j\nu})^{-1} = \cos\sqrt{\frac{\nu}{2}}\cosh\sqrt{\frac{\nu}{2}} - j\sin\sqrt{\frac{\nu}{2}}\sinh\sqrt{\frac{\nu}{2}} \quad (3)$$

The nonlinear phase shift is approximated by [4]

III. NONLINEAR PHASE NOISE VARINACE DUE TO CROSS-PHASE MODULATION

The Calculation of the nonlinear phase noise variance as a function of frequency is based on the two-channel pump-probe model of [5] and later extended to multi-span WDM systems.

A. Calculation of phase noise variance

For simplicity the model have two WDM channel, and co-propagating OOK channels are assumed to be co-polarized to cover the worst case scenario. The phase variance as a function of frequency separation between the channels [6] is:

$$\sigma^2_{xpm}(\Delta\lambda) = 4\int_{-1/T}^{1/T} \Phi_{P_2(f)} \left|H_{12}(f)\right|^2 \sin^2(\pi f T)df \,, \qquad (4)$$

where

$$\Phi_{P_2} = 2P_0 S_{sp1} + 2S_{sp1} v_{opt} \qquad (5)$$

is the noise power spectral density of the pump channel, $\sin^2(\pi fT)$ is phase transfer function of receiver delay line interferometer, S_{sp1} is the spectral density of the amplifier noise, V_{opt} is the optical filter bandwidth and P_0 is the lunched power of the pump channel. The integration is reduced from $\pm\infty$ to $\pm 1/T$ by taking into account only the phase noise over a bandwidth confined within the bit-rate. The dependence of the variance of Eq. (4) on $\Delta\lambda$ is originated from the dependence of the amplitude-phase conversion function of the fiber $H_{12}(f)$ on $\Delta\lambda$ and can be expressed from [7] and [8]as

$$H_{12}(f) = 2\gamma \frac{1 - e^{-\alpha L + j2\pi f d_{12} L}}{\alpha - j2\pi f d_{12}}, \qquad (6)$$

where γ is the nonlinear coupling coefficient, α is fiber attenuation coefficient, L is the fiber length, $d_{12} \approx D\Delta\lambda$ is the relative walk-off between two channels with wavelength separation of $\Delta\lambda$ and dispersion coefficient D.

B. Multi span WDM system

In WDM system with (2M + 1)-channels, for the worse case of the center channel, the XPM contribution has to be summed over all neighboring channels [8], the nonlinear phase noise variance in the first span is equal to:

$$\sigma^2_{XPM,1} = \sum_{k=1}^{M} \sigma^2_{XPM,0}(k\Delta\lambda) \qquad (7)$$

In case of full inline dispersion compensation - the worst case, the noise contribution sums up coherently and for a system with N fiber spans, the variance of the overall XPM-induced nonlinear phase noise depends on the method of dispersion compensation.

$$\sigma^2_{XPM,tot} = N^2\sigma^2_{XPM,1} + (n-1)^2\sigma^2_{XPM,2} + \ldots + \sigma^2_{XPM,N} \qquad (8)$$

When the whole spans have the same configuration and same noise variance, the total phase noise is

$$\sigma^2_{XPM,tot} = \frac{1}{6}N(N+1)(2N+1)\sigma^2_{XPM}\,. \qquad (9)$$

C. *The model*

The need of a simulation testbed is of strong interests for convenience of design, investigation and verification on benefits and shortcomings of advanced modulation formats on the transmission system. Simulation semi-analytical model is constructed to investigate the DPSK-OOK system and obtain the results. The block diagram of the simulation model utilized for this purpose is shown in Figure 1 and the eye diagram at the receiver on figure 2.

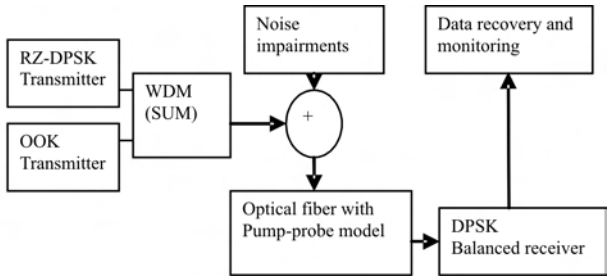


Fig. 1. Block diagram of simulated model

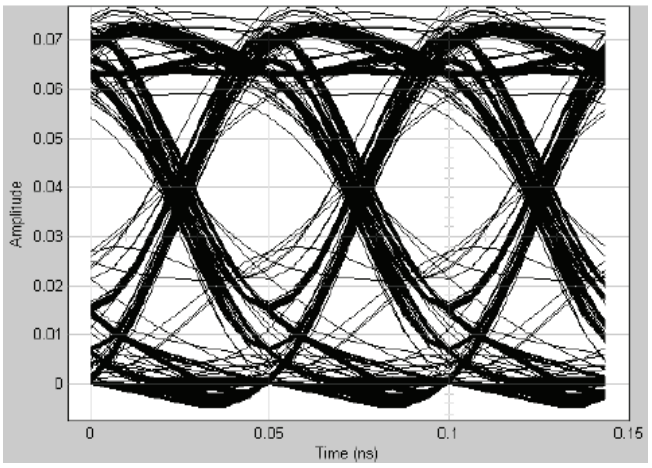


Fig. 2. Simulated eye diagram of RZ-optical pulses

IV. DISCUSSION AND RESULTS

One challenge in numerical simulations is to provide a reliable estimate of the BER. Due to limited computation time, a typical simulation programs uses only hundreds of bits and therefore, the BER is usually not counted directly but estimated by evaluating the statistics fluctuations in the received signal. It is a standard practice in numerical simulations of OOK systems such fluctuations to be often characterized by a Q-factor and BER to be obtained through the complementary error function (erfc). Figure 3 shows error probabilities as a function of SNR for WDM DPSK systems obtained through formula of equation (1).

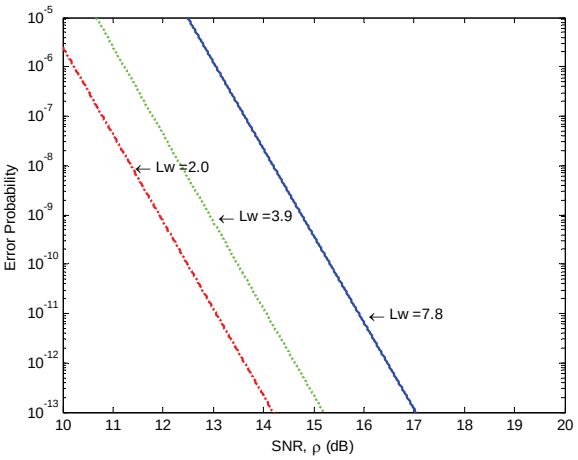


Fig. 3. Error probability as a function of SNR

Normally 10-Gb/s systems using non-zero dispersion-shifted fiber, 50-GHz spacing (NZDSF) and D = 4 ps/km/nm has a walk-off length Lw=T/d12 of 62.5 km. 40-Gb/s systems have a walk-off length of 7.8 km (100 GHz spacing, ($\Delta\lambda$ = 0.8 nm) and D = 4ps/km/nm). The walk-off length of figure 3 forms a geometric series corresponds to typical 10 and 40-Gb/s systems in NZDSF and standard single-mode fiber (SSMF) with dispersion coefficients of D = 4 and 16 ps/km/nm, respectively. The characteristic function of Eq. (2) assumes a dispersionless fiber. With fiber dispersion, due to walk-off effect, the nonlinear phase noise caused by cross-phase modulation should approximately have Gaussian distribution. The SNR penalty as a function of nonlinear phase shift is shown on figure. 4. For a perfect dispersion compensation such that XPM-induced nonlinear phase noise adds coherently, XPM-induced nonlinear phase noise gives the same SNR penalty as SPM- induced nonlinear phase noise when the walk-off length is about L_w = 7,8 km for $<\Phi_{NL}>$ less than 1 rad. With the same mean nonlinear phase shift, the nonlinear phase noise induced SNR penalty of 10-Gb/s systems is smaller than that for the corresponding 40-Gb/s systems.

