POLISH ACADEMY OF SCIENCE
COMMITTEE FOR ELECTRONICS AND TELECOMMUNICATIONS

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

VOLUME 55 – No 3

WARSAW 2009

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY
Quarterly of Polish Academy of Sciences

Address of Editorial Office
Nowowiejska Street 15/19, 00-665 Warsaw, Poland
Institute of Telecommunications, room 484, 472

Email: etq@tele.pw.edu.pl
tel. +48 22 234 7330

Ark. Wyd. 11,64 Ark. druk. 9,31	Podpisano do druku w listopadzie 2009 r.
Papier offset, kl. III 80 g. B-1	Druk ukończono w listopadzie 2009 r.

Publishing
Warszawska Drukarnia Naukowa PAN
00-656 Warszawa, ul. Śniadeckich 8
Tel./fax 628-87-77

IMPORTANT MESSAGE FOR THE AUTHORS

The Editorial Board during their meeting on the 18th of January 2006 authorized the Editorial Office to introduce the following changes:

1. PUBLISHING THE ARTICLES IN ENGLISH LANGUAGE ONLY

Starting from No 1'2007 of E&T Quarterly, all the articles will be published in English only.

Each article prepared in English must be supplemented with a thorough summary in Polish (e.g. 2 pages), including the essential formulas, tables, diagrams etc. The Polish summary must be written on a separate page. The articles will be reviewed and their English correctness will be verified.

2. COVERING THE PUBLISHING EXPENSES BY AUTHORS

Starting from No'2007 of E&T Quarterly, a principle of publishing articles against payment is introduced, assuming non-profit making editorial office. According to the principle the authors or institutions employing them, will have to cover the expenses in amount of 760 PLN for each publishing sheet. The above amount will be used to supplement the limited financial means received from PAS for publishing; particularly to increase the capacity of next E&T Quaterly volumes and verify the English correctness of articles. It is neccessary to increase the capacity of E&T Quarterly volumes due to growing number of received articles, which delays their publishing.

In case of authors written request to accelerate the publishing of an article, the fee will amount to 1500 PLN for each publishing sheet.

In justifiable cases presented in writing, the editorial staff may decide to relieve authors from basic payment, either partially or fully. The payment must be made by bank transfer into account of Warsaw Science Publishers The account number: Bank Zachodni WBK S.A. Warszawa Nr 94 1090 1883 0000 0001 0588 2816 with additional note: "For Electronics and Telecommunications Quarterly".

Editors

Dear Authors,

Electronics and Telecommunications Quarterly continues tradition of the "Rozprawy Elektrotechniczne" quarterly established 55 years ago.

The E&T Quarterly is a periodical of Electronics and Telecommunications Committee of Polish Academy of Science. It is published by Warsaw Science Publishers of PAS. The Quarterly is a scientific periodical where articles presenting the results of original, theoretical, experimental and reviewed works are published. They consider widely recognised aspects of modern electronics, telecommunications, microelectronics, optoelectronics, radioelectronics and medical electronics.

The authors are outstanding scientists, well-known experienced specialists as well as young researchers — mainly candidates for a doctor's degree.

The articles present original approaches to problems, interesting research results, critical estimation of theories and methods, discuss current state or progress in a given branch of technology and describe development prospects. The manner of writing mathematical parts of articles complies with IEC (International Electronics Commision) and ISO (International Organization of Standardization) standards.

All the articles published in E&T Quarterly are reviewed by known, domestic specialists which ensures that the publications are recognized as author's scientific output. The publishing of research work results completed within the framework of *Ministerstwo Nauki I Szkolnictwa Wyższego* GRANTs meets one of the requirements for those works.

The periodical is distributed among all those who deal with electronics and telecommunications in national scientific centres, as well as in numeral foreign institutions. Moreover it is subscribed by many specialists and libraries.

Each author is entitled to free of charge 20 copies of article, which allows for easier distribution to persons and institutions domestic and abroad, individually chosen by the author. The fact that the articles are published in English makes the quarterly even more accessible.

The articles received are published within half a year if the cooperation between author and the editorial staff is efficient. Instructions for authors concerning the form of publications are included in every volume of the quarterly; they may also be obtained in editorial office.

The articles may be submitted to the editorial office personally or by post; the editorial office address is shown on editorial page in each volume.

Editors

CONTENTS

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 403–415

New Hypercomplex Analytic Signals and Fourier Transforms in Cayley-Dickson Algebras

KAJETANA MARTA SNOPEK

Institute of Radioelectronics
Faculty of Electronics and Information Technology, Warsaw University of Technology
e-mail: K.Snopek@ire.pw.edu.pl

Received 2009.06.15
Authorized 2009.08.17

In the paper, the definitions of the new hypercomplex Fourier transform and the new hypercomplex analytic signal are recalled. They have been derived basing on Cayley-Dickson algebra multiplication rules. Two other definitions of multidimensional Fourier transforms are presented, i.e., the classical one and the Clifford FT. The relation between them for 3-D real signals is derived. The differences between three complex/hypercomplex FTs are shown regarding the even-odd spectrum components in the 3-D case and norms/pseudonorms of 3-D analytic signals.

Keywords: hypercomplex signals, Cayley-Dickson algebras, Clifford algebras, Quaternion Fourier Transform, analytic signals, complex delta distribution

1. INTRODUCTION

In the last years, the hypercomplex Fourier analysis has found numerous applications, especially in colour image processing [1]-[4]. The hypercomplex signals defined in the 19th century, [5], [6], get more and more interest due to the development of theory of analytic signals [7], [8].

This paper presents a short survey on hypercomplex algebras with some new contributions. In Chapters 2 and 3, the definitions of Cayley-Dickson algebras of order 2^n and Clifford algebras are recalled and multiplication rules of imaginary units in both cases are presented. In Chapter 4, the 2-D and 3-D Clifford FTs are compared to the classical FT. The Chapters 5 and 6 present definitions of the new hypercomplex FT and the new hypercomplex analytic signal [9]. In 3-D case it has the form of an octonionic signal. Its norm is equivalent to the norm of 1st-octant Hahn's analytic signal [7].

2. CAYLEY-DICKSON ALGEBRAS

The Cayley-Dickson construction [10] is used to define the family of algebras of order 2^n over the real numbers field R. Operations in Cayley-Dickson algebras are defined as shown in Table 1. Since any complex number $z \in \mathrm{C}$: $z = a + bi, i^2 = -1$, can be represented by the ordered pair (a, b): $a, b \in \mathrm{R}$, the complex numbers (C) form the Cayley-Dickson algebra of order 2 over R.

Applying the Cayley-Dickson construction for complex numbers we get the algebra of order 4 of *quaternions* (H) [5], [11] and then starting with quaternions – the 8^{th}-order algebra of *octonions* (O) [11], [12]. The 16^{th}-order algebra immediately following the octonions is called the *sedenions* (S) [13].

Let us denote that applying successively the Cayley-Dickson construction, we get algebras with "worse" properties. First we lose commutativity (quaternions) and associativity (octonions) and finally we lose the division algebra property (sedenions).

Table 1

Operations in Cayley-Dickson Algebras

Property	Operation
summation	$(a,b) + (c,d) = (a + c, b + d)$
additive identity	(0, 0)
reversibility	$-(a,b) + (-a,-b)$
conjugation	$(a,b)^+ = (a^+, -b)$
multication	$(a,b)(c,d) = (ac - d^+b, ad + c^+b)$
multiplicative identity	(1,0)
norm	$\lvert(a,b)\rvert = \sqrt{(a,b)^+(a,b)} = \sqrt{(a^+a + bb^+, ab - ba)} = \sqrt{\lvert a\rvert^2 + \lvert b\rvert^2}$
invertibility	$(a,b)^{-1} = \left(\frac{a}{\lvert a\rvert^2 + \lvert b\rvert^2}, -\frac{b}{\lvert a\rvert^2 + \lvert b\rvert^2}\right)$

2.1. ALGEBRA OF QUATERNIONS H

The quaternions were discovered by Hamilton in 1843 [5]. They are the generalization of complex numbers forming a non-commutative 4^{th}-order algebra over R. A *quaternion* $q \in \mathrm{H}$, [10], [11], is defined as the ordered pair of complex numbers $z_1, z_2 \in \mathrm{C}$:

$$q = (z_1, z_2) = ((a,b),(c,d)) \tag{1}$$

or equivalently as

$$q = a + bi + cj + dk,\ a, b, c, d \in R \tag{2}$$

where *1, i, j, k* are imaginary units with multiplication properties shown in Table 2. Remark that some authors replace *i, j, k* with e_1, e_2 and e_3 and the definition (2) gets the form

$$q = a + be_1 + ce_2 + de_3. \tag{3}$$

The conjugate of (3) is

$$q^* = a - be_1 - ce_2 - de_3 \tag{4}$$

and its modulus (norm):

$$\|q\| = \sqrt{qq^*} = \sqrt{q^*q} = \sqrt{a^2 + b^2 + c^2 + d^2}. \tag{5}$$

Notice that the square of the modulus (5) agrees with the standard Euclidean norm. The quaternions (3) with $a = 0$ are refered as *pure* quaternions and those with $|q| = 1$ as *unit* (*unitary*) quaternions.

2.2. ALGEBRA OF OCTONIONS O

Octonions are a non-associative and a non-commutative algebra over R. They were discovered in 1843 by Graves who called them *octaves*. Independently, in 1845 they were introduced by Cayley. That is why octonions are sometimes called *Cayley numbers* or *Cayley octaves*.

An *octonion* $o \in$ O, [9]-[11], is defined as the ordered pair of quaternions $q_1, q_2 \in$ H:

$$o = (q_1, q_2) = (((a, b), (c, d)), ((e, f), (g, h))) \tag{6}$$

where $q_1 = a + be_1 + ce_2 + de_3$, and $q_2 = e + fe_1 + ge_2 + he_3$. Octonions form a 8-D vector space over R, i.e.,

$$o = a + be_1 + ce_2 + de_3 + ee_4 + fe_5 + ge_6 + he_7 \tag{7}$$

where $1, e_1, e_2, e_3, e_4, e_5, e_6$ and e_7 are multiplied as shown in Table 2. The conjugate of (7) is

$$o^* = a - be_1 - ce_2 - de_3 - ee_4 - fe_5 - ge_6 - he_7 \tag{8}$$

and its norm:

$$|o| = \sqrt{a^2 + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + h^2}. \tag{9}$$

Again. we see that the square of the norm agrees with the standard Euclidean norm.

Table 2

MULTIPLICATION RULES IN CAYLEY-DICKSON ALGEBRAS

		quaternions			*octonions*				*sedenions*							
×	**1**	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}	e_{12}	e_{13}	e_{14}	e_{15}
1	1	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}	e_{12}	e_{13}	e_{14}	e_{15}
e_1	e_1	-1	e_3	$-e_2$	e_5	$-e_4$	$-e_7$	e_6	e_9	$-e_8$	$-e_{11}$	e_{10}	$-e_{13}$	e_{12}	e_{15}	$-e_{14}$
e_2	e_2	$-e_3$	-1	e_1	e_6	e_7	$-e_4$	$-e_5$	e_{10}	e_{11}	$-e_8$	$-e_9$	$-e_{14}$	$-e_{15}$	e_{12}	e_{13}
e_3	e_3	e_2	$-e_1$	-1	e_7	$-e_6$	e_5	$-e_4$	e_{11}	$-e_{10}$	e_9	$-e_8$	$-e_{15}$	e_{14}	$-e_{13}$	e_{12}
e_4	e_4	$-e_5$	$-e_6$	$-e_7$	-1	e_1	e_2	e_3	e_{12}	e_{13}	e_{14}	e_{15}	$-e_8$	$-e_9$	$-e_{10}$	$-e_{11}$
e_5	e_5	e_4	$-e_7$	e_6	$-e_1$	-1	$-e_3$	e_2	e_{13}	$-e_{12}$	e_{15}	$-e_{14}$	e_9	$-e_8$	e_{11}	$-e_{10}$
e_6	e_6	e_7	e_4	$-e_5$	$-e_2$	e_3	-1	$-e_1$	e_{14}	$-e_{15}$	$-e_{12}$	e_{13}	e_{10}	$-e_{11}$	$-e_8$	e_9
e_7	e_7	$-e_6$	e_5	e_4	$-e_3$	$-e_2$	e_1	-1	e_{15}	e_{14}	$-e_{13}$	$-e_{12}$	e_{11}	e_{10}	$-e_9$	$-e_8$
e_8	e_8	$-e_9$	$-e_{10}$	$-e_{11}$	$-e_{12}$	$-e_{13}$	$-e_{14}$	$-e_{15}$	-1	e_1	e_2	e_3	e_4	e_5	e_6	e_7
e_9	e_9	e_8	$-e_{11}$	e_{10}	$-e_{13}$	e_{12}	e_{15}	$-e_{14}$	$-e_1$	-1	$-e_3$	e_2	$-e_5$	e_4	e_7	$-e_6$
e_{10}	e_{10}	e_{11}	e_8	$-e_9$	$-e_{14}$	$-e_{15}$	e_{12}	e_{13}	$-e_2$	e_3	-1	$-e_1$	$-e_6$	$-e_7$	e_4	e_5
e_{11}	e_{11}	$-e_{10}$	e_9	e_8	$-e_{15}$	e_{14}	$-e_{13}$	e_{12}	$-e_3$	$-e_2$	e_1	-1	$-e_7$	e_6	$-e_5$	e_4
e_{12}	e_{12}	e_{13}	e_{14}	e_{15}	e_8	$-e_9$	$-e_{10}$	$-e_{11}$	$-e_4$	e_5	e_6	e_7	-1	$-e_1$	$-e_2$	$-e_3$
e_{13}	e_{13}	$-e_{12}$	e_{15}	$-e_{14}$	e_9	e_8	e_{11}	$-e_{10}$	$-e_5$	$-e_4$	e_7	$-e_6$	e_1	-1	e_3	$-e_2$
e_{14}	e_{14}	$-e_{15}$	$-e_{12}$	e_{13}	e_{10}	$-e_{11}$	e_8	e_9	$-e_6$	$-e_7$	$-e_4$	e_5	e_2	$-e_3$	-1	e_1
e_{bf15}	e_{15}	e_{14}	$-e_{13}$	$-e_{12}$	e_{11}	e_{10}	$-e_9$	e_8	e_7	e_6	$-e_5$	$-e_4$	e_3	e_2	$-e_1$	-1

2.3. ALGEBRA OF SEDENIONS S

Sedenions are a non-division algebra over R. A *sedenion* $s \in S$ [10], [13] is defined as the ordered pair of octonions $o_1, o_2 \in O$: $s = (o_1, o_2)$. Sedenions form a 16-D vector space over R and the imaginary units: $1, e_i, i = 1, ..., 15$ have multiplication properties as shown in Table 2.

2.4. CAYLEY-DICKSON HYPERCOMPLEX SIGNALS

As it has been shown above, all complex and hypercomplex numbers, i.e., quaternions, octonions and sedenions can be written in a general form:

$$\psi = r_0 + r_1 e_1 + r_2 e_2 + \ldots + r_n e_n \tag{10}$$

where $r_1 \in R$, $i = 0,...,n$ and $1, e_i$ $(i = 1,...,n)$ are imaginary units forming the base of a vector space V. If all r_i are replaced by functions $r_i(x)$, we get the definition of a complex/hypercomplex signal $\psi(x)$.

3. CLIFFORD ALGEBRA

The Clifford algebra [14] is an associative algebra with $e_i^2 = 1$ or $e_i^2 = -1$ and $e_i e_j = -e_j e_i$ for $i \neq j$. The basis of the Clifford algebra is a set: $\{e_{i_1} e_{i_2} \ldots e_{i_k} : 1 \leq i_1 < i_2 < \ldots < i_k \leq n, 0 \leq k \leq n\}$. Let us denote with p, the number of elements of the basis of Clifford algebra such as $e_i^2 = 1$ and with q – the number of elements with $e_i^2 = -1$. The Clifford algebra will be denoted with $Cl_{p,q}$ (R). So, $Cl_{0,1}$ (R) is the algebra of complex numbers – C, $Cl_{1,0}$ (R) – algebra of double numbers, $Cl_{0,2}$ (R) – algebra of quaternions H and $Cl_{0,3}$ (R) – algebra of split-biquaternions [6], [15]. Remark that algebras of higher orders (e.g., O and S) are not more Clifford algebras because they are not associative.

3.1. CLIFFORD ALGEBRA OF SPLIT-BIQUATERNIONS $CL_{0,3}$ (R)

The basis of $Cl_{0,3}$ (R) is a set: $\{1, e_1, e_2, e_3, e_1e_2, e_1e_3, e_2e_3, e_1e_2e_3\}$, $\omega = e_1e_2e_3$ and the multiplication rules are as follows (see Table 3):

$$e_i e_j = \begin{cases} -1, i = j \\ -e_j e_i, i \neq j \end{cases}$$

and

$$(e_1e_2)^2 = (e_2e_3)^2 = (e_1e_3)^2 = -1,$$
$$\omega^2 = (e_1e_2e_3)^2 = 1.$$

A split-biquaternion $b_q \in Cl_{0,3}$ (R) can be written in a general form as

$$b_q = r_0 + r_1e_1 + r_2e_2 + r_3e_3 + r_4(e_1e_2) + r_5(e_1e_3) + r_6(e_2e_3) + r_7(e_1e_2e_3) \qquad (11)$$

where $r_i \in$ R, $i = 0,..., 7$. Let us note that the sub-algebra spanned by four elements $\{1, i = e_1, j = e_2, k = e_1e_2\}$ is the division ring of Hamilton's quaternions, H $\equiv Cl_{0,2}$ (R).

Table 3

Multiplication rules in $Cl_{0,3}$ (R)

×	$\boldsymbol{1}$	$\boldsymbol{e_1}$	$\boldsymbol{e_2}$	$\boldsymbol{e_3}$	$\boldsymbol{e_1e_2}$	$\boldsymbol{e_1e_3}$	$\boldsymbol{e_2e_3}$	$\boldsymbol{\omega}$
$\boldsymbol{1}$	1	e_1	e_2	e_3	e_1e_2	e_1e_3	e_2e_3	ω
$\boldsymbol{e_1}$	e_1	-1	e_1e_2	e_1e_3	$-e_2$	$-e_3$	ω	$-e_2e_3$
$\boldsymbol{e_2}$	e_2	$-e_1e_2$	-1	e_2e_3	e_1	ω	$-e_3$	e_1e_3
$\boldsymbol{e_3}$	e_3	$-e_1e_3$	$-e_2e_{13}$	-1	$-\omega$	e_1	e_2	e_1e_2
$\boldsymbol{e_1e_2}$	e_1e_2	e_2	$-e_1$	ω	-1	e_2e_3	$-e_1e_3$	$-e_3$
$\boldsymbol{e_1e_3}$	e_1e_3	e_3	ω	$-e_1$	$-e_2e_3$	-1	e_1e_2	$-e_2$
$\boldsymbol{e_2e_3}$	e_2e_3	$-\omega$	e_3	$-e_2$	e_1e_3	$-e_1e_2$	-1	e_1
$\boldsymbol{\omega}$	ω	e_2e_3	$-e_1e_3$	$-e_1e_2$	e_3	e_2	$-e_1$	1

The conjugate of (11) is

$$b_q = r_0 - r_1e_1 - r_2e_2 - r_3e_3 - r_4(e_1e_2) - r_5(e_1e_3) - r_6(e_2e_3) - r_7(e_1e_2e_3) \tag{12}$$

and its norm:

$$\|b_q\| = \sqrt{b_q \cdot b_q^*} = \sqrt{b_q^* \cdot b_q}. \tag{13}$$

4. CLIFFORD FOURIER TRANSFORM

For an n-dimensional signal $u(\boldsymbol{x})$ the classical n-D complex Fourier transform U is given by the integral

$$U(\boldsymbol{f}) = \int_{\mathbf{R}^n} u(\boldsymbol{x}) \prod_{i=1}^{n} \exp\,(-e_1 2\pi f_i x_i)\, d^n\boldsymbol{x} \tag{14}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{f} = (f_1, f_2, \ldots, f_n)$ (we denote the imaginary unit in the exponent with e_1).

The Clifford Fourier transform F_C of $u(\boldsymbol{x})$ has been defined in [8], [14], [16] in the form:

$$F_C(\boldsymbol{f}) = \int_{\mathbf{R}^n} u(\boldsymbol{x}) \prod_{i=1}^{n} \exp\,(-e_i 2\pi f_i x_i)\, d^n\boldsymbol{x} \tag{15}$$

where $e_1, e_2, \ldots, e_n$ are basis vectors of the Clifford algebra. Let us note that for $n = 1$, F_C is the classical 1-D complex Fourier transform given by (14).

The inverse Clifford Fourier transform is defined by the integral:

$$\mathcal{F}_C^{-1}\{F_C(\boldsymbol{f})\} = \int_{\mathbf{R}^n} F_C(\boldsymbol{f}) \prod_{i=0}^{n-1} \exp\,(e_{n-i} 2\pi f_{n-i} x_{n-i})\, d^n\boldsymbol{f}. \tag{16}$$

Note, that the order of exponents in (15) and (16) can not be changed because the multiplication of higher order hypercomplex numbers is not commutative. For some authors, the definition (15) refers to the so called Right Clifford Fourier transform (e.g.[4]). If the order of $u(x)$ and the product $\prod_{i=1}^{n}(\cdot)$ in (15) is interchanged, we get the definition of the Left Clifford Fourier transform (the same remark applies to (16)). The choice of the order is a matter of convention. Here, we will apply the order as in (15) and (16). Let us analyze the case of 2-D and 3-D signals.

- **Case n = 2**

According to (15), the Clifford Fourier transform of a 2-D real signal $u(x)$, $x = (x_1, x_2)$ is

$$F_C(\boldsymbol{f}) = \int_{\mathbf{R}^2} u(\boldsymbol{x}) \prod_{k=1}^{2} \exp\,(-e_k 2\pi f_k x_k)\, d^2\boldsymbol{x} \tag{17}$$

where $f = (f_1, f_2)$. It is identical to the Quaternionic Fourier Transform (QFT) (denoted U_q) defined in [17] as

$$U_q(\boldsymbol{f}) = \int_{\mathbf{R}^2} u(\boldsymbol{x})\, e^{-e_1 2\pi f_1 x_1} e^{-e_2 2\pi f_2 x_2} d^2\boldsymbol{x}. \tag{18}$$

The relation between QFT and 2-D FT of a real 2-D signal $u(\boldsymbol{x})$ (14) has been derived in [18] and has the form

$$U_q(\boldsymbol{f}) = \tfrac{1}{2}(1 + e_3)\, U(f_1, f_2) + \tfrac{1}{2}(1 - e_3)\, U(f_1, -f_2). \tag{19}$$

- **Case $n = 3$**

Basing on (15), the Clifford Fourier transform of a 3-D signal $u(\boldsymbol{x})$, $\boldsymbol{x} = (x_1, x_2, x_3)$ is defined as

$$F_C(\boldsymbol{f}) = \int_{\mathbf{R}^3} u(\boldsymbol{x}) \prod_{k=1}^{3} \exp\,(-e_k 2\pi f_k x_k)\, d^3\boldsymbol{x}. \tag{20}$$

Its properties have been studied in detail in [19]. The relation between (20) and the 3-D FT of the 3-D real signal (14) has been derived in the form:

$$\begin{aligned} F_C(\boldsymbol{f}) = {} & \tfrac{1}{4}(1 - e_1 - e_2 + e_3)\, U(f_1, f_2, f_3) + \tfrac{1}{4}(1 + e_1 - e_2 - e_3)\, U(f_1, -f_2, f_3) \\ & + \tfrac{1}{4}(1 + e_1 + e_2 + e_3)\, U(f_1, f_2, -f_3) + \tfrac{1}{4}(1 - e_1 + e_2 - e_3)\, U(f_1, -f_2, -f_3). \end{aligned} \tag{21}$$

The formulas (19) and (21) are very useful in practice. They permit to derive the Clifford Fourier transforms directly from the standard FTs.

5. THE NEW HYPERCOMPLEX FOURIER TRANSFORM

Let us recall the notion of the *new n-dimensional hypercomplex Fourier transform* defined in [9] in the form:

$$F_{CD}(\boldsymbol{f}) = \int_{\mathbf{R}^n} u(\boldsymbol{x}) \prod_{i=1}^{n} \exp\,(-e_k 2\pi f_i x_i)\, d^n\boldsymbol{x} \tag{22}$$

with $k = 2^{i-1}$. Notice that for $n = 1$ all transforms (14), (15) and (22) are the same.

The *inverse new hypercomplex FT* is defined as:

$$\mathcal{F}_{CD}^{-1}\{F_{CD}(\boldsymbol{f})\} = \int_{\mathbf{R}^n} F_{CD}(\boldsymbol{f}) \prod_{i=0}^{n-1} \exp\,(e_k 2\pi f_{n-i} x_{n-i})\, d^n\boldsymbol{f}, \quad k = 2^{n-i}. \tag{23}$$

Let us formulate the definitions (22) and (23) for the 3-D and 4-D real signals $u(x)$.

- **Case n = 3**

The formulas (20) and (21) for the 3-D real signal $u(x) = u(x_1, x_2, x_3)$ are:

$$\begin{aligned} F_{CD}(\boldsymbol{f}) &= \int_{\mathbf{R}^3} u(\boldsymbol{x})\, e^{-e_1 2\pi f_1 x_1} e^{-e_2 2\pi f_2 x_2} e^{-e_4 2\pi f_3 x_3} d^3\boldsymbol{x} \\ \mathcal{F}_{CD}^{-1}\{F_{CD}(\boldsymbol{f})\} &= \int_{\mathbf{R}^3} F_{CD}(\boldsymbol{f})\, e^{e_4 2\pi f_3 x_3} e^{e_2 2\pi f_2 x_2} e^{e_1 2\pi f_1 x_1} d^3\boldsymbol{f} \end{aligned} \tag{24}$$

- **Case n = 4**

The new hypercomplex FTs (22) and (23) of a 4-D signal $u(\boldsymbol{x}) = u(x_1, x_2, x_3, x_4)$ are:

$$F_{CD}(\boldsymbol{f}) = \int_{\mathbf{R}^4} u(\boldsymbol{x})\, e^{-e_1 2\pi f_1 x_1} e^{-e_2 2\pi f_2 x_2} e^{-e_4 2\pi f_3 x_3} e^{-e_8 2\pi f_4 x_4} d^4\boldsymbol{x}.$$
$$\mathcal{F}_{CD}^{-1}\{F_{CD}(\boldsymbol{f})\} = \int_{\mathbf{R}^4} F_{CD}(\boldsymbol{f})\, e^{e_8 2\pi f_4 x_4} e^{e_4 2\pi f_3 x_3} e^{e_2 2\pi f_2 x_2} e^{e_1 2\pi f_1 x_1} d^4\boldsymbol{f}. \tag{25}$$

5.1. SPECTRUM COMPONENTS IN 3-D

The 3-D real signal $u(x_1, x_2, x_3)$ can be written as a sum of its even (subscript e) and odd (subscript o) components w.r.t. variables x_1, x_2 and x_3 [19]:

$$u(\boldsymbol{x}) = u_{eee} + u_{eeo} + u_{eoe} + u_{eoo} + u_{oee} + u_{oeo} + u_{ooe} + u_{ooo} \tag{26}$$

All components have been defined in [7], [20]. Here, we recall that for example:

$$u_{eee}(\boldsymbol{x}) = \frac{1}{16}\left[u(x_1, x_2, x_3) + u(-x_1, x_2, x_3) + u(x_1, -x_2, x_3) + u(-x_1, -x_2, x_3) + u(x_1, x_2, -x_3)\right.$$
$$\left. + u(-x_1, x_2, -x_3) + u(x_1, -x_2, -x_3) + u(-x_1, -x_2, -x_3)\right]. \tag{27}$$

It is easily shown that the 3-D complex FT (14) of the signal (26) has a form of the sum:

$$F(\boldsymbol{f}) = U_{eee} - U_{ooe} - U_{oeo} - U_{eoo} + (-U_{oee} - U_{eoe} - U_{eeo} + U_{ooo})\, e_1 \tag{28}$$

where, e.g.,

$$U_{eee}(\boldsymbol{f}) = \int_{\mathbf{R}^3} u_{eee}(x) \cos(2\pi f_1 x_1) \cos(2\pi f_2 x_2) \cos(2\pi f_3 x_3)\, d^3x, \tag{29}$$

$$U_{ooe}(\boldsymbol{f}) = \int_{\mathbf{R}^3} u_{ooe}(x) \sin(2\pi f_1 x_1) \sin(2\pi f_2 x_2) \cos(2\pi f_3 x_3)\, d^3x, \tag{30}$$

etc. The 3-D Clifford Fourier transform (15) of the signal (26) is

$$F_C(\boldsymbol{f}) = U_{eee} - U_{oee}e_1 - U_{eoe}e_2 + U_{ooe}(e_1e_2) - U_{eeo}e_3 + U_{oeo}(e_1e_3) + U_{eoo}(e_2e_3) - U_{ooo}(e_1e_2e_3). \tag{31}$$

Similarly, we obtain the new hypercomplex FT (22) decomposition into the sum of even-odd components:

$$F_{CD}(\boldsymbol{f}) = U_{eee} - U_{oee}e_1 - U_{eoe}e_2 + U_{ooe}e_3 - U_{eeo}e_4 + U_{oeo}e_5 + U_{eoo}e_6 - U_{ooo}e_7. \tag{32}$$

6. COMPLEX (HYPERCOMPLEX) ANALYTIC SIGNALS

6.1. FREQUENCY DOMAIN DEFINITIONS

Let us recall the frequency domain definition of the n-D analytic signal [7] defined as the inverse FT of a single-orthant Fourier spectrum:

$$\psi(\boldsymbol{x}) = \mathcal{F}^{-1}\{\mathbf{1}(\boldsymbol{f}) \cdot U(\mathbf{f})\} \tag{33}$$

where the n-D unit step operator is

$$\mathbf{1}(\boldsymbol{f}) = \prod_{i=1}^{n} (0.5 + 0.5\mathrm{sgn}(f_i)) \tag{34}$$

and $U(\boldsymbol{f})$ is given by (14). Analogously, the frequency domain definition of the n-D Clifford analytic signal [8] is

$$\psi_C(\boldsymbol{x}) = \mathcal{F}_C^{-1}\{\mathbf{1}(\boldsymbol{f}) \cdot F_C(\boldsymbol{f})\}. \tag{35}$$

The new hypercomplex analytic signal in the frequency domain has been defined in [9] as the inverse new hypercomplex FT (23) as follows:

$$\psi_{CD}(\boldsymbol{x}) = \mathcal{F}_{CD}^{-1}\{\mathbf{1}(\boldsymbol{f}) \cdot F_{CD}(\boldsymbol{f})\}. \tag{36}$$

6.2. SIGNAL DOMAIN DEFINITIONS

The signal domain definition of the complex (hypercomplex) analytic signal has the form of the convolution of the signal $u(\boldsymbol{x})$ with the complex (hypercomplex) delta distribution $\Psi^{\delta}(\boldsymbol{x})$ [20], i.e.,

$$\psi(\boldsymbol{x}) = u(\boldsymbol{x}) * \Psi^{\delta}(\boldsymbol{x}). \tag{37}$$

Let us recall three definitions of the n-D complex/hypercomplex delta distributions for all studied cases. Introducing the notation $\theta(x_i) = \dfrac{1}{\pi x_i}$ we have for the classical analytic signal [20]:

$$\Psi^{\delta}(\boldsymbol{x}) = \prod_{i=1}^{n} [\delta(x_i) + \theta(x_i) e_1] \tag{38}$$

and for the Clifford analytic signal [8]:

$$\Psi_C^{\delta}(\boldsymbol{x}) = \prod_{i=1}^{n} [\delta(x_i) + \theta(x_i) e_i]. \tag{39}$$

Similarly, for the new hypercomplex analytic signal [9] we have:

$$\Psi^{\delta}_{CD}(\boldsymbol{x}) = \prod_{i=1}^{n} [\delta(x_i) + \theta(x_i)\, e_k], \quad k = 2^{i-1}. \tag{40}$$

Let us study in detail the case of all 3-D FTs defined by (14), (15) and (22).

• **Case n=3**

Let us recall the notations introduced in [7]:

$$\begin{aligned}
v(\boldsymbol{x}) &= u(\boldsymbol{x}) * \theta(x_1)\,\theta(x_2)\,\theta(x_3),\\
v_1(\boldsymbol{x}) &= u(\boldsymbol{x}) * \delta(x_2)\,\delta(x_3)\,\theta(x_1),\\
v_2(\boldsymbol{x}) &= u(\boldsymbol{x}) * \delta(x_1)\,\delta(x_3)\,\theta(x_2),\\
v_3(\boldsymbol{x}) &= u(\boldsymbol{x}) * \delta(x_1)\,\delta(x_2)\,\theta(x_3),\\
v_{12}(\boldsymbol{x}) &= u(\boldsymbol{x}) * \delta(x_3)\,\theta(x_1)\,\theta(x_2),\\
v_{13}(\boldsymbol{x}) &= u(\boldsymbol{x}) * \delta(x_2)\,\theta(x_1)\,\theta(x_3),\\
v_{23}(\boldsymbol{x}) &= u(\boldsymbol{x}) * \delta(x_1)\,\theta(x_2)\,\theta(x_3)
\end{aligned} \tag{41}$$

where v is the total Hilbert transform of $u(\boldsymbol{x})$, v_i (i = 1, 2, 3) are the 1st-order partial Hilbert transforms w.r.t. variable x_i and v_{ij} (i, j = 1, 2, 3, $i < j$) are the 2nd-order partial Hilbert transforms w.r.t. variables x_i and x_j.

We get three signal domain definitions of complex (hypercomplex) analytic signals with the spectrum in the first octant of the 3-D space. The classical analytic signal derived using (37), (38) and (41) is

$$\psi(\boldsymbol{x}) = u - v_{12} - v_{13} - v_{23} + (v_1 + v_2 + v_3 - v)\, e_1. \tag{42}$$

Similarly, introducing into (37) the definitions (39) and (41) we get the definition of the Clifford analytic signal in the form of a *split-biquaternionic signal* (11):

$$\psi_C(x) = u + v_1 e_1 + v_2 e_2 + v_3 e_3 + v_{12}(e_1 e_2) + v_{13}(e_1 e_3) + v_{23}(e_2 e_3) + v(e_1 e_2 e_3). \tag{43}$$

Analogously, applying (37), (40) and (41) we get the definition of the *new hypercomplex analytic signal* in the form of an octonionic signal (7):

$$\psi_{CD}(\boldsymbol{x}) = u + v_1 e_1 + v_2 e_2 + v_{12} e_3 + v_3 e_4 + v_{13} e_5 + v_{23} e_6 + v e_7. \tag{44}$$

6.3. NORMS/PSEUDONORMS OF 3-D HYPERCOMPLEX ANALYTIC SIGNALS

The norms expressed as modulus of complex/hypercomplex signals are defined as:

$$\|\psi\| = \sqrt{\psi \cdot \psi^*} = \sqrt{\psi^* \cdot \psi}, \tag{45}$$

$$\|\psi_C\| = \sqrt{\psi_C \cdot \psi_C^*} = \sqrt{\psi_C^* \cdot \psi_C}, \tag{46}$$

$$\|\psi_{CD}\| = \sqrt{\psi_{CD} \cdot \psi_{CD}^*} = \sqrt{\psi_{CD}^* \cdot \psi_{CD}}. \tag{47}$$

where the signals conjugate to (42)-(44) respectively are:

$$\psi^*(\boldsymbol{x}) = u - v_{12} - v_{13} - v_{23} - (v_1 + v_2 + v_3 - v)\, e_1, \tag{48}$$

$$\psi_C^*(\boldsymbol{x}) = u - v_1 e_1 - v_2 e_2 - v_3 e_3 - v_{12}(e_1 e_2) - v_{13}(e_1 e_3) - v_{23}(e_2 e_3) - v(e_1 e_2 e_3), \tag{49}$$

$$\psi_{CD}^*(\boldsymbol{x}) = u - v_1 e_1 - v_2 e_2 - v_{12} e_3 - v_3 e_4 - v_{13} e_5 - v_{23} e_6 - v e_7. \tag{50}$$

Since the energy E_ψ of a signal $\psi(x)$ can be expressed as the square of a norm [21]:

$$E_\psi = \|\psi\|_{L_2}^2, \psi \in L_2(-\infty, \infty). \tag{51}$$

we can compare the energies of complex/hypercomplex analytic signals (42)-(44). We have $E_\psi = E_{CD}$ since

$$\|\psi\|^2 = \|\psi_{CD}\|^2 = u^2 + v_{12}^2 + v_{13}^2 + v_{23}^2 + v_1^2 + v_2^2 + v_3^2 + v^2. \tag{52}$$

We see that the expression (52) agrees with the standard Euclidean norm. Differently, for the Clifford analytic signal we have (note the minus sign in the sum):

$$\|\psi_C\|^2 = u^2 + v_{12}^2 + v_{13}^2 + v_{23}^2 + v_1^2 + v_2^2 + v_3^2 - v^2. \tag{53}$$

The expression (56) is a semi-norm (pseudo-norm) [22], [23].

7. SUMMARY

The definitions of the new hypercomplex Fourier transform and of the new hypercomplex analytic signal have been recalled. It has been shown that in the 3-D case, it has a form of an octonionic signal and its norm agrees with the standard Euclidean norm. Differently, the 3-D Clifford analytic signal has a form of a split-biquaternionic signal. The square of its modulus is a pseudo-Euclidean norm. In the paper other two transforms have been presented, i.e., the classical FT and Clifford FT. The formula relating them in 3-D has been derived. All transforms have been compared w.r.t. their even-odd components in the frequency space.

8. REFERENCES

1. S. J. Sangwine: *Fourier transforms of colour images using quaternion or hypercomplex numbers*, Electron. Lett., vol. 32, no. 21, pp. 1979-1980, Oct. 1996
2. T. Bülow: *Hypercomplex spectral signal representation for the processing and analysis of images*. In: Bericht Nr. 99-3, Institut f ur Informatik und Praktische Mathematik, Christian-Albrechts-Universität Kiel, Aug. 1999.
3. T. A. Ell, S. J. Sangwine: *Hypercomplex Fourier Transforms of Color Images*, IEEE Trans. Image Processing, vol. 16, no.1, pp. 22-35, January 2007
4. D. S. Alexiadis, G. D. Sergiadis: *Estimation of Motions in Color Image Sequences Using Hypercomplex Fourier Transforms*, IEEE Trans. Image Processing, vol. 18, no. 1, pp. 168-187, January 2009
5. W. R. Hamilton: *On quaternions*, Proc. Royal Irish Academy, vol. 3, pp. 1-16, 1847
6. W. K. *Clifford, Preliminary sketch of biquaternions*, Proc. London Math. Soc., pp. 381-396, 1873
7. S. L. Hahn: *Hilbert Transforms in Signal Processing*, Artech House Inc., 1996
8. T. Bülow, G. Sommer: *The Hypercomplex Signal – A Novel Extension of the Analytic Signal to the Multidimensional Case*, IEEE Trans. Signal Processing, vol. 49, no. 11, pp. 2844-2852, Nov. 2001
9. K. M. Snopek: *New n-dimensional Hypercomplex Fourier Transform Inspired by The Cayley-Dickson Algebra Multiplication Rules*, IEEE Signal Proc. Lett. (submitted)
10. J. H. Conway, R. K. Guy: *Cayley Numbers. In: The Book of Numbers*, New York, Springer-Verlag, pp. 234-235, 1996. Available: http://www.maths.tcd.ie/pub/HistMath/People/Hamilton/papers.html
11. J. Conway, D. Smith: *On Quaternions and Octonions: Their Geometry, Arithmetic and Symmetry*, A. K. Peters Ltd., 2003
12. J. C. Baez: *The octonions*, Bull. Amer. Math. Soc., vol. 39, no. 2, pp. 145-205, 2001. Available: http://math.ucr.edu/home/baez/octonions
13. K. Imaeda, M. Imaeda: *Sedenions: algebra and analysis*, Applied Mathematics and Computation, vol. 115, pp. 77-88, 2000
14. G. Sommer (ed.): *Geometric Computing with Clifford Algebras*, Springer-Verlag, 2001
15. S-C. Pei, J-J. Ding: *Quaternions and biquaternions for symmetric Markov-chain system analysis*, pp. 1337-1341, Proc. EUSIPCO 2007
16. T. Bülow, M. Felsberg, G. Sommer: *Non-Commutative Hypercomplex Fourier Transforms of Multidimensional Signals*. In: *Geometric Computing with Clifford Algebra*, G. Sommer, ed., Berlin: Springer-Verlag, 2001
17. T. A. Ell: *Hypercomplex Spectral Transforms*, Ph.D. dissertation, Univ. Minnesota, Minneapolis, 1992
18. S. L. Hahn, K. M. Snopek: *Comparison of Properties of Analytic, Quaternionic and Monogenic 2-D Signals*, WSEAS Transactions on Computers, Issue 3, vol. 3, pp. 602-611, July 2004
19. J. Ebling, G. Scheuermann: *Clifford Fourier Transform on Vector Fields*, IEEE Trans. Visualization and Computer Graphics, vol. 11, no. 4, pp. 469-479, July/August 2005
20. S. L. Hahn: *The N-dimensional complex delta distribution*, IEEE Trans. Signal Processing, vol. 44, no. 7, pp. 1833-1837, July 1996
21 S. L. Hahn: *Complex Signals with Single-orthant Spectra as Boundary Distributions of Multidimensional Analytic Functions*, Bull. Polish Academy of Sciences, Technical Sciences, vol. 51, no. 2, pp. 155-161, 2003
22. A. Mertins: Signal Analysis, John Wiley&Sons Ltd, 1999
23. N. Bourbaki: *Topological vector spaces*, Springer-Verlag, 1987
24. W. Rudin: *Functional analysis*, McGraw-Hill, 1979

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 417–433

DAVINCI Codes as Example of High Performance Non-Binary LDPC

ŁUKASZ KIEDROWSKI, HENRYK GIERSZAL, WITOLD HOŁUBOWICZ, ADAM FLIZIKOWSKI

Adam Mickiewicz University, Umultowska 85, 61-614 Poznań
ITTI, Rubież 46/C5, 61-612 Poznań
e-mail: kiedrowski@itti.com.pl, gierszal@amu.edu.pl

Received 2009.06.03
Authorized 2009.08.06

The article presents a new family of non-binary LDPC which has been elaborated within a framework of DAVINCI project. They are based on a finite field GF(64). We compare their error rate to other channel coding techniques. Results prove that DAVINCI codes allow improving the coding performance.

Keywords: Non-Binary LDPC, channel coding, WER, Galois Field, Extended-Min-Sum algorithm

1. INTRODUCTION

Low Density Parity Check Codes (LDPC) are a class of linear error correcting block codes introduced by Gallager in 1962 [1] and rediscovered by MacKay and Neal [2]. Despite their simple construction they have excellent performance. Recent improvements of LDPC have allowed them to surpass the performance of turbo codes. LDPC codes and turbo codes are among the known near Shannon limit codes that can achieve very low bit error rates for low Signal-to-Noise Ratio (SNR) applications [3][4].

LDPC codes are defined in terms of a sparse parity check matrix, in which most of the entries are zero and only a small fraction are nonzero values. Each codeword satisfies a number of linear constraints and each symbol of the codeword participates in a small number of constraints. For binary LDPC (B-LDPC) in the receiver each bit of codeword is connected with a number of parity check nodes (operations are modulo

2). In the case of errorless transmission at the output of each parity check nodes there are zeros for all codewords.

The constructions, description of an iterative probabilistic decoding algorithm and theory provided by Gallager [1] go beyond what is known today for turbo codes. Arriving before the computing power that was to prove their effectiveness they were largely forgotten until the rediscovery by MacKay and Neal [8].

Now it is well known that binary low density parity check codes achieve rates close to the channel capacity for very long codeword lengths [4], and more and more LDPC solutions have been proposed in standards (DVB[1], Wi-MAX[2], etc.). In terms of performance, binary LDPC codes start to show their weaknesses when the codeword length is small or moderate, or when higher order modulation is used for transmission.

For these cases advancements to LDPC codes include improvements in terms of non-binary versions of the codes and codes having variable number of non zero values in the parity check matrix [13]. The non-binary LDPC (NB-LDPC) codes involve encoding the messages using symbols from a finite field with more than two elements, resulting in each parity check becoming complex but decoding remaining tractable. The non-binary codes have an alternative representation as binary codes but the decoding algorithm is not equivalent to the binary algorithm.

Works in the above area are led within a FP7 project DAVINCI[3] (Design And Versatile Implementation of Nonbinary wireless Communications based on Innovative LDPC codes). The main aim of the project is to demonstrate the outstanding performance improvements and industrial feasibility of pioneering non-binary LDPC codes as well as adapted link level technologies for next generation wireless communications IMT-Advanced[4], such as IEEE 802.16m or 3GPP LTE-Advanced[5]. Authors applied elaborated DAVINCI codes into a complete telecommunication chain in order to compare their performance vs. other channel coding techniques.

2. DAVINCI SYSTEM

In this section, a review of the fundamental system-independent physical layer-related aspects, parameters and configurations is performed. The aim of this review is the identification of the main commonalities among the air interfaces of the three main IMT-advanced candidate schemes, i.e., IEEE 802.16e/m, WINNER[6] and LTE.

The information in this paragraph are used as general references for all the relevant tasks foreseen within DAVINCI project, especially within the Link Level Evaluation and

[1] Digital Video Broadcasting
[2] Worldwide Interoperability for Microwave Access
[3] FP7 STREP - INFSO-ICT-216203; www.ict-davinci-codes.eu
[4] International Mobile Telecommunications
[5] Long Term Evolution
[6] Wireless World Initiative New Radio

System Level Evaluation tasks. In fact, the identification of a common configuration of the physical layer can help the assessment of some eventual performance improvements caused by the enhanced technologies to be developed within DAVINCI both in terms of link level algorithms and channel coding / decoding schemes, and to be potentially adopted in future B3G or 4G wireless broadband transmission systems.

The performance obtained with the DAVINCI air interface, which are starting from the common aspects of the air interfaces of the three IMT-candidate schemes, will be compared with that obtainable by each candidate system equipped with its own link level technologies and coding schemes. In such a way, it will be possible to evaluate the impact of new DAVINCI technologies within the three systems, and then to highlight the advantages over existing schemes already adopted by the standards. The comparison among the three systems is not one of the objectives to be reached by the project.

Finally, it is worth saying that the assessment of the eventual performance improvements as mentioned above, will be performed by resorting to ad-hoc computer simulations.

In Table 1 the aspects are summarized which are common to the three air interfaces and then can be basically considered as system-independent. The information reported in that table [21] constitutes the starting point for the design of the optimized DAVINCI air interface.

Table 1

Common system-independent parameters / configurations / aspects

Operating RF frequencies [GHz]	• Less than 6
Operating bandwidths [MHz]	• 1.25, 5, 10, 20
Duplex schemes	• FDD, TDD
Terminal speed and mobility	• From 0 km/h (stationary) up to very high mobility (≥ 120 km/h)
Transmission type	• OFDM
FFT size	• 2048
Multiple access	• Based on OFDMA
CP (Cyclic Prefix) length	• ≤ 1/8 of OFDMA symbol duration
Modulation	• QPSK, 16-QAM, 64-QAM
ACM (Adaptive Coding Modulation)	• Supported
Frame duration	• ≥ 5 ms
Frame structure	• Equivalent to the following structure: frame, subframe, burst (or slot)
Pilot symbols within the frame	• Supported
Synchronization sequence in the frame	• Available as preamble (or midamble)
Carrier / Timing synchronization	• Split in coarse and fine recovery tasks • DA (Data Aided) over preamble (midamble) and/or pilot symbols or NDA (Non Data Aided or Blind) and based on the correlation between cyclic prefixes
Channel estimation	• DA (Data Aided), based on preamble (midamble) and/or pilot symbols
Channel equalization	• At least ZF (Zero-Forcing) correction over pilot symbols and first order interpolation of channel estimates between consecutive pilots
Presence of ARQ / H-ARQ	• H-ARQ with CC (Chase Combining) • H-ARQ with IR (Incremental Redundancy) optional
Number of ARQ / H-ARQ retransmission	• ≤ 4
Multiple antennas	• 2 antennas at the BS • 2 antennas at the MS
MIMO (Multiple Input Multiple Output) schemes	• Spatial Multiplexing • Transmit Diversity

source: [21]

3. DAVINCI CODES

3.1. CODING

An NB-LDPC code is defined by a very sparse random parity check matrix H, whose components belong to a finite field GF(q) (F_q). The matrix H consists of M rows and N columns; the code rate is defined by $r \leq (N-M)/N$. Decoding algorithms of LDPC codes are iterative message passing decoders based on a Tanner (or factor) graph representation of the matrix H [5]. In general, an LDPC code has a factor graph consisting of N variable nodes (i.e. for codewords) and M parity check nodes with various degrees (i.e. different number of edges in the graph connected to variable nodes or parity check nodes).

Davey and MacKay [6] showed that LDPC codes over GF(q) achieve superior performance to that of binary LDPC codes. In case of non-binary codes the H matrix can take values from the finite field GF(q). Again as for binary LDPC codes, $H\underline{x}=0$ (where $\underline{x}$ is a codeword), but the presence of elements from the GF(q) produces more stringent checks on the codewords of non-binary LDPC codes. Any non-zero value at H(i, j) indicates that there is an edge existing between ith row (variable node) and jth column (parity check node). For NB-LDPC codes there is a weight on each edge of the Tanner graph. This weight is the matrix entry in the parity check matrix and is chosen from the finite field GF(q). So now for nodes defined over GF(q), a parity check m would require a following expression:

$$\sum_{j \in \mathrm{N}(m)} a_{mj} \otimes x_j = 0 \tag{1}$$

where N(m) is the set of variable nodes (representing possible codewords $x_j = [x_0\, x_1 \ldots x_{N-1}]_j$ connected to the parity check m and $a_{mj} \in$ GF(q), $a_{mj} \neq 0$. All operations are done in GF(q). MacKay showed that going from binary to non-binary field may reduce the number of cycles in the Tanner graph. It is indicated in Fig. 1.

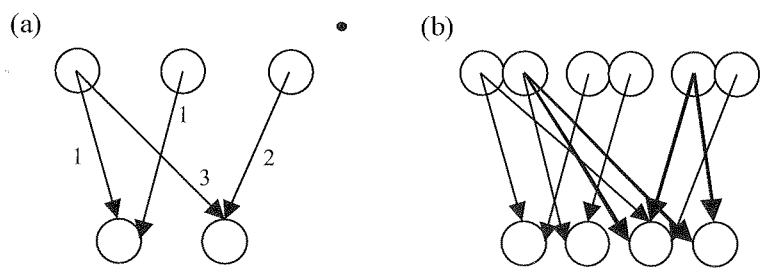


Fig. 1. Comparison of graph structure of (a) non-binary and (b) binary LDPC

From the above graph it can be seen that for the binary case there exists a short cycle of length equal 4 which is absent in the non-binary case. This is one of the

main reasons why we expect LDPC codes over non-binary fields to perform better than LDPC codes over the binary field.

MacKay and Davey [6] also show that the average entropy of messages passed in the graph for non-binary LDPC codes falls faster than the average entropy of messages passed in graph for binary LDPC codes only for certain mean column weight. Although the procedure provided in [6] finds the mean column weight where codes over GF(q) would outperform codes over GF(2) it does not give any insight into the convergence properties of the decoding algorithm.

However moving onto GF(q), $q > 2$, increases the state space of each node in the decoding graph by decoding over GF(q). In other words increasing the field order is comparable to increasing the memory of a convolutional code.

In the framework of the DAVINCI project, we use ultra-sparse non-binary LDPC codes designed in a Galois field GF(q) of order $q = 64$, which corresponds to the largest modulation order considered for wireless communications (i.e., 64-QAM[7]) in this project. The non-binary LDPC codes are described by a Tanner graph with regular and constant connection degree, with $d_v = 2$ edges at the variable node side, and varying parity-check connection d_c depending on the desired code rate. To each edge, a non-zero value belonging to the Galois field GF(64) is assigned, in order to define non-binary parity check equations. The choice of the nonzero values is especially important to obtain good performance and requires an optimization strategy. This type of non-binary LDPC code is also referred to as cycle codes and has two main advantages:

- Regular codes with $d_v = 2$ are very sparse and the corresponding Tanner graphs have very large girths compared to usual binary code graphs. As a consequence, iterative decoders show very good performance, especially at small to moderate codeword lengths. For example, the girth of a binary irregular LDPC code with length $N = 848$ bits and rate $r = 1/2$ is at most $g_b = 6$, while the girth of a NB-LDPC code with same parameters is $g_{nb} = 14$ when a good graph construction is used [15][18].
- As for the code design, it has been shown in the literature [19] that the finite length optimisation of non-binary cycle codes can be decomposed into two steps: (i) first build a Tanner graph with the maximum possible girth and the minimum number of cycles with minimal length, then list all 'short' cycles and the combination of short cycles which define the smallest trapping sets, (ii) optimise iteratively the choice of the non-zero values on the edges of the cycles, such that the local binary minimum distance computed on the set of cycles and trapping sets is maximized. This optimisation procedure allows to gain performance both in the waterfall and the error floor region, compared to a random choice of the Tanner graph structure and of the non-zero values assignment.

As for binary decoders, there are two possible representations for messages: probability weights vectors or LDR (Log-Density-Ratios) vectors. The use of the LDR form for

[7] Quadrature Amplitude Modulation

messages has been advised by many authors who proposed practical LDPC decoders. The LDR values, which represent real reliability measures on the bits or the symbols are less sensitive to quantization errors due to the finite precision coding of the messages [20]. Also, LDR measures operate in the logarithm domain, which avoids complicated operations (in terms of hardware implementation) like multiplications or divisions.

3.2. DECODING

Improvement of the coding performance using NB-LDPC is related to a greater computing power needed for the decoding due to increased decoding complexity. As in all practical coding schemes, an important feature is the complexity/performance trade-off, it is very important to try to reduce the decoding complexity of non-binary LDPC codes, especially for high order fields GF(64). The base iterative decoder of non-binary LDPC codes is the Belief Propagation (BP) decoder over the Tanner graph representation of the code. The main difference of the binary BP decoder is that for GF(q) LDPC codes, the messages from variable nodes to check nodes and from check nodes to variable nodes are defined by q probability weights, or $q-1$ log-density-ratios. As a result, the decoder complexity scales as $O(q^2)$ per check node [14], which is too complex for practical applications.

Computing the check node in the Fourier-domain reduces the complexity to $O(q \log q)$ per check node [15][16], but adapting the Fourier-domain decoder to practical implementation is tedious due to complicated operators like exponentials or real multiplications.

Recently, sub-optimum decoders based on the generalization of the Min-Sum (MS) decoder have been developed [7][17]. One of them is Extended-Min-Sum (EMS) algorithm [7] proposed for NB-LDPC codes [7][8][9]. A particularity of this algorithm is that it takes into account the memory problem of the non-binary LDPC decoders, together with a significant complexity reduction per decoding iteration.

The core idea of the Extended Min-Sum (EMS) decoder is to use only a limited number of LLR (Log-Likelihood Ratio) values $n_m << q$ both for the storage of messages, and for the computation of symbol and check nodes. In the decoder the vector messages of the decoder (represented by LDR vector) is truncated to a limited number n_m of values in order to reduce the memory requirements. This algorithm promises the best complexity/performance trade-off for LDPC codes in high order fields, and the complexity scales as $O(n_m \log(n_m))$ with $n_m << q$. The performance loss compared to BP decoding is small (around 0.1 dB) to negligible, depending on the decoder complexity which is tuned by the value of n_m.

The NB-LDPC iterative decoding algorithms are characterized by three main steps corresponding to the different nodes depicted in Fig. 2: (i) the variable node update, (ii) the permutation of the messages due to non zeros values in the matrix H and (iii) the check node update which is the bottleneck of the decoder complexity, since the BP

operation at the check node is a convolution of the input messages, which makes the computational complexity grow in $O(q^2)$ with a straightforward implementation.

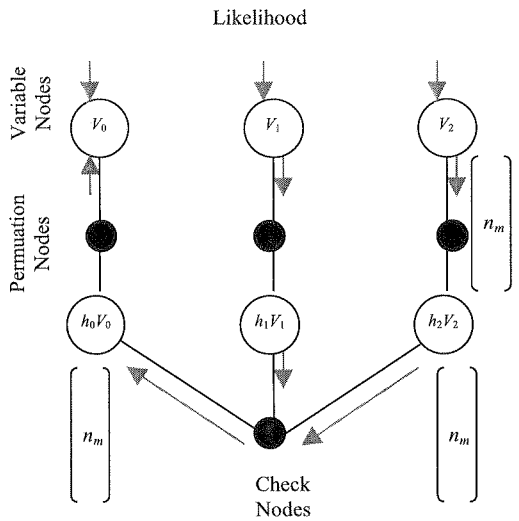


Fig. 2. Tanner graph structure of a parity check node of degree $d_c = 3$ for a non-binary LDPC code

Although interesting in terms of memory and computation reduction, the truncation of messages from $q-1$ to n_m values obviously looses potentially valuable information that leads to performance degradation on the error rate curves. This loss of performance could be mitigated by using a proper compensation of the information that has been truncated. Before truncation to n_m entries which are assumed to be the largest reliability values, the values in a message are sorted in decreasing order. Because the concern is the development of low complexity decoders, we have chosen to compensate the $(q-n_m)$ truncated values with a single scalar value γ, which is the simplest model one can use.

3.3. MODULATION FOR Q-ARY CHANNEL CODES

There are basically two possibilities to combine modulation and coding: coded modulation (e.g. Trellis Coded Modulation (TCM)) and Bit-Interleaved Coded Modulation (BICM). While the former considers both operations jointly and is therefore theoretically superior for many cases, the latter is preferred in nearly all wireless communication systems due to its lower complexity. Additionally, it has been shown that even theoretically, the performance loss of BICM is insignificant [22].

For these reasons and since BICM is also used in all three reference systems, this scheme which is depicted in Fig. 3, is also applied to the q-ary channel codes.

The channel encoder and decoder are described in [23]. The message $\mathbf{u} \in \mathrm{F}_q^K$ is encoded into a codeword $\mathbf{c} = (c_0, c_1, ..., c_{N-1}) \in \mathrm{F}_q^N$, which is interleaved and mapped to QAM symbols.

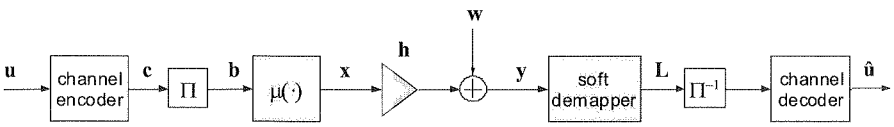


Fig. 3. Transmitter and receiver structure for BICM

Mapping

The mapping function $\boldsymbol{\mu}(\cdot)$ [23] is responsible for assigning symbols out of a M-QAM constellation to the interleaved code symbols which are taken out of a Galois field of order q. Since the cardinality of both sets is generally not identical, we have to gather m_1 code symbols and map them onto m_2 QAM symbols:

$$\boldsymbol{\mu} : \mathrm{F}_q^{m_1} \rightarrow A_x^{m_2}$$

In order to have a bijective mapping, the number of elements on both sides must be equal, i.e. m_1 code symbols out of F_q are mapped onto m_2 M-QAM symbols, such that:

$$q^{m_1} = M^{m_2}, \quad m_1, m_2 \in \mathrm{N}$$

$$m_1 = \frac{\mathrm{lcm}\,(\mathrm{ld}M, \mathrm{ld}q)}{\mathrm{ld}q}, \quad m_2 = \frac{\mathrm{lcm}\,(\mathrm{ld}M, \mathrm{ld}q)}{\mathrm{ld}M}$$

where $\mathrm{ld}(\cdot)$ is $\log_2(\cdot)$, lcm stands for *least common multiple* and $M = |\mathrm{A}_x|$ is the number of constellation points.

For common values of q and M, this gives the values for (m_1, m_2) as denoted in the following Table 2.

Table 2

Values for m_1 and m_2 for different modulations and Galois field orders

$M \backslash q$	64	256
2	(1,6)	(1,8)
4	(1,3)	(1,4)
16	(2,3)	(1,2)
64	(1,1)	(3,4)
256	(4,3)	(1,1)

The mapping function hence gathers m_1 code symbols to $\mathbf{b} = (b_0, ..., b_{m_1-1})$ and maps them onto m_2 QAM symbols:

$$\mathbf{x} = (x_0, \ldots, x_{m_2-1}) = \boldsymbol{\mu}(\mathbf{b}) = (\mu_0(\mathbf{b}), \ldots, \mu_{m_2-1}(\mathbf{b}))$$

For the following, we only consider the special case of the usual mappings from binary vectors to QAM symbols. This is possible because the code symbols $c_i \in \mathrm{F}_q$ can be represented by their binary images $(\overline{c}_{i,0}, \overline{c}_{i,1}, ..., \overline{c}_{i,p-1})$ where $p = \mathrm{ld}\, q$ and $\overline{c}_{i,j} \in \mathrm{F}_2$. We denote by $\chi : \mathrm{F}_2^{\mathrm{ld}M} \rightarrow \mathrm{A}_x$ the mapping function, which assigns to each bit vector of length ld M a QAM symbol out of the constellation A_x.

Demapping

For each code symbol $b_n \in \mathrm{F}_q$, a vector of q APP[8] L values $\mathbf{L}_n = (L_{n,0}, L_{n,1},...,L_{n,q-1})$ has to be calculated, [23] with:

$$L_{n,k} = \ln \frac{P[b_n = \alpha_k|\mathbf{y}]}{P[b_n = \alpha_0|\mathbf{y}]}, \qquad \begin{matrix} n \in \{0, 1, \ldots, m_1 - 1\} \\ k \in \mathrm{Z}_q \end{matrix}$$

where α_k are the Galois field elements, i.e. $\mathrm{F}_q = \{\alpha_0, \alpha_1,...,\alpha_{q-1}\}$, and $\mathrm{Z}_q = \{0, 1,..,q-1\}$ is a set of integers from 0 to $q - 1$.

Since generally more than one code symbol is involved in the mapping, in order to calculate this vector we require a marginalization:

$$L_{n,k} = \ln \frac{\sum\limits_{\mathbf{b} \in \mathrm{B}_n^k} P[\mathbf{b}|\mathbf{y}]}{\sum\limits_{\mathbf{b} \in \mathrm{B}_n^0} P[\mathbf{b}|\mathbf{y}]} = \ln \frac{\sum\limits_{\mathbf{b} \in \mathrm{B}_n^k} p(\mathbf{y}|\mathbf{b})\, P[\mathbf{b}]}{\sum\limits_{\mathbf{b} \in \mathrm{B}_n^0} p(\mathbf{y}|\mathbf{b})\, P[\mathbf{b}]}$$

where $\mathrm{B}_n^k = \{\mathbf{b} : b_n = \alpha_k\}$ is the set of all code symbol vectors $\mathbf{b} = (b_0,...,b_{m_1-1})$ with the n-th component fixed to α_k.

We assume that all code vectors are equiprobable, i.e. $P[\mathbf{b}] = q^{-m_1}$, and the channel is memoryless, i.e., $p(\mathbf{y}|\mathbf{b}) = \prod\limits_{i=0}^{m_2-1} p(y_i|\mathbf{b})$. Then:

$$L_{n,k} = \ln \frac{\sum\limits_{\mathbf{b} \in \mathrm{B}_n^k} \prod\limits_{i=0}^{m_2-1} p(y_i|\mathbf{b})}{\sum\limits_{\mathbf{b} \in \mathrm{B}_n^0} \prod\limits_{i=0}^{m_2-1} p(y_i|\mathbf{b})}$$

For a flat fading channel, given by:

$$y_i = h_i \cdot x_i + w_i, i = 0, 1, \ldots, m_2 - 1$$
$$x_i = \mu(\mathbf{b}), w_i \sim \mathrm{CN}(0, N_0)$$

[8] A Posteriori Probability

where h_i is a channel coefficient, and CN(0,N_0) is a noise sample of complex Gaussian distribution with mean equal 0 and variance N_0, we have the conditional probability density function (pdf) of the received signal:

$$p(y_i|\mathbf{b}) = p(y_i|x_i) = \frac{1}{\pi N_0} \exp\left(-\frac{|y_i - h_i x_i|^2}{N_0}\right)$$

which leads to the LLR vector:

$$L_{n,k} = \ln \frac{\sum\limits_{\mathbf{b}\in \mathrm{B}_n^k} \prod\limits_{i=0}^{m_2-1} \exp\left(-\frac{|y_i - h_i\mu(\mathbf{b})|^2}{N_0}\right)}{\sum\limits_{\mathbf{b}\in \mathrm{B}_n^0} \prod\limits_{i=0}^{m_2-1} \exp\left(-\frac{|y_i - h_i\mu(\mathbf{b})|^2}{N_0}\right)} =$$

$$= \ln \frac{\sum\limits_{\mathbf{b}\in \mathrm{B}_n^k} \exp\left(-\sum\limits_{i=0}^{m_2-1} \frac{|y_i - h_i\mu(\mathbf{b})|^2}{N_0}\right)}{\sum\limits_{\mathbf{b}\in \mathrm{B}_n^0} \exp\left(-\sum\limits_{i=0}^{m_2-1} \frac{|y_i - h_i\mu(\mathbf{b})|^2}{N_0}\right)} \quad \begin{matrix} n = 0, 1, \ldots, m_1 - 1 \\ k = 0, 1, \ldots, q - 1 \end{matrix}$$

The cardinality of the set for the first summation is $|\mathrm{B}_n^k| = q^{m_1-1}$. Since the denominator does not depend on k, we can compute only the first term and then normalize such that $L_{n,0} = 0$:

$$L_{n,k} = a_{n,k} - b_n, \quad a_{n,k} = \ln\left(\sum_{\mathbf{b}\in \mathrm{B}_n^k} \exp\left(-\sum_{i=0}^{m_2-1} \frac{|y_i - h_i \cdot \mu_i(\mathbf{b})|^2}{N_0}\right)\right) \tag{2}$$

In the following, we assume $q = 64$ and $p = \mathrm{ld}\, q = 6$, and we denote by $(\alpha_{k,0}, \alpha_{k,1},\ldots,\alpha_{k,p-1})$ the binary image of the GF element α_k.

For 64 QAM modulation, the mapping and demapping functions are especially simple since one code symbol corresponds to one QAM symbol, i.e. $m_1 = m_2 = 1$ and hence the formula (2) simplifies to:

$$a_k = -\frac{1}{N_0}|y - h \cdot \mu(\alpha_k)|^2$$

For higher-order modulations, very similar expressions can be derived from the general formula (2) [23]. It should be noted here, that so far no efforts have been undertaken to optimize these mappings.

Truncated and sorted LLR vector

The decoder expects as input parameter a truncated and sorted version of the LLR vector defined above [23]. Given the LLR vector as $\mathbf{a} = (\alpha_0, \alpha_1,\ldots,\alpha_{q-1})$, we first sort these values in descending order such that:

$$b_k = \alpha_{n(k)} \text{ with } b_0 \geq b_1 \ldots \geq b_{q-1}$$

Next, we keep only the greatest n_m values and subtract a constant such that the last value is zero:

$$\mathbf{b} = \begin{pmatrix} b_0 - b_{n_m-1} \\ b_1 - b_{n_m-1} \\ \vdots \\ 0 \end{pmatrix}$$

There is one such vector for each code symbol, i.e. for each codeword, there are N such vectors. Together with the truncated LLR vector, the permutation $(\pi(0), \pi(1),...,\pi(n_m - 1))$ is fed into the decoder.

4. SIMULATIONS RESULTS

In this section results of simulation research are presented that allow comparing the performance of DAVINCI codes in a function of their parameters and in relation to other channel coding techniques. In the physical layer there are BPSK or 64-QAM modulation schemes and AWGN[9] channel. The maximum number of iterations in the algorithm has been fixed to 20, and a stopping criterion based on the syndrome check is used.

In Fig. 4 WER (Word Error Rate) is presented for DAVINCI codes of different codeword lengths K_{bin} and rate $r = 1/2$. The longer codewords, the better performance can be obviously obtained.

The same kind of results was obtained for 64-QAM modulation scheme (Fig. 5). The gain between $K_{\text{bin}} = 576$ and 1152 is about 0.4 dB while for BPSK only ca. 0.2 dB.

If the code rate is increased, the performance of DAVINCI codes decreases, however it can be compensated using longer codewords (Fig. 6). The loss in Eb/N0 is about 1 dB for code rate 3/4 vs. 1/2. $K_{\text{bin}} = r \cdot N \cdot \text{ld}(q)$.

[9] Additive White Gaussian Noise

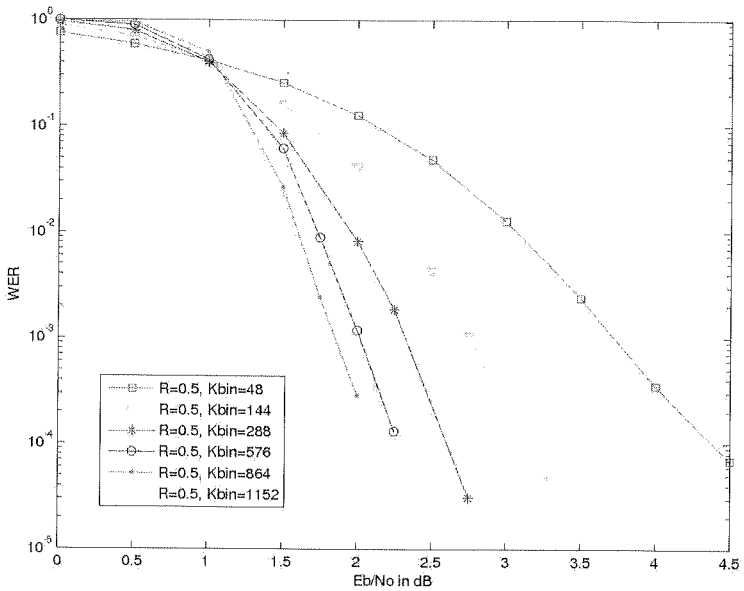


Fig. 4. WER in a function of $E_b/N0$ for DAVINCI codes of different codeword lengths for BPSK

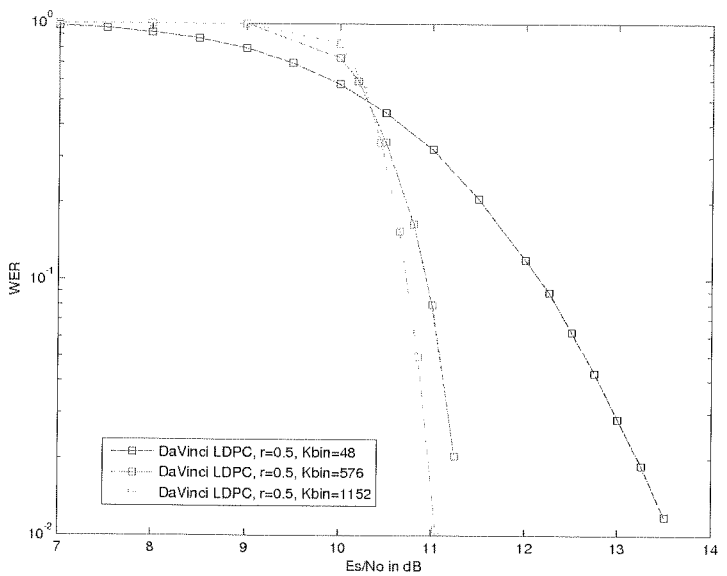


Fig. 5. WER in a function of $E_b/N0$ for DAVINCI codes of different codeword lengths for 64-QAM

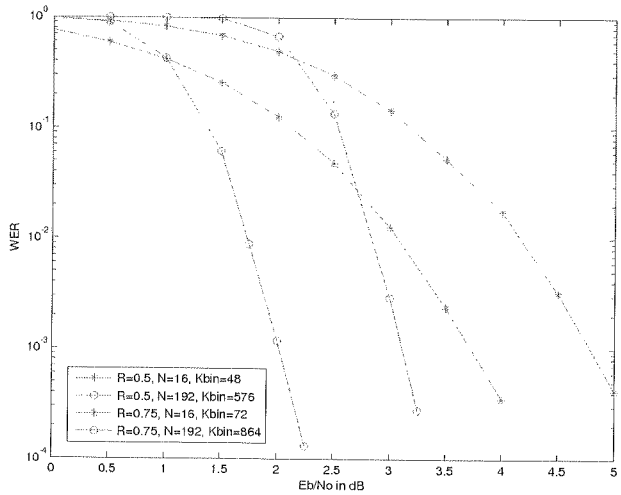


Fig. 6. WER in a function of $E_b/N0$ for DAVINCI codes of different rates

Comparing Block Turbo Codes (BTC) with DAVINCI ones, the latter ensures better performance for the same codeword lengths (Fig. 7). The gain is at least 0.7 dB.

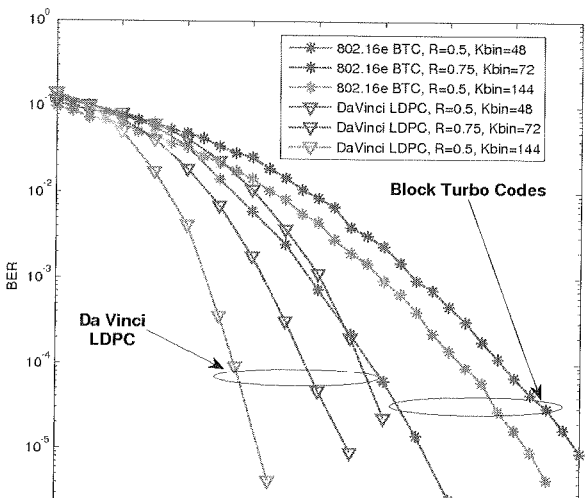


Fig. 7. BER in a function of $E_b/N0$ for DAVINCI codes and BTC

In Fig. 8 DAVINCI codes and another coding technique are compared. The maximal number of iterations i_{max} in decoding algorithms was equal:

- 30 for B LDPC used in WiMAX (decoding algorithm is Sum-Product – SP or Min-Sum – MS) [11];
- 20 for DAVINCI codes (decoding algorithm is Extended-Min-Sum – EMS).

For longer codeword lengths a gain of DAVINCI codes over B LDPC with an MS decoder reaches 0.5 dB. This result is obtained for 20 iterations in the EMS decoder instead of 100 iterations in the MS decoder. If a SP decoder is used, BER is almost equivalent for DAVINCI's EMS decoder and 802.11e SP decoder in spite of five times less iterations in the EMS decoder, which furthermore has a less complexity.

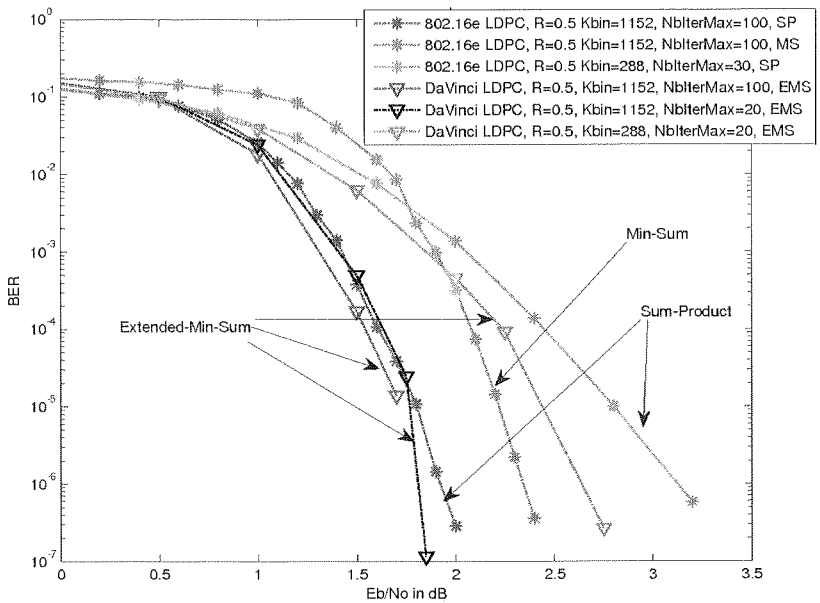


Fig. 8. WER in a function of E_b/N0 for DAVINCI codes, BTC and Wi-MAX B-LDPC

5. CONCLUSIONS

Channel coding based on NB-LDPC becomes an attractive solution for mobile radio systems. They can be seen as a patent-free competitor for turbo codes.

Using the truncated messages, an efficient implementation of the Extended-Min-Sum decoder can be proposed, which starts to be reasonable enough to compete with binary decoders. The performance of the low complexity algorithm with proper compensation is quite good with respect to the important complexity reduction.

The DAVINCI codes were shown to outperform the advanced reference codes of all IMT-Advanced candidates. The gain achieved by DAVINCI codes was found to vary with respect to the codeword length, coding rate and modulation order. The gain increases with any increase of the coding rate, or modulation order. On the other hand, the gain increases when the codeword length decreases. This leads to the conclusion

that DAVINCI codes are very promising solutions to achieve high spectral efficiency even in the challenging scenarios of short codeword lengths.
Iterative decoding lets improve coding performance, especially for short and medium codeword lengths. Limiting 10-times number of iterations a loss on the coding gain is about 1.5 dB.