We emphasize here that (1) is based on the assumption of Gaussian approximation for the distribution of the XPM-induced nonlinear phase. Because XPM-induced nonlinear phase noise is generated by the interaction of many bits or WDM channels, the Gaussian approximation is valid and XPM-induced degradation on DWDM system performance is not a strong function of number of channels. Therefore, 5-channel modeling system can be sufficient for system investigation on the effects of XPM The XPM effect can be integrated into the propagation model with ease. The mean nonlinear phase is approximated by $<\Phi_{NL}>\approx N\gamma L_{eff}P_0$ and the walk-off length is the main constraint for nonlinear interaction between two or multiple pulses when the XPM effects occur. More specifically, XPM can be considered to have no effects when the faster moving pulse completely walks through the slower moving pulses.

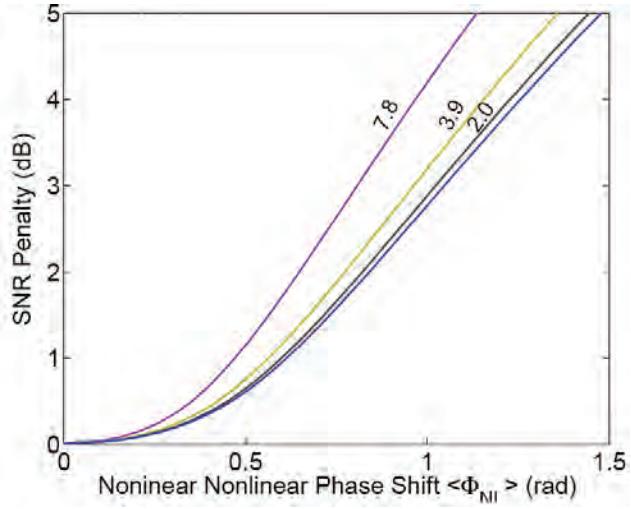


Fig. 4. SNR penalty as a function of nonlinear phase shift<ΦNL>

Figure 5 shows the signal to noise (SNR) penalty for DPSK (DQPSK) probe channel. The SNR penalty versus lunched power per channel for 5 channel system(100GHz channel separation) with 4 neighboring OOK channels When the co-propagating DPSK channels yield very low additional penalties on the probe channel, the penalty caused by co-propagating OOK channels increase fast with increasing lunch power P_0. It is important and this fact can be explained with signal dependent part of the noise variance has the most significant XPM contribution for OOK channel and not so important for the DPSK and DQPS channels. We have to confess with the increasing of the lunch power also increases absolute SNR and a settlement by compromise must be taken. The optimum operation point of every optical system is where linear and nonlinear impairments balance each other.

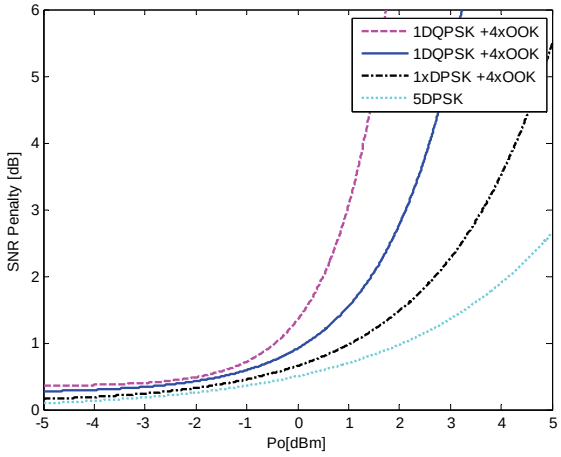


Fig. 5. SNR penalty versus lunched power

V. Conclusion

We have investigate the performance degradation of DPSK systems through the represented theoretical model for the influence of XPM induced nonlinear phase noise from co-propagating intensity modulated channels in the multi-span

WDM systems. The performance of the system was measured with different nonlinear phase shift. The BER curves dependence of XPM penalty is based on the amplitude phase conversion function of fiber, delay interferometer function and ASE noise.

The represented theoretical model for the influence of XPM induced nonlinear phase noise from co-propagating intensity modulated optical channels has demonstrated that for low penalties, it is beneficial DPSK channels to use higher rate than OOK. However, 40-Gb/s systems have four times the bandwidth and require four times the power of the corresponding 10-Gb/s systems having the same SNR.

REFERENCES

[1] G. P. Agrawal "Nonlinear fiber optics" , AT&T, Academic Press, San Diego, 1998

[2] Ho, K.-P. Compensation improvement of DPSK signals with nonlinear phase noise. *IEEE Photon. Technology Letters* vol. 15 (2003), no. 9, pp. 1216-1218.

[3] Nicholson, G. Probability of error for optical heterodyne DPSK system with quantum phase noise. *Electron. Lett.*, vol. 20 (1984), no 24, pp. 1005-1007.

[4] Ho, K.-P. Probability density of nonlinear phase noise. *J. Opt. Soc. Amer. B*, vol. 20 (2003), no. 9, pp. 1875-1879.

[5] Chiang, T.-K., Kagi, N., Marhic, M. E., and Kazovsky, L. G. "Cross-phase modulation in fiber links with multiple optical amplifiers and dispersion compensators. *J. Lightwave Technology*, vol. 14 (1996), no. 3, pp. 249-260.

[6] Hoon Kim, "Cross-phase-Modulation-Induced Nonlinear Phase Noise in WDM Direct-Detection DPSK Systems" Journal of Lightwave Technology, vol. 21, no. 8, pp. 1770-1774, August 2003.

[7] Chiang, T.-K., Kagi, N., Fong, T. K., Marhic, M. E., and Kazovsky, L. G. (1994). Cross-phase modulation in dispersive fibers: theoretical and experimental investigation of the impact of modulation frequency. IEEE Photon. Technol. Letters, vol. 6 no. 6, pp. 733-736.

[8] Chiang, T.-K., Kagi, N., Marhic, M. E., and Kazovsky, L. G. (1996). Cross-phase modulation in fiber links with multiple optical amplifiers and dispersion compensators. J. Lightwave Technol., vol.14, no. 3, pp.249 -260

[9] Ho, K.-P. Performance degradation of phase-modulated systems with nonlinear phase noise. *IEEE Photon. Technology Letters*, vol. 15 (2003)., no 9, pp. 1213-1215.

Comparative Analysis of Cooperative Diversity and Classical Spatial Diversity

Bojan Dimitrijević[1], Jasmina Spasić[2] and Zorica Nikolić[3]

Abstract – **In this paper is presented comparative analysis of systems in which spatial diversity is achieved differently. The performances of systems with classical spatial diversity are compared with cooperative diversity systems. For cooperative diversity systems, performance are evaluated in both symmetric and asymmetric interuser-channel cases.**

Keywords **– Spatial diversity, cooperative diversity.**

I. INTRODUCTION

In wireless networks, signal fading arising from multipath propagation is a particularly severe form of interference that can be mitigated through the use of diversity techniques. Space, or multi-antenna, diversity techniques are particularly attractive for implementation. Cooperative diversity is a method that enables single antenna mobiles in a multi-user environment to share their antennas and generate a virtual multiple-antenna transmitter that allows them to achieve spatial diversity.

II. SPATIAL DIVERSITY

Mobile users' data rate and quality of service are limited by the fact that they experience severe variations in signal attenuation, thereby necessitating the use of some type of diversity. Spatial diversity relies on the principle that signals transmitted from geographically separated transmitters, and/or to geographically separated receivers, experience fading that is independent. In this paper is analyzed system with DPSK modulation and two diversity branches.

Bit error probability (BEP) for DPSK modulated signal at output of a receiver with L branches is given by [1]:

$$P_b = \frac{1}{2^{2L-1}(L-1)!(1+\gamma_d)^L}\sum_{i=0}^{L-1} b_i (L-1+i)!\left(\frac{\gamma_d}{\gamma_d+1}\right)^i \quad (1)$$

where γ_d is signal-to-noise ratio (SNR), and:

$$b_i = \frac{1}{i!}\sum_{n=0}^{L-1-i}\binom{2L-1}{n}.$$

III. COOPERATIVE DIVERSITY

Classical multiple-input multiple-output systems exploit spatial diversity through multiple transmit and/or receive antennas. However, due to the size of a mobile and carrier frequency constraints, achieving spatial diversity through multiple antennas may not be possible. Cooperative diversity is an alternative way to achieve diversity with a single antenna at the mobile. As there are many users in the wireless system, cooperative diversity allows sharing the resources among cooperative users and each user can serve as a relay for the transmissions of other users' information. Diversity is achieved by the relay transmissions by other users in the system.