6. REFERENCES

1. R. G. Gallager: Low density parity check codes, Ph.D. thesis, MIT, Cambridge, Mass., September 1960
2. D. J. C. MacKay and R.M. Neal: Good codes based on very sparse matrices; in: Colin Boyd, Cryptography and Coding. 5th IMA conference, number 1025 in Lecture Notes in Computer Science, pp. 100-111, Springer, Berlin/Heidelberg, 1995
3. R. G. Gallager: Low-Density Parity-Check Codes, IEEE Transactions on Information Theory, pp. 21-28, Jan. 1962
4. C. Berrou, A. Glavieux and P. Thitimajshima: Near Shannon limit error-correcting coding and decoding: Turbo-codes, Proc. ICC93, Geneve, Switzerland, pp. 1064-1070, May 1993
5. R. M. Tanner: A Recursive Approach to Low Complexity Codes, IEEE Trans. Inform. Theory, vol. 27, pp. 533-547, 1981
6. M. Davey and D. MacKay: Low density parity check codes over GF (q), Information Theory Workshop, Killarney, Ireland, pp. 70-71, 1998
7. A. Voicila, D. Declercq, F. Verdier, M. Fossorier and P. Urard: Low-complexity decoding algorithm for non-binary LDPC codes in high order fields, IEEE Transactions on Communications, 2008
8. A. Voicila, D. Declercq, F. Verdier, M. Fossorier and P. Urard: Low complexity, low memory EMS algorithm for non-binary LDPC codes, Proc. of ICC'07, Glasgow, UK, June 2007
9. A. Voicila, F. Verdier, D. Declercq, M. Fossorier and P. Urard: Architecture of a low-complexity non-binary LDPC decoder for high order fields, Proc. of ISCIT'07, Sydney, Australia, October 2007
10. Madhurima Potluri: Distributed Source Coding using Non Binary LDPC, Master Thesis, Bechelor of Engineering, Visweswaraiach Technological University, India, 2005
11. The Coded Modulation Library, http://www.iterativesolutions.com
12. T. J. Richardson, M. A. Shokrollahi and R. L. Urbanke: Design of Capacity-Approaching Low-Density Parity Check Codes, IEEE Trans. Inform. Theory, vol. 47, pp. 619-637, Feb. 2001
13. Matthew C. Davey: Error-correction using low density parity check codes, Ph.D. thesis, Cambridge
14. H. S. H. Wymeersch and M. Moeneclaey: Log-domain decoding of LDPC codes over GF(q), Proc. IEEE ICC, Paris, France, pp. 772-776, June 2004
15. X.-Y. Hu and E. Eleftheriou: Binary representation of cycle Tanner-graph GF(2^q) codes, Proc. IEEE ICC, Paris, France, pp. 528-532, June 2004
16. L. Barnault and D. Declercq: Fast decoding algorithm for LDPC over GF(2^q), Proc. IEEE ITW, Paris, France, 2003
17. D. Declercq and M. Fossorier: Decoding algorithms for nonbinary LDPC codes over GF(q), IEEE Trans. on Commun., vol. 55, pp. 633-643, April 2007
18. D. D. A. Venkiah and C. Poulliat: Design of cages with a randomized progressive edge growth algorithm, IEEE Commun. Letters, vol. 12, pp. 301-303, April 2008
19. M. F. C. Poulliat and D. Declercq: Design of regular (2,dc)-LDPC codes over GF(q) using their binary images, IEEE Trans. on Commun., vol. 56, pp. 1626-1635, October 2008
20. L. Ping and W. K. Leung: Decoding low density parity check codes with finite quantization bits, IEEE Commun. Lett., 4(2):pp. 62-64, February 2000

21 INFSCO-ICT-216203 DAVINCI, IR2.1.2 v1.0, State of Art (SoA): Targeted Systems Comparison / Overview, 30 June 2008
22 G. Caire, G. Taricco, E. Biglieri: Bit-interleaved coded modulation, IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 927-946, May 1998
23 INFSO-ICT-216203 DAVINCI, D4.1 Benchmark Non Binary LDPC codes: Report and Software (C-ANSI), July 2008

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 435–451

Log-MAP Decoding of Turbo Codes and Turbo Trellis-Coded Modulation Using Piecewise-Linear Approximations of the max* Operator

MICHAŁ SYBIS*, PIOTR TYCZKA*, STYLIANOS PAPAHARALABOS**, P. TAKIS MATHIOPOULOS**, GUIDO MASERA***, and MAURIZIO MARTINA***

*Chair of Wireless Communications, Poznań University of Technology, Poznań, Poland
e-mail: msybis@et.put.poznan.pl, tyczka@et.put.poznan.pl
**Institute for Space Applications and Remote Sensing, National Observatory of Athens, Athens, Greece
e-mail: spapaha@space.noa.gr, mathio@space.noa.gr
***VLSI Lab, Dipartimento di Elettronica, Politecnico di Torino, Torino, Italy
e-mail: guido.masera@polito.it, maurizio.martina@polito.it

Received 2009.06.16
Authorized 2009.08.18

A novel approach for approximating the *max** operator, used in the Log-MAP decoding algorithm, is presented. In this approximation it is not necessary to use any additional correction term, as previous approximation methods require. Several decoding algorithms are obtained for both turbo and turbo trellis-coded modulation (TTCM) codes with optimal and near-optimal bit error rate (BER) performance depending on a single parameter, namely the number of piecewise-linear (PWL) approximation terms used. The decoding complexity is estimated, showing the practical implementation advantages of the proposed method against the Log-MAP algorithm.

Keywords: Turbo codes, TTCM, decoding algorithms, Log-MAP algorithm, *max** operator, reduced-complexity algorithms

1. INTRODUCTION

Introduction of turbo codes in 1993 [1] and their impressive near Shannon-limit performance in an additive white Gaussian noise (AWGN) channel have triggered a lot of research devoted to this promising error-correcting coding technique. Turbo coding has become a part of the mainstream of telecommunication theory and practice. Consequently, turbo codes have been proposed for application in various communication

systems, such as deep space, cellular mobile and satellite communication networks. They have also been included in the specifications for the third-generation cellular standards, such as UMTS and cdma2000. Moreover, the turbo (iterative) decoding principles have found widespread applications beyond error control, including detection, interference suppression, equalization and synchronization.

Turbo codes were originally proposed for binary modulation (BPSK). One of their extensions towards the use of higher order modulations (e.g. 8-PSK, 16 QAM) and hence improving the overall bandwidth efficiency, is turbo trellis-coded modulation (TTCM) [2]. Essentially, TTCM scheme possesses a turbo encoder (parallel concatenation) structure with trellis-coded modulation (TCM) codes [3] as component codes. The performance of TTCM schemes with good component codes exceeds – at comparable complexity – both that of classical TCM with Ungerboeck codes and of turbo codes combined with M-ary two-dimensional modulations using Gray mapping. This makes TTCM a promising bandwidth-efficient channel coding scheme for various applications.

Turbo and turbo TCM codes can be decoded iteratively using the maximum *a posteriori* probability (MAP) algorithm [4] in the component decoders. In practical approach, however, a MAP decoder operating in the log domain is usually considered. Such a MAP algorithm is referred to as Log-MAP [5] [6] and in terms of performance it is almost equivalent to the optimal MAP, yet much more convenient for implementation.

The inherent feature of the Log-MAP decoding is the exact calculation of the max^* function using the Jacobian logarithm [7]. This calculation mainly determines the complexity of the Log-MAP algorithm. In recent years, several approaches aiming at simplification of the max^* operator for decoding turbo codes have been proposed [8]-[12]. Most of these methods reduce computational effort of the Log-MAP algorithm by using simpler approximations of the, so-called, *correction term* in the max^* operation. It should be noted, however, that in the open literature very little attention has been paid to reduced-complexity algorithms for turbo TCM codes.

In this paper, we consider Log-MAP decoding of turbo and turbo TCM codes employing novel approximations of max^* function based on a certain number of piecewise-linear (PWL) terms [13]. In contrast to all previously published methods [8]-[12], in this approach the max^* operator is favorably approximated as a whole and there is no need to use any correction terms. Performance evaluation results show the near-optimal and essentially optimal performance of the approximations. In our efforts to reduce further decoding complexity, we investigate a simplification of the algorithm in the case of turbo TCM codes [14].

The paper is organized as follows. Section 2 reviews turbo codes and turbo TCM. Next, in Section 3 Log-MAP decoding algorithm is presented for both schemes. In Section 4 novel max^* approximations are introduced and modified Log-MAP algorithms are formulated. A technique for reducing computational effort of the proposed algorithms is also described in this Section. Simulation results studying the performance of turbo and turbo TCM codes with the novel approximation for the max^* operator,

as well as decoding complexity comparisons are presented in Section 5. Conclusions are given in Section 6.

2. A BRIEF REVIEW OF TURBO CODES AND TURBO TCM

2.1. TURBO CODES

A turbo encoder consists of two parallel concatenated recursive systematic convolutional (RSC) component encoders separated by an interleaver (Fig. 1). The turbo code interleaver permutes the information bit sequence and the output is passed to the second RSC encoder. In order to increase the code rate over 1/3, the turbo encoder usually makes use of the puncturer.

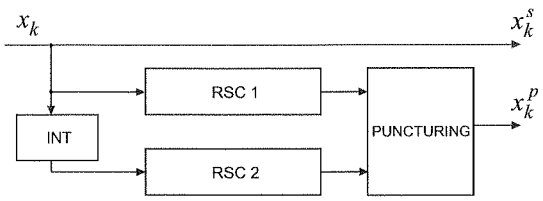


Fig. 1. Turbo encoder

The turbo decoding principle is based on the iterative process between two *soft-input soft-output* (SISO) decoders. As shown in Fig. 2, the input of the first SISO decoder is fed by the *a priori* information of bits and the received channel values corresponding to systematic and parity bits of the first encoder. It then produces a soft output value. Next, the so-called extrinsic information is used as *a priori* information for the subsequent SISO module. The extrinsic information of the latter is deinterleaved and looped back to the first SISO decoder, which finishes one round of the iterative process. After a certain number of iterations, hard decisions are taken from the second decoder output.

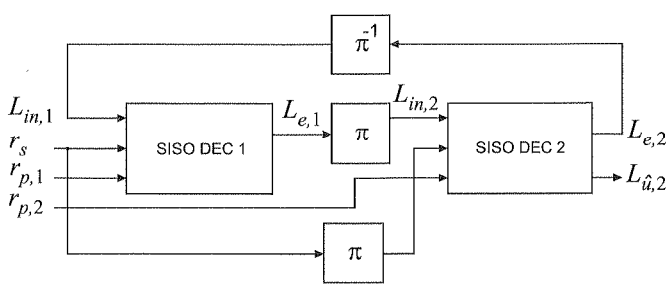


Fig. 2. Structure of a turbo decoder

2.2. TURBO TCM CODES

TTCM is a combination of trellis-coded modulation with the concept of turbo coding. Parallel concatenation of TTCM coding is seen as a direct extension of classical

turbo coding. Its block diagram is shown in Fig. 3. In this solution, two TCM encoders are deployed that work in parallel. The input data to the first TCM encoder is a given data sequence. The input data to the second TCM encoder is the same data sequence but after interleaving. The final sequence of transmitted symbols is generated by selecting symbols alternately from the two encoders (output symbols of the second one may be deinterleaved but this operation can be carried out in the decoder as well). Those symbols are subsequently mapped on the M-ary phase shift keying (M-PSK) or M-ary quadrature amplitude modulation (M-QAM) constellation points and shaped with the transmit filter before the signal is transmitted.

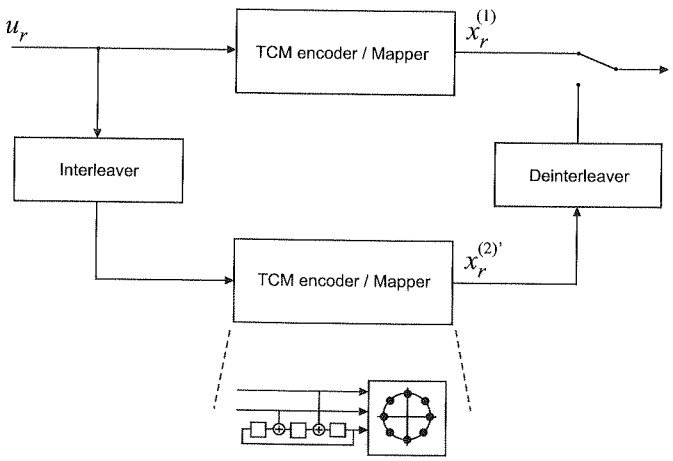


Fig. 3. Parallel TTCM encoder with 8PSK mapping

A TTCM decoder is structured analogously to the classical turbo decoder and is depicted in Fig 4. The decoder operates on the received sequence of noisy symbols y. This sequence is alternately fed to the first and second SISO decoders which are related to the first and second encoders, respectively. Similarly to turbo codes, the decoding process is performed iteratively and the SISO decoders employ a MAP-like algorithm suitably modified for nonbinary trellises [15].

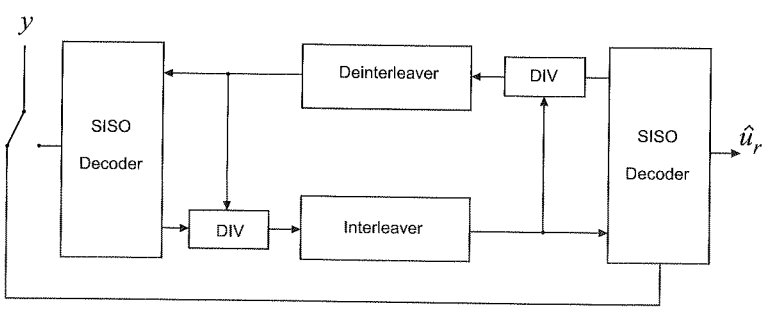


Fig. 4. Block diagram of parallel TTCM decoder

3. LOG-MAP DECODING ALGORITHM

3.1. TURBO CODES

Let us consider the Log-MAP algorithm for turbo decoding. Assume that u is an information block of N bits. This block is turbo encoded, binary phase-shift keying (BPSK) modulated, with possible values ± 1, and transmitted through an additive white Gaussian noise (AWGN) channel. After being corrupted by noise, the block y is assumed to be received. Then, soft-demodulation is performed before the turbo decoder estimates the transmitted block of bits. Consider a trellis transition that occurs from a state M', at time instant $k-1$, to a state M at time instant k. The forward recursion $\bar{\alpha}_k$ and the backward recursion $\bar{\beta}_k$ can be computed recursively, as

$$\begin{aligned}\bar{\alpha}_k(M) &= \ln \sum_{M'} \exp\{\bar{\alpha}_{k-1}(M') + \bar{\gamma}_k(M', M)\} = \\ &= \max_{M'}{}^{*}\{\bar{\alpha}_{k-1}(M') + \bar{\gamma}_k(M', M)\}\end{aligned} \tag{1}$$

$$\begin{aligned}\bar{\beta}_{k-1}(M') &= \ln \sum_{M} \exp\{\bar{\beta}_k(M) + \bar{\gamma}_k(M', M)\} = \\ &= \max_{M}{}^{*}\{\bar{\beta}_k(M) + \bar{\gamma}_k(M', M)\},\end{aligned} \tag{2}$$

where $\bar{\gamma}_k$ is the branch metric associated with the corresponding trellis transition and with the *a priori* information in the iterative process. Note that both $\bar{\alpha}_k$ and $\bar{\beta}_k$ have to be initialized first [1].

The max^* operation is defined as [7]

$$\max{}^{*}(x_1, x_2) = \ln(\exp x_1 + \exp x_2) \tag{3}$$

and is calculated using the Jacobian logarithm

$$\begin{aligned}\max{}^{*}(x_1, x_2) &= \max(x_1, x_2) + \ln\{1 + \exp(-|x_2 - x_1|)\} = \\ &= \max(x_1, x_2) + f_c(|x_2 - x_1|),\end{aligned} \tag{4}$$

where $f_c(.)$ is a correction function.

The decoder's soft-output value of the transmitted bit u_k, in terms of the log-likelihood ratio (LLR), is computed as

$$\begin{aligned}L(\hat{u}_k) &= \ln\left(\frac{\sum\limits_{(M',M)\, u_k=+1} \exp\{\bar{\alpha}_{k-1}(M') + \bar{\gamma}_k(M', M) + \bar{\beta}_k(M)\}}{\sum\limits_{(M',M)\, u_k=-1} \exp(\bar{\alpha}_{k-1}(M') + \bar{\gamma}_k(M', M) + \bar{\beta}_k(M)\}}\right) = \\ &= \max_{(M',M)\, u_k=+1}{}^{*} \{\bar{\alpha}_{k-1}(M') + \bar{\gamma}_k(M', M) + \bar{\beta}_k(M)\} \\ &\quad - \max_{(M',M)\, u_k=-1}{}^{*} \{\bar{\alpha}_{k-1}(M') + \bar{\gamma}_k(M', M) + \bar{\beta}_k(M)\},\end{aligned} \tag{5}$$

In the Log-MAP algorithm, the correction function $f_c(.)$ is practically implemented with a look-up table (LUT) with eight values. If the correcting value of the LUT is omitted, then the Log-MAP algorithm simplifies to the Max-Log-MAP algorithm [5], [6].

3.2. TURBO TCM CODES

Similar to turbo codes, the Log-MAP gives the logarithm of probability $\ln P(d_k|\underline{y})$ for each data symbol d_k, and for all k, where $y = (y_1,..., y_N)$ is the received sequence of N noisy TCM symbols. The soft output of the Log-MAP algorithm is computed as

$$\ln P(d_k = i|y) = \ln \sum_{M} \sum_{M'} \exp\{\overline{\alpha}_{k-1}(M') + \overline{\gamma}_i(y_k, M', M) + \overline{\beta}_k(M)\} \tag{6}$$

where

$$\begin{aligned}\overline{\alpha}_k(M) = \ln \alpha_k(M) = \ln \sum_{M'} \exp\{\overline{\alpha}_{k-1}(M') + \overline{\gamma}_T(y_k, M', M)\} \\ - \ln \sum_{M} \sum_{M'} \exp\{\overline{\alpha}_{k-1}(M') + \overline{\gamma}_T(y_k, M', M)\}\end{aligned} \tag{7}$$

$$\begin{aligned}\overline{\beta}_k(M) = \ln \beta_k(M) = \ln \sum_{M''} \exp\{\overline{\gamma}_T(y_{k+1}, M, M'') + \overline{\beta}_{k+1}(M'')\} \\ - \ln \sum_{M''} \sum_{M} \exp\{\overline{\alpha}_k(M) + \overline{\gamma}_T(y_{k+1}, M, M'')\}\end{aligned} \tag{8}$$

are the logarithms of the forward and backward variables and $S_{k+1} = M$ is the state at step k+1. The remaining variables in (6)-(8) are defined as follows

$$\overline{\gamma}_i(y_k, M', M) = \ln \gamma_i(y_k, M', M) \tag{9}$$

$$\overline{\gamma}_T(y_k, M', M) = \ln \gamma_T(y_k, M', M) = \ln \sum_{i=0}^{2^m-1} \gamma_i(y_k, M', M) \tag{10}$$

where $\gamma_i\ (y_k, M', M)$ denotes the branch transition probability for the step k in the trellis, i.e.

$$\gamma_i(y_k, M', M) = p(d_k = i, y_k, S_k = M|S_{k-1} = M') \tag{11}$$

4. NOVEL max* APPROXIMATIONS FOR LOG-MAP DECODING

In recent work [13], dealing with the convex log-sum-exp (*lse*) function, the authors presented a constructive algorithm for finding the best PWL approximation terms of

the bivariate *lse* function. Noting that the bivariate *lse* function is equivalent to the max^* operator defined in (3), it is straightforward to conclude that the approximations derived in [13] can be applied, for the first time, to Log-MAP decoding of turbo and TTCM codes.

A novel approximation for the max^* operator is obtained using *max* operator and a small number of PWL terms. Hence, the novel approximation can be expressed as follows

$$\max^*(x_1, x_2) \approx \max(\kappa_1 x_1 + \lambda_1 x_2 + \mu_1, \ldots, \kappa_i x_1 + \lambda_i x_2 + \mu_i) \tag{12}$$

where κ_i, λ_i, and μ_i are real positive values and $i \geq 2$. The exact number of PWL terms depends on the approximation error resulting from the original bivariate *lse* function. Note that approximation (12) does not use any additional correction term.

The best PWL approximations of the max^* operator with different number of terms, denoted with r, are shown in Table 1. The approximation error reduces in the order of $\sqrt{2}/r^2$ [13]. It is easy to notice from Table 1 that in case of turbo/turbo TCM decoding, the $r = 2$ approximation is identical to the Max-Log-MAP algorithm.

Table 1

Best approximations of the max^* operation using a certain number of PWL terms and the *max* operation [13]

No. of terms (r)	Resulting $\max^*$ approximation	No. of max ops.	Appr. error
2	$\max(x_1, x_2)$	1	0.693
3	$\max\begin{pmatrix} x_1, \\ 0.5 \cdot x_1 + 0.5 \cdot x_2, \\ x_2 \end{pmatrix}$	2	0.223
4	$\max\begin{pmatrix} x_1, \\ 0.271 \cdot x_1 + 0.729 \cdot x_2 + 0.584, \\ 0.729 \cdot x_1 + 0.271 \cdot x_2 + 0.584, \\ x_2 \end{pmatrix}$	3	0.109
5	$\max\begin{pmatrix} x_1, \\ 0.167 \cdot x_1 + 0.833 \cdot x_2 + 0.45, \\ 0.5 \cdot x_1 + 0.5 \cdot x_2 + 0.693, \\ 0.833 \cdot x_1 + 0.167 \cdot x_2 + 0.45, \\ x_2 \end{pmatrix}$	4	0.065

Inspection of Log-MAP equations from Section 3 reveals that for some of them, i.e. (5) and (6)-(8), it is necessary to compute the expression of $\ln(\exp x_1 + \ldots + \exp x_n)$ with $n > 2$. To obtain max^* operation for more than two arguments, recursive computations must be performed. For example, considering $n = 3$ (three arguments), it yields

$$\max^*(x_1, x_2, x_3) = \max^*\{\max^*(x_1, x_2), x_3\}. \tag{13}$$

It implies that the max^* approximations from Table 1 need to be successively applied n-1 times. In order to reduce computational effort of modified Log-MAP algorithm

with novel max^* operator, the *Two Largest* (TL) approach can be adopted [14]. The idea is to find the two largest values of x_i and then apply the single max^* approximation with these two values as arguments. This simplified decoding algorithm will also be investigated in computer simulations.

5. SIMULATION RESULTS AND DECODING EFFORT COMPARISONS

To evaluate performance of Log-MAP decoding with novel max^* approximations from Table 1, computer simulations were carried out for both turbo code and turbo TCM schemes in the AWGN channel with the one-sided noise power spectral density N_0. We also applied scaling of the extrinsic information (as in [16]) for additional improving the BER performance.

5.1. TURBO CODE RESULTS

A 16-state turbo code with coding rate R=1/2 and generator polynomials (1, 33/23) in octal form representing the feedforward and backward polynomials, respectively, was examined. An information sequence (frame) of N = 1000 bits and a pseudo-random interleaver were assumed. A maximum of 10 decoding iterations at the receiver were performed. BER performance results for the best approximations of the max^* operator and different number of PWL terms (denoted with r) are depicted in Figs. 5 and 6.

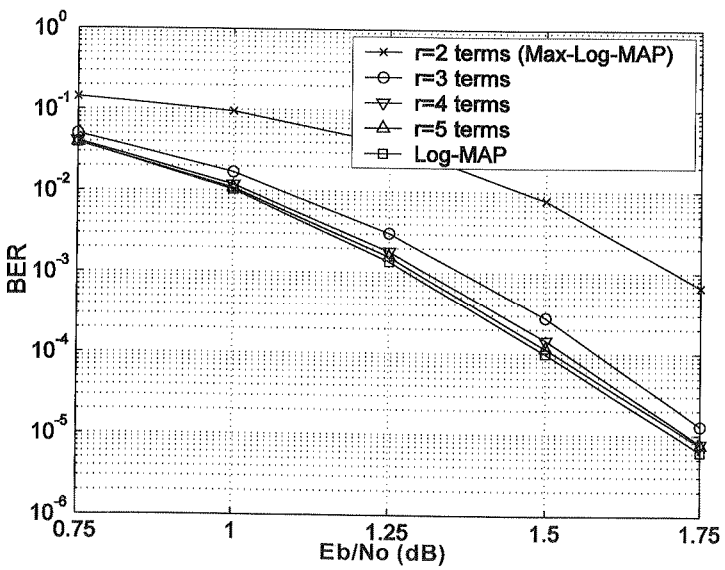


Fig. 5. Turbo code BER performance with the best approximations of the max^* operator and different number of PWL terms, denoted with r. N = 1000 bits, BPSK modulation, AWGN channel, and 10 iterations

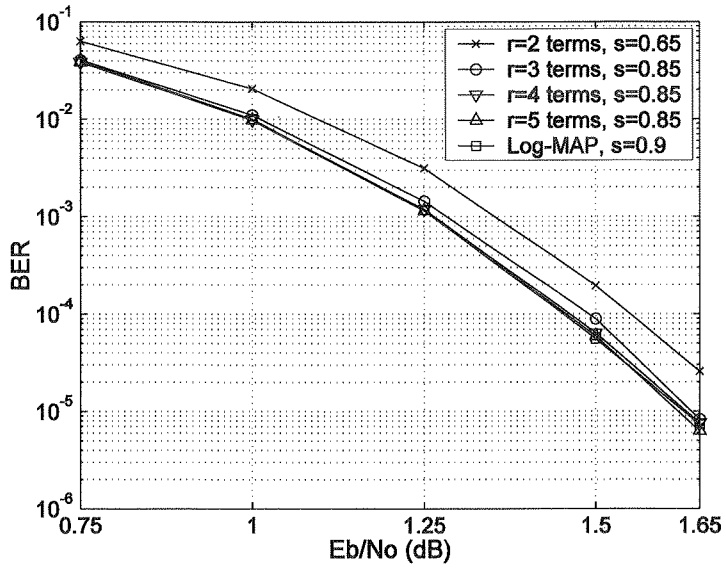


Fig. 6. As in Fig. 5 but with the extrinsic information scaled by a factor of s

As shown in Fig. 5, the $r = 5$ approximation achieves almost identical Log-MAP performance. The performance degradation of the $r = 4$ approximation with respect to the Log-MAP algorithm is less than 0.1 dB at BER of 10^{-5}. The $r = 3$ approximation has performance degradation of approximately 0.1 dB at the same BER value. When scaling is used (Fig. 6), it is expected that all algorithms improve the BER performance. In particular, the scaling factor, denoted with s, had the value of 0.65 for Max-Log-MAP, 0.9 for Log-MAP, and 0.85 for the rest of the algorithms. Note that for a fair comparison, scaling has been applied in the Log-MAP algorithm. As shown in Fig. 6, both the $r = 5$ and $r = 4$ approximations achieve identical Log-MAP performance. The $r = 3$ approximation degrades less than 0.1 dB at BER of 10^{-5}.

5.2. TURBO TCM RESULTS

The proposed max^* approximations were investigated in TTCM schemes with two parallel concatenated rate-3/4 8-state Ungerboeck's TCM encoders and 16-QAM signals [2]. An S-random interleaver with spreading factor $S = 13$ was assumed. Two TTCM schemes with block sizes of $N = 684$ and 5000 symbols, respectively, were examined. For comparison purposes (as it was also done for a turbo code in Section 5.1), the conventional Log-MAP (with LUT storing eight values) and Max-Log-MAP algorithms were also evaluated. Additionally, for improving the BER performance, the algorithms with scaling of the extrinsic information were also investigated. For both TTCM schemes (i.e. for both block sizes) the best found value of the scaling factor s is 0.85 for all algorithms, except for Max-Log-MAP, which is scaled by $s = 0.75$. The number of iterations in TTCM decoder was 8 for all algorithms. Performance evaluation

results are shown in Figs. 7 - 10. In all the Figures with TTCM performance, modified Log-MAP algorithm with r PWL terms is denoted as LM-r whereas (conventional) Log-MAP and Max-Log-MAP algorithms are denoted as LM and MLM, respectively.

As it can be seen from Fig. 7, both the $r = 4$ and $r = 5$ approximations achieve almost identical to Log-MAP performance. The $r = 3$ approximation has a very small performance degradation with respect to Log-MAP, i.e. less than 0.1 dB at BER of 10^{-4}. From Fig. 8 it is noticed that when scaling is applied, the algorithms improve their BER performance and all approximations offer practically identical Log-MAP performance. Similar conclusions can be drawn from Figs. 9 and 10 for TTCM transmission with longer block of symbols.

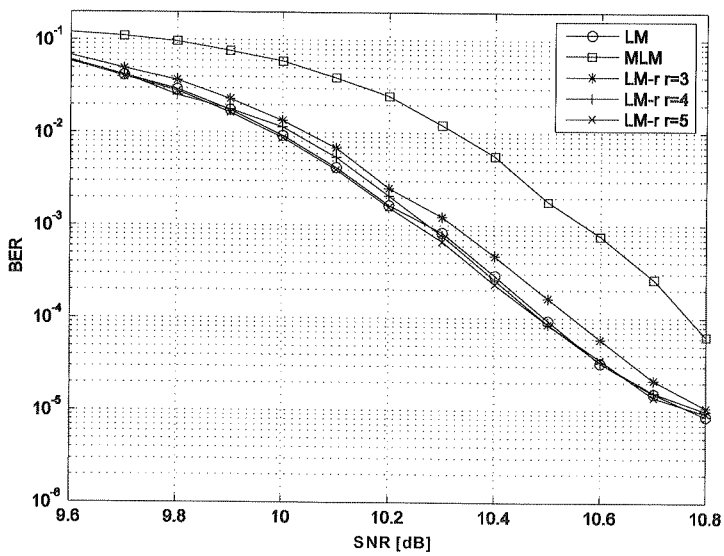


Fig. 7. TTCM code BER performance with the best approximations of the max^* operator and different number of PWL terms, denoted with r. N=684 symbols, 16-QAM modulation, AWGN channel, and 8 iterations

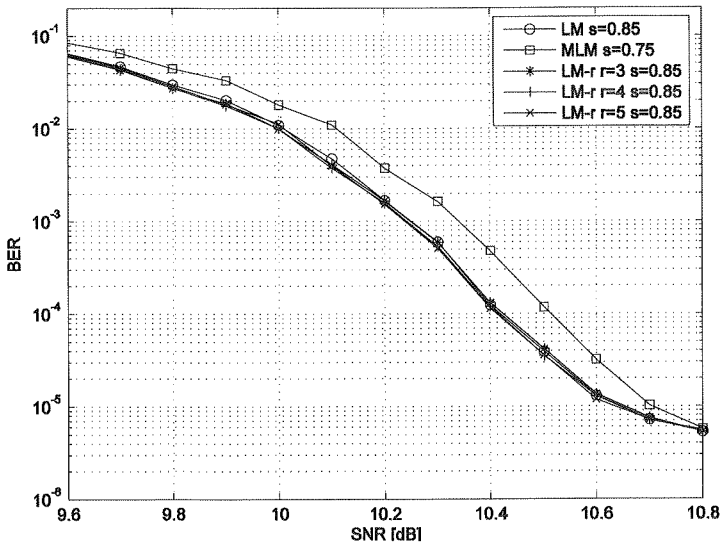


Fig. 8. As in Fig. 7 but with the extrinsic information scaled by a factor of s

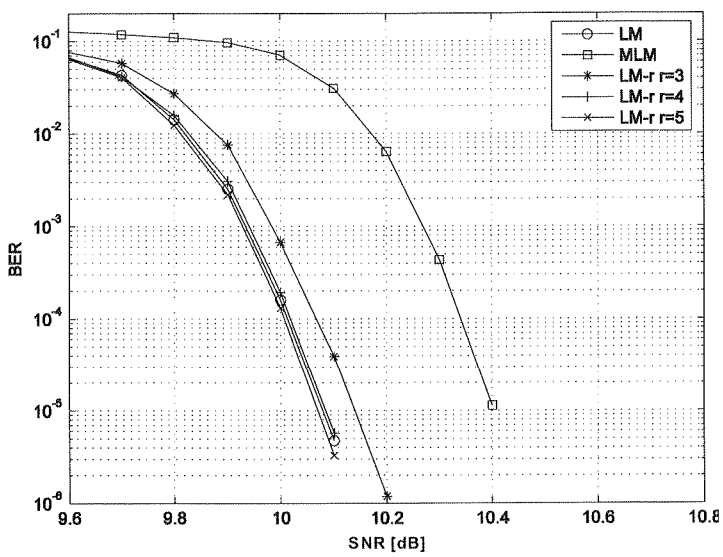


Fig. 9. TTCM code BER performance with the best approximations of the max^* operator and different number of PWL terms, denoted with r. N=5000 symbols, 16-QAM modulation, AWGN channel, and 8 iterations

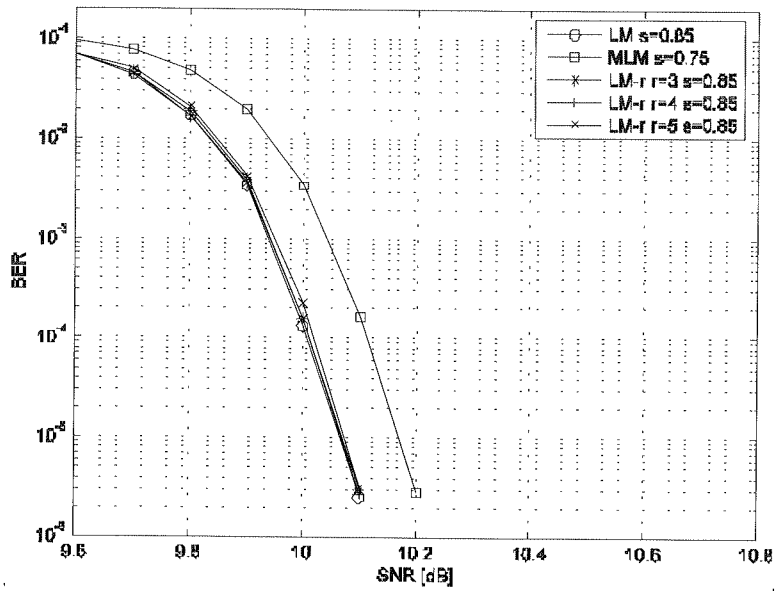


Fig. 10. As in Fig. 9 but with the extrinsic information scaled by a factor of s

5.3. SIMPLIFIED DECODING OF TURBO TCM CODES

Simulated BER performance results for the TTCM schemes with modified Log-MAP algorithm using novel max^* operator and TL approach are shown in Figs. 11-14.

The analysis of the results obtained for simplified decoding using only two largest values in max^* operation shows that performance degradation with respect to Log-MAP is very small for both symbol block sizes. The r=4 and r=5 approximations are inferior to Log-MAP only of less than 0.1 dB for BER< 10^{-4}. However, as can be seen from Figs. 12 and 14, the use of proper scaling suppresses this gap for all values of r. Consequently, the LM-TL algorithms with scaling achieve near-optimal performance, yet having reduced complexity.

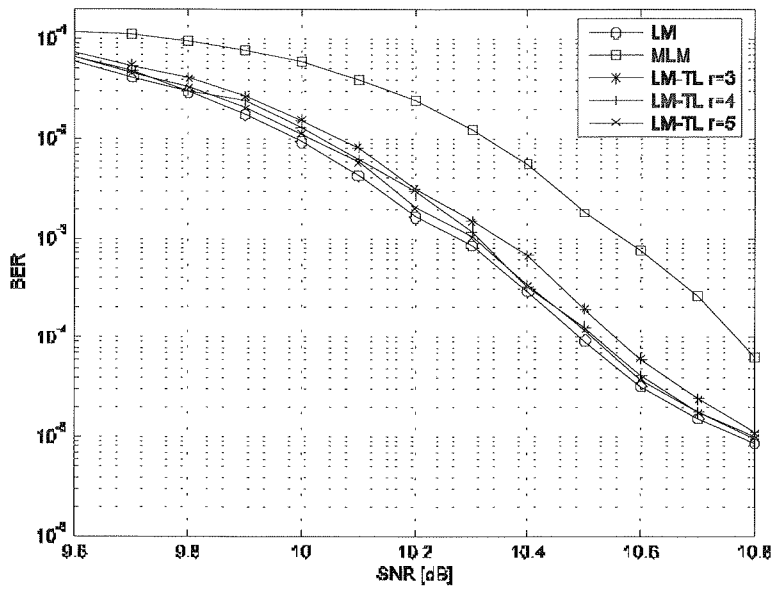


Fig. 11. TTCM code BER performance with the best approximations of the max^* operator, different number of PWL terms, denoted with r, and TL approach. N=684 symbols, 16-QAM modulation, AWGN channel, and 8 iterations

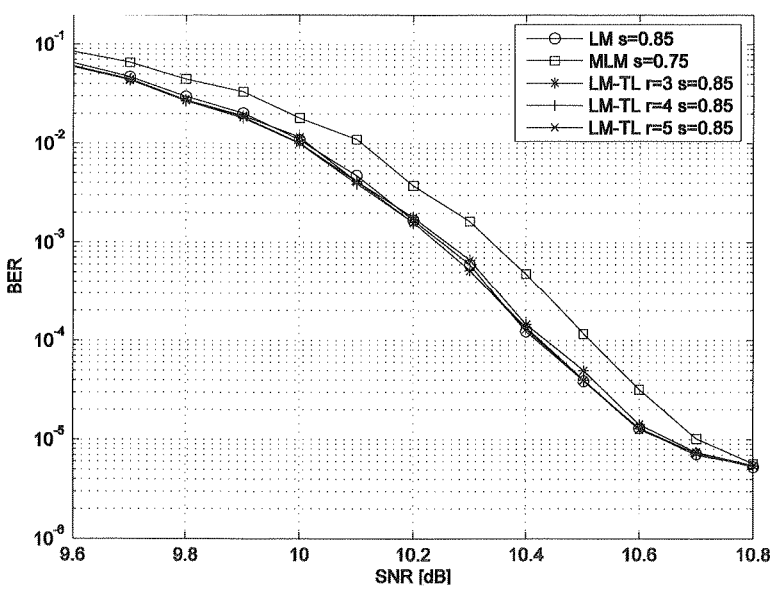


Fig. 12. As in Fig. 11 but with the extrinsic information scaled by a factor of s

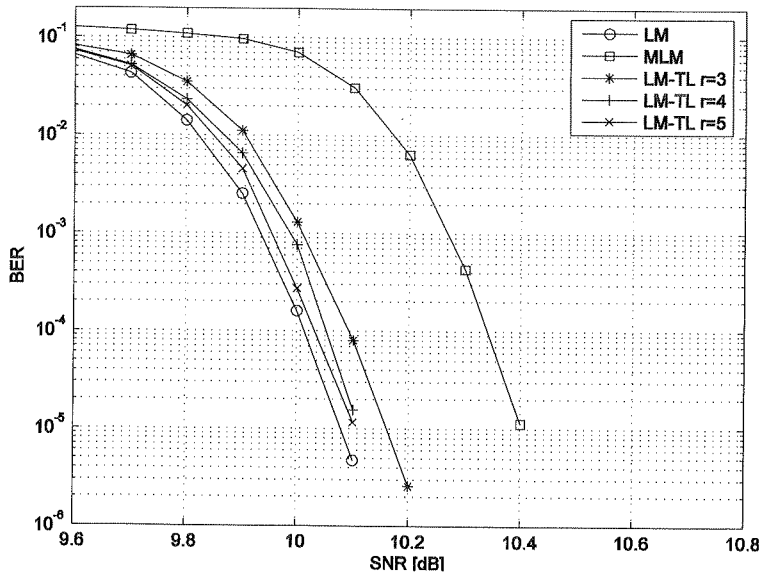


Fig. 13. TTCM code BER performance with the best approximations of the max^* operator, different number of PWL terms, denoted with r, and TL approach. N=5000 symbols, 16-QAM modulation, AWGN channel, and 8 iterations

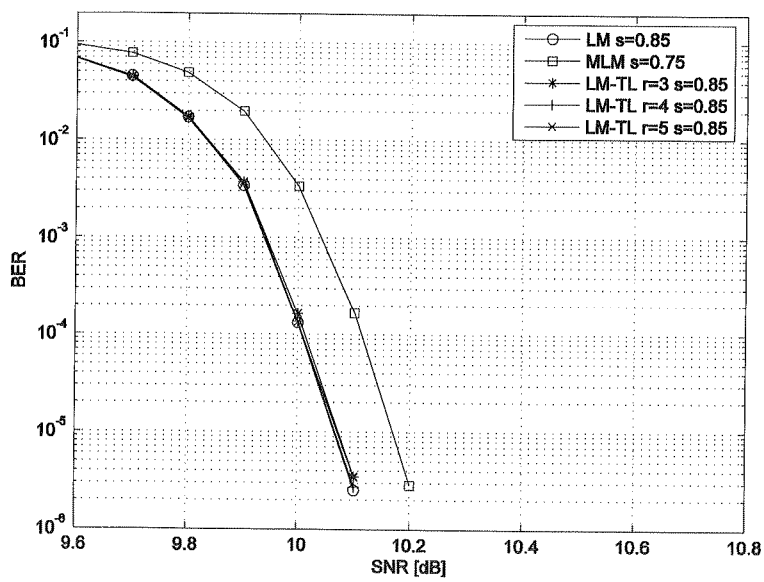


Fig. 14. As in Fig. 13 but with the extrinsic information scaled by a factor of s

5.4. COMPLEXITY ISSUES OF DECODING ALGORITHMS

Synthesizable VHDL descriptions have been produced for the proposed max^* approximations, as well as for Log-MAP (eight entry LUT) and Max-Log-MAP algorithms in turbo decoding. In order to derive fair comparisons, the same area optimization effort of the synthesis tool must be guaranteed for all cases. For this purpose, although all considered implementations of the max^* are pure combinational architectures, registers have been placed at the architecture inputs and output: this allows setting a unique clock frequency constraint for all considered cases, f_{CK} = 200 MHz. Synthesis results obtained in terms of area occupied by the combinational part on a 130 nm standard cell CMOS technology are given in Table 2, for metrics represented with 8, 10 and 12 bits, respectively. These results show that the proposed approximations have a low implementation cost. Particularly, the occupied area with r = 3 is 25% smaller than that required by the Log-MAP. However, the r = 5 approximation, which is not shown in Table 2, occupies an area greater than that required by the Log-MAP.