In this paper is presented the two-user cooperative diversity system, which uses a differential modulation scheme. Both users communicate to the same destination (e.g., base station, BS). Both users send information in the same frequency band, therefore, multiple access is done by time-division multiple access (TDMA). The transmission is half-duplex. The cyclic redundancy check (CRC) bits are used to detect whether the information has been received correctly.

A. System model

The paper is analyzing system in which two cooperative users are transmitting their information to the BS. Each user and the receiver at BS has single antenna. Both users share the same frequency band and the mobile of each user cannot transmit and receive a signal at the same time. The TDMA scheme is shown in Fig. 1. A total time frame is divided into three time intervals. During the first time interval, first user sends its information. The second time interval is reserved for the second user information. The third time interval is used by both users for relaying each other's message to the BS.

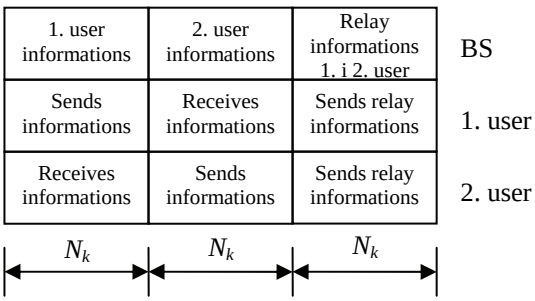


Fig. 1. TDMA scheme for analyzed system.

[1]Bojan Dimitrijević is with the Faculty of Electronic Engineering, Niš, Serbia, E-mail: dbojan@elfak.ni.ac.yu

[2]Jasmina Spasic is scholar of Ministry of science and environmental protection, Serbia, E-mail: mima@elfak.ni.ac.yu

[3]Zorica Nikolic is with the Faculty of Electronic Engineering, Niš, Serbia, E-mail: zora@ni.ac.yu

The channels between users (interuser channels) and user-BS channels are mutually independent, with frequency flat feding. One time frame consists of N symbol intervals, and each user, in one time interval, sends information in $N_k = N/3-1$ symbol intervals.

In the first time interval, the first user mobile transmits its own message and the BS receiver and the second user mobile receive the signal. At the BS receiver, the received signal corresponding to the direct transmission $y_{1d}[n]$ is given as

$$y_{1d}[n] = a_{1d}x_1[n] + w_d[n] \quad (2)$$

where $x_1[n]$ is a symbol transmitted from the first user at the nth symbol interval, $n = 0, 1, \ldots, (N/3) - 1$. During this time interval, at the second user mobile, the received signal $y_{12}[n]$ is written as

$$y_{12}[n] = a_{12}x_1[n] + w_2[n]. \quad (3)$$

In the second time interval, the roles of the first and the second user are switched. Therefore, at the BS receiver and at the first user mobile, the respective received signals $y_{2d}[n]$ and $y_{21}[n]$ are

$$y_{2d}[n] = a_{2d}x_2[n] + w_d[n],$$

$$y_{21}[n] = a_{21}x_2[n] + w_1[n]. \quad (4)$$

where $x_2[n]$ is a symbol transmitted from the second user at the nth symbol interval, $n = N/3, (N/3) + 1, \ldots, (2N/3) - 1$.

In the third time interval, both user mobiles act as relays and transmit the relay signals at the same time. The signal arriving at the BS is a linear combination of signals from both transmission paths, which can be written as

$$y_d[n] = a_{1d}x_{1r}[n] + a_{2d}x_{2r}[n] + w_d[n] \quad (5)$$

where $x_{1r}[n]$, $x_{2r}[n]$ are relayed symbols transmitted from the first and the second user mobile, respectively, at the nth symbol interval $n = 2N/3, (2N/3) + 1, \ldots, N - 1$.

In Eq. (2)-(5), the coefficients, a_{ij}, $i,j \in \{1, 2, d\}$, are associated with each transmission path from i to j. They represent channel gains, which take into account fading and shadowing, and incorporate transmission signal energies. They are modeled as independent zero-mean complex Gaussian random variables. The subscripts "1" and "2" represent the first and the second user mobiles and the subscript "d" represents the BS. Since the channels are fixed during the multiple-access time frame, the time variable n is omitted in $a_{ij}[n]$. The $w_i[n]$, $i \in \{1, 2, d\}$ representing additive white noise and interference, are modeled as independent zero-mean complex Gaussian random variables with variances N_i and $i \in \{1, 2, d\}$, respectively. The signal-to-noise ratios (SNR) associated with each link shown in Fig. 2 are defined as

$$\gamma_d = E\left[\left|a_{1d}\right|^2\right]/N_d = E\left[\left|a_{2d}\right|^2\right]/N_d, \quad (6a)$$

$$\gamma_{k_1} = E\left[\left|a_{12}\right|^2\right]/N_2, \quad \gamma_{k_2} = E\left[\left|a_{21}\right|^2\right]/N_1. \quad (6b)$$

The γ_{k_1}, γ_{k_2} are referred to as interuser SNRs. For simplicity in the comparison with the no-cooperation case, both SNRs in the user-BS links are set to be equal to γ_d.

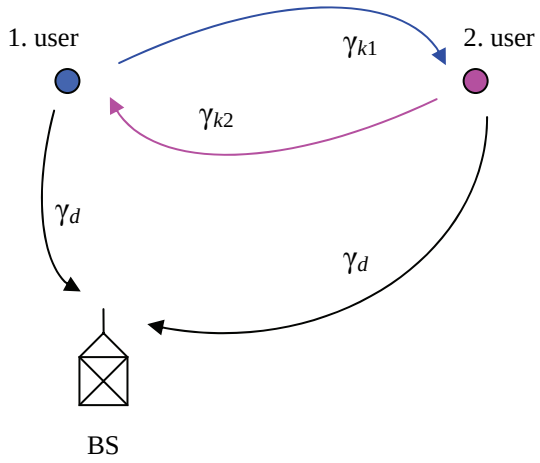


Fig. 2 System with cooperative diversity.

If the quality of interuser channels is low, several relay symbols will be in error. These relay symbols in error result in performance degradation even when the direct transmission channels are good. Therefore, it turns out that it is better not to transmit anything from the relay mobile when the interuser channels are not good. The decision to transmit the relay signals or not depends on the amplitudes of the fading gains between the source mobile and the relay mobile. If these values fall below a certain threshold, the relay mobile is not allowed to transmit the relay signal but instead transmits its own information signal. The CRC bits are added before differential encoding at the source mobile. Then, at the relay mobile, the receiver performs a CRC check after differential decoding to see if the whole frame has been received correctly. If the frame is received correctly, the relay mobile is allowed to perform differential re-encoding and forward the relay signal, otherwise, it does nothing. For the purpose of this paper, 16-bit CRC is applied and its error detection is perfect.

IV. Performance Analysis

As a performance measure in this paper is used BEP. The analysis here includes both the symmetric and asymmetric interuser cases.

When the cooperative users have perfect information about each other's transmission sequence, there will be no misleading relay information transmitted. This ideal case of the perfect interuser information represents the limiting performance of cooperative diversity. In the perfect interuser case, the BEP does not depend on the interuser channel. It depends on how the destination decodes the signals.

For the analyzed system, it might happen that one of the mobiles stops its relay transmission (one user does not have perfect information of the other). Therefore, the computation of the BEP is divided into two cases.

1) Full Relaying: Full relaying occurs when both user mobiles transmit their relay signals. For a single receive antenna at the destination, the BEP is given by [3]:

$$P_{b,\ full\ releying} = \frac{1}{2}\left[1-\sqrt{\frac{\gamma_d}{\gamma_d+6}}\left(\frac{\gamma_d+9}{\gamma_d+6}\right)\right]. \quad (7)$$

γ_d is defined in Eq. (6a).

2) Single Relaying: When one of the user mobiles stops its relay transmission, the destination receives only the relay signal from the other user. For a single receive antenna at the destination, the BEP is given by [3]:

$$P_{b,\ si\ ngle\ releying} = \frac{1}{2}\left[1-\sqrt{\frac{\gamma_d}{\gamma_d+2}}\left(\frac{\gamma_d+3}{\gamma_d+2}\right)\right]. \quad (8)$$

γ_d is defined in Eq. (6a).