Table 2

Occupied area (square μm) on a 130 nm standard cell technology

Metrics representation	Max-Log-MAP	Log-MAP	Proposed (r=3)	Proposed (r=4)
8 bits	250.133	1022.72	768.553	1054.996
10 bits	312.666	1276.888	980.359	1301.094
12 bits	373.182	1613.76	1198.217	1694.448

For TTCM schemes, decoding effort comparison for considered algorithms has been shown in Table 3. Table 3 contains the required number of operations (i.e. addition, multiplication, comparison, bit shift, conversion to integer and assignment) per decoding step in the implementation of the algorithms for the investigated TTCM scheme. It can easily be noticed in Table 3 that most of the algorithms are less complex than the conventional Log-MAP. In particular, decoding with only two largest values in the novel max^* approximations offers smaller number of operations per decoding step as compared to Log-MAP for all values of r under consideration. The reduction is significant and is about 35%, 25% and 17% for r=3, 4 and 5 PWL terms, respectively. Taking into account that near-optimal performance can be achieved with this simplified method, the TL approach combined with novel max^* approximations seems to be an attractive modification of the Log-MAP for decoding TTCM codes. It should also be noted that decoding complexity increase due to scaling is minimal and thus can be neglected.

Table 3

Complexity comparison of decoding algorithms per decoding step in TTCM scheme

Algorithm	***LM***	***LM-r, r=3***	***LM-TL, r=3***	***LM-r, r=4***
Additions	680	680	392	1016
Multiplication	0	0	0	672
Comparisons	357	357	357	525
Bit shifts	168	168	24	0
Conversion to integer	168	0	0	0
Assignment	233	281	269	449
Overall	**1606**	**1486**	**1042**	**2662**
Algorithm	***LM-TL, r=4***	***LM-r, r=5***	***LM-TL, r=5***	***MLM***
Additions	440	1352	488	344
Multiplication	96	672	96	0
Comparisons	381	693	405	189
Bit shifts	0	168	24	0
Conversion to integer	0	0	0	0
Assignment	293	617	317	173
Overall	**1210**	**3502**	**1330**	**706**

6. CONCLUSIONS

We have presented novel approximations for the max^* operation for reduced complexity decoding of turbo and TTCM codes. In this approach, the max^* operation makes use of a certain number of PWL terms and deploys the *max* operator. Computer-based simulation results for both turbo and TTCM codes have shown the near-optimal performance for the investigated algorithms against the Log-MAP decoding. Furthermore, complexity estimation has revealed that the proposed method is also attractive from the implementation point of view.

7. ACKNOWLEDGMENT

This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract no. 216715).

8. REFERENCES

1. C. Berrou, A. Glavieux and P. Thitimajshima: *Near Shannon limit error correcting coding and decoding: Turbo codes*. Proc. IEEE Int. Conf. Commun. (ICC), pp. 1064-1070, Geneva, Switzerland, 1993
2. P. Robertson and T. Wörz: *Bandwidth-efficient turbo trellis-coded modulation using punctured component codes*, IEEE J. Sel. Areas Commun., vol. 16, pp. 206-218, Feb. 1998

3. G. Ungerboeck: *Channel coding with multilevel/phase signals*, IEEE Trans. Inform. Theory, vol. IT-28, pp. 55-67, Jan. 1982
4. L. R Bahl, J. Cocke, F. Jelinek and J. Raviv: *Optimal decoding of linear codes for minimizing symbol error rate*, IEEE Trans. Inf. Theory, IT-20, vol. 3, pp. 284-287, 1974
5. P. Robertson, E. Villebrun, and P. Hoeher: *A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain*, Proc. IEEE Int. Conf. Commun. (ICC), pp. 1009-1013, Seattle, USA, Jun. 1995
6. P. Robertson, P. Hoeher, and E. Villebrun: *Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding*, European. Trans. Telecomm., vol. 8, pp. 119-125, Mar.-Apr. 1997
7. A. J. Viterbi: *An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes*, IEEE J. Sel. Areas Commun., vol. 16, pp. 260-264, Feb. 1998
8. W. J. Gross and P. G. Gulak: *Simplified MAP algorithm suitable for implementation of turbo decoders*, Electron. Lett., vol. 34, no. 16, pp. 1577-1578, Aug. 1998
9. J. F. Cheng and T. Ottosson: *Linearly approximated Log-MAP algorithms for turbo decoding*, Proc. IEEE Vehic. Techn. Conf. (VTC), Spring, pp. 2252-2256, Tokyo, Japan, May 2000
10. B. Classon, K. Blankenship, and V. Desai: *Channel coding for 4G systems with adaptive modulation and coding*, IEEE Wireless Commun. Mag., vol. 9, no. 2, pp. 8-13, Apr. 2002
11. H. Wang, H. Yang, and D. Yang: *Improved Log-MAP decoding algorithm for turbo-like codes*, IEEE Commun. Letters, vol.10, no.3, pp. 186- 188, Mar. 2006
12. S. Talakoub, L. Sabeti, B. Shahrrava, M. Ahmadi: *An improved Max-Log-MAP algorithm for turbo decoding and turbo equalization*, IEEE Trans. Instrum. Meas., vol.56, no.3, pp. 1058-1063, Jun. 2007
13. K. L. Hsiung, S. J. Kim, and S. Boyd: *Tractable approximate robust geometric programming*, Springer Optim. Eng. Journal, vol. 9, no. 2, pp. 95-118, Jun. 2008
14. M. Sybis, P. Tyczka, S. Papaharalabos, and P. T. Mathiopoulos: *Reduced-complexity algorithms for near-optimal decoding of turbo TCM codes*, Electron. Lett., vol. 45, no. 5, pp. 278-280, Feb. 2009
15. C. B. Schlegel and L. C. Perez: *Trellis and Turbo Coding*, Wiley and Sons, Inc., 2004
16. J. Vogt and A. Finger: *Improving the Max-Log-MAP turbo decoder*, IEE Electron. Lett., vol. 36, no. 23, pp. 1937-1939, Nov. 2000

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 453–470

Functional Decomposition System Dedicated to Multi-Output Boolean Functions

PAWEŁ MORAWIECKI

Kielce University of Commerce, Institute of Computer Science
Peryferyjna 15, 25-562 Kielce, Poland
e-mail: pawelm@wsh-kielce.edu.pl

Received 2009.05.28
Authorized 2009.07.31

The functional decomposition has important applications in many fields of modern engineering and science (FPGA synthesis, information systems, neural networks and many others). In this paper, a new method for functional decomposition is presented. The proposed approach is dedicated to the multi-output boolean functions. It uses the concept of dividing the complex function into single-output functions and utilizing common information between these functions. Additionally, it allows the multi-output function to be represented as a set of separate truth tables for each output, which might be very beneficial for time and memory usage of the system. To test the proposed solutions, a prototype tool was implemented and the results are presented in the paper.

Keywords: boolean function, logic synthesis, functional decomposition

1. INTRODUCTION

Functional decomposition consists of breaking down a complex system of discrete functions into a network of smaller and relatively independent cooperating subsystems in such a way that the original system's behaviour is preserved, while some constraints are satisfied and some objectives are optimised. The motivation for using functional decomposition in system analysis and design is to reduce the complexity of the problem by divide-and-conquer paradigm and to find an appropriate network of coherent subsystems: a system is decomposed into a set of smaller subsystems, such that each of them is easier to analyse, understand or synthesise.

One of the functional decomposition applications is a hardware mapping. It is especially useful on case of FPGA (*Field Programmable Gate Array*) programmable

architectures composed of logic cell components based on Look-Up Tables. Logic cell can implement any function of given input variables. It can be programmed to perform the function of basic logic gates such as AND, XOR, or more complex combinational functions such as decoders or mathematical functions.

If a function is going to be realized by FPGA device it can be divided into logic blocks (i.e. small sub-functions) that can cooperate in such way that the original system's behaviour is preserved. If logic block is of the size of a single logic cell then such a logic design can be directly mapped into FPGA architecture. The functional decomposition provides such design methodology, useful for FPGA devices.

Decomposition-like synthesis methods are not limited only to logic synthesis of digital circuits. The strong motivation for developing decomposition techniques comes recently from modern research areas such as pattern recognition, knowledge discovery and machine learning in artificial intelligence [12]. The quality of function decomposition can be described by the number of logic cells (sub-functions) or the number of levels. The fewer logic cells, the better decomposition.

Logic circuits usually have many outputs. The functional decomposition in such case has additional advantage, since it allows extracting shared logic. However the synthesis of multi-output circuit requires such representation of multiple-output function that allows efficient creation of sub-circuits common to all of the outputs. There have been proposed methods to represent multiple-output functions by using compact binary decision diagrams (BDDs). One of the first methods is a multi-terminal binary decision diagram (MTBDD) [22]. Unfortunately, MTBDDs tend to be too large to construct. The second method is a binary decision diagram (BDD) for the characteristic function (CF) of the multiple-output function. The advantage of the CF is its small evaluation time. CFs are used in logic simulation and multi-level logic optimization [21]. The third method is a shared binary decision diagram (SBDD) [22]. In many cases, SBDDs are smaller than corresponding MTBDDs and BDDs for CFs. Recently some other types of BDDs have been proposed and used in the prototype decomposition systems [14,15].

A very promising decomposition alternative based on the blanket calculus approach to circuit synthesis has been proposed in [2]. This synthesis process does not use the technology independent optimization phase, because this would destroy the design-freedom represented by "*don't cares*" of the original function specification and construct the network without any relation to the actual synthesis target. The functional decomposition is applied directly to the original function specified by truth table, in order to use its whole design-freedom to directly construct a feasible network of LUTs optimized for a given FPGA. Moreover, the concept of parallel decomposition was introduced and effectively applied in the so-called balanced decomposition method implemented in university synthesis system DEMAIN [8]. Based on the input variable analysis of each single output of a multi-output function F, parallel decomposition separates F into two or more sub-functions, each of which has as its inputs and outputs a subset of the original inputs and outputs of F. Although the decomposition method uses a mix of both the classic functional decomposition and parallel decomposition,

the crucial process in the whole mapping is the functional decomposition, which (in contrast to parallel) is called serial decomposition. Using both procedures leads to very good results [7]. However, the decomposition strategy used in DEMAIN brings some limitations. Once the parallel decomposition is applied, two new sub-functions are being decomposed separately, till the end of the whole decomposition process. Because of this, the common information between two sub-functions is lost and can not be utilized. What we mean by common information here is that some inputs are the same in sub-functions. Additionally, for a given input variables partition, the sub-functions may give the similar incompatibility graphs (explained in more detail in next sections).

Functional decomposition methods based on BDD, as well as on blanket calculus require the decomposed multi-output function to be represented as single BDD or truth table. This makes the decomposition of function sets a difficult task, when single functions depend on different sets of input variables or are specified by different input patterns.

The main idea of the proposed decomposition method is to keep the possibility of utilizing common information between sub-functions through the whole multi-level decomposition process . Initially, the multi-output function is divided into single-output functions and then for some of them a shared G-block is constructed. Once the G-block is constructed, all the single output functions can take part in the next step of the decomposition process. Therefore, functions are not taken apart as in Demain decomposition strategy.

2. BASIC DEFINITIONS AND THEOREMS

Functional decomposition relies on partitioning a switching function into a network of two smaller and independent co-operating sub-functions, in such a way that the original system's behaviour is preserved (Fig. 1).

The set X of function's input variable is partitioned into two subsets: *free variables* U *and bound variables* V, such that $U \cup V = X$. Let us Assume that the input variables $x_1,...,x_n$ have been relabelled in such a way that:

$$U = \{x_1, ..., x_r\} \text{ and}$$
$$V = \{x_{n-s+1}, ..., x_n\}.$$

Consequently, for an n-tuple x, the first r components are denoted by x^U, and the last s components, by x^V.

Definition 1:
Let F be a Boolean function, with $n > 0$ inputs and $m > 0$ outputs, and let (U, V) be as above. Assume that F is specified by a set $\mathcal{F}$ of the function's cubes. Let G be a function with s inputs and p outputs, and let H be a function with $r + p$ inputs and m

outputs. The pair (G, H) represents a serial decomposition of F with respect to (U, V), if for every minterm b relevant to F, $G(b^V)$ is defined, $G(b^V) \in \{0, 1\}^p$, and $F(b) = H(b^U, G(b^V))$. G and H are called blocks of the decomposition.

If $U \cap V = \emptyset$, a decomposition is called disjoint. In case there is a variables set $C(U \cap V = C)$, a decomposition is called nondisjoint. It means that there may be a variable set which is part of inputs for both G and H block.

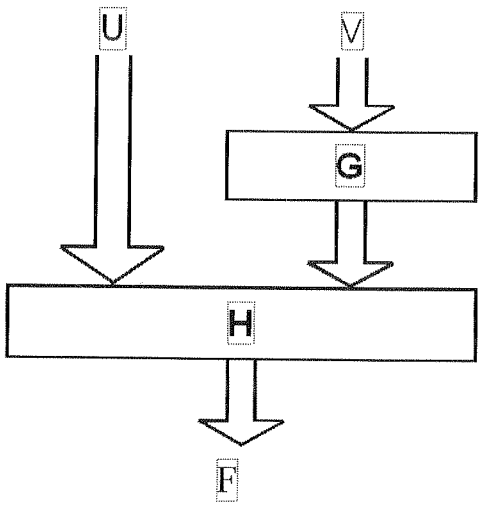


Fig. 1. Schematic representation of the serial functional decomposition

The existence of functional decomposition can be formulated using many different notions (decomposition charts, blanket algebra theorems [2], binary decision diagrams [4]). The presented method can be clearly described using decomposition charts and graphs.

Definition 2:
Decomposition chart is a matrix $2^{|U|} \times 2^{|V|}$. The chart has as many columns as minterms induced by bound-set variables and as many rows as minterms induced induced by free-set variables. Each cell in the matrix represents a certain value of f for the corresponding minterm of input variables. Each column (row) is labaled by a minterm of bound-set (free-set) variables.

Example 1:
Let F be a boolean function described by a truth table (Table 1). Let $U = \{c, d, e\}$ and $V = \{a, b\}$. Having U and V sets, a decomposition chart can be constructed (Table 2).

The next step to verify the existence of serial decomposition is to construct an incompatibility graph. Each graph vertex corresponds to the one column from the decomposition chart. Two vertices are connected by an edge only if the vertices (columns)

Table 1

Truth table of function F

a	b	c	d	e	y
0	0	0	0	0	1
0	1	0	0	0	0
0	1	0	0	1	1
1	0	0	0	1	0
0	0	0	1	0	1
1	0	0	1	0	0
1	1	0	1	1	1
0	1	1	0	0	0
1	0	1	0	1	1
1	0	1	1	0	1
0	1	1	1	1	1

Table 2

Decomposition chart based on Table 1

ab\cde	000	001	010	011	100	101	110	111
00	1	–	1	–	–	–	–	–
01	0	1	–	–	0	–	1	–
10	–	0	0	–	–	1	–	1
11	–	–	–	1	–	–	–	–

are incompatible. Columns are incompatible when they have different values at least in a one row. For example, '100' column is incompatible with '110' column since they have different values in the second row.

The incompatibility graph is coloured with graph colouring procedure. Each vertex is assigned such a colour, that no two vertices connected by an edge have the same colour assigned. The minimal number of colours is called a graph chromatic number. The next step is to encode the colours. Each colour is assigned the unique binary code. The number of bits used to encode the colours is equal to the number of outputs from the G block. The introduced notions allow to formulate the following theorem:

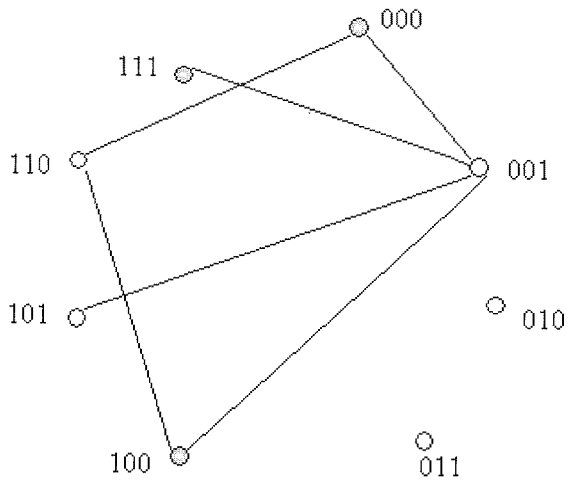


Fig. 2. Incompatibility graph for the decomposition chart (Table 2)

Theorem 1:
Let F be a boolean function and U and V be sub-sets of its input variables. Let Γ be the incompatibility graph based on a given variables partitions. Let k be the chromatic number of Γ graph and v the number of variables in V set.
If $\lceil log_2 k \rceil < v$ then function F has a serial decomposition for given U and V sets.

Merging incompatibility graphs

Let us assume, that there are two sub-functions and each sub-function has a serial decomposition for a given partition of input variables. When two incompatibility graphs have the same vertices (V set is the same for both decompositions) it is possible to merge these graphs into one. This is done by marking all the edges from both graphs into one, new graph. This new graph is the same as it would be constructed for the function consisting of these two sub-functions (variables partitions have to be the same). Example 2 shows the procedure of merging two graphs. The concept of incompatibility graphs can be also described using different formalism – blanket algebra[13]. Computationally it is equal to the description based on decomposition charts.

Example 2:
There are two incompatibility graphs (Fig. 3). To construct the merged graph, all the edges have been marked on the new graph (Fig. 4).
For the precise description of the main algorithm, it is useful to formulate two more theorems based on the concept of graphs merging and Theorem 1.

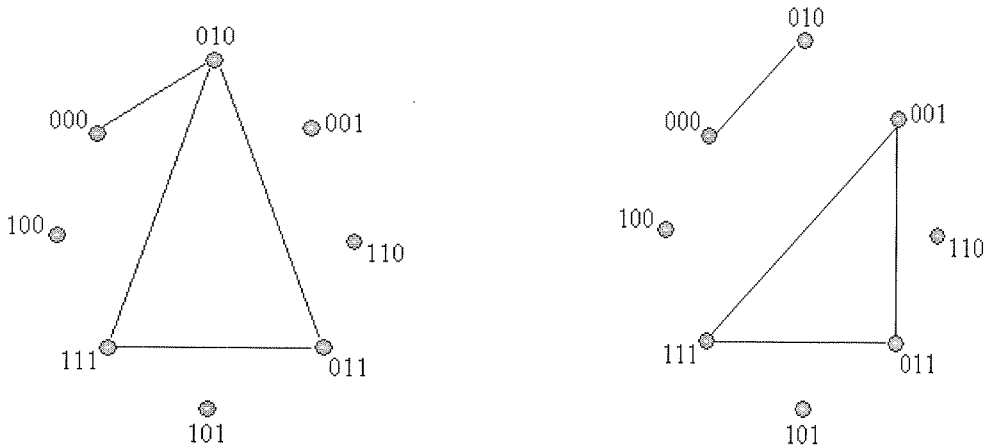


Fig. 3. Two incompatibility graphs

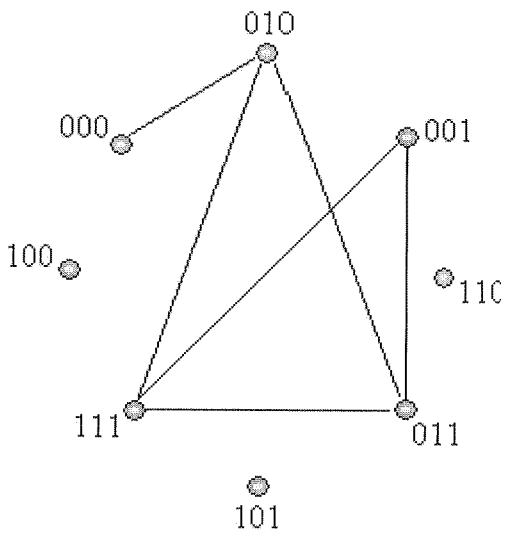


Fig. 4. Incompatibility graph built by merging graphs from Figure 3

Theorem 2:
It is given a set of graphs $\Gamma_1 ... \Gamma_i$ and Γ_S graph built by merging $\Gamma_1 ... \Gamma_i$ graphs. If Γ_S graph is correctly coloured (each pair of vertices connected by an edge has a different colour) then such colouring is also correct for each graph $\Gamma_1 ... \Gamma_i$.

Theorem 3:
It is given an incompatibility graph Γ_S built by merging graphs $\Gamma_1 ... \Gamma_i$. Its chromatic number equals c. Graphs $\Gamma_1 ... \Gamma_i$ refer to serial decompositions of functions $F_1 ... F_i$, each decomposition with the same V set. Let v be a number of variables in V set.

If $\lceil log_2 c \rceil < v$ then there is such a function G that for each function F_i exists a serial decomposition $F_i = H_i\ (U_i, G(V))$.

3. SERIAL DECOMPOSITION with SHARED G FUNCTION

There are many methods for decomposition of multi-output Boolean functions. These methods are mostly developed for cube or BDD-based representation of logic circuit. However, the common characteristic of these methods is the requirement, that decomposed multi-output Boolean function is represented by single truth table in case of cube representation or single BDD. In case of BDD, this causes the necessity of development of special techniques that allow for efficient representation of many functions with single BDD. Since Boolean functions can be not fully specified (can have *don't cares*) and each function can depend on different input variables, it is difficult to find efficient solution to this problem [19].

In case of decomposition methods based on cube representation it is easier, since to represent multi-output Boolean function espresso format can be used. However cube representation can be very inefficient when many functions which depend on different input variables have to be decomposed. Here a method for decomposition of Boolean function sets based on cube representation is presented.

The logic multi-output circuit of n input variables X can be represented as a set of Boolean functions $f_1(X_1),\ldots, f_m(X_m)$, where X_i is sub-set of X. In general case each single function may depend on different subset of input variables and can be described by different set of cubes. Application of serial functional decomposition based on blanket calculus is impossible in this case. We propose the modified serial functional decomposition algorithm that allows decomposing function sets described by separate truth tables, even if each single function depends on different subset of input variables and truth tables consist of different sets of cubes.

When decomposing a set of Boolean functions $f_1,\ldots, f_m$ it is necessary to find such sub-function G that satisfies decomposition condition for each function:

$$f_1 = h_1(U_1, G(V)),$$
$$f_2 = h_2(U_2, G(V)),$$
$$\ldots$$
$$f_m = h_m(U_m, G(V)).$$

The presented algorithm can be applied to the set of single-output functions. First such subset of single-output functions has to be selected that has common subset of input variables. They are potential candidates to construct a shared G function. Decomposition process consists of a few steps. First an appropriate input support V has to be selected for block G (input variable partitioning). In case when functions are specified by different truth tables, such a set V has to be selected that it is subset of

input variables of each decomposed function (V must consist of variables common to all decomposed function). Once the variables partition (V and U sets) is determined the incompatibility graphs for each sub-function taking part in the decomposition process are constructed. Next, by merging the incompatibility graphs, such a shared G block (function) is computed, that satisfies Theorem 3.

Example 3:
Let us consider two sub-functions F_1 and F_2 described by Table 3 and Table 4. Let us assume the functions have to be decomposed into smaller functions with 3 inputs and 1 output.

Table 3

Truth table of function F_1

a	b	c	e	g	y
0	0	0	0	0	1
0	1	0	0	0	0
0	1	0	0	1	1
1	0	0	0	1	0
0	0	0	1	0	1
0	0	0	1	0	1
1	0	0	1	0	0
1	1	0	1	1	1
0	1	1	0	0	0
1	0	1	0	1	1
1	0	1	1	0	1
0	1	1	1	1	1

Table 4

Truth table of function F_2

a	b	c	e	g	y
0	0	0	0	0	1
0	0	0	1	0	0
0	0	0	1	1	1
0	1	0	0	1	0
0	0	1	0	0	1
0	1	1	0	0	0
0	1	1	1	1	1
1	0	0	1	0	0
1	1	0	0	1	1
1	1	1	0	0	1
1	1	0	0	0	1
1	0	0	1	1	0
1	0	1	1	1	1

For the functions described by Tables 3 and 4, the following variables partitions have been chosen:

$$U = \{a, b\}, V = \{c, e, g\} \text{ for the first function (Table3)}$$

$$U = \{d, f\}, V = \{c, e, g\} \text{ for the second function (Table4)}$$

Next step is to construct decomposition charts and incompatibility graphs for both sub-functions.

Table 5

Decomposition chart for the function from Tab. 3

	ceg							
ab	000	001	010	011	100	101	110	111
00	1	–	1	–	–	–	–	–
01	0	1	–	–	0	–	1	–
10	–	0	0	–	–	1	–	1
11	–	–	–	1	–	–	–	–

Table 6

Decomposition chart for the function from Tab. 4

	ceg							
df	000	001	010	011	100	101	110	111
00	1	–	1	–	–	–	–	–
01	0	1	–	–	0	0	1	–
10	–	0	0	–	1	1	–	1
11	–	–	–	1	–	–	–	–

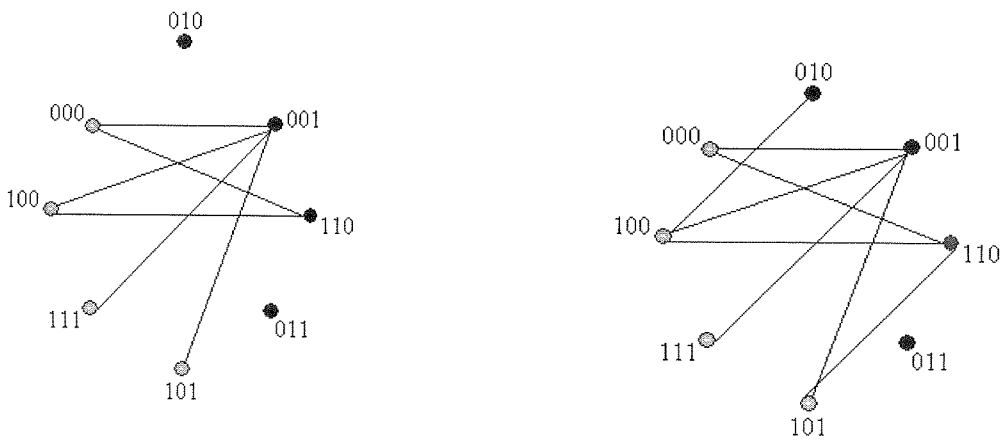


Fig. 5. Incompatibility graphs for the decomposition charts (Tab. 5 and Tab. 6)

It can be noticed that both graphs have the chromatic number equal to 2 so the colours could be encoded on one bit. These two functions could be decomposed separately and the result would be four logic cells, each cell with 3 inputs and 1 output. However, the functions have some common information and it could be utilized to produce better results. Variables c, e and g are common so the construction of one G block is possible.

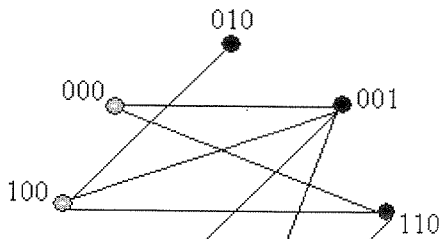


Fig. 6. Incompatibility graph constructed by merging graphs from

The merged graph (Fig. 6) has a chromatic number equal to 2 so only one bit is needed to encode the graph. Therefore, these two functions can be decomposed into three logic cells. Figure 4 shows the decomposition scheme.

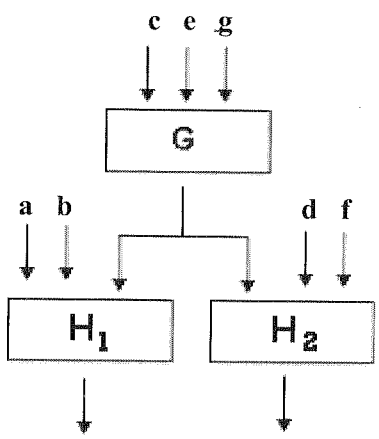


Fig. 7. Decomposition with shared G function

The presented method can be used with the functions which are described by separate truth tables with different sets of input variables (only a subset of variables has to be common to construct a shared G function). In the systems where the function is described by one structure (truth table or decision diagram) such a flexible approach is not provided.

4. FUNCTIONAL DECOMPOSITION SYSTEM

Usually the process of decomposition is recursive. Once the G and H block are created, the H block is being decomposed again. The process stops when H block has satisfactory size. Therefore, to build a complete decomposition the algorithm described in the Section 3 has to be used recursively. However, to construct an efficient decomposition system, more problems have to be taken into consideration. Before the designed algorithm is applied, one has to decide which sub-functions are taken to create a common G block and which variables should be chosen.

Firstly a user has to specify the number of input and output of G block. The system has been tested on typical values for FPGA (like 4/1 or 5/2 blocks). Then the multi-output function is divided into single-output functions. After this the main iteration is executed.

The iteration is repeated until all functions have satisfactory size (specified by the user at the beginning).

One iteration of the proposed system consists of the following procedures:

1. selection of core-function (one of the single-output functions). In the tests, the function with the fewest number of inputs was chosen.
2. generation of 3 to 5 different variables partitions (U and V sets). If the number of inputs is less than 12, the exhaustive search is used. Otherwise, the evolutionary algorithm is applied [16].
3. for each generated variable partition VP_i do
   a. select functions which potentially can be used to create a shared G function with previously chosen core-function. Such a candidate function among its input variables has to have all the variables which are in V set of VP_i.
   b. According to Theorem 3 check if selected functions can be used to construct a shared G function
   c. calculate the usefulness of variable partition VP_i with equation: usefulness = A/B, where:
      A – number of functions used in constructing a shared G function
      B – number of outputs of the shared G function
4. choose variable partition with the highest usefulness and construct a shared G function

The presented system was implemented in a prototype application and some experiments were performed. The functions from the standard benchmark set were used in test (Table 7). Also the binary to binary-coded decimal transcoder (bin2bcd) was used in some experiments. The application (called DSG – decomposition with shared G block) was compared with leading academic systems SIS, Sawada [17], Irma2FPGA [18], DEMAIN [8]. Table 8 and 9 show the results of the experiments.

Table 7

Functions used in tests

Function	Inputs	Outputs	Rows
rd73	7	3	141
Z4ml	7	4	128
e rd84	8	4	256
Ex5	8	63	390
9sym	9	1	87
Sao2	10	4	137
Alu2	10	6	391
Ex1010	10	10	1823
Plan	13	25	125
Misex3	14	14	2549
Misex3c	14	14	1566
kirk	16	10	304
pdc	16	40	3083
Duke2	22	29	405
Misex2	25	18	100
Vg2	25	8	304
E64	65	65	327
B2bcd_10	10	16	1024
B2bcd_11	11	16	2048
B2bcd_12	12	16	4096
B2bcd_13	13	16	8192

Table 8

Comparison of decomposition systems. Decomposition for logic cells with 5 inputs/1output

Function	SAWDADA	IRMA2FPGAS	SIS 1.2	DSG
9sym	7	7	7	6
kirk	–*	37	56	43
Rd84	7	9	10	10
Rd73	6	7	6	6
Misex2	36	36	38	41
Misex3	213	222	517	194
Sao2	21	25	35	24
Vg2	21	25	40	41
Duke2	152	162	159	256
Alu2	48	39	84	60
Z4ml	5	5	5	5
Pdc	–	174	155	92
Ex1010	–	285	686	221
Ex5	–	139	103	109
Σ		**1172**	**1901**	**1108**

*– results not published

Table 7

Table 8 shows that DSG obtained better results than three other academic systems. The results are 5,5% better than IRMA2FPGA's and 42% better than SIS results. There are no summary results for SAWADA as some results (especially for bigger functions) were not published. DSG system was designed especially for multi-outputs functions. For such examples (ex5, pdc, ex1010, misex3, e64) the results are considerably better than these obtained by IRMA2FPGA and SIS (25% and 58% better).

Table 9

Comparison with DEMAIN. Decomposition for architecture XC2000

Function	DEMAIN	DSG
5xp1	18	19
9sym	11	10
Clip	27	27
Misex1	16	17
Rd73	10	10
Plan	360	357
Sao2	37	41
Z4	7	7
E64	86	86
Σ	**572**	**574**

Table 10

Comparison with dekBDD. Decomposition for architecture XC3000

Function	dekBDD	DSG
kirk	67	42
rd84	7	10
rd73	5	6
Misex2	28	41
Misex3	288	194
Misex3c	252	110
Sao2	22	24
Vg2	55	41
duke2	208	256
alu2	52	60
pdc	99	92
Ex1010	588	221
Ex5p	101	109
E64	86	80
B2bcd_10	58	51
B2bcd_12	148	99
Σ	**2064**	**1436**

Table 8

t

Table 9 shows the comparison with DEMAIN. The results of DSG are slightly worse (0,3%) but it has to be stressed that this comparison consists of only small

functions. It is because DEMAIN uses exhaustive search of variables partition and cannot deal with big functions (over 20 inputs). E64 function may seem a big function but in fact it is a set of small single-output functions which can be easily obtained by a parallel decomposition. Therefore, it is acceptable for DEMAIN.

Table 10 shows the comparison with dekBDD [3] The results of DSG are clearly better. The difference is especially noticeable for multi-output functions specified by many rows (over 1000). For such functions DSG results are nearly 100% better than dekBDD's.

Table 11 shows the comparison with the commercial system Quartus II developed by Altera [20]. For Quartus II each function was specified in VHDL language using 'case' for each vector of truth table. DSG's results are more than two times better.

Table 11

Comparison with Quartus II. Decomposition for architecture Flex10k

Function	Quartus	DSG
5xp1	29	19
9sym	89	10
clip	263	27
Sao2	66	41
Misex2	35	47
Pdc	226	130
B2bcd_11	263	154
B2bcd_13	672	224
Σ	**1643**	**652**

5. CONCLUSIONS

Logic circuits usually have many outputs. There are many methods for decomposition of multi-output Boolean functions. These methods are mostly developed for cube or BDD-based representation of logic circuit. However, the common characteristic of these methods is requirement, that decomposed multi-output Boolean function is represented by single truth table in case of cube representation or single BDD.

The proposed decomposition method allows decomposing multi-output Boolean functions presented as set of single-output functions described by separate truth tables. Decomposed function may depend on different input variables and can be described by different set of cubes. It uses the concept of creating the incompatibility graphs for each single output function and then merging them into one graph. This approach allows to utilize common information between subfunctions through the whole decomposition process. The additional advantage of the method is that computing the

incompatibility graphs for single output functions and merging them is less memory and time consuming than the computation for the whole function. More details were presented in the papers [5], [6].

The presented system (DSG) combined with evolutionary algorithm (for searching good variables partition) gives promising results. Its advantage can be clearly seen especially for big, multi-output functions.

6. ACKNOWLEDGMENT

This paper was supported by Ministry of Science and Higer Education financial grant for years 2006-2009 (Grant No. SINGAPUR/31/2006).

7. REFERENCES

1. T. Luba, H. Selvaraj, M. Nowicka, A. Kraśniewski: Balanced multilevel decomposition and its applications in FPGA-based synthesis, In: Logic and Architecture Synthesis (G.Saucier, A.Mignotte ed.), Chapman&Hall, 1995
2. J. A. Brzozowski and T. Łuba: Decomposition of Boolean Functions Specified by Cubes, Journal of Multiple-Valued Logic and Soft Computing, Vol. 9, pp. 377-417, Old City Publishing, Inc., Philadelphia, 2003
3. A. Opara: Dekompozycyjne metody syntezy układów kombinacyjnych wykorzystujące binarne diagramy decyzyjne. Praca doktorska, Politechnika Śląska, 2009
4. C. W. Chang and M. Marek-Sadowska: Finding maximal symmetric groups variables in incompletely specified Boolean functions. International Workshop on Logic Synthesis, Lake Tahoe, June 1999
5. P. Morawiecki, M. Rawski: Input Variable Partition Method in Functional Decomposition based on Shannon Expansion, Konferencja Reprogramowalne Układy Cyfrowe, RUC Szczecin 2007
6. M. Rawski, P. Morawiecki, H. Selvaraj: Decomposition of Combinational Circuits Described by Large Truth Tables, Proceedings of Eighteenth International Conference on Systems Engineering, pp. 401-406, Coventry, United Kingdom, September 5-7 2006
7. M. Nowicka, M. Rawski, T.Łuba: FPGA-based decomposition of boolean functions. Algorithms and implementation. Proceeding of the 6th International Conference on Advanced Computer Systems, 1999
8. M. Nowicka: Zrównoważona metoda odwzorowania technologicznego dla układów FPGA. Praca doktorska, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska, 1999
9. M. Rawski, L. Jóźwiak, M. Nowicka, T. Luba: Non-Disjoint Decomposition of Boolean Functions and Its Application in FPGA-oriented Technology Mapping, Proc. of the EUROMICRO'97 Conference, Budapest, Hungary, Sept. 1-4, 1997, pp.24-30, IEEE Computer Society Press, 1997
10. T. Luba, H. Selvaraj: A General Approach to Boolean Function Decomposition and its Applications in FPGA-based Synthesis. VLSI Design. Special Issue on Decompositions in VLSI Design, vol. 3, Nos. 3-4, 289-300, 1995
11. S. C. Chang, M. Marek-Sadowska, and T. T. Hwang: Technology Mapping for TL FPGAs Based on Decomposition of Binary Decision Diagrams, IEEE Trans. on CAD, vol.15 No. 10, pp. 1226-1236, October, 1996
12. C. Files, M. Perkowski: Multi-Valued Functional Decomposition as a Machine Learning Method, ISMVL'98, pp. 173-178, Fukuoka, Japan, 1998

13. M. Rawski: Decomposition of Boolean Function Sets, Kwartalnik Elektronika I Telekomunikacja II'2007
14. A. Opara, D. Kania: Synteza wielowyjściowych układów logicznych prowadząca do wykorzystania wspólnych bloków logicznych. Pomiary Automatyka Kontrola, 7, 2007
15. A. Opara, D. Kania: Wykorzystanie pseudo-mtbdd w dekompozycji zespołu funkcji. Pomiary Automatyka Kontrola, 8, 2008
16. P. Morawiecki: Zastosowanie algorytmów ewolucyjnych przy doborze zmiennych w dekompozycji szeregowej. Praca inżynierska. Wydział Elektroniki i Technik Informacyjnych. Politechnika Warszawska, 2003
17. H. Sawada, T. Suyama, A. Nagoya: Logic synthesis for look-up table based fpgas using functional decomposition and boolean resubstitution. IEICE Transactions on Information and Systems, 1997
18. A. Chojnacki: Effective and Efficient Circuit Sythesis for LUT FPGAs Based on Functional Decomposition and Information Relationship Measures. Praca doktorska, Technische Universiteit Eindhoven, 2004
19. T. Sasao, M. Matsuura, Y. Iguchi, and S. Nagayama: Compact BDD Representations for Multiple-Output Functions and Their Application. ISMVL pp. 207-212, 2001
20. www.altera.com
21. G. De Micheli: Synthesis and Optimization of Digital Circuits. McGraw-Hill, 1994
22. T. Sasao, and M. Fujita: Representations of Discrete Functions. Kluwer Academic Publishers, 1996

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 471–483

Performance of Distributed Multi-Stage Virtual MIMO Systems with Random Position of Ancillary Nodes

ALBERTO ZANELLA*, CHIARA BURATTI+

*WILAB, IEIIT – National Research Council (CNR)
Viale Risorgimento, 2 40136 Bologna, Italy
+WiLAB, DEIS – University of Bologna,
Viale Risorgimento, 2 40136 Bologna, Italy
email: alberto.zanella@cnr.it, c.buratti@unibo.it

Received 2009.05.28
Authorized 2009.07.20

The performance of two-hop decode-and-forward virtual multiple-input-multiple-output systems with random location of nodes, is investigated. We suppose that source, relay and destination nodes create clusters with ancillary nodes which are spatially distributed according to a Poisson point process. Random fluctuations are accounted for in the channel model which considers both slow and fast fading. By assuming that the application requires a minimum amount of rate, we evaluate the system performance in terms of cumulative density function of the required signal-to-noise ratio. The use of orthogonal space-time-block codes is also considered and the role played by connectivity parameters is investigated.

Keywords: multiple-input multiple-output (MIMO), distributed MIMO, Poisson point processes, outage capacity

1. INTRODUCTION

Wireless systems based on the concept of virtual (or distributed) multiple-input multiple-output (V-MIMO) have attracted the interest of research community in these years, owing to their capability to exploit the inherent diversity offered by the presence of spatially separated devices [1, 2, 3]. The main characteristic of a V-MIMO system is the possibility to create clusters of cooperating nodes (composed by a main node and ancillary nodes), which are usually denoted as virtual antenna arrays (VAAs) [1, 3]. A short range communication system can be used for data transmission within the

cluster (intra-VAA communication), whereas a long range communication system is required to transmit data from cluster to cluster (inter-VAAs communication) [1, 2, 3].

In [1] and [2] the ergodic capacity of a V-MIMO system for single and two-hop ad-hoc network scenarios, respectively, is derived. [3] considers a multi-hop V-MIMO and derives an optimal resource allocation strategy, in terms of fractional bandwidth and power allocation to each relay. In [4] the diversity gain achieved by a V-MIMO is investigated in realistic indoor propagation environment. In more recent works the concept of V-MIMO has been applied to wireless sensor networks, where the co-operating devices are sensors, equipped with a single antenna element [5]. In [6] an approximate expression for the end-to-end outage probability for a multi-hop distributed MIMO system, where orthogonal space-time block codes (OSTBC) are used for transmission, is provided. Also an efficient near-optimal power allocation approach with low complexity is proposed. This near-optimum solution leads to the result that the same power is assigned to each node of a VAA. In [7] cooperative diversity in cellular networks is studied. It is shown that the use of VAA improves performance not only in terms of maximum mutual information but also of outage probability. Also a distributed method to select multiple relay mobile terminals (MTs) in opportunistic way is proposed. A novel cooperative diversity system model is proposed in [8], and compared to conventional two-hop cooperative systems, under the condition that all terminals are equipped with only one antenna. Both source and destination nodes request the presence of nodes around them to create VAAs based on some selection criterion.

In the papers [1]-[8], the topology of the network is assumed to be fixed and the performance is usually related to a particular configuration of nodes. There exist few works related to connectivity aspects in MIMO systems in the context of ad-hoc networks. In [9] the performance of some spatial diversity techniques including maximal ratio combing MIMO are investigated. In [10] a probabilistic analysis of achievable capacity on individual links in the case of random topology is derived. The paper assumes that a pair of nodes are connected only if their bi-directional capacity is larger than a given threshold. Although Poisson fields of nodes are studied in several works (i.e., [9, 11]), very few papers consider the effect of randomness of nodes on the mechanism of creation of the VAA. One of the first attempts to quantify the effect of connectivity on the performance of V-MIMO systems can be found in [12], where the ancillary nodes are supposed to be distributed according to a Poisson point process (PPP) [13].

In this paper we investigate a decode-and-forward two-hop V-MIMO scenario in which nodes are distributed according to a PPP. We consider ergodic channels and the role of connectivity parameters on the distribution of the signal-to-noise ratio (SNR) needed to obtain a rate. The distribution of SNR is quantified in terms of the cumulative density function (CDF). These results are an extension of those in [12], which were related to the analysis of the maximum capacity for a fixed value of SNR. The use of orthogonal space-time-block codes (OSTBC) [14] is also considered and

their performance is compared with the theoretical capacity (in the Shannon sense) achievable by the V-MIMO system.

2. SYSTEM DESCRIPTION

Throughout the article vectors and matrices are indicated by bold, I is the identity matrix, $|A|$ denotes the determinant of the matrix A, $\mathrm{Tr}\{A\}$ is the trace of A, and $\|A\|$ is the Frobenius norm of A. $\{a_{i,j}\}_{i,j=1,\ldots,M}$ is an $M \times M$ matrix with elements $a_{i,j} = \{A\}_{i,j}$, $\dagger$ is the operator of conjugation and transposition. Also, $\mathsf{E}\{\cdot\}$ denotes expectation, and $\mathsf{P}\{E\}$ denotes the probability of the event E.

In our considered scenario, we suppose that the ancillary nodes are spatially distributed in A_S, A_R and A_D according to a PPP [13]. With such model the probability of having one node in the infinitesimal area δA is $\eta\delta A$, where η denotes the nodes' density [13]. As a general case, nodes densities in the three areas may be different: we denote as η_S, η_R and η_S, the densities of the ancillary nodes distributed around the source, the relay and the destination, respectively.

2.1. VAA CONCEPT

We consider a two-hop V-MIMO system (see Figure 1), where a source node has to transmit data to a destination node via a relay node. In the following, source, relay and destination nodes will be denoted as *main* nodes. A number of *ancillary* nodes are located in three areas, A_S, A_R and A_D, around the main nodes. In particular, ancillary nodes are distributed over the areas according to a Poisson point process (PPP). We assume that nodes work in a half-duplex mode and that a decode and forward strategy is implemented at the relay [3]. Random fluctuations of the wireless channel as well as a distance-dependent deterministic path-loss are accounted for in our radio channel model. In our scenario, the main nodes can cooperate only with their ancillary nodes which guarantee a feasible quality of the link. Owing to the randomness nature of the channel, the number of transmit and receive antennas is a random variable and a certain outage probability there exists.

The main nodes transmit a query to the ancillary nodes, by using the short-range radio interface. Owing to propagation conditions, only a subset of the ancillary nodes can really cooperate with the main nodes. The number of nodes which actually communicate with source, relay and destination is denoted by n_S, n_R and n_D, respectively, and are called *cooperating* nodes. We also assume that the distances source-relay and relay-destination are much larger then the distance between a main node and its cooperating nodes. So that the short-range radio interface can be used only to transmit/receive data to/from the main node and its cooperating nodes (intra-VAA communication). The communication in the two-hop case is performed according to the following steps: (i) The source transmits data to the n_S cooperating nodes; (ii) the $n_\mathsf{S} + 1$ nodes of the

s-VAA transmit data toward the relay through the V-MIMO channel, using the high rate interface (inter-VAA communication); (iii) the $n_R + 1$ nodes of the r-VAA cooperate to decode the received data and forward it toward the destination; (iv) the $n_D + 1$ nodes of the d-VAA receive data from the r-VAA and cooperate to decode it. The maximum number of cooperating nodes that the main nodes can actually handle is obviously limited by their hardware equipment, we denote this number by M_S, M_R and M_D for source, relay and destination, respectively. Note that our analysis can be applied regardless the criterion for the selection of the cooperating nodes which has been chosen.

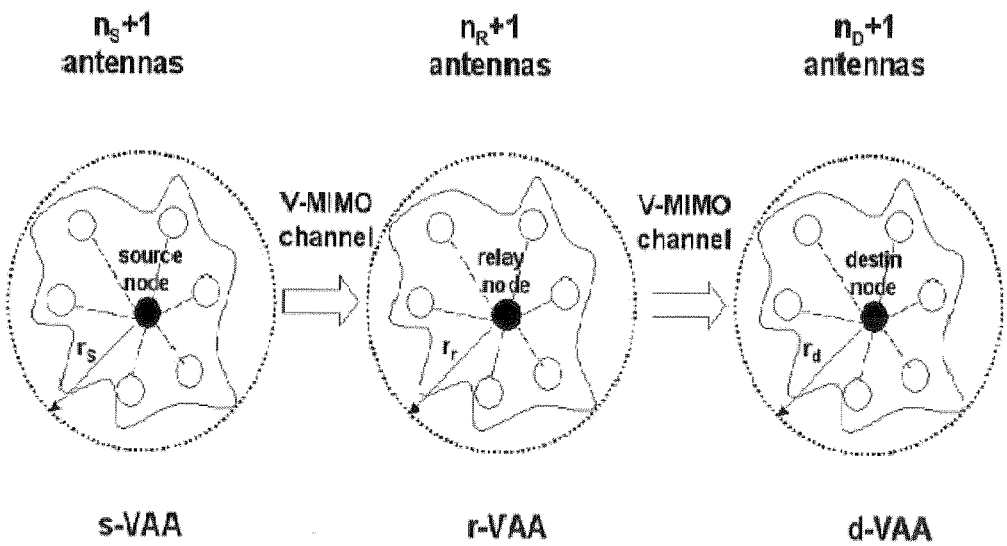


Fig. 1. The Virtual MIMO communication system

2.2. CHANNEL MODEL

The channel model we consider in the paper takes power loss, due to propagation effects including a distance-dependent path loss, the slow and fast channel fluctuations, into account. We assume that the ratio between the transmited power, P_T, and the received power, P_R, is given by

$$\frac{P_T}{P_R} = \frac{k \cdot d^{\beta} \cdot s}{f}, \tag{1}$$

where k is the propagation coefficient, d is the distance from the transmitter and the receiver, β is the attenuation coefficient which commonly ranges from 2 to 5, finally, s and f are the long-term (shadowing) and the short-term (fast) fading components,

respectively. Shadowing is assumed to be log-normally distributed with standard deviation σ, and Rayleigh fading is considered, therefore f is exponentially distributed with unitary mean.

We define $L = k \cdot d^{\beta} \cdot s$ as the averaged (with respect to fast fading) loss (in linear scale). By introducing the logarithmic scale, we obtain

$$L[dB] = k_0 + k_1 \ln d + S[dB], \tag{2}$$

where $k_0 = 10\ log_{10} k$, $k_1 = \beta \dfrac{10}{\ln 10}$, and S [dB] is a Gaussian random variable, with zero mean and variance σ^2.

Note that, for each air interface (intra-VAA and inter-VAA) we could have different power transmission (P_T) and propagation parameters (k_0, k_1, σ).

2.3. CONNECTIVITY ASPECTS

We assume that ancillary nodes are randomly distributed around main nodes. Owing to the random position of ancillary nodes and channel fluctuation effects, the number of cooperating nodes at the main nodes is not deterministic. This is true regardless the connectivity model we are considering [12]. As a result, the number of cooperating nodes is a random variable (r.v.) whose statistical properties depend on the connectivity models we are using and on the spatial distribution of nodes. In particular, when the position of ancillary nodes is distributed according to a PPP, the number of communicating nodes is always a Poisson r.v. regardless the communication system used [13]. Therefore, the probability distribution of n is

$$\mathrm{P}\{n = n_1\} = P(n_1, N) = \frac{N^{n_1}}{n_1!} e^{-N}, \tag{3}$$

where $N = \mathrm{E}\{n\}$ depends on the connectivity model chosen.

In this paper, we assume that two nodes are connected, when the loss $L < L_{\mathrm{th}}$, where L_{th} represents the maximum loss tolerable by the communication system, and d is the distance between the two nodes. The threshold L_{th} depends on transmit power and receiver sensitivity.

When the areas A_S, A_R and A_D are circular (with center in the main nodes) with radius r_S, r_R and r_D, the mean value of the number of nodes in A_S for which $L < L_{\mathrm{th}}$, is denoted by N_S, and can be written as [15]

$$N_S = \eta_S \pi \left[e^{2L_{\mathrm{th}}/k_1 - 2k_0/k_1 + 2\sigma^2/k_1^2} + \Psi\left(\frac{L_{\mathrm{th}} - k_0}{\sigma}, \frac{k_1}{\sigma}, r_S \right) \right], \tag{4}$$

where

$$\Psi(a_1, b_1, r) = r^2 \Phi(a_1 - b_1 \ln r) - e^{\frac{2a_1}{b_1} + \frac{2}{b_1^2}} \Phi\left(a_1 - b_1 \ln r + \frac{2}{b_1} \right) \tag{5}$$

and $\Phi(x) = 1/(\sqrt{2\pi}) \int_{-\infty}^{x} e^{-u^2/2} du$; $a_1 = (L_{th} - k_0)/\sigma$, and $b_1 = k_1/\sigma$.

N_R and N_D (i.e. the mean number of nodes in A_R and A_D for which $L < L_{th}$) can be easily obtained from (4) by using the couple of values (η_R, r_R) or (η_D, r_D) instead of (η_S, r_S). Finally, the parameters k_0, k_1, and L_{th} in (4) and (5) refer to intra-VAA transmission.

3. CAPACITY OF A V-MIMO SYSTEM

We assume a VAA has perfect knowledge of the channel state at the receiver, whereas the transmitter knows only the average loss (path-loss and shadowing) [1, 2, 3]. Under such hypothesis, the received signal at the $\ell^{\underline{th}}$ hop can be written as

$$y_\ell = \sqrt{P_\ell H_\ell b_\ell + n_\ell} \tag{6}$$

where y_ℓ is a ($n_R = 1$) (for $\ell = 1$) or ($n_D + 1$) -dimensional (for $\ell = 2$) vector. P_ℓ, H_ℓ, b_ℓ, n_ℓ are the averaged (over fast fading) power received by a given node in the $\ell^{\underline{th}}$ link, the fast fading channel matrix, the transmitted symbol vector and the thermal noise vector, respectively. We assume $\mathsf{E}\{b_\ell b_\ell^\dagger\} = I$, and $\mathsf{E}\{n_\ell \cdot n_\ell^\dagger\} = \sigma_N^2 I$, where σ_N^2 is the thermal noise power per antenna element. We consider a flat uncorrelated Rayleigh environment so that the elements of H_ℓ, $h_{i,j}^{(\ell)}$, can be modelled by a collection of i.i.d complex-valued Gaussian r.v.'s having $\mathsf{E}\{h_{i,j}^{(\ell)}\} = 0$ and unitary mean $\mathsf{E}\{|h_{i,j}^{(\ell)}|^2\} = 1$. Since the distances source-relay and relay-destination are much larger then the distance between a main node and its cooperating nodes, the averaged power (P_ℓ) received by a node does not depend on the specific transmit node.

We consider ergodic channels, and the performance is quantified by evaluating two metrics: i) the ergodic capacity (also known as Shannon capacity), that is the maximum rate achievable on a given channel, independent of the signal processing at either end, and ii) the maximum (ergodic) rate obtainable by using OSTBC [16].

3.1. ERGODIC CAPACITY

The ergodic capacity for the $\ell^{\underline{th}}$ link with n_1 and n_2 ancillary nodes at the transmit and receive side, respectively, can be expressed by using the following well known expression (see i.e. [17])

$$\bar{C}_{n_1,n_2}(\rho) = \mathsf{E}\left\{\sum_{i=1}^{n_{min}} log_2 |I + \rho H_\ell H_\ell^\dagger|\right\} = n_{min}\mathsf{E}\left\{log_2\left(1 + \rho\lambda^{(\ell)}\right)\right\} \tag{7}$$

where $n_{min} = 1 + \min\{n_1, n_2\}$, $\lambda^{(\ell)}$ is the unordered eigenvalue of $H_\ell H_\ell^\dagger$ and ρ is the signal-to-noise ratio. Several closed-form expressions are available for (7) in the case of Rayleigh fading (see e.g., [18]) and are not shown here for brevity. The ergodic

capacity in the two-hop case, $\overline{C}^{(2)}$, is the minimum between the mean capacity of the first link (from the source to the relay) and that of the second link (from the relay to the destination) [2]. Therefore, by assuming n_S, n_R and n_D cooperating nodes, the source-destination ergodic capacity can be written as

$$\bar{C}^{(2)}_{n_S,n_R,n_D} = \frac{1}{2} min\left\{\bar{C}^{(1)}_{n_S,n_R}(\rho_1), \bar{C}^{(1)}_{n_R,n_D}(\rho_2)\right\}, \tag{8}$$

where the term 1/2 reflects the fact that half of the resources (in the time or frequency axes) are spent for the transmission from source to relay and half for the transmission from relay to destination. $\overline{C}^{(1)}_{n_1,n_2}(\rho_\ell)$ is the ergodic capacity of a MIMO channel with $n_1 + 1$ transmit (the main node plus n_1 cooperating nodes) and $n_2 + 1$ receive antennas and ρ_ℓ is the signal-to-noise ratio, defined as $\rho_\ell = P_\ell/\sigma^2_N$.

3.2. ARCHIEVABLE RATE OF OSTBC

In the case of OSTBC, the corresponding achievable rate is [16]

$$\bar{R}^{(1)}_{n_1,n_2}(\rho) = R_L \mathsf{E}\left\{log_2\left(1 + \frac{\rho\|H_\ell\|^2}{R_L}\right)\right\} = R_L \mathsf{E}\left\{log_2\left(1 + \frac{\rho t^{(\ell)}}{R_L}\right)\right\} \tag{9}$$

where R_L is the rate loss introduced by the OSTBC, and $t^{(\ell)} = \mathrm{Tr}\{H_\ell H_\ell^\dagger\}$. In the case of uncorrelated Rayleigh fading, $t^{(\ell)}$ is a central chi-squared r.v., whose probability density function (p.d.f.) is

$$f_{t^{(\ell)}}(x) = \begin{cases} \dfrac{1}{(n_1(n_2+1)+n_2)!} x^{n_1(n_2+1)+n_2} e^{-x} & \text{if } x \geq 0 \\ 0 & otherwise. \end{cases} \tag{10}$$

It is worth noting that (9) can be written in closed – form by using (10) together with [19, eq. (78)].

The achievable rate in the two-hop case, $\bar{R}^{(2)}$, can be obtained from (8) by replacing $\bar{C}^{(1)}_{n_1,n_2}(\cdot)$ with $\bar{R}^{(1)}_{n_1,n_2}(\cdot)$.

4. CDF OF THE SIGNAL-TO-NOISE RATIO

If we fix at R_0 the required rate for a given application, the random value of the number of nodes makes the required signal-to-noise ratio a random variable too. To quantify the amount of SNR (and therefore transmit power), we can consider the CDF of ρ_1 and ρ_2, which can be written as

$$F_{\rho_1}(\rho_0) = \sum_{s=0}^{M_S} \sum_{r=0}^{M_R} Q(s, N_S) Q(r, N_R) I\left(\rho_{1,s,r}, \rho_0\right) \tag{11}$$

and

$$F_{\rho_2}(\rho_0) = \sum_{r=0}^{M_R} \sum_{d=0}^{M_D} Q(r, N_R) Q(d, N_D) I(\rho_{2,r,d}, \rho_0) \tag{12}$$

where $\rho_{1,s,r}$ denotes the SNR of the first link with n_S and n_R ancillary nodes at the transmitter and receiver side and $\rho_{2,r,d}$ denotes the SNR of the second link with n_R and n_D ancillary nodes at the transmitter and receiver side. Finally, the indicator function $I(x, y)$ is defined as

$$I(x, y) = \begin{cases} 1 & x < y \\ 0 & otherwise. \end{cases} \tag{13}$$

Since the number of cooperating nodes is limited to M_S, M_D, and M_R, n_S, n_R and n_D do not have Poisson distribution but their distribution can be easily obtained from (3) as

$$Q(s, N_S) = \begin{cases} P(s, N_S) & for s < M_S \\ 1 - \sum_{l=0}^{M_S - 1} P(l, N_S) & for s = M_S \end{cases} \tag{14}$$

equivalent expressions can be written for $Q(r, N_R)$ and $Q(d, N_D)$.

To evaluate (11) and (12), we need to solve (7) and (9) in the variable ρ. Unfortunately neither (7) nor (9) appear to be invertible, and we have to resort to the use of bounds.

In the case of ergodic capacity, when a single-hop communication is considered, we can use the following lower bound [20]

$$\bar{C}^{(1)}_{n_1,n_2}(\rho) \geq n_{\min} log_2 \left(1 + \rho \, exp\left(\frac{1}{n_{\min}} \left(\sum_{k=0}^{n_{\min}-1} \psi(n_{\max} - k)\right)\right)\right) \tag{15}$$

where $n_{max} = 1 + \max\{n_1, n_2\}$ and $\psi(\cdot)$ is the digamma function

$$\psi(n) = \sum_{k=1}^{n-1} \frac{1}{k} - \gamma_{EM} \tag{16}$$

where $\gamma_{EM} = 0.577...$ is the Euler-Mascheroni constant. (15) can now be solved in terms of ρ by obtaining the following inequality

$$\rho_{n_1,n_2}(R_0) \leq \frac{2^{\frac{R_0}{n_{\min}}} - 1}{B(n_{\min}, n_{\max})} \tag{17}$$

where

$$B(n_{\min}, n_{\max}) = exp\left(\frac{1}{n_{\min}} \sum_{k=0}^{n_{\min}} \psi(n_{\max} - k)\right). \tag{18}$$

It is worth noting that (17) can be used as a tight approximation for the required signal-to-noise ratio.

In the case of OSTBC, by applying the Jensen's inequality to the expression (9) (valid for the single-hop case), we can write

$$\bar{R}^{(1)}_{n_1,n_2}(\rho) \le R_{\mathrm{L}}\, log_2\left(\mathsf{E}\left\{1+\frac{\rho t^{(\ell)}}{R_{\mathrm{L}}}\right\}\right) = R_{\mathrm{L}}\, log_2\left(1+\frac{\rho\, \mathsf{E}\left\{t^{(\ell)}\right\}}{R_{\mathrm{L}}}\right) = R_{\mathrm{L}}\, log_2\left(1+\frac{\rho(n_1+1)(n_2+1)}{R_{\mathrm{L}}}\right) \tag{19}$$

where we have used the following equality to obtain (19)

$$\mathsf{E}\left\{t^{(\ell)}\right\} = \mathsf{E}\left\{\sum_{i=1} \lambda_i^{(\ell)}\right\} = E\left\{\sum_{i=1}^{n_2+1} \sum_{j=1}^{n_1+1} |h_{i,j}^{(\ell)}|^2\right\} = (n_1+1)(n_2+1). \tag{20}$$

Now, a lower bound for SNR in the case of OSTB can be easily obtained as

$$\rho_{n_1,n_2}(R_0) \ge \frac{R_{\mathrm{L}}\left(2^{R_0/R_{\mathrm{L}}} - 1\right)}{(n_1+1)(n_2+1)}. \tag{21}$$

Note that, in the two-hop communication case two different SNRs at the relay and destination can be obtained: $\rho_{1,s,r}$ and $\rho_{2,s,r}$. Moreover, in the two-hop case we have $\rho_{1,s,r} = \rho_{n_1=s,n_2=r}(2R_0)$ and $\rho_{2,s,r} = \rho_{n_1=r,n_2=d}(2R_0)$, where ρ_{n_1,n_2} is given by eq. (17) in the ergodic capacity case, and by eq. (21) in the OSTBC case.

5. NUMERICAL RESULTS

In this Section the behavior of the cumulative functions $F_{\rho_1}(\rho_0)$ and $F_{\rho_2}(\rho_0)$ is shown by varying different scenario parameters and considering the cases of ergodic capacity and STBCs. Results are obtained by setting, if not otherwise specified, the following parameters: $r_{\mathrm{S}} = r_{\mathrm{R}} = R_{\mathrm{D}} = 10$[m], $\sigma = 4$ dB; $\sigma^2_{\mathrm{N}} = 8 \cdot 10^{-15}$ [W]. We consider two different channel models for intra-VAA and inter-VAAs communication. In the first case, we set $k_0 = 41$ dB, $k_1 = 13.03$ ($\beta = 3$), and $L_{\mathrm{th}} = 92$ dB; whereas we set $k_0 = 15$ dB and $k_1 = 17.37$ ($\beta = 4$) for the inter-VAAs transmissions (L_{th} is not fixed in this case, since we assume that the s-VAA and the r-VAA so that the r-VAA and the d-VAA are always connected).

Figures 2 and 3 are related to the ergodic capacity cas and have been obtained by setting $M_{\mathrm{S}} = M_{\mathrm{R}} = M_{\mathrm{D}} = 10$ and $R_0 = 3$ [bit/s/Hz]. In Fig. 2 we set $\eta_{\mathrm{S}} = \eta_{\mathrm{D}} = 10^{-3}$ [m^{-2}], meaning that the distribution of the SNRs for

the two links becomes identical
(i.e., $F_{\rho_1}(\rho_0) = F_{\rho_2}(\rho_0)$). The different curves are obtained by varying η_R. As expected, the curves shift on the left by increasing η_R, since the SNR required to reach the target value of R_0 decreases. This can be easily explained, by observing that by increasing η_R the average number of antennas at the transmitter (second hop) or at the receiver (first hop) increases and the corresponding required SNR decreases. Note that in the Figure when $\eta_R = 10^{-4}$ or $5 \cdot 10^{-4}[m^{-2}]$ $F_{\rho_1}(\rho_0)$ does not reach the value 1 since values of ρ_0 larger that 30 [dB] are required.

Almost the same behavior could be observed in Fig. 3, obtained by setting $\eta_S = 10^{-3}$ $[m^{-2}]$ and $\eta_D = 5 \cdot 10^{-5}[m^{-2}]$ and by varying η_R. In this case $F_{\rho_1}(\rho_0)$ and $F_{\rho_2}(\rho_0)$ are different and, once again, by increasing the ancillary nodes densities curves moves to the left. Moreover $F_{\rho_1}(\rho_0)$ outperforms $F_{\rho_2}(\rho_0)$ in all cases since $\eta_S > \eta_D$.

In the Figures 4 and 5 the ergodic capacity case is compared with the OSTBC case. In both Figures we set $R_0 = 2.5$ [bit/s/Hz] and $\eta_S = \eta_D = 10^{-3}$ $[m^{-2}]$ and we vary η_R.

In Figure 4 we set $R_L = 1/2$ and $M_S = M_R = M_D = 7$, because half-rate STBCs exist only when the number of transmit antennas is in the range [2,...,8] (meaning a maximum number of ancillary nodes participating to the communication equal to 7) [14]. As we can see there is a notable worsening in the performance degradation with OSTBC, meaning that performance of half-rate OSTBCs are quite far from theoretical bounds.

Finally, in Figure 5, the case of OSTBCs with $R_L = 3/4$ has been accounted for. Since, as stated in [14] 3/4 rate OSTBC could be realised only when the number of transmit antennas is equal to 3 and 4, we assume that nodes use a different code rate, according to the number of ancillary nodes forming the VVAs. In particular, when only one ancillary node is connected to the source or the relay main nodes an half-rate code is used ($R_L = 1/2$); when, instead, two or three ancillary nodes are connected a 3/4 -rate code is used. We also set $M_S = M_R = M_D = 3$ to limit the number of transmitting nodes to 4.

From the Figure we can see that the use of a 3/4 – rate code improves the performance considerably if compared with half rate OSTBCs.

6. CONCLUSION

In this paper we have investigated the performance of a two-hop decode-and-forward virtual multiple-input-multiple-output system with random location of nodes. The channel model we consider takes both slow and fast fading into account. System performance has been evaluated in terms of CDF of the SNR required to obtain a given rate. Results show that ancillary nodes densities strongly affect performance and the increase of these densities improves the connectivity of the network, therefore the CDF. The comparison between orthogonal space-time-block codes and the ergodic

capacity shows that, by increasing the code rate, the distance between OSTBCs and the theoretical bounds decreases.

7. ACKNOWLEGMENTS

This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++ (contract n. 216715).

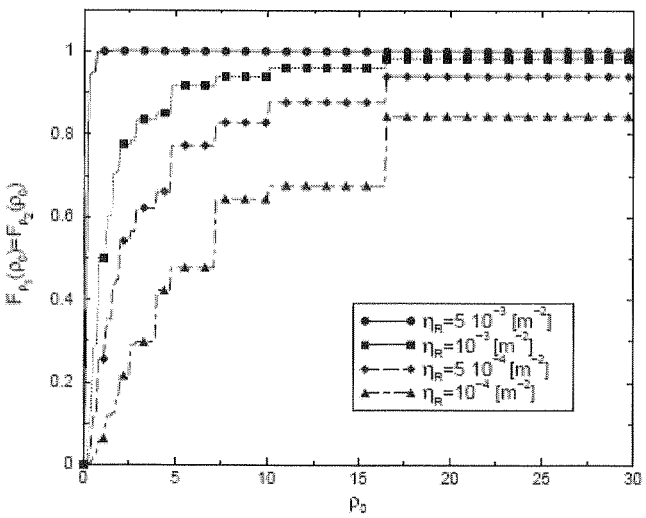


Fig. 2. $F_{\rho_1}(\rho_0) = F_{\rho_2}(\rho_0)$ when varying η_R, for the ergodic capacity case

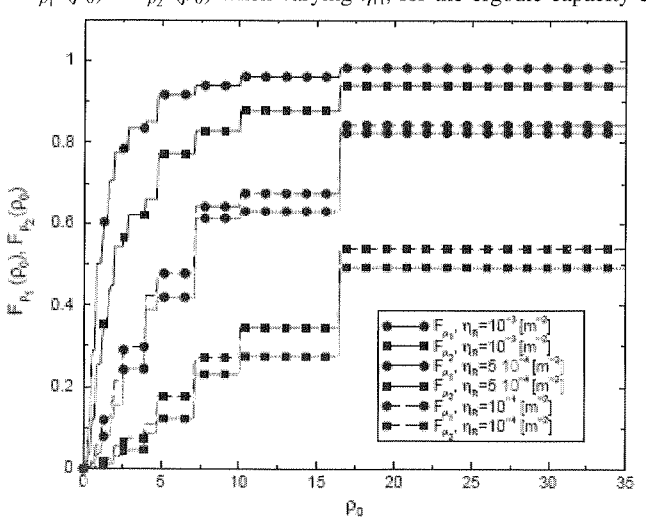


Fig. 3. $F_{\rho_1}(\rho_0)$ and $F_{\rho_2}(\rho_0)$ when varying η_R, for the ergodic capacity case

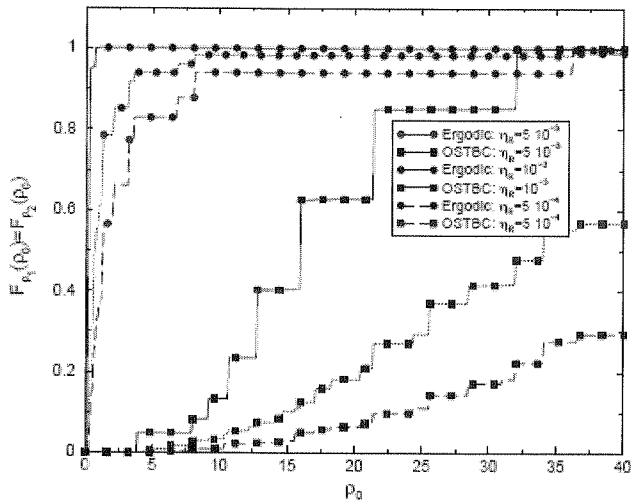


Fig. 4. $F_{\rho_1}(\rho_0) = F_{\rho_2}(\rho_0)$ when varying η_R, for the OSTBC case, having set $R_L = 1/2$. The values of η_R are expressed in $[m^{-2}]$

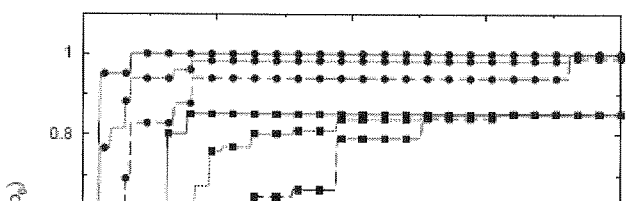


Fig. 5. $F_{\rho_1}(\rho_0) = F_{\rho_2}(\rho_0)$ when varying η_R, for the OSTBC with variable R_L. The values of η_R are expressed in $[m^{-2}]$

8. REFERENCES

1. M. Dohler, J. Dominguez, H. Aghvami: Link capacity analysis for virtual antenna arrays, in Proc of Vehicular Technology Conference, VTC 2002-Fall., 24-28 Sept. 2002
2. M. Dohler, A. Gkelias, H. Aghvami: 2-hop distributed MIMO communication system, Electronics Letters, vol. 39, n. 18, pp. 1350-1351, 4 Sept. 2003
3. M. Dohler, A. Gkelias, H. Aghvami: A resource allocation strategy for distributed MIMO multi-hop communication systems, IEEE Communications Letters, vol. 8, n. 2, pp. 99-101, Feb. 2004
4. H. Zhang and H. Dai: On the capacity of distributed MIMO systems, in Proc. of the Conf. on Information Sciences and Systems (CISS 2004), Princeton University, Princeton, NJ, March 2004
5. S. Jayaweera: V-BLAST Virtual MIMO for distributed wireless sensor networks, IEEE Trans. on Communications, vol. 55, n. 10, pp. 1867-1872, Oct. 2007
6. Y. Lang, D. Wubben, and K.D. Kammeyer: Efficient power allocation for outage restricted asymmetric distributed MIMO multi-hop networks, 19th Int. Symp. on Personal, Indoor and Mobile Radio Communications, (PIMRC 2008), Cannes, France, Sept. 2008
7. P. Zhang, F. Wang, L. Tu, K. Li, S. Diouba, I. Khider: Opportunistic virtual antenna array with optimal relay mobile terminals selection, 4th Int. Conf. on Wireless Communications, Networking and Mobile Computing (WiCOM '08), 12-14 Oct. 2008
8. W. Deng, X. Gao: Three-hop cooperative diversity system and symbol error rate analysis, Int. Symp. on Intelligent Signal Processing and Communication Systems (ISPACS 2007), Nov. 28 2007-Dec. 1 2007
9. R. H. Y. Louie, I. B. Collings, M. R. McKay: Analysis of dense ad hoc networks with spatial diversity, in Proc. of GLOBECOM 2007, 26-30 Nov., Washington, DC, 2007
10. H. Jafarkhani, H. Yousefizadeh, J. Kazemitabar: Capacity-based connectivity of MIMO fading ad-hoc networks, in Proc. of GLOBECOM '05, vol. 5, 2-2 Dec. 2005
11. K. Stamatiou, J. G. Proakis, J. R. Zeidler: Evaluation of MIMO techniques in FH-MA ad hoc networks, in Proc. of GLOBECOM 2007, 26-30 Nov., Washington, DC, 2007
12. C. Buratti, A. Zanella: Capacity analysis of two-hop virtual MIMO systems in a Poisson field of nodes, Veh. Tech. Conf. (IEEE VTC2009 – Spring), Barcelona, Spain, F, 26-29 April 2009
13. R. Verdone, D. Dardari, G. Mazzini, A. Conti: Wireless sensor and actuator networks, Elsevier, first ed., 2008
14. V. Tarokn, H. Jafarkhani and A. R. Calderbank: Space-time block codes from orthogonal designs, IEEE Trans. on Information Theory, Vol. 45, n. 5, pp 1456-1467, July 1999
15. J. Orriss, S.K. Barton: Probability distributions for the number of radio transceivers which can communicate with one another, IEEE Trans. on Communications, vol. 51, n. 4, Apr. 2003
16. E. G. Larson, P. Stoica: Space-time block coding for wireless communications,Cambridge Univ. Press, 2003
17. E. Telatar: Capacity of multi-antenna Gaussian channels, Europ. Trans. on Telecomm., vol. 10, pp. 585–595, Nov.-Dec. 1999
18. A. Zanella, M. Chiani, M. Z. Win: A General framework for the distribution of the eigenvalues of Wishart matrices, Int. Conf. on Communications (IEEE ICC 2008), Beijing, China, 19-23 May 2008
19. M. S. Alouini, A. Goldsmith, Capacity of Rayleigh fading channels under different adaptive transmission and diversity techniques, IEEE Trans. on Veh. Technology, vol. 48, pp. 1165-1181, July 1999
20. O. Oyman, R. U. Nabar, H. Bolcskei, A. J. Paulraj: Characterizing the statistical properties of mutual information in MIMO channels, IEEE Trans. on Signal Processing, vol. 51, n. 11, pp. 2784-2795, Nov. 2003

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 485–491

Automatic Audio Content Identification

TOMASZ MĄKA

West Pomeranian University of Technology
ul. Zolnierska 49, 71-210 Szczecin, Poland
e-mail: tmaka@wi.zut.edu.pl

Received 2009.06.01
Authorized 2009.08.08

This paper addresses the problem of automatic audio content identification. In order to determine regions of speech, music and silence in audio stream, the fusion of feature contours and their envelopes has been used. Additionally, a voicing detector and four class music genre identification stage have been incorporated into classification system. To minimize boundary errors of different audio regions, a smoothed envelope of feature contours has been proposed. Experimental results show that using proposed scheme, makes it possible to achieve acceptable classification rates for audio data segmentation. In result, this approach can be applied to the content type dependent multimedia processing.

Keywords: audio classification, feature contours, envelope of feature contours, music genre identification, SVM

1. INTRODUCTION

Audio content identification can improve many multimedia processing tasks. The retrieval and indexing systems, multimedia databases, speech recognition and universal audio compression systems use audio content type information extensively [1]. The multimedia system may improve its performance significantly by using such information. Therefore, it is important to use discriminator to distinguish between expected signal type and the rest of audio content [2], [3]. The most frequently designed discriminators determine two classes of audio: speech and music [1]. These two types have different properties, and can be distinguished with good results in many cases [4]. However, there are some situations where analyzed audio stream contains long transitions between speech and music, background noise, environmental sounds, etc. In such cases robust identification of speech and music regions is quite difficult and

stimulates to develop a new audio features, classifiers and modeling methods. One of the applications where audio identification can improve efficiency is audio coding. Typical speech and audio codecs obtain high compression ratio along with good perceptual quality. Speech codecs are designed to operate on speech signals and cannot compress audio signals with as good quality as for speech. The same situation is for audio codecs and speech signals [5]. It is important to use separate coders for each type of audio signals or content-based universal audio coding [6], [7]. The main approach to universal audio coding utilizes multi-mode compression based on signal classification [7]. Coder selection is performed after short-time signal type classification and can lead to frequent switching between coders. Furthermore, coders operate on different frame sizes and buffers, so such coding has to use additional control mechanisms [5]. Another issue is connected with the classification process. Identifying type of signal based on single or several frames of audio stream often complicate the decision process (the decision threshold changes over the time). Although this approach can cope with the real-time constraints, it may decrease classification rate. Taking these issues into account, in this paper we propose a non real time approach of generating audio content structure. For example, applying content type properties to generate context flags, can improve compression ratio in lossless audio coding [8].

2. AUDIO CLASSIFICATION

The accuracy of classification process depends on feature set and classification method. In our approach the input signal is segmented into frames with quasi-stationary signal. Then, for each frame a set of features is calculated giving a feature contour. In the next step the contour is normalized into [0; 1] range and the discrimination function is applied. Our system can determine speech, music and silence signal along with voicing type for speech regions and four-class genre classification for music regions. The architecture is depicted in Fig. 1. After studying the properties of many time and frequency domain based features, we have selected those, where contours have similar trends for each signal class.

Since the most approaches for speech determination are based on zero-crossing rate (ZCR) and short-time energy [1] we used these features to determine between speech and music. Due to the discrimination errors, we replaced short-time energy with cepstrum [9] energy and used it as a new feature. Cepstrum energy (CE) is calculated for each frame using formula (1), where: $x(m)$ – input signal, N – frame size, $W_N^k = e^{-j\frac{2\pi k}{N}}$. In the result a cepstrum energy contour is calculated.

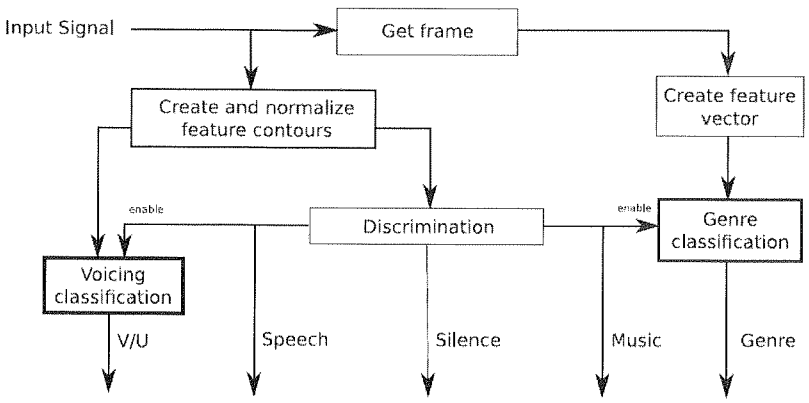


Fig. 1. Audio classification system architecture

$$CE = \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{n=0}^{N-1}\log\left|\sum_{m=0}^{N-1}x(m)\cdot W_N^{mn}\right|\cdot W_N^{-nk}\right)^2, \tag{1}$$

Next, the envelope of contour is determined (Fig. 2b) by applying envelope follower [10] (with 1ms attack and 15ms release time) to cepstrum energy contour, then smoothed using one pole IIR low pass filter with $\alpha = 0.95$ [11]. Using this feature, an average gain about 10 per cent has been obtained due to better speech region detection.