A user mobile transmits the relay signal when the whole frame has been received correctly. This implies that no symbols in error will be transmitted in the relay signal. The probability of a user mobile transmitting the relay signal is $(1 - P_{\mathrm{DPSK},k_i})$, where P_{DPSK,k_i} is the frame error probability of DPSK belonging to the *i*th user mobile. The frame error probability of DPSK is given by [3]:

$$P_{\mathrm{DPSK},k_i} = 1-\left(\frac{1}{2}\right)^N\left[1+\sum_{k=1}^{N}\prod_{l=1}^{k}\frac{N+1-l}{l+\dfrac{l}{\gamma_{k_i}}}\right]. \quad (9)$$

The BEP for the first user transmission can be derived considering the following four situations.

1) *No users transmit a relay signal*: The probability of this case is $P_{\mathrm{DPSK},k_1}\cdot P_{\mathrm{DPSK},k_2}$. Bit errors occurring at the destination are from DPSK decoding of the direct transmission signal belonging to the first user.
2) *Only the second user transmits a relay signal*: The probability of this case is $(1 - P_{\mathrm{DPSK},k_1})\cdot P_{\mathrm{DPSK},k_2}$. Bit errors occurring at the destination are from differential decoding with single relaying.
3) *Only the first user transmits a relay signal:* The probability of this case is $P_{\mathrm{DPSK},k_1}\cdot(1 - P_{\mathrm{DPSK},k_2})$. Bit errors occurring at the destination are from DPSK decoding similar to the first case.
4) *Both users transmit a relay signal:* The probability of this case is $(1 - P_{\mathrm{DPSK},k_1})\cdot(1 - P_{\mathrm{DPSK},k_2})$. Bit errors occurring at the destination are from differential decoding with full relaying.

The BEP is found by averaging BEPs of the above four cases. This can be written as

$$\begin{aligned} P_b &= P_{\mathrm{DPSK},k_1}\cdot P_{\mathrm{DPSK},k_2}\cdot P_{b,\mathrm{DPSK}} \\ &+ (1-P_{\mathrm{DPSK},k_1})\cdot P_{\mathrm{DPSK},k_2}\cdot P_{b,\ si\ ngle\ releying} \\ &+ P_{\mathrm{DPSK},k_1}\cdot(1-P_{\mathrm{DPSK},k_2})\cdot P_{b,\mathrm{DPSK}} \\ &+ (1-P_{\mathrm{DPSK},k_1})\cdot(1-P_{\mathrm{DPSK},k_2}) \\ &\cdot P_{b,\ full\ relaying} \end{aligned} \quad (10)$$

where $P_{b,\ \mathrm{DPSK}} = 1/(1 + 2\gamma_d)$ is a conventional BEP of DPSK under Rayleigh-fading channels. When the $\gamma_{k_1} = \gamma_{k_2} = \gamma_k$, (10) reduces to

$$\begin{aligned} P_b &= P_{\mathrm{DPSK},k}\cdot P_{b,\mathrm{DPSK}} + (1-P_{\mathrm{DPSK},k})\cdot P_{\mathrm{DPSK},k} \\ &\cdot P_{b,\ si\ ngle\ releying} + (1-P_{\mathrm{DPSK},k})^2\cdot P_{b,\ full\ relaying} \end{aligned} \quad (11)$$

In order to compare cooperative diversity and classical spatial diversity with a different rate loss due to a reference symbol and CRC bits, let us define the pure information bit SNR as

$$\mathrm{SNR} = \frac{\gamma_d}{R}\cdot\frac{N+1}{N+1-\alpha} \quad (12)$$

where $R = 1$, $\alpha = 0$ for DPSK and $R = 2/3$, $\alpha = 1$ + no. of CRC bits for the cooperative diversity. The frame length (N + 1) is 131 symbol intervals which accommodate 130 information symbols.

V. Numerical Results

A. Symmetrical interuser channel

The performances of systems with spatial diversity and cooperative diversity are shown in Fig. 3.

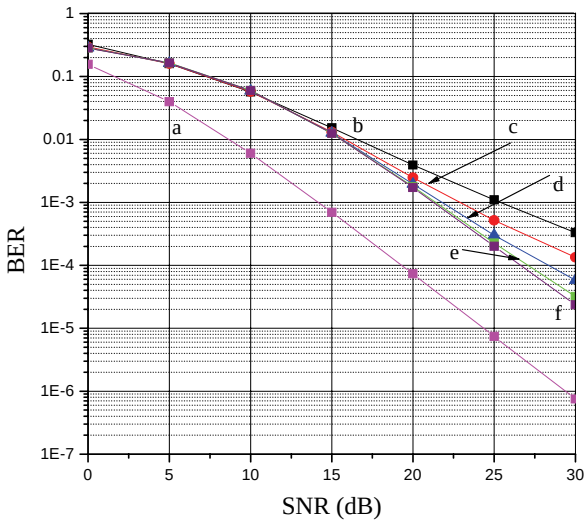


Fig. 3. Performances of systems with spatial diversity and cooperative diversity under symmetric interuser channels. a) system with spatial diversity, cooperative diversity: b) γ_k = 10 dB, c) γ_k = 15 dB, d) γ_k = 20 dB, e) γ_k = 25 dB and f) γ_k = 30 dB.

For cooperative diversity is considered case of symmetrical interuser channels, when performances of both users are identical. In Fig. 3, line (a) represents performance of system with spatial diversity with two branches. Lines (b), (c), (d), (e) and (f) defines performance of cooperative diversity for different interuser channel.

A. Asymmetrical interuser channel

Fig. 4 shows performance of systems with spatial diversity and cooperative diversity, when the interuser channels are asymmetric and in that case each user has different performance. This situation might occur when one user experiences local interference, while the other does not. For example, one user might be surrounded by local interfering devices or mobiles which do not affect the other user since the distance is too far. The asymmetric interuser case can also occur when power control is used and due to different distances from the base station, the users' transmit powers could be different.

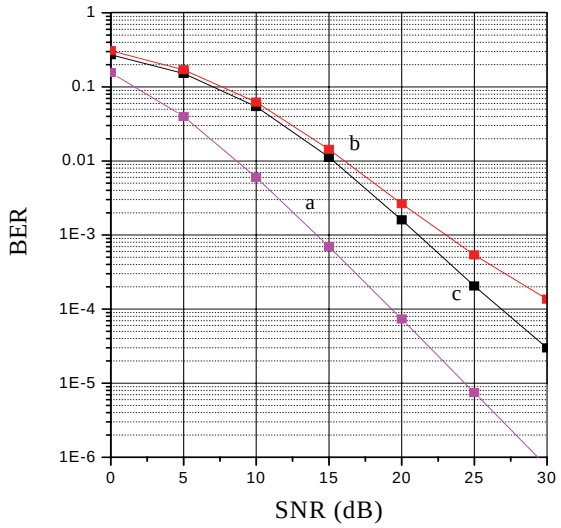


Fig. 4. Performances of systems with spatial diversity and cooperative diversity under asymmetric interuser channels. a) system with spatial diversity, cooperative diversity γ_{k1} = 25 dB and γ_{k2} = 15 dB: b) second user, c) first user.

In Fig. 4, line (a) represents performance of system with spatial diversity with two branches. The lines (b) and (c) defines performance of second and first user, respective, which are cooperating.

As seen on Fig. 3-4, system with classical spatial diversity shows better performance in whole area of interest. Although, in situations where exists some limitations concerning number of available antennas, cooperative diversity offers solution for minimizing effects of fading.

VI. CONCLUSION

In this paper are presented two ways of achieving spatial diversity. Classical spatial diversity considerable improves performances, but in systems where classical implementation of spatial diversity is not applicable due to various limitations, cooperative diversity provides an effective way for a collection of wireless users to relay signals for one another in order to exploit benefits of spatial diversity in the channel.

REFERENCES

[1] T. Eng, N. Kong, L. B. Milstein, "Comparison of Diversity Combining Techniques for Rayleigh-Fading Channels," IEEE Trans. Commun., vol. 44, pp. 1117-1129, September 1996.

[2] A. Sendonaris, E. Erkip i B. Aazhang, "User cooperation diversity—Part I: System description," IEEE Trans. Commun., vol. 51, Nov. 2003, pp. 1927-1938.

[3] P. Tarasak, H. Minn i V. K. Bhargava, "Differential Modulation for Two-User Cooperative Diversity Systems", IEEE Journal on Selected Areas in Communications, vol.23, no.9, Sept. 2005.