Voicing detection exploits a hard-decision approach with two features: spectrum tilt and low-band to full-band energy ratio [12]. Although this technique works fine in most cases, it highly depends on the speech content (noisy, mixed, etc.). The discrimination technique is shown in Fig. 3 where $C_0(n)$ denotes ZCR contour, $C_1(n)$ – short time energy contour, $C_2(n)$ – spectrum tilt contour, $C_3(n)$ – lowband to full-band energy ratio contour, $E_0(n)$ – envelope of cepstrum energy contour, $\alpha_{st} = 0.1$, $\alpha_{sp} = 0.5$, $\alpha_{sv} = 0.4$ and N denotes the number of frames of the analyzed signal.

The genre identification is performed on input signal using 14th dimensional vectors of time and frequency domain features calculated per every frame. Extracted feature vector is then classified employing the support vector machine linear classifier. The majority of selected features shown in Tab. 1 are widely used in the music genre classification systems [13], [14].

3. EXPERIMENTAL RESULTS

In order to verify our classification scheme, we have prepared two databases with audio material recorded at 44.1kHz sampling rate (feature contours and vectors have been generated using 23ms long frames with 50% overlap).

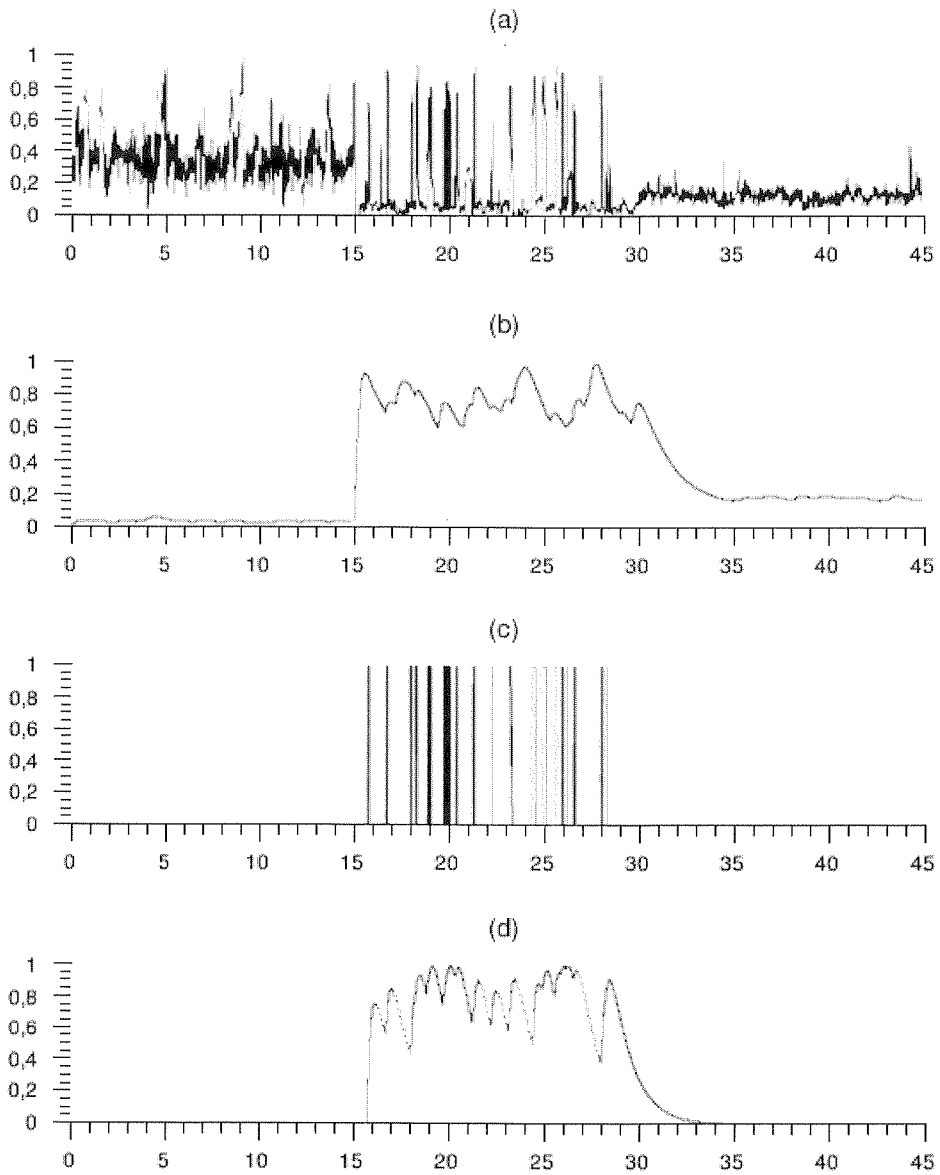


Fig. 2. Speech detection process: ZCR contour (a), envelope of cepstrum energy contour (b), detected speech (c), envelope of speech region (d) (analyzed signal contains three parts: 0-14s music, 15-29s speech, 30-44s environmental noise)

Algorithm 1 Discrimination algorithm

for $n = 1$ to N **do**
  if ($C_0(n) < \alpha_{st}$ and $C_1(n) < \alpha_{st}$) **then**
    $Type(n) \leftarrow SILENCE$
  else
    if ($C_0(n) > \alpha_{sp}$ and $E_0(n) > \alpha_{sp}$) **then**
      $Type(n) \leftarrow SPEECH$
      if ($C_2(n) > \alpha_{sv}$ and $C_3(n) > \alpha_{sv}$) **then**
        $SpeechType(n) \leftarrow VOICED$
      else
        $SpeechType(n) \leftarrow UNVOICED$
      end if
    else
      $Type(n) \leftarrow MUSIC$
      $Genre(n) \leftarrow GetMusicGenre(n)$
    end if
  end if
end for

Fig. 3. Discrimination algorithm

Table 1

Feature vector characteristic

Time domain	Frequency domain
• ZCR [1]	• Spectrum Flux [14]
• Irregularity [15]	• Centroid [14]
• Log-energy [9]	• Rolloff-85 [14]
• Spectrum tilt [12]	• Bandwidth [3]
• Pre-emphasized energy [12]	• Sub-band power (4 sub-bands) [3]
• Low-band to full-band energy ratio [12]	

The first database contains 25 audio clips with three kinds of sounds: music (different genres), speech (male, female and children) and environmental noise (street, office, city center, bus, etc.). Each piece is 15 seconds long and has normalized volume. Performed tests show an almost 92 percent correct classification ratio for the speech detection based on ZCR and envelope of cepstrum energy contours. An example of speech detection process for one clip from this database is depicted in Fig. 2. The second database has 80 music clips, each from 15 to 30 seconds long. These fragments

have been extracted at random positions from 80 full length songs in following genres: classical, disco, jazz and metal. The music genre classification stage utilizes the SVM (support vector machine) classifier using libSVM [16] with RBF (radial basis function) kernel. The results of each genre identification rate have been presented in Tab. 2. As can be seen, the lowest ratio was obtained for jazz and the highest for disco genre. Overall hit ratio for this database is about 74 per cent – 16621 of 22172 frames have been classified correctly.

Table 2

Genre classification results

Genre	Classical	Disco	Jazz	Metal
Hit ratio	76%	84%	67%	78%

4. CONCLUSIONS

In the paper, the audio content discrimination algorithm based on fusion of feature contours and its envelopes was introduced. By replacing short time energy feature with cepstrum energy feature, the performance of speech discriminator increases for many speech signals. Also, applying the envelope follower for feature contours leads to decreasing discrimination errors. The results obtained for music genre classifier are promising, but overall performance depends on audio structure and recording quality. Proposed approach is suitable for analysis of audio content structure where using envelope along with constant threshold level leads to increase of accuracy within region boundaries detection and simplifies decision process. Thus, these parameters can be used for tracking audio content type changes. Such technique may be employed in audio coding, where using well defined context at the data modeling stage can boost coding efficiency [8].

The research work presented in this paper was sponsored by Polish Ministry of Science and Higher Education (years 2007-2010).

5. REFERENCES

1. A. I. Al-Shoshan: Speech and Music Classification and Separation: A Review, J. King Saud Univ., Eng. Sci., vol. 19 (1), pp. 95-133, Riyadh, 2006
2. K. El-Maleh, M. Klein, G. Petrucci, P. Kabal: Speech/Music Discrimination for Multimedia Applications, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing: ICASSP'00, 2000
3. S. Z. Li: Content-Based Audio Classification and Retrieval Using the Nearest Feature Line Method, IEEE Transactions On Speech And Audio Processing, vol. 8, no. 5, Spetember, 2000

4. S. Yuan-Yuan and W. Xue and S. Bin: Several Features for Discrimination Between Vocal Sounds and Other Environmental Sounds, XII European Signal Processing Conference – EUSIPCO'04, September 6-10, Austria, 2004
5. S. A. Ramprashad: A multimode transform predictive coder (MTCP) for speech and audio, IEEE Workshop on Speech Coding for Telecommunication, pp. 10-12, Finland, 1999
6. B. Bessette, R. Lefebvre, R. Salami: Universal Speech/Audio Coding Using Hybrid ACELP/TCX Techniques, IEEE Int. Conf. on Acoustics, Speech and Signal Processing: ICASSP'05, 2005
7. L. Tancerel, S. Ragot, R. Lefebvre: Speech/Music Discrimination for Universal Audio Coding, 20th Biennal Symposium on Communications, Queen's University, Kingston, Canada, May 28-31, 2000
8. M. Hans and R. W. Schafer: Lossless Compression of Digital Audio, IEEE Signal Processing Magazine, July, 2001
9. A. M. Peinado and J. C. Segura: Speech Recognition Over Digital Channels: Robustness and Standards, John Wiley & Sons, Ltd., 2006
10. P. R. Cook: Real Sound Synthesis for Interactive Applications, A K Peters, Ltd., 2002
11. J. O. Smith III, Introduction To Digital Filters With Audio Applications, W3K Publishing, 2008
12. A. M. Kondoz, Digital Speech – Coding for Low Bit Rate Communication Systems, John Wiley & Sons, Ltd., 2004
13. J. J. Aucouturier and F. Pachet, Representing Music Genre: A State of the Art, Journal of New Music Research, vol. 32, no. 1, pp. 83-93, 2003
14. G. Tzanetakis and P. Cook, Musical Genre Classification of Audio Signals, IEEE Transactions On Speech And Audio Processing, vol. 10, no. 5, July, 2002
15. J. Krimphoff and S. McAdams and S. Winsberg: Characterization of the timbre of complex sounds. 2. Acoustic analysis and psychophysical quantification, J. de Physique, 4(C5), pp. 625-628, 1994
16. C. C. Chang and C. J Lin: LIBSVM: a library for support vector machines, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 493–499

A Contribution to the Discrete Logarithm Problem*

ANDRZEJ PASZKIEWICZ

Warsaw University of Technology
e-mail: anpa@tele.pw.edu.pl

Received 2009.06.21
Authorized 2009.08.25

In this paper we consider a common, fundamental problem of all cryptographic schemes based on the discrete logarithm modulo a prime. This is the choice a base of the modular exponent function or other words a generator of a prime order multiplicative group. We construct a class of primes for which its least generators can be extremely large.

Keywords: The discrete logarithm problem, multiplicative group modulo a prime, the least quadratic non-residue, the least primitive root

1. INTRODUCTION

Let p be an odd prime. It is well known that the multiplicative group $GF^*(p)$ is cyclic [3]. Mathematicians working in number theory call the generators of that group primitive roots modulo a prime. The least generator $g(p)$ of the group $GF^*(p)$ is called the least primitive root modulo p. Primitive roots modulo p play important role in several cryptographic schemes, such as Diffie-Hellman and ElGamal key exchanging schemes [4]. This is because the exponentiation function $y = g^x \pmod p$ is a bijection for g being generator modulo p. It means that for a given generator g modulo p and a natural argument x there exists a reverse function $x = g^y \pmod p$ which allows to determine y if we know x, g and p. This function we call discrete logarithm and write $y = \log_g x$.

Having g, p and a natural number x, it is quite easy to find the result of exponentiation $y = g^x \pmod p$, but if we know y, g and p problem of determining y becomes hard, excluding some special cases. It means that the huge theory of approximations

* This paper was supported by the Ministry of Science and Higher Education of Poland – research grant no. N517 003 32/0583 for 2007-2010

developed for real and complex functions does not work in the case of discrete modular functions, especially for the discrete logarithm function. For this reason such transformations are suitable for many cryptographic schemes.

To construct a unique discrete logarithm function modulo p, where p an odd prime, we need to find any generator g modulo p. A generator g is a number which has maximal multiplicative order modulo p. This gives us an easy way to determine generators by the method of exhaustive search, starting from the smallest numbers and checking the consecutive in the case of failure. This is reasonable approach because there exist natural densities for sets of such primes having their least primitive roots equal to small natural numbers [1], [5]. For the initial natural numbers 2, 3, 5, 6, and 7 these densities are given by the following formulas

$$
\begin{aligned}
D(2) &= \Delta_1, \\
D(3) &= \Delta_1 - \Delta_2, \\
d(5) &= \frac{20}{19}\Delta_1 - \frac{200}{91}\Delta_2 + \frac{500}{439}\Delta_3, \\
D(6) &= \Delta_1 - \frac{282}{91}\Delta_2 + \frac{1000}{439}\Delta_3, \\
D(7) &= \Delta_1 - \left(4 + \frac{9}{91} + \frac{5}{281}\right)\Delta_2 + \left(6 + \frac{183}{439} + \frac{4826}{67585} + \frac{147193}{296698115}\right)\Delta_3 \\
&\quad - \left(3 + \frac{1107}{2131} + \frac{71825}{1290197} + \frac{26503425}{2749409807}\right)\Delta_4
\end{aligned}
$$

where

$$
\Delta_m = \prod_p \left(1 - \frac{1}{p-1}\left(1 - \left(1 - \frac{1}{p}\right)^m\right)\right) \text{ and } m = 1, 2, \ldots
$$

The values Δ_m can be calculated with high accuracy. For m=1,2,3, and 4 we have: $\Delta_1 = 0,373955$, $\Delta_2 = 0,147349$, $\Delta_3 = 0,060821$, $\Delta_4 = 0,026107$. For the densities given by the above formulas we have $D(2)$=0,373955, $D(3)$=0,226606, $D(5)$=0,139065, $D(6)$=0,055881, $D(7)$=0,068702. That means that more than $D(2)+D(3)+D(5)+D(6)+D(7)$=0,864208 of all primes (above 86,42%) have their least prime generators just the initial integers being not squares and not exceeding the number 7. Despite of that we can find sometimes hard instances of (large) primes which have very large its least generators. That is the topic of the next paragraph.

2. PRIMES WITH LARGE MULTIPLICATIVE GENERATORS

For a given prime number p denote $n_2(p)$ its least quadratic non-residue. By a quadratic non-residue we mean a number a for which the modular equation $x^2 \equiv a \pmod{p}$ has no solutions. Such smallest a is just called least quadratic

non-residue modulo p. Suitable tool for checking whether given a is quadratic residue modulo a prime or not is the Legendre symbol [3].

Let k be any natural number. Consider the set of numbers $q_k(x)$ of the following form $q_k(x) = 8\, p_2 \ldots p_k \cdot x + 1$, where natural number x is a parameter and $p_2, p_3, \ldots, p_k$ are consecutive primes. By Dirichlet's theorem on arithmetic progressions this set contains infinitely many prime numbers. Let x be such a number for which $q_k(x)$ is prime. The following theorem holds

THEOREM. *For each prime number $q_k(x)$ defined above its least quadratic non-residue is greater than p_k.*

Proof. Let i be any natural number between 2 and k. We show that the Legendre symbol $\left(\frac{p_i}{q_k}\right) = 1$. By the reciprocity low for Legendre symbols we have $\left(\frac{p_i}{q_k}\right) = (-1)^{\frac{p_i-1}{2}\cdot\frac{q_k-1}{2}}\left(\frac{q_k}{p_i}\right)$ and because $\frac{p_i-1}{2}\cdot\frac{q_k-1}{2}$ is even we obtain $\left(\frac{p_i}{q_k}\right)=\left(\frac{q_k}{p_i}\right)=\left(\frac{8\, p_2 \ldots p_i \ldots P_k, x+1}{p_i}\right)=\left(\frac{1}{p_i}\right)=1$. Now it suffices to show that $\left(\frac{2}{q_k}\right) = 1$, but this is obvious. Now by Gauss rule for Legendre symbols $\left(\frac{2}{q_k}\right) = (-1)^{\frac{q_k^2-1}{8}} = (-1)^{\frac{(q_k-1)\cdot(q_k+1)}{8}} = (-1)^{p_2\ldots p_k\cdot x\cdot(q_k+1)}$. Remarking that $q_k + 1$ is even we see that $\left(\frac{2}{q_k}\right) = 1$ which finishes our proof.

From the Theorem we have the following

COROLLARY. *For each prime number $q_k(x)$ defined above its least primitive root is greater than p_k.*

For proof it suffices to remark that a primitive root modulo a prime should also be quadratic non-residue. The above Corollary gives us an effective tool to construct prime numbers with arbitrary large least primitive roots.

3. NUMERICAL EXPERIMENTS WITH PRIMES HAVING LARGE MULTIPLICATIVE GENERATORS

Let k be fixed. Denote by N_k the least x for which $q_k(x)$ defined in the paragraph 2 is prime. For each $k \leq 1000$ we found a pair: N_k such that $q_k(N_k)$ is prime and the least primitive root of $q_k(N_k)$. For the sake of brevity we will denote $q_k(N_k)$ just by q_k and the least primitive root $g(q_k(N_k))$ simple by g_k. W illustrate our investigations by a serial of 4 figures and a table.

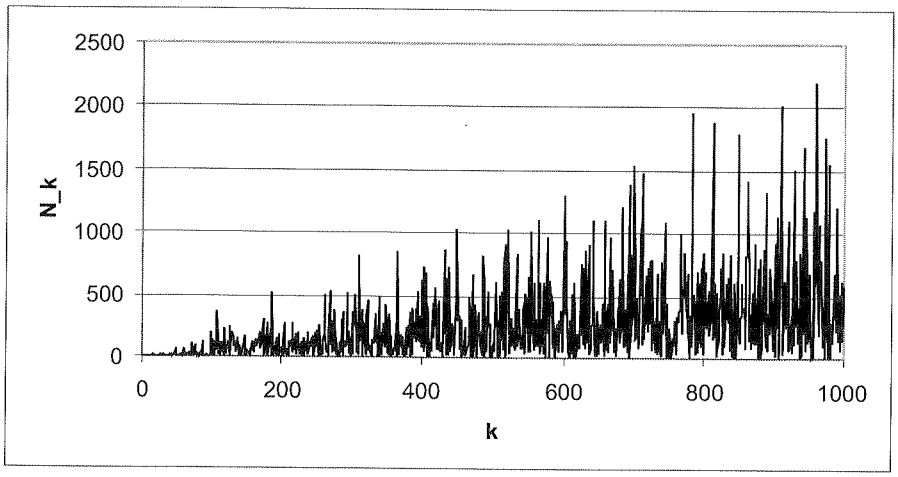


Fig. 1. Numbers N_k for which q_k are prime (linear scale)

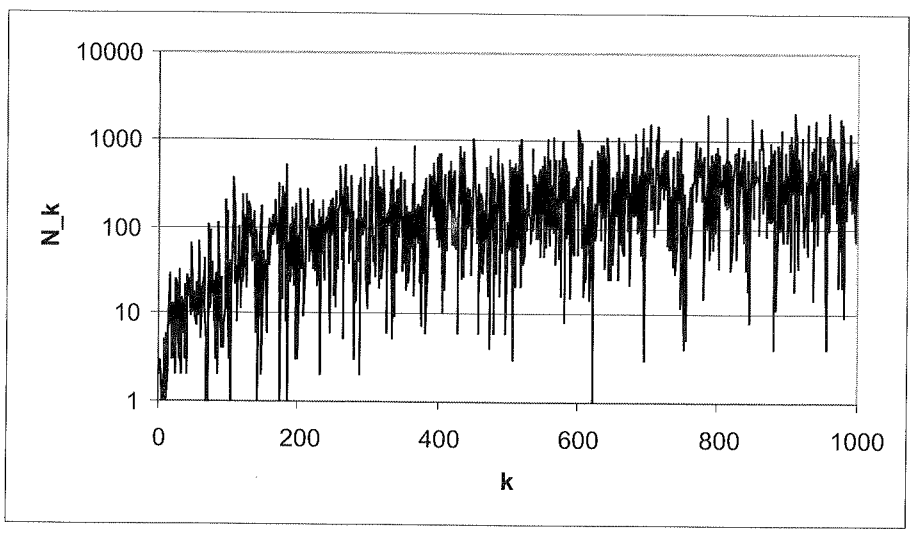


Fig. 2. Numbers N_k for which q_k are prime (logarithmic scale)

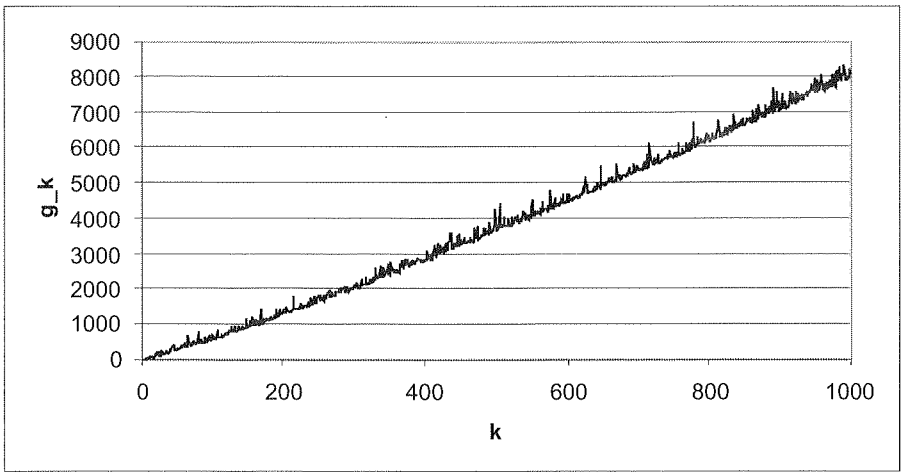


Fig. 3. The least primitive roots g_k

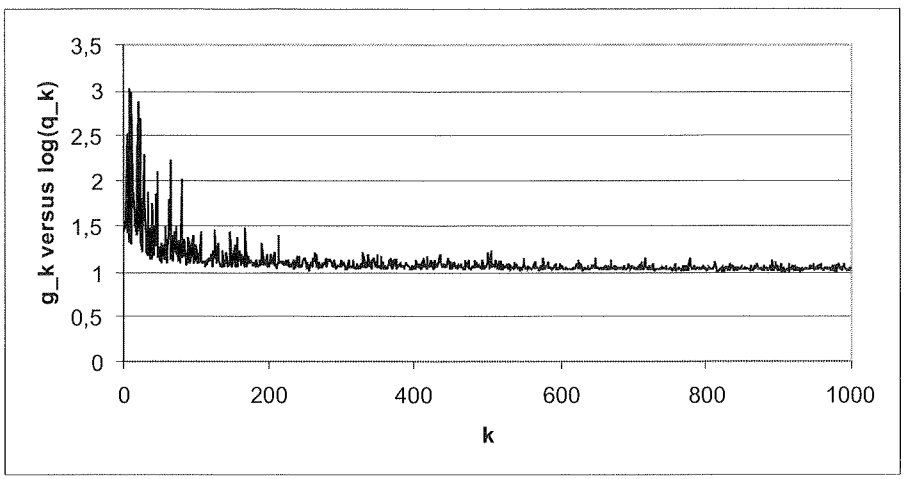


Fig. 4. The least primitive roots g_k versus natural logarithm of q_k

Table 1

Primes with very large least primitive roots

k	p_k	N_k	g_k
95	499	11	613
168	997	79	1103
249	1499	78	1597
303	1999	353	2161
367	2477	27	2617
430	2999	867	3079
489	3499	16	3571
550	3989	127	4523
610	4493	23	4567
669	4999	212	5021
725	5483	769	5521
783	5987	344	6317
842	6491	18	6689
900	6997	131	7019
950	7499	605	7523
1007	7993	546	8111
1059	8467	778	563
1117	8999	498	9043
1177	9497	320	9631
1229	9973	655	10427
2262	19997	223	20107

4. CONCLUSIONS

The most difficult part of our investigation was rigorous primality testing of large primes of the form $q_k\ (x) = 8\ p_2 ... p_k \cdot x + 1$ for growing k. The nice feature is that the factorization of $q_k\ (x)$ - 1 is known in advance and can be omitted. Finding a primitive root of a number $q_k\ (x)$ is equivalent to checking its primality, so it is a good test for the correctness of computation. As it can be seen (Fig. 1 and Fig. 2) many initial numbers in arithmetic progressions $q_k\ (x)$ are consecutive composite numbers. This is because the quantity N_k grows. For example if k=961, then corresponding to it N_k = 2205. Few of the numbers N_k = 1 (see Fig. 2 in logarithmic scale for these arguments where zero values appear). Figures 3 and 4 show the growth rate of g_k as a function of k. It can be seen that in our construction the growth rate of g_k is not much faster

than logarithmic of $q_k(x)$. The best recent unconditional result is due to Graham and Ringrose and it differs from our result by a triple iterated factor of $q_k(x)$ [2].

The complexity of computations grows very fast. Even modular exponentiation of large numbers takes much time, so for some part of computations we used FFT algorithm for fast multiplication. The last two rows of the table 3, where in the last column we store very large least primitive roots, were, for example, computed with the usage of FFT. It is no wonder, because we have to do with really huge numbers. The first one of them has its least primitive root equal to 10427 and the second one 20107.

5. REFERENCES

1. P. D. T. A. Elliott, L. Murata: On the average value of the least primitive root modulo p, J. London Math. Soc. 56, pp. 435-454, 1997
2. S. W. Graham, C. J. Ringrose. Lower bounds for least quadratic non-residues, Analytic Number Theory, Proceedings in Honor of Paul T. Bateman, Progress in Mathematics 85, pp. 269-309, Birkhäuser, Boston 1990
3. K. Ireland, M. Rosen, A Classical Introduction to modern Number Theory, Springer-Verlag, New York, Heidelberg, Berlin, 1982, Graduate Texts in Mathematics, 87
4. A. Menezes et all, Handbook of Applied Cryptography, CRC Press, Boca Raton, 1997
5. A. Paszkiewicz: On the least prime primitive root modulo a prime, Math. Comp. v. 71, no. 239, pp. 1307-1321, 2002

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 501–513

Realization of CARE and EuCARD Projects in ISE-WUT Accelerator and FEL Research, Development and Applications in Europe

RYSZARD S. ROMANIUK

Institute of Electronic Systems,
Warsaw University of Technology
R.Romaniuk@ise.pw.edu.pl

Received 2009.06.09
Authorized 2009.08.25

There are described coordinating actions of the accelerator science in Europe in 2003-2009. The actions embrace basic science, as well as development and applications. The accelerator research was not coordinated in Europe at a global scale but was rather concentrated in a few centers owning large infrastructure. These centers include: CERN, DESY, GSI, INFN, LAL, PSI etc. Such coordinating actions enable a lot of positive processes including new possibilities for research centers in this country. It is much easier for them to extend, deepen or even start from the beginning their activities in the field of the accelerator technology. This field includes also free electron lasers. There are described two European framework projects CARE and EuCARD on accelerator technology, their extent and the participation of the Institute of Electronic Systems ISE WUT in them.

Keywords: Accelerator science, accelerator technology, synchrotron, free electron laser, elementary particles, electronic and photonic systems, European framework programs

1. INTRODUCTION

The Accelerator Technology is one of these research branches, which involves design, construction, commissioning and then exploitation, during some confined time, a large or even immense, and very complicated technical infrastructure. This period, which is an effective lifetime of a large accelerator lasts around 20-25 years. It is possible, in the future, with the increased pace of development, that this period will shorten. On the other hand researchers tend to build even bigger and more complicated

machines for more money, which will prevent this tendency. Smaller accelerators, of advanced construction, for technical and medical applications are built as a consequence, and parallel product, of intense research work on big accelerators. The size of research infrastructure is imposed by bigger requirements for energy and intensity of the accelerator beam. The biggest research accelerators, due to very high costs of construction and exploitation, are only a few and are almost only built as a result of wide international cooperation. The technologies created for the biggest accelerators, including mechanics, chemistry, electronics, photonics, material engineering, etc. are transferred very fast to the industry, including power engineering, nuclear, geology and mining, medicine, safety, environment engineering, etc. To gain a direct source access to these technologies, it is necessary for the research teams from this country to participate in large research experiments including: accelerators, nuclear, power, photonics, laser, cosmic, astronomic, etc. It is also necessary to build, around such large experiments in the country, local industrial and technical consortia. A question appears now, if this participation is sufficient, which is hard to answer. This participation might be much greater, reflecting the potential of the local research community.

After several years of wider participation of the local research and technical communities from this country in the European Framework Research Programs (FP), particular research teams have gathered critical experience, allowing for deeper and more essential reflections. Some research groups from this country participated in FP5, most of them started in FP6, quite a few continue in FP7 [1]. The participation of domestic teams was and is now so numerous, that an administrative infrastructure was built to facilitate the cooperation. The infrastructure consists of a national central office for the EU FP [2], regional information offices [3] and open community, non-governmental organizations. One of such useful associations is National Committee of EU FP Coordinators – KRAB [4]. The bylaws and mission of KRAB assumes realization of the flowing tasks: exchange of experiences from project realizations, promotion of good practices for professional conduct of projects, promotion of participation in FP teams from Poland, cooperation with governmental and local community administrations, recommendation of changes in the research law to facilitate the participation in the European Projects, education, organization of conferences on European Programs. KRAB defines the most important problems to be solved, as follows: big difficulties to obtain an European Project for a lot of research teams from Poland, big disproportions between salaries of the project participants in Europe and in this country, lack of stable law solutions for realized projects, lack of professional managerial staff supporting project realization in Poland. The aggregated participation of Poland in the realization of EU FPs, together with Operational Programs (POIG, POKL and POIiS) funded by the EU, is so massive that fast changes, breaking large inertial resistance on all levels – governmental and beneficiaries, are unavoidable.

Participation in large EU FP, and such are the integrated infrastructure projects concerning the accelerator technology, requires fulfillment of a few, sometimes quite difficult for universities, conditions: presence of a national research team in a sufficien-

tly massive topical consortium which generates particular project, proposing an attractive research offer for infrastructure development (this infrastructure is very frequently outside this country) and declaration of own either financial or in-kind share. The offer has to be attractive for the consortium and for the owner of the infrastructure. In some cases, this declared share may be as high as 2/3 of the value of proposed work package or task and typically is around 1/2. The necessity to declare such high own financial share prevents a lot of national teams to participate in big infrastructural undertaking, despite the support by directed national co-financing grants. The EU FP projects are now realized in the FC – full costs accounting frames. Previously, within the FP6, the accounting was AC – additional costs. This requires from the domestic universities to maintain complicated multilevel bookkeeping documentation, for which they are hardly prepared. Fulfilling of these requirements burdens not only the administration of the university but also the research leader of the project. It is evident, that some research leaders, who are prepared to undertake the effort, avoid it assuming the lack of appropriate support from the side of administration services of the institution.

The research teams from Poland are participating in a number of scientific and technical projects concerning construction and/or development of the accelerator or nuclear infrastructure in Europe (like E-XFEL, FAIR, ITER). These programs are coordinated on several levels by governmental agencies of the biggest European countries, by non-governmental organizations, also by the European Commission. The main coordinator on the European level in this topical area is ESGARD [5] – European Steering Group for Accelerator R&D. The first project realized within FP6 was CARE [6-8], which run during the period 2004-2008. A continuator of this project is EuCARD [9-12] European Coordination of Accelerator R&D, scheduled within FP7 for the period of 2009-2013. All European accelerator infrastructures participate in the project as well as many other institutions like universities having accelerator research groups, including Warsaw University of Technology. The area of activities of a team from ISE WUT in CARE and EuCARD is SRF – superconducting radio frequency [13-14]. The teams and persons from ISE WUT participate also in different related projects concerning accelerators, astro-particle physics, elementary particles realized previously like ZEUS/HERA and realized now like ALBA, CMS/LHC, SPS, FLASH, E-XFEL, ILC, PITZ, CBM, Chandrayaan satellite, Pi-of-the-Sky, etc. The people involved directly in technical work are mainly M.Sc. and Ph.D. students, which work under the supervision of university faculties.

2. APPLICATIONS OF ACCELERATOR TECHNOLOGY

The current possibilities of broad applications of accelerator technology are a direct consequence of large research projects and experiments in Europe like: TESLA, HERA, LHC, CLIC and worldwide like: ILC, CEBAF, SLAC, SNS, KEK, carried out from a few decades, and resulting in essential achievements in fundamental and technical areas. Chosen European centers of accelerator infrastructure are gathered in

table 1. The research in HEP concentrates today in a few main directions: recreation of the initial conditions during the BB – big bang; search for very rare events; observations of the universe; gathering of immense data sets enabling reliable statistical analyses. Four types of large, complex and costly accelerator infrastructures are required to realize these aims. These embrace: accelerators of big energies and big intensities of the beams; accelerator detectors; management and gigantic data processing systems (zetta-byte data sets 10^{21}); large auxiliary infrastructure, for example underground. The development of each kind of these infrastructures translates nearly immediately to industrial applications, from which the whole society may profit. These applications include: medicine, safe nuclear power utilities, new materials, information technologies, Internet, and in the future: making the Moon and the Mars our home, and other space technologies.

Table 1

European accelerator infrastructures (chosen objects)

Laboratory	Accelerator	Description
ALBA Barcelona	ALBA	Synchrotron light source
BESSY Berlin	BESSY	Synchrotron light source
CCLRC – RAL Didcot	ISIS, MICE	Accelerator complex: muons and neutrons
CERN Geneve	LHC Detectors CLIC, SLHC	Proton accelerator complex, Neutrino beam, Ion accelerator, Two beam electron linac, SRF laboratory
DESY Hamburg	FLASH, PITZ, E-XFEL, TESLA	Superconducting electron linac, Electron injectors, FEL, Synchrotron,
FZ Rossendorf	ELBE	Linear electron accelerator
GSI Darmstadt	SIS, ESR, FAIR, UNILAC	Accelerator complex for heavy ions, Beam tests laboratory
INFN Frascati	SPARC DAFNE	FEL, Synchrotron source
LAL IN2P3 Orsay	PHIL	Electron injector, European synchrotron

In the area of fundamental research, the accelerator technology infrastructure is used to find answers to the following questions concerning the energy: are there any other laws of physics?, what is dark energy?, are there additional dimensions?, do all the forces originate from a single one?, and questions concerning the elementary particles: why there are so many elementary particles?, what is dark mass and how it can be generated?, what is the nature of neutrinos?, and also questions concerning the

universe: how it started into being?, what has happened to the antimatter? The research accelerators for HEP are developing into bigger energies, bigger intensities and smaller dimensions of the beam. The further development of accelerator technology requires access to test facilities: accelerated beams, high field magnets and superconductive resonant cavities of the highest quality. The development of the biggest accelerators for HEP gave birth to a number of other research accelerator families for new generation of light sources of the highest intensities [15-16], muon and neutron sources, neutrino beams [17], and then medical accelerators for cancer therapy, as well as industrial ones for ion implantation, material engineering, welding and cutting with an electron beam, radioisotope production, ion transmutation, nondestructive testing and safety. Certain accelerator devices of appropriate power are used additionally for transmutation processes of nuclear power reactor waste. Table 2 gathers an arbitrary estimate concerning the worldwide, industrial accelerator markets in 2009.

Table 2

Industrial market for accelerators (predictions for 2009) in [pcs.] and [M€]

Application	Number of working systems [pcs.]	Numer of systems sold in the last year [pcs.]	Total value of trade [M€]	Range of price for a single system [M€]
Medical therapy	> 10 000	> 600	> 2000	2,0 – 5,0
Ion implantation	> 12 000	500	> 1500	1,0 – 3,0
Cutting and welding with electron beam	> 5 000	> 100	> 200	0,5 – 2,5
Irradiators with e and X	> 2 000	> 100	> 150	0,2 – 5,0
Radiosotope production and PET	> 650	> 50	> 100	1,0 – 3,0
Nondestructive testing and safety systems	> 1 000	> 200	> 100	0,3 – 2,0
Ion beams analysis and AMS	> 250	> 30	> 30	0,4 – 1,5
Neutron generators	> 1 100	> 50	> 30	0,1 – 3,0
Isotope transmurtation, transmutation of nuclear reactor waste (nuclear power stations)	> 100			Similar to the production of radioisotopes
FEL	single			v.high
Other	> 1 000			Relat.small
Together (plus other industrial applications)	> 35 000	> 2 000	> 5 000	0,1 – 5,0
Research: Accelerator Energy Recovery; Accelerator I-FEL; Plasma –laser accelerators	single			Very high

3. EUROPEAN ACCELERATOR PROJECTS: CARE AND EuCARD

Project CARE embraced the following, so called, Joint Research Activities JRA: SRF [13] – superconducting radio frequency, PHIN [18] – photon injectors, HIPPI [19] – proton injectors, NED [20] – high field magnets from Nb_3Sn, and research networks: ELAN [21] – European electron linear accelerators, BENE [22] – beams for neutrino experiments, HHH [23] – hadron beams of high energies and high intensities. As a result of CARE realization, a large development step was obtained in the following areas: quality of superconducting RF cavities for ultimately high fields – fig.1, control and tuning of cavities, diagnostics and control of the beams, parameters of couples and high field magnets, injector construction, surface preparation and quality improvement of accelerator materials, etc. The pioneering project CARE initiated and cooperated with a series of other European undertakings and common activities in accelerator technology, like: EuroTeV [24], EuroLEAP [25], EURISOL [26], CLIC [27], SLHC-PP [28].

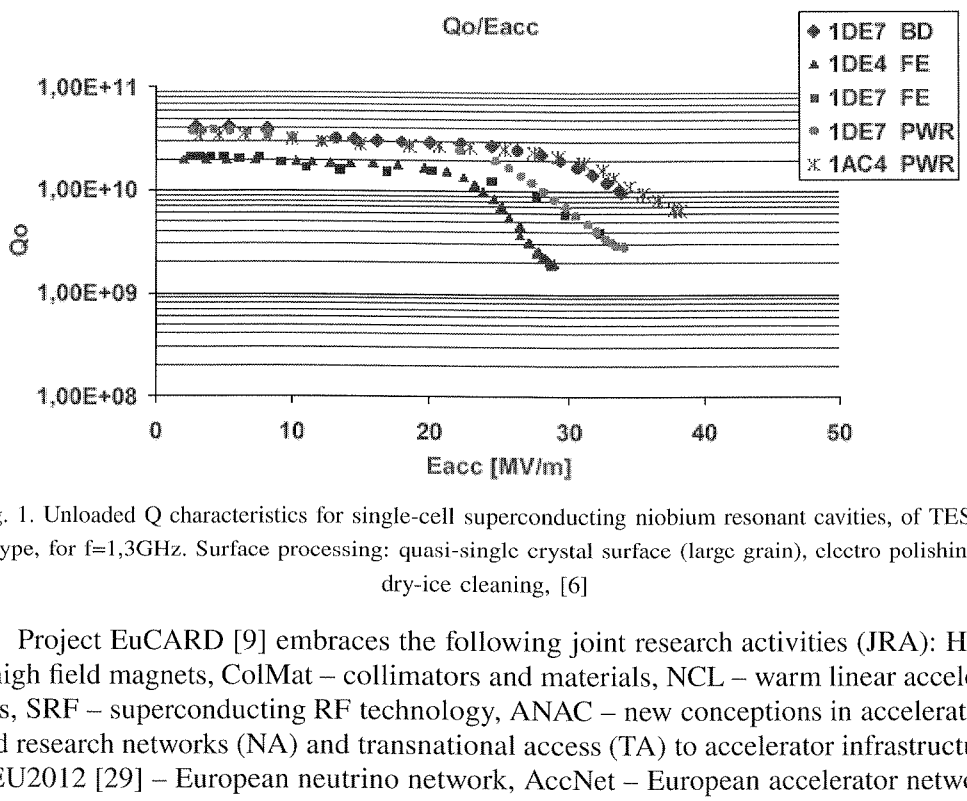


Fig. 1. Unloaded Q characteristics for single-cell superconducting niobium resonant cavities, of TESLA type, for f=1,3GHz. Surface processing: quasi-single crystal surface (large grain), electro polishing, dry-ice cleaning, [6]

Project EuCARD [9] embraces the following joint research activities (JRA): HFM – high field magnets, ColMat – collimators and materials, NCL – warm linear accelerators, SRF – superconducting RF technology, ANAC – new conceptions in accelerators, and research networks (NA) and transnational access (TA) to accelerator infrastructure: NEU2012 [29] – European neutrino network, AccNet – European accelerator network, HiRadMat – research access to proton and ion beams in SPS CERN, and MICE [30] – research access to muon and neutron beams in RAL [31]. The accelerator network (and neutrino network) and a good organization of facilitated, transnational access to the beams are a sort of a backbone of the EuCARD project. Without a considerably

extended access to the existing beams for wider research community, international, European and domestic, it is impossible to justify the realization of such expensive projects like EuCARD. Are we going to use this facilitated access?

4. ISW WUT PARTICIPATION IN CARE AND EuCARD PROJECTS

The participation of research teams from ISE WUT concerned a realization of a few tasks including: design, construction, testing, application and commissioning of advanced electronic and photonic systems for accelerator technology, in close co-operation with the DESY – a host of this infrastructure. The other coworkers in this effort were domestic institutions: IPJ-Świerk, Łódź Uni. of Technology, Wrocław Uni. of Technology, Institute of Experimental Physics, Warsaw University, and Institute of Nuclear Physics, Kraków.

There were carried out the following tasks: modelling and measurements of electronics and photonics degradation mechanisms in VLSI chip based systems, in particular memory modules and FPGA circuits, all working in high ionizing radiation fields. A laboratory set-up was constructed for testing of the kinds of errors appearing commonly under the influence of highly energetic EM radiation and corpuscular beams, in a system managed by an FPGA chip. There were tested mitigation methods based on introduction to the system hardware redundancy and algorithmic redundancy. Adverse environment in an accelerator channel includes: highly energetic electromagnetic radiation and secondary, relatively low corpuscular radiation, like neutrons. The biggest sensitivity to the neutron radiation show SRAM memories and FET transistors. A remotely controlled monitoring system for radiation level was mounted in FLASH tunnel. The system was oriented towards checking the influence of radiation on the FPGA based circuit. The system was integrated with DOOCS which enabled on-line analysis including radiation measurements, watching the effect of accumulation of radiation energy deposits and harmful changes in electronics with power of dose. There were also carried out numerous measurements of radiation hardness of electronics in the Linac 2 in DESY.

Noise phenomena were investigated in a frequency down converter with the intermediate frequency IF=250kHz, used in analog-digital control feedback loop for superconducting resonant Tesla cavity. The measurements showed the noise level of around two orders of magnitude larger than calculated from the specification of the key components used for building of the device. The reasons were identified as external noise sources and errors in the down converter design. The electromagnetic environment of down converter work is highly noisy via the presence in the vicinity the high power supplies, klystrons, modulators, waveguides, cabling, digital electronics, etc. A classification of noise sources was done taking into account their influence on the analog electronics. A new solution of low-noise down converter was proposed, with the IF=81MHz. The linearity of the device was measured for the new IF on the level of 10^{-3}. The quality requirements for vector sum control of many resonant cavities

were estimated. Integration of the down converter was considered in a form of a single chip.

A few models of multilayer system PCBs were designed and manufactured. They work in motherboard-daughterboard configuration and contain 8 channels of ADC, four channels of DAC, FPGA and DSP chips and a few channels of gigabit optical links. The noise level of working digital system was measured and equal to 1 – 4 mV rms. This noise level of the first solution of PCB was not satisfactory. It was lowered in the next PCB designs. Control programming for the PCBs was designed and implemented. It consisted of software layer and firmware layer. The system was tested in different conditions of laboratory work, and then during exploitation conditions accessible in the accelerator infrastructure. In particular, remote tests of electronic equipment, software and algorithms were done in level Chechia thermostat test stand and in multicavity (containing a single criomodule without a RG gun) MTS test hall.

Climatic tests were designed and performed, as well as exploitation tests of optical fiber, stabilized frequency distribution system. The necessity to apply such a system stems from large dimensions of accelerator infrastructure, of the order of hundred m and even km. The control, measurement and safety electronics is distributed over distant locations around the engineering infrastructure. There were carried interferometric measurements with the usage of optical phase shifter. The methods to lower the phase noise of master oscillator were introduced. The necessary phase stability was initially below 1ps for the distance of a few km. Next this value was lowered to 100fs and 1ps for analysis time 100ms and 1000s respectively.

There were carried out works on accelerator control software including a single cavity and simultaneously multiple cavities control. The control algorithm based on mechano-electrical model of the cavity, written in MatLab. This algorithm was subject to optimization. The software was written for work and communication management of the system. User panels were prepared for operator communications. The user system bases on DOOCS – a virtual control environments used popularly in DESY. Other cavity control tests were performed with the usage of alternative solutions like EPICS and simply web browser.

An algorithm was designed and implemented based on object identification and a full closed feedback loop. Tests of remote control of cavities mounted in the MTS in DESY from the VME system mounted in ISE WUT were successfully done. A sub-system of auto-calibration was proposed during multichannel vector sum control. A module with eight cavities is controlled with closed feedback loop (FB). The FB is enhanced by the direct control feed forward (FF). The FPGA controller performs procedure in agreement with the data in the set point tables: Feed-Forward, Set-Point, Corrector-Unit. Nonlinearities and deterministic perturbations are compensated by FF table for open loop. The correction for closed loop (tuning) is performed by Complex Gain module in the Correction Block. The circuit contains also a module for klystron linearization. The amplification in the loop was around 300 during ACC1 control of FLASH accelerator. The adaptive control algorithm was applied for FB and FF

work modes, according to process identification. The applied control method for a crio-module is useful for repeatable deterministic work conditions of the accelerator, what was confirmed experimentally. The following parameters of field stabilization in the cavity were obtained: relative amplitude stabilization - 10^{-4}, phase stabilization approx. 2 10^{-4} rad.

New control algorithms were elaborated for superconducting cavities and for the RF gun. The algorithms use fully the extended hardware resources of the new, third generation of LLRF control system SIMCON. In the control system for power generated by the RF gun the following stabilities were obtained during the first run: phase – better than 1°, amplitude – better than 1%. The control system of the RF gun can not measure the field value in the gun cavity. The SIMCON system, consisting of hardware and software, was used for the control and work stabilization of the RF gun in FLASH accelerator for FEL operation of SASE type. For a pulse of the length which is not bigger than 100μs, it is enough to use a common PI controller. Longer RF pulses, and repetitive sets of pulses require additionally the AFF control (adaptive feed forward). SIMCON system of the third generation was used to control a few versions of the RF gun. The control software for RF gun was written in VHDL and implemented in FPGA circuit of fit resources. A DOOCS server was prepared for the hardware-software control system. An improved stability of phase and amplitude was achieved. There were performed tests of superconducting linac control, via introduction of field gradient dispersion between different cavities. The measurement results show that, via fitting of the loaded Q and phase of the input signal at preset detuning of the cavities during the pulse, the gradient dispersion in cavities of the order of 5% may be tolerated. In a system with arbitrary programmable, nonstationary, changing during the high power RF pulse, detuning of an individual cavity, the gradient dispersion may be fully compensated. Then, each cavity works with its own maximum of the field gradient. The following conditions: available maximum power, range of loaded Q tuning in a cavity, phase of forward wave and cavity detuning, impose a boundary value for the gradient dispersion which is equal to ±20%. A few % of the average gradient value is lost due to mechanical resonances of the cavities.

The chosen architecture version of the individual loop in the LLRF control system has a large influence on the costs. Such a loop is replicated in the system hundred times. During the whole process of the LLRF system design there were considered all the time the possibilities to reduce the overall costs. There are a few ways to reduce the costs. The ultimate confinement for system cost reduction is considerable narrowing of the control system functionalities, which cannot be accepted by the machine operators. One of the cost reduction paths leads via the functional integration of the whole LLRF system circuitry. The frequency down-converters may be possibly integrated with the ADC circuits and with the initial data processing in the FPGA, where a value for a partial vector sum is calculated. Such an input (front-end) signal processing circuit from the cavity may be positioned close to the individual criomodules, as so called front-end RF patch panel. This circuit is then connected with the rest of the system

by means of gigabit optical links. This architecture eliminates expensive connections with the usage of thermally stabilized RF cables of very high quality.

Other potential methods to lower the LLRF system costs are: automation of control procedures in the widest extent by using of state machines and knowledge data base about the system; application of industrial standards for hardware and software; reduction of the number of separately transmitted signals; as deep system digitalization as possible; standardized signal multiplexing all over the system; design and testing of the system in own laboratory. One of the important system cost reduction options, with a clear look out to the future, is exchange of the now prevailing VME standard with the new one ATCA (or its version μTCA). ATCA system is now widely introduced in the telecommunications.

An important component in the design and testing process of the complex LLRF system is evaluation of its reliability. The reliability of an individual RF station was evaluated. This evaluation was extended to the whole LLRF system. The following assumptions were taken: MTBF for a single VME crate is around 10^5, the station consists of 10 crates, the system consists of a few tens of stations, a few crates in each station is critical for system performance, the system is redundant on the level of hardware and control method (algorithm), with the possibility to switch the work mode between closed loop and direct deterministic control. At these assumptions, it is estimated that a single LLRF system halt caused by either hardware or software breakdown appears less than one time during a year.

The third generation LLRF system SIMCON, ver.3 bases on double FPGA circuit (Fig.2.). One of FPGA chips fulfills logical functions while the other does system tasks (performs the control algorithm). The FPGA chips have inbuilt DSP blocks to facilitate fast floating point calculations. The VME 6U PCB has 10 wideband analog inputs 270MHz, ± 1V, fit to 50 Ω, 4 DAC channels, a few trigger I/O ports, reference clock input of the stability 0,3ps rms, static and dynamic memories, interfaces for optical gigalinks, and Ethernet input. The system was tested in Chechia horizontal thermostat and in the TTF accelerator. There were prepared cotrol algorithms in MatLab and DOOCS server for the SIMCON 3. VHDL firmware was written for SIMCON 3 PCB. The results of control and measurements of phase and amplitude for eight cavities of ACC1 criomodule of FLASH laser, obtained with the aid of SIMCON 3, were presented in Fig.3. A separate board of optical gigabit signal concentrator for SIMCON system was designed, manufactured and tested. The board called SIMCON 4 has eight optical giga links and Flash memory card pocket for storing the configuration data. It enables the system to work stand alone. The data concentrator serves for data exchange between other SIMCON boards and VME controller. This board was tested as a candidate for a universal node of a distributed, multichannel control measurement version of the LLRF system.

The LLRF control system for accelerator is under further development. This development is now continued in the EuCARD project in the following directions: optimization of software and algorithms, modularization of the software for low-level LLA

and high-level HLA parts, introduction of coupled FPGA-DSP processors, change to standardized, intelligent, telecom grade, system platform ATCA with the IPMS functionality.

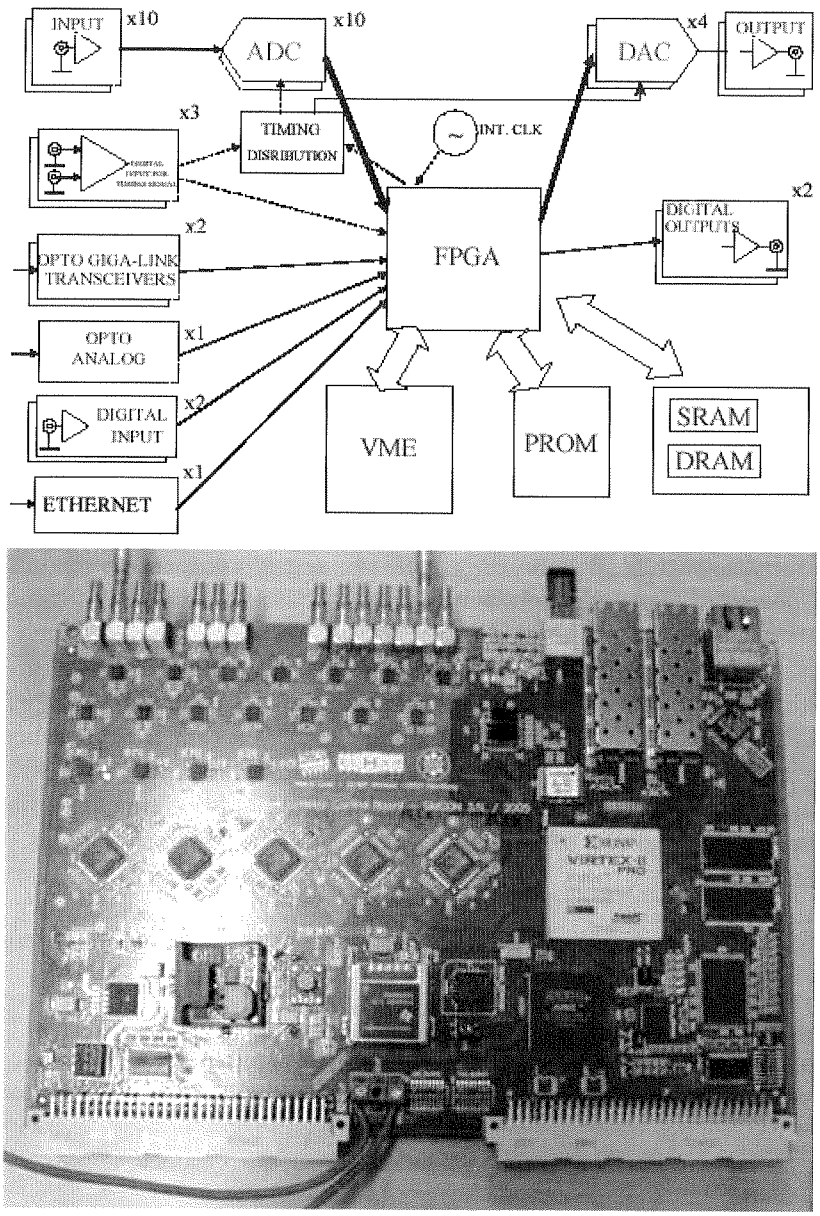


Fig. 2. SIMCON 3.1. PCB and its block diagram

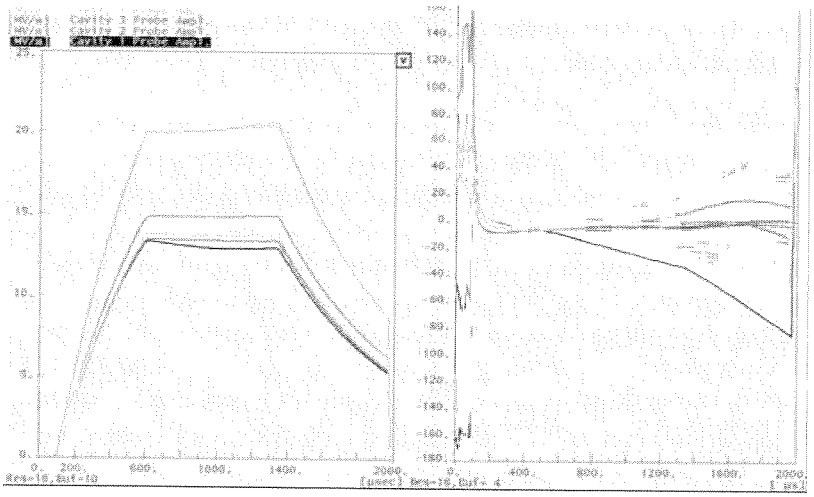


Fig. 3. Amplitude and phase of signals in eight cavities of ACC1 module of FLASH accelerator

5. ROLE AND EXPECTATIONS FOR EuCARD PROJECT

The EuCARD Project belongs to a group related to building a large, common European Research Infrastructure ERA. The expectations of the European Commission is that this infrastructure will be, in a positive sense, competitive against to other regions, especially Japan and USA. These ambitious assumptions are moderated by the current conditions of the world wide crisis. These economic conditions may influence the pace of realization of the biggest accelerator projects requiring the aggregation of financing from many various sources. Participation in such a big pan-European project like EuCARD is a large challenge and simultaneously is a chance for smaller research teams, especially from countries not possessing an own, large accelerator infrastructure. One of these chances is to overcome a critical threshold of building own infrastructure, leading directly to extension of industrial applications of new technologies.

6. REFERENCES

1. EC FP7 [http://cordis.europa.eu/fp7/home_en.html]
2. KPK [http://www.kpk.gov.pl]
3. UPK EPB PW [http://www.pw.edu.pl/Uczelnia/Struktura-uczelni-Sklad-osobowy/Jednostki-organizacyjne-podlegle-Rektorowi/Centrum-Wspolpracy-Miedzynarodowej/Uczelniany-Punkt-Kontaktowy-Europejskich-Programow-Badawczych-UPK]
4. KRAB [http://www.if.pw.edu.pl/~krab]
5. ESGARD http://esgard.lal.in2p3.fr/
6. CARE Project [http://care.lal.in2p3.fr]
7. CARE 08 Annual Meeting [http://indico.cern.ch/conferenceDisplay.py?confId=36153]
8. CARE Publications [http://care.lal.in2p3.fr/Publications]
9. EuCARD [https://eucard.web.cern.ch/EuCARD]

10. EuCARD kick-off meeting, CERN, 05.12.2008, [http://indico.cern.ch/sessionDisplay.py?sessionId=16&slotId=0&confId=36153#2008-12-05]
11. EuCARD 1st Governing Board Meeting, CERN, 01.04.2009, (CERN NICE account login required) [http://indico.cern.ch/conferenceDisplay.py?confId=54248]
12. EuCARD 1st Steering Committee Meeting, 03.04.2009, (CERN NICE account login required) [http://indico.cern.ch/conferenceDisplay.py?confId=55073]
13. CARE JRA SRF DESY [http://jra-srf.desy.de]
14. EuCARD WP10 SRF Superconducting Radio Frequency kick-off Meeting, DESY, 24.03.2009, [https://indico.desy.de/conferenceDisplay.py?confId=1879]
15. FLASH [http://flash.desy.de] Free Electron Laser, Hamburg
16. Photon Science [http://hasylab.desy.de/facilities/flash/index_eng.html]
17. ISIS [http://www.isis.rl.ac.uk] European Spallation Neutron Source
18. PHIN [http://www.infn.it/phin] Photon Injectors
19. HIPPI [http://mgt-hippi.web.cern.ch/mgt-hippi] High Intensity Pulsed Proton Injectors
20. NED [http://lt.tnw.utwente.nl/research/HCS/Projects/CARE-NED] Next European Dipole
21. ELAN [http://elan.desy.de] Electron Linear Accelerator Network, Superconducting Linac Technology
22. BENE [http://bene.web.cern.ch/bene] Beams for European Neutrino Experiments
23. HHH [http://care-hhh.web.cern.ch/care-hhh] High Energy, High Intensity, Hadron Beams
24. EuroTeV [http://www.eurotev.org] European Design Study towards a Global TeV Linear Colllider
25. EuroLEAP [http://www.laser-electron-acceleration-plasma.eu] European Laser Electron controlled Acceleration in Plasmas to GeV energy range
26. EURISOL [http://www.ganil.fr/eurisol] European Isotope Separation On-Line
27. CLIC [http://clic-study.web.cern.ch/clic-study] Compact Linear Collider
28. SLHC [http://info-slhc-pp.web.cern.ch/info-slhc-pp] Large Hadron Collider upgrade
29. NEU2012 [http://bene.web.cern.ch/bene/NEU2012.htm] Neutrino Beams for Europe in 2012
30. MICE [http://mice.iit.edu] Muon Ionization Cooling Experiment
31. RAL STFC [http://www.scitech.ac.uk] Rutherford Appleton Lab
32. ALBA [http://www.cells.es] Synchrotron Barcelona
33. CBM [http://www.gsi.de/fair/experiments/CBM] Compressed Barionic Matter
34. FAIR [http://www.gsi.de/fair] Antiproton and Ion Research
35. E-XFEL [http://xfel.desy.de] European X-Ray Laser

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 515–524

Helical Resonator for Measurements of Parameters of Dielectrics at 25 MHz

MATEUSZ ŻUKOCIŃSKI*, ADAM ABRAMOWICZ^

Warsaw University of Technology
Institute of Electronic Systems
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
**M.Zukocinski@elka.pw.edu.pl*
^aabr@ise.pw.edu.pl

Received 2009.06.09
Authorized 2009.08.10

This paper presents a helical resonator for measurement of complex permittivity of dielectric materials in form of thin sheets at 25 MHz. The indirect method of permittivity reconstruction is investigated. The full-wave 3D electromagnetic simulator is used to analyse the structure, create experiment model and build lookup database. The results of experiments are used to verify the possibility of accurate resonant measurement method at such low frequency.

Keywords: measurements of dielectrics, helical resonator, resonant frequency, quality factor

1. INTRODUCTION

The need of accurate characterisation of dielectrics is one of the key points of the electromagnetic metrology. The knowledge of complex permittivity $\varepsilon_r = \varepsilon_r' (1 - jtg\delta)$ and its changes caused by elusive effects appearing in development and manufacturing process is critical in quality control, optimization of material composition and production process of electronic materials [1]. Capacitive and transmission line methods used today give accuracy of several percent at most [1] what becomes insufficient in modern material applications. That is why this project is focused on development of a new accurate method for permittivity measurement of thin dielectric sheets like PCB substrates.

The choice of frequency of presented resonator was imposed by lack of sufficient measurement methods in range of 10 – 50 MHz. Furthermore signals of 20 – 30 MHz

are intensively used in many areas of science and technology like electronics, medicine, biology, geoscience and physics.

The resonant methods provide the best accuracy of dielectric measurements as can be seen in [2,3]. However, this accuracy greatly depends on the quality factor therefore these methods require high Q resonant structures. This can be accomplished in RF and microwave resonators easily but it is difficult to build a high quality resonator in the several MHz frequencies region. For these reasons we have investigated the properties of helical resonators as the measurement device [4]. The helical resonators were applied for many decades and are still considered to be a preferred solution in lower bands of communications as resonators and filters [5,9]. In these frequency bands they are relatively small and at the same time have relatively high value of unloaded Q-factor.

The indirect method of permittivity reconstruction (also known as retromodelling) is used here. Development of indirect material measurement techniques has become possible thanks to the fast growth in electromagnetic simulation software and in computer clock speed in recent years [6,7]. In this method a database (lookup table or lookup chart) is created by multiple electromagnetic simulations of the measurement system with different dielectric specimens of pre-defined geometry and distinct dielectric parameters. Then the best match is sought for an actually measured response of real-life sample. In other words the dielectric parameters are reconstructed by comparison of measured and simulated S-parameter curves. The advantage of this kind of method is that the one does not need to know field expressions for particular structure to find measured quantities. This fact opens up prospect of use in measurements structures which accurate solutions remain unknown.

2. HELICAL RESONATOR DESIGN

Considered here helical resonator is a quarterwave helical line, short-circuited at one end, placed in a cylindrical metal housing (see Fig. 1 and 2). At the beginning the helical resonator of the resonant frequency of the dominant mode equal to 25 MHz has been designed on basis of well-known in literature experimental equations [8,9]. Those equations concern the case of helical resonator of special proportions, with shields of square or circular cross-sections, which allow yield the highest Q for the given resonant frequency and shield size. The obtained value of quality factor, using copper as the material of the helix and the shield, has been about 2900. Then, this designed structure has been analysed in an electromagnetic FDTD simulator to search and understand properties of such helical resonator. The structure shown in Fig. 1 has been analysed using QuickWave-3D – three dimensional electromagnetic simulator based on FDTD method [10]. Those simulations have also served for optimization of the resonator in order to obtain efficient and simple structure for measuring dielectric properties. Simulations have shown that the E-field strength is maximum at the open-end of helix, between the helix and the cover. This fact guaranties that when dielectric specimen in

the form of sheet is placed in this gap, the structure sensitivity to dielectric properties of the specimen will be high.

Results of computer analysis have proven high usability and efficiency of the considered structure of helical resonator and spurred further works [11]. A prototype resonator has been built. It is presented in Fig. 1, 2a and 3. The helical line has been made out of brass tube and later on plated with silver to increase quality factor by reducing skin effect losses. Cylindrical shield of the resonator is made from copper. Transmission coupling is done by two simple antennas made as wire sections soldered to N-type connectors in the shield. PTFT supports have been used to minimize mechanical vibrations of the helical. The outer dimensions of the resonator are: diameter D = 292 mm, height H = 326 mm. The sample of dielectric sheet (Fig. 2b) has got circular shape with diameter equal to the diameter of resonator shield, and thickness t = 2 mm which corresponds to typical thickness of PCB substrates available on the market.

3. MEASUREMENTS

The prototype resonator has been submitted to several experiments in set-up presented in Fig. 3. The resonator has been tested using Vector Network Analyser, resonant frequencies and 3dB bandwidths (Q-factor) have been investigated under different conditions. The measurements of a set of samples of different dielectrics and PCB substrates have been done as well. Those measurements allowed to analyse properties of the real structure, to find its weaknesses and then to introduce appropriate improvements like silver plating of the helical line, PTFT supports or coupling adjustments. Sensitivity to the dimension of the air gap between open end of the helical and the lid, where samples are placed, has been observed. This gap size (nominally 2 mm) is subject to mechanical changes. Measurement of this gap is the main factor limiting calibration accuracy and accuracy of the whole ε_r' extraction method as it is explained later in this paper.

The resonator is quarterwave type and the infinite number of only odd resonances exist in its frequency response. Figure 4 shows the transmission characteristic $|S_{21}|$ of the fundamental mode of helical resonator with PTFT sample inside. The measured quality factor with PTFT is $Q_{PTFT} = 2017$ at $f_r \approx 25$ MHz, and for the empty resonator $Q_{empty} = 2150$ at $f_{r1} \approx 26.2$ MHz. Next modes of empty resonator appear at $f_{r2} \approx 88$ MHz with $Q \approx 2000$, $f_{r3} \approx 141$ MHz with $Q \approx 1300$, $f_{r4} \approx 191$ MHz with $Q \approx 1000$.

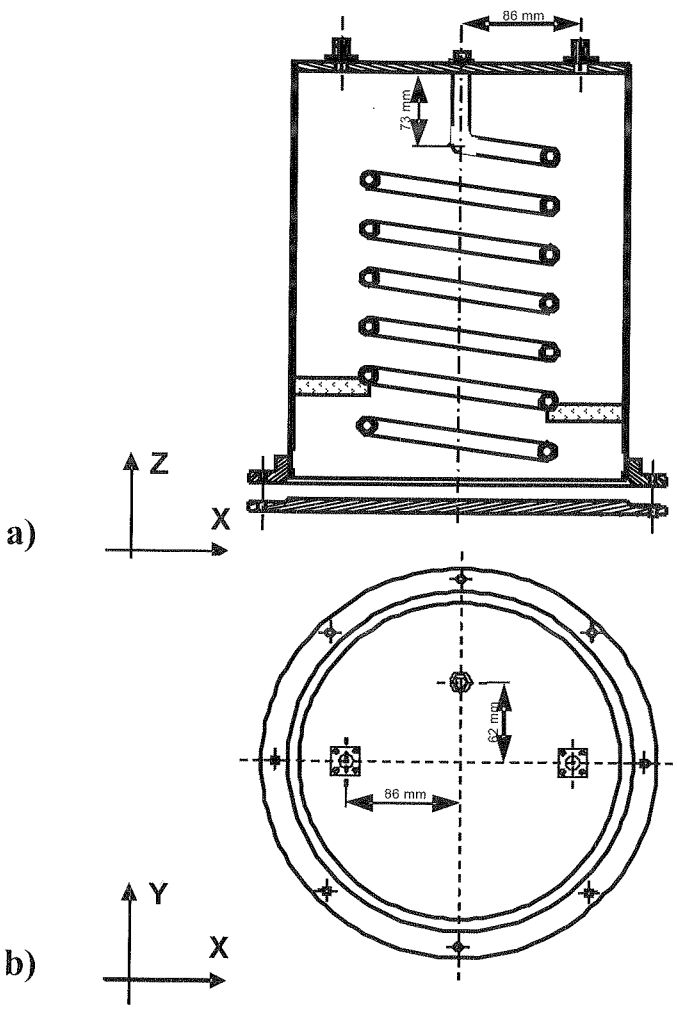
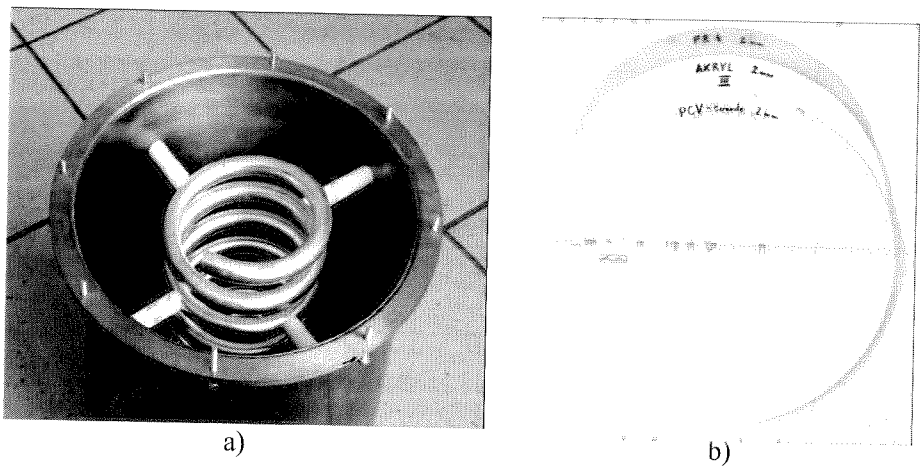


Fig. 1. Assembly drawing of helical resonator structure: a) vertical cross-section; b) cylinder lid side



a) b)

Fig. 2. a) Resonator inside, view from open end of helical where dielectric samples are placed. White PTFT supports limit mechanical vibrations of the structure. Helical line is silver plated; b) Dielectric sheet samples (diameter of each circular sample is comparable to resonator cylinder dimension)

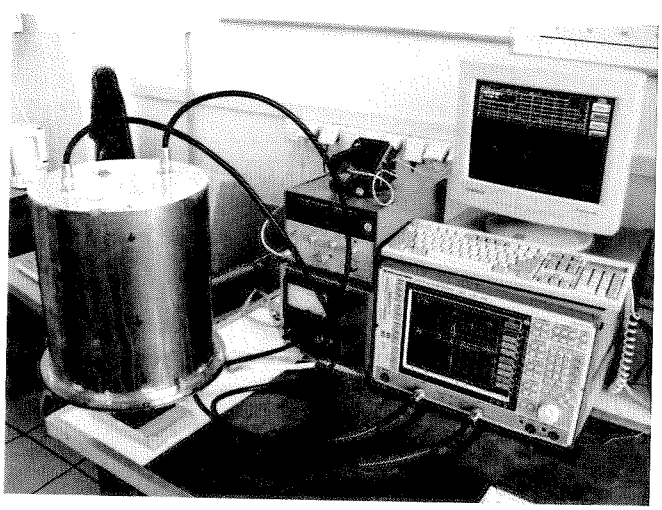


Fig. 3. Helical resonator prototype in measurement set-up. The resonator is coupled to network analyser Rohde&Schwarz ZVRE 9 kHz ÷ 4 GHz

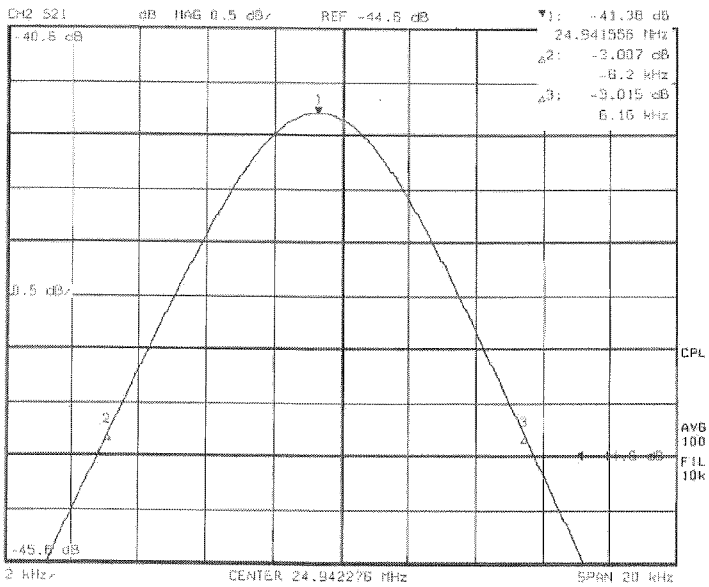


Fig. 4. Fundamental mode of the helical resonator with PTFT sample inside (f_r = 24.941556 MHz, Q = 2017)

4. STRUCTURE ANALYSIS

To reflect real-life structure of the prototype accurate computer model of resonator has been built and used in FDTD QuickWave-3D simulator. New dimensions as well as all crucial details of the prototype have been included. Additionally, the model has been tested and optimized for the best accuracy and speed of calculations. Field distributions of resonant modes have been calculated and optimal place for the sample, where high electric field density exists, have been recognized [11,12]. Appropriate FDTD mesh size and fine mesh volumes in regions with high gradients of electromagnetic field (open end of helical and sample regions) have been set up.

For such a model several series of simulations have been performed. Comparison of results from measurements and from simulations allowed us to verify the model. Simulations done for samples of 2 mm thickness and series of ε_r' values have given the resonant frequency shift chart (Fig. 5). The frequency decreases from 26.2 MHz in empty resonator (sample of ε_r' = 1) to 22.6 MHz for ε_r' = 10. The total frequency change, in this range of permittivity, is 3.6 MHz and this relation is non linear. Calculation of derivative from Fig. 5 has shown that the sensitivity of the resonator to 1% change of sample ε_r' is 19.0 kHz for ε_r' = 1, and 17.3 kHz for ε_r' = 10. This result assures high accuracy of resonant frequency determination.

Similar analysis has been done for samples of different thickness. The aim was to find relation that makes it possible to include sample thickness in calculations of ε_r'. Figure 6 presents a correction factor chart that describes influence of sample thickness on the result of resonant frequency measurement.

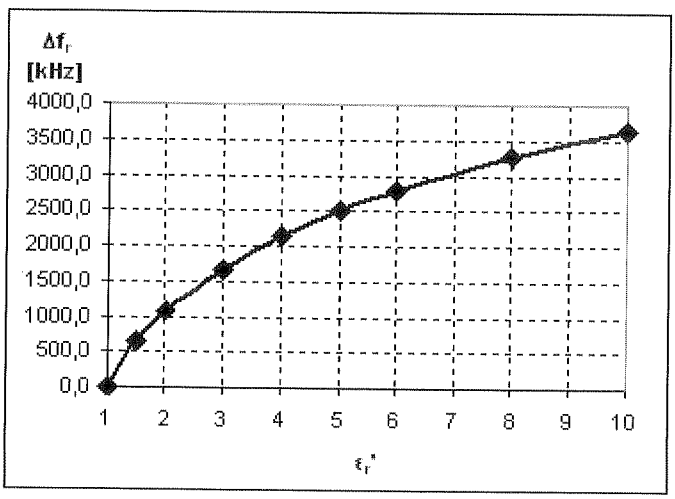


Fig. 5. Resonant frequency shift curve (related to empty resonator, $\varepsilon_r' = 1$)

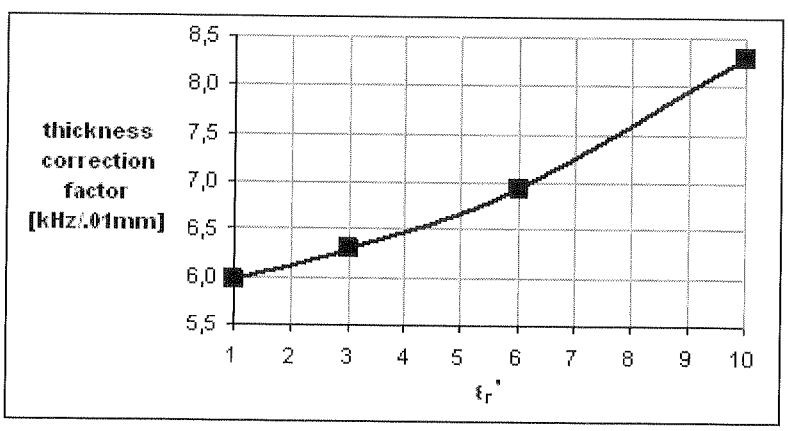


Fig. 6. Sample thickness correction factor with respect to 2.0 mm thick sample

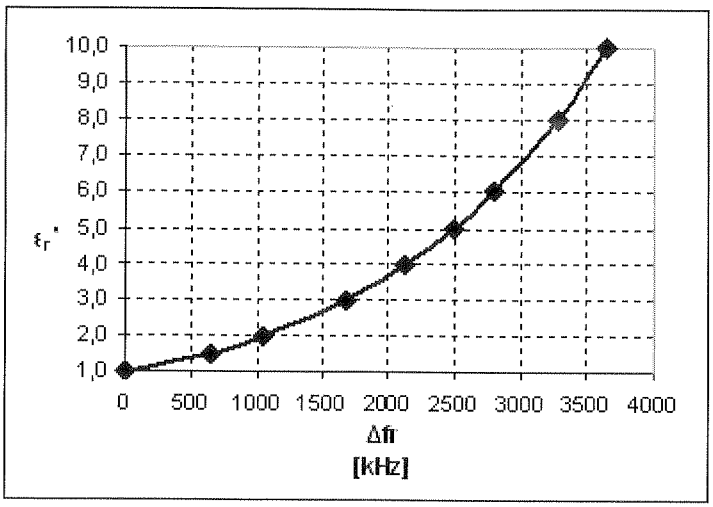


Fig. 7. Lookup curve for extraction of ε_r' (with respect to empty resonator, $\varepsilon_r' = 1$)

The inverse of relation from Fig. 5 is a measurement lookup chart shown in Fig. 7. For discrete simulation points of relative permittivity and frequency shift an approximation was applied giving following convenient exponential formula:

$$\varepsilon_r' = \exp\left(\frac{\Delta f_r + 6,238}{1573,9}\right) \tag{1}$$

where: ε_r' – relative dielectric permittivity, Δf_r – shift of resonant frequency with respect to empty resonator.

Frequency shift Δf_r can be calculated having a result of resonant frequency measurement of a real sample (f_r), a value of thickness correction factor (Δf_{corr} from Fig. 6) and calibration measurement (f_{r0}):

$$\Delta f_r = f_{r0} - f_r - \Delta f_{corr} \tag{2}$$

Calibration means a measurement of resonant frequency of the resonator with sample of known permittivity and thickness. Here, sample of $\varepsilon_r' = 1$ and thickness t = 2 mm have been assumed as fiducial sample for calibration.