[4] J. N. Laneman, D. N. C. Tse i G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," IEEE Trans. Inf. Theory, vol. 50, no. 12, Dec. 2004, pp. 3062-3080.

[5] T. E. Hunter and A. Nosratinia, "Cooperation diversity through coding," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2002, p. 220.

Voice Over WLAN 802.11g in Presence of AWGN and Rayleigh Fading

Igor Angelovski[1] and Liljana Gavrilovska[2]

Abstract - **In this paper we investigate the quality of voice communication over 802.11g WLANs (Wireless Local Area Networks) taking realistic channel models into accounts. We show that the assessment of voice over WLANs (VoWLANs) performance is overly optimistic if fading is ignored when compared with systems where fading is taking into account. Via Matlab simulation, we derive the voice performance in terms of bit error rate (BER) and packet error rate (PER) for 54 Mbps data rate specified in the IEEE 802.11g ERP-OFDM (Extended Rate Physical Layer - Orthogonal Frequency Division Multiplexing) operational mode. More precisely, we evaluate BER and PER in IEEE 802.11g ERP-OFDM operation mode versus signal-to-noise ratio (SNR) in the presence of AWGN (additive white Gaussian noise) and one-path Rayleigh fading. The results obtained from the simulation show the benefit of implementation of pilots and long training sequence, especially for Rayleigh fading compensation.**

Keywords – **VoWLANs, BER, PER, IEEE 802.11g ERP-OFDM, AWGN, Rayleigh fading.**

I. INTRODUCTION

IEEE 802.11 based Wireless Local Area Networks (WLANs) are becoming popular in home, enterprise and public access areas primarily due to their low cost, simplicity of installation and high data rates. While WLANs continue to be predominantly data centric, there is growing interest in using WLANs for voice, especially in enterprise markets. Since conventional WLANs have been designed for packet data, communicating voice over WLANs has its own challenges. For example, packet voice communication is sensitive to delay, but relatively less sensitive to packet losses compared to communication of data packet. In addition, multiple choices of speech codec, the packetization interval and the PHY layer bit rate are available.

IEEE 802.11g standard boosts the wireless LANs speed to 54 Mbps using orthogonal frequency division multiplex (OFDM). OFDM is adopted as the mandatory modulation in the IEEE 802.11g WLANs specification [1] for a further high-speed physical layer (PHY) extension to the 802.11b standard in the 2.4 GHz Industrial-Scientific-Medical (ISM) band. This specification is backward compatible with the widely deployed IEEE 802.11b and defines multiple PHY operational modes. Our analysis consider ERP-OFDM operational mode.

[1] Igor Angelovski is with the Faculty of Electrical Engineering – Skopje, University "Ss. Cyril and Methodius", igorangelovski@yahoo.com

[2]Liljana Gavrilovska is with the Faculty of Electrical Engineering – Skopje, University "Ss. Cyril and Methodius", liljana@etf.ukim.edu.mk

Wireless channel manifest signal variation in the form of path-loss, shadowing and small-scale fading. Path-loss is the distance-dependent attenuation in signal power. Shadowing is the large-scale variation in signal power due to large objects (objects with sizes larger than the wavelength) in the environment. Small-scale fading is the rapid variation in signal envelope due to the transmitted signal undergoing reflection(s) and scattering leading to multiples copies with different amplitudes, phases and delays summing up at the receiver [2].

This paper studies the impact of one-path Rayleigh fading and AWGN on the voice performance over IEEE802.11g ERP-OFDM operation mode. Using appropriate channel model, we evaluate BER and PER in the terms of SNR, which is defined as the ration of the desired power signal to the noise power. SNR indicates the reliability of the link between the transmitter and receiver, especially in environments with strong influences from different electronic devices.

The rest of the paper is organized as follows. Section 2 details related work. Section 3 describes the system model and frame structure used for simulation. Section 4 gives the simulation parameters and results. Section 5 comprises the conclusion.

II. RELATED WORK

Generally, research pertaining to VoWLANs has been centered on the effects that packet delay and packet loss occurring at the MAC layer and higher layers have on voice performance. In most of these studies, errors in transmission over the wireless channel are either ignored or represented by simplified models which do not necessarily reflect realistic wireless environments [3], [4].

There is prior literature on the performance of the physical layer of WLANs in realistic wireless channels, taking into account fading. In [5], [6], [7] the authors simulated the physical layers of HIPERLAN/2 and IEEE 802.11a systems (which are very similar) in fading channels using the ETSI channel models [8], and they obtained the average packet error rate (PER) by averaging over the channel fading realizations. Using the average PER, the impact of various ETSI channel models on the performance of these systems was studied. In [9] the average throughput for both systems, derived from the average PER, is used for link adaptation purposes.

Lampe et al. simulated the HIPERLAN/2 system using the ETSI channel models and showed that using average PER for link adaptation can lead to "fatal misadaption" [10]. Therefore, the authors proposed that the PER corresponding to

the particular fading realization experienced should be used for link adaptation, and not the average.

Olufunmilola et al. study the impact of fading in the wireless channel on packet loss and the performance of VoIP over IEEE802.11a using a realistic channel model. VoIP performance is measured in terms of the user- perceived quality over a speech period comprising a number of VoIP packets. They consider VoIP stationary users that are subjected to a particular fading realization; this leads to a particular value for PER which in turn translates to a mean opinion score (MOS) value representing the voice quality for calls under that fading realization. By considering a large number of fading realizations, they derive a distribution for voice quality across users.

III. SIMULATION MODEL

This section describes the end-to-end simulation model shown in Fig. 1. The model contains all elements used in IEEE 802.11g WLAN network. The both transmit and receive side are explained. The major blocks are Scrambler, Encoder, Puncturer, Interleaver, and Modulator. In addition, we describe the frame structure, coding and modulation techniques used in IEEE 802.11g ERP-OFDM operational mode. The simulation of this model is performed in Matlab.

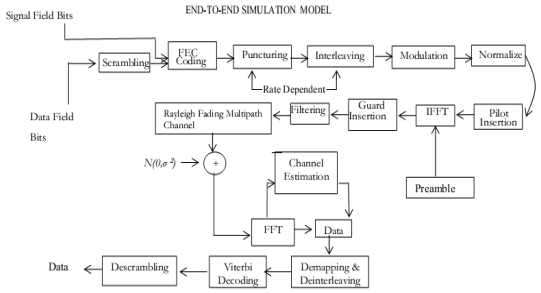


Fig.1 End-to-end simulation model

Simulation model elements: *Scrambler* is a block that prevents long runs of 1s and 0s in the input data being input to the remainder of the modulation process. The data are scrambled with pseudo random sequence of length 127.

The *Convolutional Encoder* block encodes a sequence of binary input vectors to produce a sequence of binary output vectors. This block can process multiple symbols at time. If the encoder takes k input bit streams (that is, can receive 2^k possible input symbols), then this block's input vector length is L*k for some positive integer L. Similarly, if the encoder produces n output bit streams (that is, can produce 2^n possible output symbols), then this block's output vector length is L*n. The input can be a sample-based vector with L=1, or a frame based column vector with any positive integer for L.

Puncturing is a procedure for omitting some of the encoded bits in the transmitter (thus reducing the number of transmitted bits and increasing the coding rate) and inserting a dummy "zero" metric into the convolutional decoder on the receive side in place of the omitted bits.

Interleaver is the next important block. The interleaver is defined by a two-step permutation. The first permutation ensures that adjacent coded bits are mapped onto nonadjacent subcarriers. The second ensures that adjacent coded bits are mapped alternately onto less and more significant bits of the constellation and, thereby, long runs of low reliability (LSB) bits are avoided. Interleaving is a key component of many digital communication systems involving forward error correction (FEC) coding. Applications that store or transmit digital data require error correction to reduce the effect of spurious noise that can corrupt data. Digital communication systems designers can choose many types of error-correction codes (EECs) to reduce the effect of errors in stored or transmitted data depending upon the type of modulation desired.

OFDM is a *modulation technique* that uses a large number of parallel narrowband subcarriers instead of a single wideband carrier to transport information, i.e. OFDM is a parallel transmission scheme, where a high-rate serial data stream is split up into a set of low-rate substreams, each of which is modulated on separate subcarriers. Thereby, the bandwidth of the subcarriers becomes small compared with the coherence bandwidth of the channel, i.e. the individual subcarriers experience flat fading, which allows for simple equalization. To overcome the effect of multipath propagation, short guard interval is introduced.