The accuracy of the method is considered as a sum of partial uncertainties. The most important is calibration measurement which, assuming 0.05mm error in air gap thickness measurement (accuracy of the vernier calliper), gives ~1.5% of uncertainty. Next, the sample sheet thickness measurement error of 0.01mm (typical value given in datasheets of PCB substrate manufacturers) results in 0,3% for $\varepsilon_r' = 1$ to 0,5% in case of $\varepsilon_r' = 10$. Inaccuracy of resonant frequency measurement done by VNA is less than 0.1%. Approximation error of formula (1) given above is less than 0.1%. Influence of sample diameter is negligible. Errors of simulation and computer modelling are

cancelled by the calibration and do not add into account. According to that analysis the total accuracy of the presented method is in the range of ~2% for $\varepsilon_r' = 1$ and ~2,2% for $\varepsilon_r' = 10$. As can be seen the main factor limiting this performance is calibration (measurement of air gap between open end of the helical and the lid).

Due to high Q = 2150 the presented structure can be applied for extraction of loss factor $tg\delta$ as well. Accuracy of that measurement results from the value of quality factor and can be estimated as $\pm 5x10^{-5}$.

5. CONCLUSIONS

The helical resonator operating at ~25 MHz and of Q = 2150 for measuring dielectric properties of thin dielectric sheets has been presented. This type of resonator has got the advantage of small dimensions compared to the wavelength and high unloaded quality factor. Experiments and analysis allowed to determine properties of considered resonator and propose method of ε_r' extraction based on simulation technique. Helical structure presented here fills the gap in dielectric properties measurement methods which exists between 10 and 50 MHz. The accuracy predicted on the basis of computer analysis is ~2% in the whole considered range of permittivity ($\varepsilon_r' = 1 \div 10$). Further improvements, mainly in calibration technique and geometry of the structure can increase accuracy to the range of ~1% or less. High Q of the resonator allows measurements of loss factor as well. Consequently, it provides attractive opportunity for measurements of complex permittivity $\varepsilon_r = \varepsilon_r' (1 - jtg\delta)$ of materials of low and medium ε_r' and medium value of loss factor $tg\delta$. Proposed resonator can be redesigned and optimized to accommodate different type of samples or to do measurements of other properties, e.g. magnetic. All these results achieved in the prototype already makes this structure competitive to traditional measurement set-ups and proves effectiveness of resonant method for measurements of dielectric properties at low frequencies.

6. REFERENCES

1. Baker-Jarvis J., Janezic M. D., Riddle B., Holloway Ch. L., Paulter N. G., Blendell J. E.: Dielectric and Conductor-Loss Characterization and Measurements on Electronic Packaging Materials. National Institute of Standards and Technology, Technical Note 1520, July 2001
2. J. Krupka, A. Abramowicz, K. Derzakowski: Precise Measurements of The Complex Permittivity at Microwave Frequencies. Proc. X Union of Radio Science Symposium URSI 2002, pp. 345-353, Poznań, Poland, 2002
3. J. Krupka, R. G. Geyer, J. Baker-Jarvis, J. Ceremuga: Measurements of The Complex Permittivity of Microwave Circuit Board Substrates Using Split Dielectric Resonator and Reentrant Cavity Techniques. DMMA'96 Conference, Bath, UK, 23-26 Sept. 1996
4. W. Meyer: Helical Resonators for Measuring Dielectric Properties of Materials. IEEE Trans. on Microwave Theory Tech., vol. MTT-29, pp. 240-247, March 1981
5. J. Antoniuk, M. Żukociński, A. Abramowicz, W. Gwarek: Resonant Frequencies of Helical Resonators. Proc. Int. Conf. MIKON 2004, pp. 1044-1047, Warsaw, Poland 2004

6. Agilent Technologies: Agilent 85071E Materials Measurement Software. http://cp.literature.agilent.com/litweb/pdf/5988-9472EN.pdf
7. B. Wäppling-Raaholt, P.O. Risman: Permittivity Determination of Inhomogeneous Food Items by Retromodelling with a Degenerate Mode Cavity. Proc. of 9th Int. Conf. On Microwave and High Frequency Heating, Loughborough, UK, September 2003
8. Z. Bieńkowski: Poradnik ultrakrótkofalowca. WKiŁ, pp. 222-228, Warszawa 1988
9. P. Vizmuller: RF Design Guide: Systems, Circuits, and Equations. Artech House, pp. 239-241, 1995
10. QWED (http://www.qwed.eu): QuickWave-3D Electromagnetic Simulator V.5.0 Manual. Warsaw 2005
11. M. Żukociński, A. Abramowicz: Design of a Helical Resonator for Measurement of Dielectric Properties of Materials at 21 MHz. XI National Symposium of Radio Sciences, URSI2005, Poznań, 2005
12. R. Diehl, D. Wheatley, T. Castner: The Electromagnetic Modes of a Helical Resonator. Review of Scientific Instrument, No.11, pp. 3904-3906, Nov. 1996

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2009, 55, no 3, pp. 525–538

Photonics and Web Engineering: WILGA 2009

RYSZARD S. ROMANIUK

Institute of Electronic Systems,
Warsaw University of Technology
R.Romaniuk@ise.pw.edu.pl

Received 2009.06.15
Authorized 2009.08.18

The paper is a digest of work presented during a cyclic Ph.D. student symposium on Photonics and Web Engineering WILGA 2009. Symposium is organized by ISE PW in cooperation with professional organizations IEEE, SPIE, PSP and KEiT PAN. There are presented mainly Ph.D. and M.Sc. theses as well as achievements of young researchers. These papers, presented in such a big number, more than 250 in some years, are in certain sense a good digest of the condition of academic research capabilities in this branch of science and technology. The undertaken research subjects for Ph.D. theses in electronics is determined by the interest and research capacity (financial, laboratory and intellectual) of the young researchers and their tutors. Basically, the condition of academic electronics research depends on financing coming from applications areas. During Wilga 2009 there were organized, and thus the paper debates, the following topical sessions concerning applications of advanced electronics and photonics systems: merging of electronic systems and photonics, Internet engineering, distributed measurement systems, security in information technology, astronomy and space technology, HEP experiments, environment protection, image processing and biometry. The paper contains also more general remarks concerning the workshops organized by and for the Ph.D. students in advanced photonics and electronics systems.

Keywords: photonics, optoelectronics, Internet engineering, web engineering, advanced electronics systems, WILGA Symposium, IEEE, SPIE, Photonics Society of Poland

1. INTRODUCTION

Institute of Electronic Systems of Warsaw University of Technology (ISE PW) [1] has been organizing, for over a decade, an interdisciplinary Ph.D. student Symposium, which is known in this country and internationally as WILGA [2]. WILGA is organized two times a year since 1998, in January and May editions. The place is a small village near Warsaw on the Vistula river. There is a holiday resort owned by WUT. The Sym-

posium has gathered together more than 3500 Ph.D. and M.Sc. students. There were published more than 1500 research papers in more than 10 volumes of Proceedings of SPIE [3] in the USA and several volumes of special issues of professional journals including Electronics Monthly [4], Electronics and Telecommunications Quarterly by PAN [5], Measurement Systems and Technology by IOP. The WILGA Symposium is attended by representatives of nearly all technical universities in this country (from faculties of electronics, information technology, electrical engineering, mechatronics, technical physics) and Ph.D. students from physics faculties of universities. Each year the WILGA Symposium is attended by young researchers from abroad.

During the last decade, the WILGA Symposium on Photonics Applications has established a meaningful position on the domestic market of academic meetings. It is a pretext for young scientists for comparisons of research work conditions in different parts of the country and abroad. In WILGA there are conditions to undertake a more general discussion on the condition of particular disciplines or technical research and perspectives of their further development. These perspectives are determined by financial conditions, application requirements, and a focused interest to develop a new technology. WILGA is a sensitive mirror showing the status of a certain segment of the 'young science' on the international background. This WILGA barometer shows quite precisely a lot of positive and negative processes in this domain. There is observed a period of strong reconstruction in the area of technical sciences in this country and internationally. The most gifted young people do not choose technical sciences as a subject and aim for their professional career.

A number of research branches, and especially technical sciences, are developing in an interdisciplinary way. This interdisciplinary path concerns also, and particularly, a number of realized Ph.D. theses at electronics, photonics and their applications, as well as related disciplines. Particular neighboring disciplines start to overlap, giving rise and/or opening new research areas. One of such relatively new research areas, combining a number of subjects like solid state physics, optics, mechanics and mechatronics, electronics, material engineering, chemistry, and other ones, are 'microsystems'. A research product, as well as more and more frequently an application product of this branch are objects of a general name including (integrated micro-electro-opto mechanical systems) [6] like MEMS, MOEMS, SOC, LOC, etc. Another example of such an interdisciplinary branch, which now is a separate field of research and industry is photonics. A similar example is mechatronics. Many, nearly all, of these products of these fields can not work without a complex multi-level programming. This programming is, on one hand, strictly combined with cooperating hardware, determining directly their functionality, while on the other hand, it is connected to the global communication network, which is in most cases the Internet.

An important branch of technology is, recently a subject to intense development, integration of mechanical, optical and electronic hardware and programming. A basic theoretical as well as practical question here is which functionalities to place in the electronic hardware and which in the software. A developmental tendency in this

domain is to position in the hardware, apart from the operating system, only the following basic resources needed for the hardware to operate: calculation power, memory and logics. The whole parameterized and configurable and, thus, flexible and scalable functionality is placed in the multilevel software in a few basic kinds of specialized microprocessors: GPP – general purpose, DSP – digital signal processing (floating points calculations), FPGA – behavioral modeling in VHDL, fixed point calculations, but also DSP, and basically logics. Such an electronic node of a system, of the aggregated processing power, fit to the application needs, is connected into a network by means of broadband wireless and/or optical fiber links. The node has (apart from TCP/IP) universal, industrial I/O ports, analog and digital.

The development of design of complex functional systems, serviced by advanced electronics, is observed in several fundamental directions. Two of these were observed in WILGA 2009 works. These concerned probably two separated poles of these application processes of the modern electronics. On one hand, a big number of very narrowly specialized, and economically tailored devices is designed, optimized for usage of confined resources. This is clearly observed in cheap gadgets which do not need any versatility but optimally fulfill the imposed tasks. On the other hand, the research interest concerns big systems of extended calculation potential. During a developmental stage of a large system, due to a fast decrease of unit costs of the resources, the design of advanced electronic systems are frequently quite redundant. The nontrivial task of optimization of usage of large calculation power pays only back in the case of serial applications, which is only rarely a case in academic research.

Below, there are presented some chosen subject groups moved in WILGA 2009 works. There were also discussed some more general development problems concerning electronic and photonic systems in the following aspects: hardware and software balance and integration, design and applications, further development in the nearest future.

2. ELECTRONICS AND INFORMATION TECHNOLOGIES

The increase in calculation resources and the increase in reliability is done, in certain sense, without any excess costs from the point of view of the system designer. The rule for hardware cost reduction, with the introduction of a new generation of microprocessors, is valid not only for GPP chips but also for DSP and FPGA ones. This rule includes also the unit costs of broadband communications between chips and PCBs. This opens a way to use in academic projects the components of much greater resources at nearly the same overall costs. Availability of increased resources allows for reaching more ambitious projects, possible for realization not only in virtual environments. Other considered design aspects of electronic systems during WILGA 2009 were: system ruggedness, resistance to adverse working environments, including ionizing radiation, temperature shocks (cosmic conditions), EMI and EM compatibility in,mixed analog and digital circuits, design and realization of advanced multilayer PCBs, tc.

Another important aspect in the development of electronic systems, which found a reflection in the works of WILGA 2009, is relatively new (or renewed after a shorter or longer absence) area of applications including: telemetric networks of large extent, intelligent sensory wireless networks – including self-configurability, specialist satellite communications, radar technology, SDR and RFID, biomedical engineering, automobile and airborne technologies, and in particular homeland security. In nearly all technologically advanced societies, the universities are actively involved in the research on the security systems. It concerns particularly the research on electronics security solutions. We all hope that the situation in this country will change in favor of academic research teams. The first steps have recently been done on the level of Ministry of Research and Higher Education [7]. This organization has announced recently two series of grants in the area of national security. It will soon be reflected in realization of Ph.D. work. The first work of this kind was presented in WILGA 2009 by WUT and MUT students. One of the works, presented during a session on measurement data categorization, presented by students from Italy concerned a safe thermal method of detection of landmines.

3. PHOTONICS, OPTICAL FIBER AND LASER TECHNOLOGIES

Photonics [8] is one of the most dynamically developing branches of technical sciences. Optical fiber technology is now a solid foundation of strictly standardized, optical, telecommunications, broadband, transport networks. This area is hardly available for a number of academic laboratories in this country. There is a research carried out and development work in many technical niches around optical fiber specialist telecommunications. An optical fiber network, with links stabilized thermally and mechanically, is a good medium for time distribution systems, reference phase and frequency. Such systems are a step towards the direction of building a technical infrastructure for optical clocks or optical reference rules. The aim is to build an optical clock system with accuracy surpassing the atom clocks. An optical fiber ultra-stable link for connection between two or more atom clocks is build by a team from AGH University in Kraków (dr P.Krehlik).

Active and sensory optical fibers and photonic materials are under development in several academic research centers in this country including: Białystok Uni. of Technology [9], UMCS Lublin [10], ITME Warsaw [11], AGH Kraków and WUT. The possibility and interest in research in this area is caused by several factors, and among them: availability of instrumentation – nontelecom optical fibers manufactured by several technological groups, relatively low price of such fibers, mastering of building of optical circuits from such instrumentation fibers, and first of all a large variety of such optical fibers. Broad range of possibilities of instrumentation fiber construction has not yet exhausted all ideas, despite a very broad literature in this domain, reaching thousands in numbers.

One of the most interesting presented solutions using instrumentation fibers concerned optical capillaries. It was a number of applications to build an integrated photonic and chemical laboratory on optical capillary. The work was presented by dr M.Borecki from WUT [12]. Among other measurements the following were presented: quality factors of milk, veterinary measurements in cows, food industry - quality of edible oils and alcohols, petrochemical industry – differentiation between and quality of petrol.

Laser technology is traditionally developed on the academic level in MUT, WUT, WrUT and a number of other centers. The research teams from MUT (and from IFPiLM) participate in a number of European programs concerning creation of large European research laser infrastructures like: HIPER [13] or ELI [14]. Several domestic laboratories is building tailored demonstrators of the relevant laser technologies. The role of these costly laboratory projects is to train experts, on the doctoral level as well as development and application research.

4. COMMUNICATIONS AND LOCAL NETWORKS

A number of academic laboratories is doing research in the field of passive, transparent optical networks PONs and in the area of cost effective solutions with multimode optical fibers. A multimode optical fiber transmission with modal groups multiplexing was a subject of a recent Ph.D. thesis in the Institute of Telecommunications of WYT (prof.J.Siuzdak). The effective multiplexing factor was two or three on a distance of several hundred m.

Optical fiber CATV systems are very cost effective. They provide proper bandwidth, even with multimode fibers, and thus transmission quality, and relevant transmission distance between numerous video signal sources and the video central distribution station. A development work is carried out in the direction of using different analog and digital modulation methods and optimization of application solutions for various technical work conditions. There were presented results from the Institute of Telecom of WUT on multichannel system using a multimode fiber working in the first transmission window 0f 850nm. A video FM signal was transmitted outside the baseband of the fiber.

5. INTERNET ENGINEERING

The global network is a magnificent platform for development of a variety of different access systems: networking, measurement, telemetric, safety, multimedia, etc. A strong observed development tendency is a cooperation between many systems on the platform of Internet like: distributed measurement networks, public measurement result information, GPS location, GIS data, urban traffic monitoring, and many more. The Internet engineering embraces hardware layer and mainly software layers. The latter consists of many sub-layers in agreement with the standard OSI model.

Self-configuring, distributed measurement networks, integrated with the internet and using GPS, GIS and other systems are predicted in the near future for versatile natural environment monitoring. A lot of miniature measurement sensors has to communicate with the network backbone via wireless communications using SDR standard but also via RFID technology. There are imposed very rigid requirements concerning the power supply and power usage confinements. There are used low voltage electrets batteries with zero current supply or micro nuclear ones with high current supply.

The Internet engineering, in popular understanding, embraces mainly the user layer. In reality, these are the following components: data, network, transmission, software and firmware, deposits and data bases, services, security. These components possess complex internal structure like: data kinds and structure, transmission and acquisition of data, data transmission and distribution network, network topology and reliability, diversification of access networks for users, hardware and its configuration, TCP/IP v6 services, measurement data transmission protocols, software and OSI layers, contents and management, offered services, kinds and configuration of services, applications connected with services, programming environments, multimedia, etc. The Internet engineering embraces such problems like: availability, best effort or QoS, mobility, interactivity, standardization, identification, anonymity, confidentiality, identity confirmation, integrity, security, continuous development, and e-technologies like e-science, e-work, e-banking, e-taxes, e-trade, e-business, e-voting, e-polls, e-fun, e-society.

WILGA 2009 Symposium featured a number of topical session devoted to different aspects of Internet engineering and to the systems which are built on the basis of the global network like specialist access networks and measurement networks (prof.W.Winiecki – Institute of Radioelectronics WUT [15] and prof. T.Adamski – Institute of Electronic Systems WUT).

6. SECURITY IN INFORMATION TECHNOLOGY

The main subject of research in the area of the IT security in academic groups is a formal analysis of the problem including specialist securities and cryptology. The researched theoretical and practical problems in this area are the following: updating the applications on the administrator level, configuration of security in operational systems and browsers, distinguishing of work between admin and user, observation of the most endangered ports: 9669, 52522, 59989, 36802, 48811, 21400,40821, 1990, 41174, 12288, hacking, phishing, cryptology, cryptography and cryptanalysis in the Internet, confidentiality, identification, authorization, authentication, integrity, proof with zero knowledge, cryptology attacks, sharing of secrets, coded books, coding and codes, stream and block codes, secure protocols, secure algorithms, RC4, MD5, SHA-1, DES, RSA, DSA, IDEA, PGP, elliptic curves, modulo arithmetic, keys, PKI, digital signature, symmetric and asymmetric cryptography, quantum cryptography, coding as armor, DRM, security infrastructure, secure logging OAuth, security in wireless networks, WEP, dependence of security on platform system, distribution of quantum

key, web security, weak sides of web sites, classification of threats – STRIDE model [16], design of a safe web application, firewall configuration, security tests, security audit of web, warning services of ISP, security organizations and certificates CISA, ISACA [17], ISSAA [18], IEEE CIST [19], Internet research for security.

There is realized a number of M.Sc. and Ph.D theses in ISE WUT in the area of cryptology and web application security. Some of the works are realized in cooperation with the Institute of Informatics of WUT [20] and with NASK [21].

7. ASTRONOMY, ASTROPARTICLE PHYSICS, SPACE TECHNOLOGIES

The students at WUT interested in cosmic technologies are organized in a few research circles. One of these organizations is Students Club of Cosmic Engineering [22]. There is an effective cooperation of these groups with CBK PAN [23] and MEiL WUT Faculty [24]. These groups participate in realization of extended international programs to build mini students satellites. A few of such satellites were launched to the Earth's orbit. There was performed a number of practical experiments concerning the flight trajectory, dropping of parcels from the orbit, optical observations, measurements, etc. All of these experiments required solid resources in the form of advanced electronic measurement and control systems. These systems had to fulfill all strict technical requirements to survive the cosmic conditions of work.

The Mars Society announces each year the University Rover Challenge [25]. This year's competitions of Mars robots took place at the end of May at Utah desert. One of the rovers, called Skarabeusz (The Scarab), was from the WUT prepared by the students in cooperation with PIAP. The rover consists of a versatile car and a gripper. An obstacle race consists of a difficult terrain path, taking probes of soil for analysis and helping a 'wounded astronaut'. Apart from mechanics, the key role in robot operation has an electronic control and automation system. The development work on this rover was presented by G.Kasprowicz, a Ph.D. student of ISE WUT.

A group of M.Sc. and Ph.D. students from WUT, WU and PAN takes part in the realization of an international project of wide angle optical observations of the whole sky. The name of the project is 'Pi-of-the-Sky' [26]. The aim of the project is to detect optical flashes accompanying the GRB effects [27]. During the project realization, it has turned out that the optical and electronic apparatus is very suitable for many other measurements like cataloguing of changing stars, observations of satellite paths, discovering of meteorites, cataloguing of space debris, etc. The experiment has telescopes localized in a few places around the globe, among them in the European Southern Observatory ESO in Chile [28]. In March 2008, the Pi-of-the-Sky experiment observed one of the largest GRBs originating from the distance around 7 bln light years. The observation results were published in Nature. A number of M.Sc. and Ph.D. students from ISE WUT are participating in the design and construction of ultra-low-noise and cooled CCD cameras with programming. The WILGA 2009 Symposium featured two

'cosmic technology' sessions (prof.L.Mankiewicz, prof.G.Wrochna, prof.F.Zarnecki). There were presented around 20 papers in these subjects.

Ph.D.Students from ISE WUT cooperate with the MPS Institute [29]. They participated in design, construction, testing and fabrication of an on-board, cosmic-grade version of a near infrared spectroscope SIR. This instrument was launched in an Indian space mission Chandrayaan-1 [30]. The satellite now circles around the Moon on a polar orbit. It makes precise spectral and geodesic measurements and mapping of the lunar surface. In this topical region a Ph.D. thesis of P.Sitek is under preparation.

Other region of cooperation with the MPS Institute concerns the matrix detectors for IR spectral region designed for cosmic conditions. These detectors, working in the spectral range of 0,8 – 2,5μm, are designed for Pamela project which will be realized during one of the planned low-orbit LEO [31] missions in a few years time.

8. ENVIRONMENT PROTECTION, MEASUREMENTS OF SURFACE WATER

There is realized a number of research projects in the ISE WUT concerning measurements of the quality factors of surface water. The projects are performed within European FP6 initiative. The research projects SEWING [32] (coordinated by ISE WUT, prof.A.Filipkowski) and WARMER [33] brought very valuable practical results. Part of these results are discussed in Electronics Monthly No 8/2009. A number of papers on SEWING and WARMER projects were published in Proc.SPIE. A practical working technology demonstrator for water measurements is under final tests. The system is an advanced measurement network with data fusion and processing. The system measures typical water parameters endangered with pollution, like alkaline salts and heavy metal ions. The projects results are traditionally demonstrated during WILGA Symposium (prof.L.Opalski, prof.J.Ogrodzki). The priority of these projects is to build cost effective automatic networking systems for warning against water pollution. Sewing and Warmer projects attracted a number of gifted M.Sc. and Ph.D. students in ISE WUT.

9. KNOWLEDGE DISCOVERY FROM MEDICAL DATA BASES AND AUTOMATIC DATA CATEGORIZATION

Large bases of data of various nature are expected to contain so called undiscovered knowledge. There are however several technical conditions on such data bases. The data has to be reliable, sound and good. These may be technological data concerning the production of electronic or photonic components, data about weather changes in the particular geographical region, water quality data, and biomedical data. Research on complex mechanisms of correlation in seemingly uncorrelated data with the usage of advanced statistical and analytical tools like SVM and knowledge discovery from data mines leads in medicine to completely new conclusions.

The basic tasks for the future in this research on knowledge discovery is systematic building of spacious, credible technical as well as medical databases. Not all available data bases now fulfill these criteria, despite seemingly highly ordered data form and high quality. A large hope for further development of this research direction particularly in medicine is building large data bases. Now, there is observed, a considerable development of data classification methods using multiparameter data sets. Proper data classification, sorting and correlation methods are basic tools in this field of data processing.

WILGA 2009 Symposium featured two big, well attended sessions on this subject (prof.J.Mulawka, dr S.Jankowski). The majority of papers presented during these sessions were delivered by two persons each: an M.D. and an engineer. M.D. person presented a medical problem, while an engineer presented a technical solution using data mining methods and obtained numerical results, Finally, these results were commented by an M.D. The medical problems were presented by the representatives of the following institutions: CZD and UM in Warsaw, CZMP in Łódź and Lublin Medical University.

10. IMAGE PROCESSING

One of the topical sessions in WILGA 2009 was devoted to the image processing in general theoretical, practical and computational aspects. Some aspects concerned the biometry -recognition and/or detection of a face, a palm of hand, an eye etc. A group of Ph.D. students for the Institute of Radioelectronics of WUT, under the guidance of prof W.Skarbek [15] presented a series of papers from these subjects creating a homogeneous picture of the development of this discipline of science and technology.

A research on automatic determination of pollen level is carried out in the ISE WUT. A three year grant was realized in this subject. A tutorial was presented by dr Z.Wawrzyniak with the project results. A measurement system was designed and constructed, consisting of the following parts: a device gathering the pollen grains in a standardized way, device for pollen microscopic image acquisition, multilevel software for image processing and pollen classification. The software classified pollen and calculated the number of pollen grains in each category. A decision was calculated for the pollen levels. The system was shown to be also practical for measurements and classification of other defined air pollutants.

11. RADAR TECHNOLOGY AND DIGITAL SIGNAL PROCESSING

WILGA Symposium is traditionally a multi-conference event. Every two years Wilga has a partner conference on Radar Technology and Digital Signal Processing organized by prof K.Kulpa of ISE WUT. Every other two years the Radar Symposium is organized as a large international event under the name Radar Week. The radar

conference was traditionally organized in Wilga in other neighboring resort. This year the Radar Symposium was organized in Jachranka, a resort near Warsaw on the Zegrzynski Lake. The proceedings of Wilga and Jachranka conferences are published in a common volume of Proc.SPIE.

12. EXPERIMENTS AND PROJECTS: LHC, CMS, E-XFEL, POLFEL, ILC

Young researchers from the ISE WUT take part in a few large international research experiments in the area of synchrotron radiation, high intensity high power lasers, high energy physics, photon physics, elementary particle physics, accelerator science and technology, nuclear technology, cosmic technology, etc. The M.Sc. and Ph.D. students of ISE WUT spend frequently fellowships in DESY [34] in Hamburg (at the machines FLASH [35] and E-XFEL [36]) in CERN [37] in Geneva (at the experiments LHC [38] and CMS [39]) in Fermilab [40] in Chicago (at the experiment ILC [41], in PSI [42] in Willingen. They participated in building Spanish national synchrotron ALBA [43] by CELLS consortium.

The young researchers from ISE WUT participate in DESY in building FLASH and E-XFEL lasers. They take part in the design, construction and tests as well as in the commissioning of electronic and photonic systems for the control of superconducting linear accelerator. This accelerator is a power supply in a form of an electron beam for the FEL laser. A construction of an E-XFEL clone is planned in Poland, in a smaller scale, under the name of POLFEL. The young experts from WUT are potentially a natural crew of the future POLFEL team.

The Ph.D.students from ISE WUT take part in CERN in realization of a few large experiments:CMS, LHC Interlock, Proton Synchrotron upgrade, etc. In thie area there are realized a few M.Sc and Ph.D theses. A few Ph.D.students stay in CERN as permanent residents.

In the frames of FP6 CARE and FP7 EuCARD [44] projects the Ph.D and M.Sc students from ISE WUT carry out research on electronic and photonic systems for accelerator technology, since several years.

13. PROFESSIONAL ORGANIZATIONS: IEEE-SPIE-PSP- KEIT PAN-SEP

The professional societies play a very important role in the academic life of students. A majority of serious professional organizations of national and international extent have students chapters. WILGA Symposium was bound from the very beginning with IEEE and SPIE and their Poland Sections. Now this role for SPIE Poland Chapter plays the Photonics Society of Poland – PSP [49], a continuator of this organization. PSP and SPIE sponsors the best student paper/presentation competition during WILGA.

The international professional organizations assure the possibility of the best possible publication place for WILGA Proceedings. The publications are in the data

bases like: SPIE Digital Library [spiedl.org], AIP – America Institute of Physics, IEEE eXplore, PSP Photonics Letters [50]. WILGA publishes its works in Proc.SPIE. Electronics Monthly by SEP, and ETQ by KEiT PAN.

14. WILGA 2009

The XXIVth WILGA 2009 Symposium was organized on 25-31 May in the WILGA resort owned by Warsaw University of Technology on Vistula river near Warsaw. The Symposium lasted traditionally the whole week from Monday till Sunday. A tendency for the various research teams to come to WILGA for only a short visit deepened from a year before. The teams are either from a region of representing a particular topic. Each day WILGA saw a little different set of participants, which can be seen on attached photographs from particular topical sessions. During WILGA 2009 there were presented around 200 presentations from many academic centers in this country: Warsaw, Gdańsk, Toruń, Kraków, Rzeszów, Białystok, Lublin, Łódź.

The organizers of WILGA Symposium are Ph.D. and M.Sc. students of ISE WUT. A very effective chairman of WILGA 2009 Organizing Committee was dr Maciek Linczuk. During certain years, the students form ISE WUT are supported in organization efforts by their colleagues, members of student chapters IEEE [45] and SPIE [46] at WUT and from other universities [47]. The symposium is under the patronage of the Alma Mater and domestic professional organizations: SEP, PAN, PSP. The symposium is supported in Europe, Middle East and in Africa by the SAC IEEE of Region 8 [48]. The effects of this support are frequently observed as WILGA is attended by very nice exotic guests from distant and neighboring countries alike.

SPIE best student presentation award. Traditionally, during the WILGA Symposium there is a student paper contest sponsored by SPIE. Last year, during WILGA 2008 the competition was judged by the eminent and high representatives of SPIE headquarters in Bellingham WA, USA. The participating persons were prof.B.Culshaw from Univ. Strathclyde and dr Emery Moore from Elmonics. The winners of SPIE WILGA 2009 student paper contest are M.Sc. students (Ph.D. students were not taken into account):

- First place – Łukasz Koniusz for the presentation 'Integration of astronomical telescope and weather station with 'Pi-of-the-sky' experiments astronomical system';
- Second place – Agnieszka Zagoździńska for the presentation 'Parametrization of the componants described in VHDL';
- Third place – Łukasz Dymanowski and Kamil Lewandowski for the presentation 'Universal platform for high speed digital signal processing';
- Fourth place – Paweł Drabik for the presentation 'Component Internal Interface framework';
- Fifth place – ex aequo Stefan Korolczuk for the presentation 'Radiation results of the SEE test of Xilinx XC3S400 FPGA instances'; and Michał Bohdanowicz for presentation 'Multichannel acquisition system with photomultipliers detectors'.

Three first places are accompanied with monetary awards, an annual access to SPIE digital library (confined number of downloads), free one-year-membership in SPIE and in PSP. All winners obtain diplomas signed by the SPIE President 2008 dr Kevin Harding and congratulation letter from the President of PSP prof. T.R.Woliński. A very skilful secretary of the WILGA 2009 Award Committee was dr Ryszard Kossowski of ISE WUT.

15. WILGA 2010

Symposium WILGA 2010 will be organized in two editions – January 29-31.01 and May 24-30.05, traditionally during the whole last week of May. Information about

WILGA 2009 Topical sessions and supervisors (l to r). 1 – Program WARMER – Water management in Europe; prof.A.Filipkowski, prof.L.Opalski, prof.J.Ogrodzki; 2 – Image Processing and Biometry; prof.W.Skarbek; 3 – Cosmic Technologies and Astronomical Program Pi-of-the-Sky; prof.L.Mankiewicz, prof.G.Wrochna, prof.F.Zarnecki; 4 – Security of Internet Measurement Systems; prof.W.Winiecki, prof.T.Adamski; 5 – Photonics – Optical Fiber Engineering; prof.J.Dorosz, dr.M.Borecki, doc.K.Jędrzejewski; 6 – Knowledge discovery in medical databases; prof.J.Mulawka; 7 – LHC, CMS, FLASH, E-XFEL, Electronics for HEP; prof.J.Krolikowski, dr K.Poźniak. Brak sesji nt Data Classifiers; dr S.Jankowski; sesji Internet Engneering; prof.R.Romaniuk; sesji Communications; prof.A.Płatonow

WILGA is available on the web page http://wilga.ise.pw.edu.pl. There are organized numerable Ph.D. student conferences in this country, for researchers in technical sciences. WILGA Symposium has its own unique and valuable character worked diligently during many years of hard and positive academic work. The organizers of WILGA Symposium, who are successive generations of M.Sc. and Ph.D students of the ISE WUT, supported by IEEE and SPIE student members, warmly invite their colleagues, young researchers, together with their tutors and mentors to participate in WILGA 2010 in May.

SPIE Best Student Presentation Award Ceremony. Students with Photonics Society of Poland Officers, Faculty of Physics, WUT, 26.06.2009

16. REFERENCES

1. Instytut Systemów Elektronicznych Politechniki Warszawskiej [http://www.ise.pw.edu.pl]
2. Sympozjum WILGA, Fotonika i Inżynieria Sieci Internet [http://wilga.ise.pw.edu.pl]
3. SPIE [http://spie.org];
4. Elektronika [http://www.elektronika.orf.pl]
5. ETQ PAN [http://etq.tele.pw.edu.pl/index.php]
6. MOEMS [http://en.wikipedia.org/wiki/MOEMS]
7. MNiSzW [http://www.nauka.gov.pl];
8. Photonics [http://www.photonics.com]
9. Katedra Promieniowania Optycznego Politechniki Białostockiej [http://we.pb.edu.pl/~kpo/j/]
10. Zakład Technologii Światłowodów UMCS [http://www.umcs.lublin.pl/articles.php?aid=1484]
11. ITME [http://www.itme.edu.pl]
12. Instytut Mikroelektroniki i Optoelektroniki Politechniki Warszawskiej [http://www.imio.pw.edu.pl]
13. HIPER Laser [http://www.hiper-laser.org]
14. ELI [http://www.extreme-light-infrastructure.eu]
15. Instytut Radioelektroniki Politechniki Warszawskiej [http://www.ire.pw.edu.pl]
16. STRIDE model of threat categories [http://teck.in/stride-model-of-threat-categories.html]
17. ISACA [http://www.isaca.org.pl/];

18. ISSAA WG CS IEEE [http://issaa.org]
19. IEE Computation Intelligence Soc. [http://www.ieee-cis.org]
20. Instytut Informatyki PW [http://www.ii.pw.edu.pl];
21. NASK [http://www.nask.pl]
22. Studenckie Koło Inżynierii Kosmicznej Politechniki Warszawskiej [http://skik.pw.edu.pl/]
23. CBK PAN [http://www2.cbk.waw.pl/];
24. MEiL PW [http://www.meil.pw.edu.pl/]
25. The Mars Society; University Rover Challenge [http://www.marssociety.org/portal/c/urc/frontPage];
26. Pi od the Sky [http://grb.fuw.edu.pl/]
27. GRB [http://pl.wikipedia.org/wiki/GRB_080319B]
28. ESO [http://www.eso.org/]
29. Max Planck Institute for Solar System Research [http://www.mps.mpg.de/en/]
30. Chandrayyan-1 satellite [http://en.wikipedia.org/wiki/Chandrayaan]
31. Low-Earth orbit [http://en.wikipedia.org/wiki/Low_Earth_orbit]
32. SEWING Project. System for European Water Monitoring [http://www.sewing.mixdes.org/]
33. Project WARMER, Water Risk Management in Europe [http://www.projectwarmer.eu]
34. DESY Hamburg [http://wwwdesy.de];
35. FLASH Laser [http://flash.desy.de]
36. E-XFEL [http://xfel.desy.de/]
37. CERN Geneva [http://public.web.cern.ch]
38. LHC [http://lhc.fuw.edu.pl/]
39. CMS [http://cms.web.cern.ch/cms]
40. FERMILAB [http://www.fnal.gov]
41. ILC [http://www.linearcollider.org/
42. PSI [http://www.psi.ch/]
43. ALBA CELLS [http://www.cells.es/]
44. FP7 Project EuCARD [https://eucard.web.cern.ch/EuCARD/index.html]
45. IEEE PL [http://ieee.pl/]
46. SPIE PL [http://www.spie.pl/]
47. SPIE and OSA Politechnika Wrocławska [http://www.spie.if.pwr.wroc.pl/links.htm]
48. IEEE R8 [http://www.ewh.ieee.org/reg/8/cms/]
49. Photonics Society of Poland [http://photonics.pl]
50. Photonics Letters of Poland [http://photonics.pl/PLP]

INFORMATION FOR AUTHORS

An article published in other magazines can not be submitted for publishing in E.T.Q. The size of an article can not exceed 30 pages, 1800 character each, including figures and tables. The paper should be sent by e-mail to etq@tele.pw.edu.pl

In a separate text file following information should be sent:

- mailing address (home or office)
- phone (home and/or office)
- e-mail

In case of changing of workplace or home address authors are asked to inform the editorial staff.

Instruction for preparation of final typescript

The Title of The Paper, 14pt, Bold, Centered

NAME(S) OF THE AUTOR(S), 12PT, CENTERED, CAPITAL LETTERS

Author Affiliation(s), 11pt, Italics, centered

E-mail(s), 11pt, Italics, centered

Received (leave blank space here).

Authorized (leave blank space here).

These instructions give you basic guidelines for preparing papers for Electronics and Telecommunications Quarterly. For uniformity, MS Word 2003 is recommended for text preparation. (This is abstract; 12pt; paragraph left and right indent is 12.5mm; first line indent is an additional 12.5mm; justified).

Keywords: 5–10 Keywords; 11pt; paragraph left and right indent is 12.5mm; left justified in case of one line only, Justifie in other cases; don't punctuate

1. MAIN HEADING, 12PT, CENTERED, CAPITAL LETTERS

Prepare your paper in full-size A4 format with pages numbering. Use Times New Roman font, 12 point. Text of the paper is one-column. Set size of 25.4mm for top and bottom margin; 35mm for left and right margin. Paragraphs should be justified, using 1.5 spacing and first line indentation of 12.5mm. Leave one blank line before the main and secondary heading; before and after figures and tables, unless it is top of the next page.

All words in title should be capitalized except for conjunctions, prepositions (e.g. on, of, by, and, or, but, from, with, without, under) and definite and indefinite articles (the, a, an) unless they appear at the beginning. Formula letters must be typeset as in the text. Use the same rule for keywords.

Place figures and tables as close as possible to the first references to them in the paper. Nevertheless, avoid placing figures and tables before their first mention in the text. Place figure caption below the figures (Fig. 1). Use the abbreviation "Fig. 1," unless it is the beginning of a sentence. Figure 1 illustrates this scheme. Colored photographs or figures will not be reproduced in color.

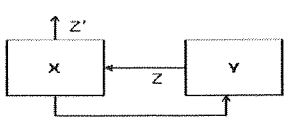


Fig. 1. Title of figure, 11pt, centered

Table headings should be above the tables (Tab. 1). Number tables consecutively with table numbers right justified.

Table 1

Title of table, 11pt, centered

Xxx	Aaa	Bbb	Ccc	Ddd
Xxx	Aaa	Bbb	Ccc	Ddd
Xxx	Aaa	Bbb	Ccc	Ddd

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (1). To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use an en dash (–) rather than a hyphen for a minus sign. Use parentheses to avoid ambiguities in denominators. Punctuate equations with commas or periods when they are part of a sentence, as in

$$a + b = c. \tag{1}$$

Number citations consecutively in square brackets [1]. Punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]. Use Reference [4] at the beginning of a sentence: "Reference [4] was the first ..."

1.1. SECONDARY HEADING, 11PT, NUMBERED, CENTERED, CAPITAL LETTERS

Avoid leaving a heading at the bottom of a page, with the subsequent text starting at the top of the next page. Use extra spacing (between earlier figures or sections) to push the heading up to the top of the same page as its text. However, do please make the fullest possible use of the text area.

2. REFERENCES

1. A. Bababa, C.Dadada: *Name of paper in italics*. In: Name of book (Name(s) of editor(s). (Ed)), page numbers. Publisher, Place of publication, 2007
2. E. Fafafa: *Name of book in italics*, page or chapter numbers if relevant. Publisher, Place of publication, 2008
3. G. Haha: *Name of paper in italics*. Name of journal, Volume no, page numbers, 2009
4. http://www.aaa.bbb.com

Additional information

Author is entitled to the twenty free-of-charge copies of article. Additional copies of the full magazine can be bought at publisher.

Author is obliged to perform the author's correction which should be accomplished within 3 days from the date of receiving the paper from the editorial office. Scanned copy (format pdf or .jpg) with correction marks placed on the margin should be returned to the editorial office personally or by e-mail: etq@tele.pw.edu.pl or by mail eventually (in that case appropriate information should be sent by e-mail to the editorial office).

In the case when the author's correction is not received on time, correction will be performed by technical editorial staff.