Frame structure: The look of IEEE 802.11g ERP-OFDM frame format is shown in Fig. 2. The short and long training sequences (STS, LTS), and 64-point Inverse Fast Fourier Transform (IFFT) are used to generate short and long training sequences of the preamble. STS is used for used for AGC convergence, diversity selection, timing acquisition, and coarse frequency acquisition in the receiver. The duration of the short training sequence is 8 μs and it consists of ten identical patterns, each 16 samples long. LTS are used for used for channel estimation and fine frequency acquisition in the receiver. The duration of the long training sequence is also 8 μs and it consists of three parts. The second and third parts are identical and are obtained directly from the 64-point IFFT of the long training sequence. The first part, however, is the second half of the second or third parts. The header or SIGNAL field is twenty-four bits long and contains the data rate and packet length information. This information is protected by one parity bit. The SIGNAL field also contains six tail bits to flush the convolutional encoder. The SIGNAL field is first protected by rate ½ convolutional code, then interleaved and finally modulated by binary phase shift keying. The SIGNAL field is then passed through the 64-point IFFT and the last one-quarter of the signal is appended to make and packet length information. This information is protected by one parity bit. The SIGNAL field is then passed through the 64 point IFFT and the last one-quarter of the signal is appended to make the length of the SIGNAL field 80 samples. The duration of the SIGNAL field is 4 μs [guard interval plus one OFDM symbol]. The DATA field contains information about seeding the descrambler, the random data and tail bits - to return the trellis in the Viterbi decoder to the all zero state.

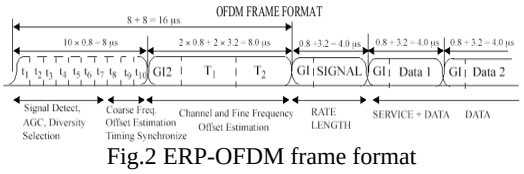


Fig.2 ERP-OFDM frame format

Coding techniques: The baseband supports three coding techniques in addition to four modulation formats. The basic coding rate is $^1/_2$ while the derived rates, through puncturing, are $^2/_3$ and $^3/_4$. The data rates vary from as low as 6 Mbps [BPSK, $^1/_2$] to as high as 54 Mbps [64-QAM, $^3/_4$]. The coding, puncturing and modulation are done according to the IEEE 802.11g section in the specification. The data is scrambled, encoded, punctured [depending on the rate], interleaved, modulated and then passed through the IFFT. The DATA field is divided into an integer number of OFDM symbols where each OFDM symbol is 4 μs in duration and 80 samples long. The scrambler is again IEEE 802.11g compliant and scrambles the information in the DATA field only. The same scrambler is used to scramble transmit data and to descramble receive data.

For the purpose of our simulation, the DATA field is composed of 5 OFDM symbols, which is modulated using 64-QAM with coding rate ¾ that gives data rate of 54Mbps. The long training sequences are used for channel estimation and DATA field compensation. The data is passed through the fast Fourier transform, and is then compensated by the estimated channel gain and phase - frequency domain equalization. The relevant data in each OFDM symbol only occupies 48 subcarriers and four subcarriers are reserved for the pilots. The pilots' subcarriers are used in order to make the coherent detection robust against frequency offsets and phase noise. The pilots shall be BPSK modulated by a pseudo binary sequence to prevent the generation of spectral lines.

The SIGNAL field is demodulated, deinterleaved and decoded using a hard decision Viterbi decoder. The DATA field on the other hand is also depunctured [depending on the rate] before decoding and finally descrambled. The pilots in SIGNAL field are used for SIGNAL field data compensation. The descrambler seed is extracted from the received signal. The performance simulation metric is the BER and PER.

IV. Simulation Parameters and Results

The model presented in Fig. 1 is able to simulate IEEE 802.11g ERP-OFDM operational mode at 54Mbps using OFDM modulation technique. Numerical values for the OFDM parameters are given in Table 1 [9]. In order to prevent intersymbol interference (ISI), a guard interval is implemented by means of a cyclic extension. Thus, each OFDM symbol is preceded by a periodic extension of the symbol itself. The total OFDM symbol duration is $T_{total}=T_g+T_U$ where T_g is the guard interval and T_U is the useful symbol duration. When the guard interval is longer than the excess delay of the radio channel, ISI is eliminated.

We introduced noise and Rayleigh fading where only amplitude is fluctuated. The Rayleigh fading raises the error rate of the received data. In this simulation we use one-path Rayleigh fading as described in [12]. Noise is modeled as additive white Gaussian noise (AWGN). The propagation characteristics can be estimated by *pilot subcarriers insertion* or by *using long training sequences* of the preamble. In this simulation, the first method is applied for SIGNAL field and the second for DATA field.

Parameter	Value
Sampling Rate (f_s)	20 MHz
Useful Symbol Duration (T_U)	3.2 μs
Guard Interval Duration (T_g)	0.8 μs
Total Symbol Duration (T_{Total})	4.0 μs
Number of data sub-carriers (N_D)	48
Number of pilot sub-carriers (N_P)	4
FFT Size	64
Sub-carrier spacing (Δ_f)	0.3125 MHz
Total Bandwidth (B)	16.875 MHz

Table 1. OFDM Parameters

Speech data generated by the encoder is packetized incurring overheads at the various layers (transport, network and MAC layers). We consider G.711 which generates speech data at the rate of 64 Kbps, and consider each packet to contain 10 ms of speech, or 80 bytes of data. At the transport layer, the RTP and UDP protocols introduce 12 byte and UDP 8 byte headers, respectively. At the network layer, the IP protocol introduces a header of 20 bytes. All headers and voice data are included in DATA field MAC frame (Fig. 2) requiring 5 OFDM symbols. This means that the length of DATA field is 135 bytes. At the MAC layer the header is 34 bytes. Hence a 10 ms voice packet in our simulation is 169 bytes long.

For a given channel model, we simulate 1000 fixed size frames, i.e. packets. To determine the system performance, we evaluate the BER and PER. BER is defined as a ratio of number of error bits and number of transmitted bits. The transmitted bits are the bits included in the DATA and SIGNAL field. Packet is defined as the number of transmitted bits in one frame unit. In our case, three OFDM symbols exist in one frame unit. The packet error occurs when at least one of the transmitted bits in one frame is erroneous.

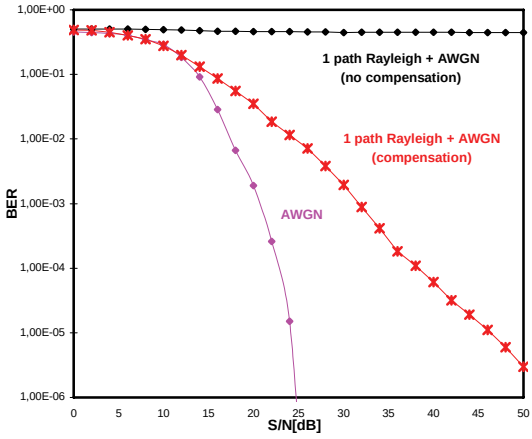


Fig.3 Bit error rate (6 waves generate fading)

Our simulation scenario includes:

- BER and PER in presence of AWGN.
- BER and PER in presence of AWGN and Rayleigh fading without compensation, i.e. we don't use the pilot symbols in SIGNAL field compensation and long training sequences DATA field compensation.
- BER and PER in presence of AWGN and Rayleigh fading with compensation.

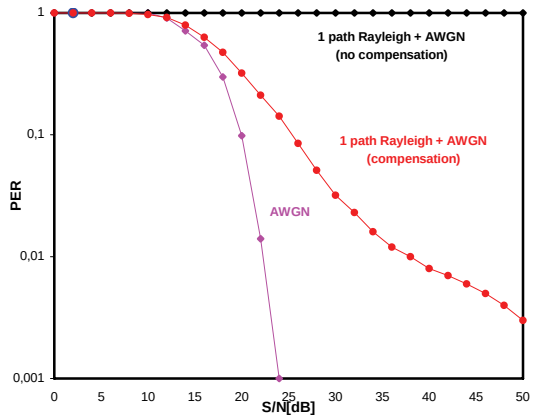


Fig.4 Packet error rate (6 waves generate fading)

The results from the simulation are shown in the Figure 3 and 4. From the curves we conclude that the BER and PER decreases when we increase the signal-to-noise ratio. When wireless channel is subjected to low SNR we obviously will have huge BER and PER. Moreover, if wireless channel is affected with Rayleigh fading and we assume no compensation we also get huge BER and PER values. When we use pilots and long training sequence for AWGN and Rayleigh fading compensation in SIGNAL and DATA field, respectively, the BER and PER experienced lower values. This shows the benefit of implementation of pilots and long training sequence, especially for Rayleigh fading compensation.

V. CONCLUSION

This simulation work was performed to derive BER and PER in order to assess the minimum SNR for good voice performance in IEEE 802.11g ERP-OFDM operational mode. We showed the impact of AWGN and Rayleigh fading in wireless channels on the performance of voice over IEEE 802.11g ERP-OFDM system and how overly optimistic the assessment of voice performance could be if fading is ignored. We also showed the benefit coming from long training sequence and pilots for fading compensation. The results obtained from the simulation suggest that the training sequences specified in the standards are adequate for near-to-perfect channel estimation in a time-invariant channel.

REFERENCES

[1]"Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications" IEEE Std 802.11g™ – 2003. Sponsor: LAN/MAN Standards Committee of the IEEE Computer Society Approved 12 June 2003 IEEE-SA Standards Board.

[2] G.L.Stuber, "Principles of Mobile Communication". *Massachusetts: Kluwer Academic Publishers, 2nd Edition*, 2001.

[3] Veeraraghavan, M.; Cocker, N.; Moors, T.; "Support of voice services in IEEE 802.11 wireless LANs" INFOCOM 2001. *Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE Proceedings.*, Volume: 1, April 22-26, 2001 Page(s): 488-497.

[4] Garg, S.; Kappes, M.; "Can I add a VoIP call?" *IEEE International Conference on Communications, 2003. ICC '03.* , Volume: 2, May 11-15, 2003 Page(s): 779-783.

[5] Armour, S.; Doufexi, A.; Nix, A.; Bull, D.; "A study of the impact of frequency selectivity on link adaptive wireless LAN systems IEEE 56th Vehicular Technology Conference", 2002. Proceedings. VTC 2002-Fall. 2002 Volume: 2, 2002 Page(s): 738-742.

[6] Doufexi, A.; Armour, S.; Nix, A.; Bull, D.; "A comparison of HIPER-LAN/2 and IEEE 802.11a physical and MAC layers" *Symposium on Communications and Vehicular Technology*, 2000. SCVT-200., 2000 Page(s): 14-20.

[7] Doufexi, A.; Armour, S.; Butler, M.; Nix, A.; Bull, D.; "A study of the performance of HIPERLAN/2 and IEEE 802.11a physical layers' *IEEE 53rd VTS Vehicular Technology Conference, 2001*. VTC 2001 Spring. Volume: 1, May 6-9, 2001 Page(s): 668-672.

[8] Medbo, J. 'Schramm, P.; "Channel models for HIPERLANS/2 in different indoor scenarios", 3ERI085B, HIPERLANS/2 ETSI/BRAN contribution, March 30, 1998.

[9] Doufexi, A.; Armour, S.; Nix, A.; Bull, D.; "A comparison of HIPER-LAN/2 and IEEE 802.11a physical and MAC layers" *Symposium on Communications and Vehicular Technology*, 2000. SCVT-200., 2000 Page(s): 14-20.

[10] Lampe, M.; Rohfing, H.; Zirwas, W.; "Misunderstandings about link adaptation for frequency selective fading channels" *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2002.Volume: 2, Page(s): 710-714.

[11] Olufunmilola, A.; Fouad A. T.;" Effect of Fading on the Performance of VoIP in IEEE 802.11a WLANs", Department of Electrical Engineering, Stanford University Stanford, CA. 94305-9515.

[12] Harada, H.; Prasad, R.;"Simulation and Software Radio for Mobile Communication", *Artech House*, 2002.

AUTHOR INDEX

A
Acevski, N., 599,
Aćimović-R.V.,27,177,245,
Acović, M., 471,
Aleksandrov, A., 711,
Aleksić, S., 227,
Aleksieva, V., 621,
Altimirski, E., 329,
Anastasov, J., 189,
Andonov, A., 879,
Andonova, A., 815,
Andreeva, P., 841,
Angelov, K. N., 329,
Angelov, K., 287, 421, 425,
Angelovski, I., 445,
Antić, D., 557,
Antolović, I., 403,
Antonov, P., 43, 869,
Antonova, V., 43, 869,
Antovski, Lj., 155,
Arapinoski, B., 683,
Arnaudov, R., 333,
Arsov, G., 645,
Arsov, Lj., 121,
Asparuhova, K., 785,
Atamian, D., 303,
Atanasković, A., 231, 391,
Atanasov, I., 241, 387,
Atanasov, N., 897,
Atanasovski, M., 587,
Atteia, G., 61,

B
Babić, Z., 403,
Baev, S., 83,
Balabanova, I., 111,
Bankov, N., 719, 723,
Bedzhev, B., 51,
Bekiarski, A., 355,
Bellanger, M., 39,
Bilkiewicz, J., 467,
Blagojević, D., 295,
Blanuša, B., 679,
Bogdanov, M., 511,
Bogdanova, S., 381,
Boiadjiev, S., 835,
Bonev, B., 203,
Borovska, P., 395, 399,
Boumbarov, O., 521,
Bozhikova, V., 861, 865,
Bungurova-Kaziovska, Z., 173,

C
Cârstea I., 579, 583,
Cârstea, D., 579, 583,
Chantov, D., 549,
Cundeva-Blajer, M., 121,
Curcievska, D., 151,
Cvetković, N., 91,
Cvetković, S., 503, 697,

D
Damyanov, D., 495,
Danković, B., 557,
Despotović, V., 629,
Diakov, D., 905,
Dimcev, V., 125, 103,
Dimić, G., 623,
Dimitrijević, A., 261,
Dimitrijević, B., 97, 441,
Dimitrijević, T., 219,
Dimitrov, A., 735,
Dimitrov, D. P., 777,
Dimitrov, D. Tz., 455, 459,
Dimitrov, K., 755, 757,
Dimitrov, V., 411,
Dimitrova, E., 807,
Dimitrova, V., 481,
Dineff, P., 667, 811,
Dizdarević, F., 185,
Djokić, B., 543,
Djorđević, N., 527,
Djorđević-Kajan, S., 403,
Djordjević, G., 135,
Djordjević-Kozarov, J., 691,
Dobrev, D., 279, 417,
Dobrikov, G., 835,
Dobrodolac, M., 873,
Dodov, N., 81, 215,
Dončov, N., 223, 391,
Dontschewa, M.,373,377,407, 693,
Dostanić, A., 471,
Dotsinsky, I., 633,
Drača, D., 275,
Draganov, I., 477,
Draganov, N., 711,
Draganov, V., 641,

E
Eferica, P., 403,
Efnusev, I., 165,

F
Fetaji, B., 609, 613, 617,
Fetaji, M., 609, 613, 617,
Filev, Ts., 517,

G
Gabrova, R., 803, 887,
Gadjeva, E., 739,
Gajda, J., 13,
Ganev, G., 671,
Gavrilovska, L., 445,
Gavrovski, C., 103,
Georgiev, A., 803, 887,
Georgiev, G., 437,
Georgieva, V. K., 905,
Georgieva, V. M., 689,
Gesheva, K., 835,
Gluhchev, G., 521,
Gmitrović, M., 309,
Gocheva-Ilieva, S., 769,
Gocić, M., 265,
Golev, K., 303,
Goleva, M., 303,
Goleva, R., 303,
Gorvits, Y., 605,
Gospodinova, D., 811,
Gradinarova, B., 605,
Grigorova, Ts., 719, 723, 785,
Guliashki, V., 39, 249,
Gusev, M., 151, 155,

H
Hristov, G., 181,
Hristov, V., 845,
Hussein, K., 61,

I
Idzkowski, A., 469,
Ilić, S., 227,
Iliev, B., 3,
Iliev, G., 47, 347,
Iliev, I. G., 199, 531,
Iliev, I. P., 769,
Iliev, M., 181,
Iliev, T., 31, 35,
Ilievski, B., 539,
Iontchev, E., 113,
Ivanova, I., 749,
Ivanova, K., 483,
Ivanović, Ž., 679,

J
Janevski, T., 165, 169, 173,
Janković, D., 691,
Janković, N., 653,
Joković, J., 219,
Jordanova, L., 279, 417, 429,
Jovanović, G., 715,
Jovanović, Z., 557,
Jović, M., 691,

K
Kafadarova, N., 815,
Kapacinović, N., 185,
Karailiev, H., 637,
Karakehayov, Z., 131,
Karakehayova, L., 161,
Karapenev, B., 341,
Karatepe, E., 291,
Karova, M., 361, 849,
Kempter, G., 693,
Koceska, N., 561,
Koceski, S., 561,
Koitchev, K., 287,421, 425, 433,
Kokolanski, Z., 103,
Kolev, I., 767,
Kolev, S., 749,
Kolev, V., 675,
Koleva, E., 767,
Korobko, I., 883,
Korsemov, Ch., 857,
Kosanović, M., 143,
Koserkova, A., 835,
Kosev, J., 645,
Kostadinova, L., 803, 887,
Kostić-Ljubisavljević, A., 177,
Kostov, M., 511,
Kostov, N., 23, 107,
Kotevski, Z., 705,
Kountchev, R., 351, 365, 369,
Kountcheva, R., 351, 365,
Kovatchev, D., 735, 807,
Krantov, N., 901,
Krasowski, A., 469,
Krastev, M., 215,
Krstić, D., 185, 649,
Kuk, K., 623,
Kunev, G., 487,
Künz, A., 407,

L
Latkoski, P., 539,
Lazarova, M., 395,
Letskovska, S., 101, 781,
Levchev, Ch., 73, 77,
Loskovska, S., 613,
Lutskanov, K., 553,

M
Maier, E., 693,
Makal, J., 467, 469,
Males-Ilić, N., 231,
Manchev, V., 739,
Mančić, Ž., 227,
Manoilov, P. G., 399,
Manoilov, P. K., 451, 463,
Marinković, Z., 727,
Marinov, E., 553,
Marinova, G. I., 39,
Marinova, G. K., 23, 107,
Marković, D., 873,
Marković, G., 27,
Marković, V., 727,
Matić, P., 679,
Matoruga, H., 591,
Meijer, G., 3,
Mekić, E., 185,
Mezei, I., 359, 499,
Mihajlović, V., 261, 403,
Mihov, G. S., 633, 773,
Mihov, S. G., 745,
Mihova, B., 739,
Miletiev, R., 113, 333,
Milić, Z., 649,
Milijić, M., 235,
Milivojević, D., 629,
Milivojević, S., 507,
Milivojević, Z., 507,
Milojković, M., 557,
Milosavljević, A., 261,
Milosavljević, Č., 565,
Milošević, B., 623,
Milovanović, B.,219,223,231,235,391,
Milović, D., 275,
Milushev, M., 901, 905,
Mir, N., 283,
Mirčevski, J., 543,

Mirchev, V., 369,
Mironov, R., 369,
Mirtchev, S., 193,
Mitić, A., 295, 299,
Mitrevski, P., 151, 155,
Mitrovski, C., 511, 657,
Mladenov, V., 751,
Mladenović, S., 535,
Mladenovski, Z., 645,
Monov, Z., 161,
Mulalić, E., 503,
Musa, S., 211, 283,

N
Najdenkoski, K., 125,
Naumović, M., 573,
Nazarkiewicz, A., 467,
Nedelchev, I., 313,
Nedelchev, M., 199, 531,
Nedelkovski, I., 705,
Nemigenchev, I., 313, 341,
Nenkov, J., 429,
Nenov, C., 487,
Nikolić, P., 649,
Nikolić, S., 503, 697,
Nikolić, T., 135,
Nikolić, V., 169,
Nikolić, Z., 441,
Nikolov, G., 731,
Nikolov, T., 253, 269,
Nikolova, N., 81,
Nikolova, Z., 491,
Nikolova-Jahn, I., 117,

O
Odalović, M., 789,
Ognjanović, V., 265,
Osmani, S., 609,
Ovtcharov, M., 47,

P
Pachamanov, R., 139,
Palović, B., 189,
Panagiev, O.,321, 325, 757,
Panajotović, A., 275,
Pandiev, I., 831,
Pankov, I., 517,
Pankov, J., 517,
Pantić, D., 653, 827,
Paraliev, M., 663,
Parashkevov, D., 793, 797,
Pavlov, M., 629,
Pavlov, S., 793,
Pavlović, N., 905,
Pencheva, E., 387,
Penev, S., 849,
Pešić, B., 827,
Pesić, T., 653,
Petkov, P., 81,
Petkova, J., 361, 849,
Petković, D., 789,
Petkovski, M., 657,
Petrov, D., 55,
Petrov, P. Tz., 701, 761,
Petrov, P., 569,
Petrova, P., 55,
Peulić, A., 471,
Pleshkova-Bekiarska, S., 495,
Popnikolova R., M., 683,
Popova, A., 477,
Popović, H., 299,
Popović, N., 543,
Popovski, B., 539,
Poulkov, V., 47,
Predić, B., 403,
Prijić, A., 827,
Prijić, Z., 827,
Pronić, O., 727,

R
Radev, D., 31, 147,
Radojičić, V., 177,
Radonjić, V., 245,
Raev, R., 257,
Raeva, I., 257,
Rahnev, P., 101, 781,
Rajković, P., 507,
Rančić, D., 261, 403, 527,
Rančić, M., 337,
Rančić, P., 91, 337,
Ranđelović, M., 391,
Randić, S., 471,
Rankovska, V., 637,
Rashkova, E., 147,
Rassovska, M., 835,
Ristić, S., 827,
Ruskov, T., 853,
Ruskova, N., 853,

S
Sadiku, M., 211,
Sadinov, S., 287, 421, 433,
Sarevska, M., 235,
Savov, M., 521,
Savov, S., 69,
Shaalan, A., 61,
Sharlandjiev, P., 835,
Simić, M., 97,
Sirakov, E., 317,
Smarkov, V., 361,
Sokolović, M., 649,
Sokullu, R., 291,
Spalević, P., 623,
Spasić, J., 381, 441
Spirovski, M., 599,
Srbinovska, M., 103,
Srećković, M., 543, 819, 823
Sroka, R., 13,
Stanchev, Z. D., 641,
Stanimir, V., 295,
Stanivuković, B. V., 873,
Stanković, Z., 235,
Stankovski, D., 147, 257,
Statev, S., 193,
Stefanov, M., 65,
Stefanova, I., 207,
Stefanović, D., 295, 299,
Stefanović, I., 299,
Stefanović, M., 185, 189, 275,
Stoeva, M., 861, 865,
Stoilkov, V., 125,
Stojčev, M., 135, 143, 715,
Stošić, B., 309,
Stoyanov, G., 483,
Stoyanov, N., 73, 77, 437,
Struharik, R., 359, 499,
Sturn, T., 373,

T
Tanchev, I., 641,
Tanev, A., 879,
Tashev, T., 51,
Tasheva, Z., 51,
Tasić, V., 629,
Taskovski, D., 381,
Temelkovski, I., 189,
Terzieva, S., 751,
Todorov, G., 671,
Todorov, V., 351, 365,
Todorova, M., 595,
Toshev, A., 87,
Toshev, H., 857,
Toshkova, D., 569,
Traikovski, G., 433,
Trpezanovski, Lj., 587,
Tsankov, B., 139,
Tsenov, A., 253,
Tsenov, G., 751,
Tzanev, Tz., 675,
Tzeneva, R., 667,
Tzvetkova, S., 675,

V
Valchanov, H., 853,
Valchev, V., 731, 735, 807,
Van den Bossche, A., 731,
Varbanov, G., 487,
Varbanova, N., 425,
Vasileva, M., 595,
Vassileva, S., 841,
Vassileva, V., 69,
Velchev, Y., 749, 891, 895,
Veličković, V., 189,
Velikov, N., 23,
Veselić, B., 565,
Veselinović, B., 819, 823,
Veselinović, I., 819, 823,
Vesković, S., 535,
Vlajnić, M., 823,
Vukosavić, S., 679,

W
Weber, H., 407,

Y
Yakimov, P., 751,
Yordanov, A., 773,
Yordanova, S., 23, 107,
Yudov, D., 731, 735,

Z
Zaharieva-Stoyanova, E., 523,
Zapryanov, G., 745,
Zerbe, V., 901,
Zhekov, Z. S., 553,
Zildžo, H., 591,
Zu-yao Chang Xiujun Li, 3,