POLISH ACADEMY OF SCIENCE
COMMITTEE FOR ELECTRONICS AND TELECOMMUNICATIONS

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

VOLUME 54 – No 1

WARSAW 2008

Addres of Editorial Office
00-665 Warszawa, ul. Nowowiejska 15/19 Politechnika, pok. 470
Instytut Telekomunikacji, Gmach im. prof. JANUSZA GROSZKOWSKIEGO

Editor-on-duty
Mondays and Wednesdays
From 2pm to 4pm
Phone number: (022) 234 77 37

Telephone numbers
Editor-in-Chief: (022) 825 15 80; (022) 234 73 30
Responsible Secretary: 0500044131
www.tele.pw.edu.pl/keit

Ark. wyd. 8,0 Ark. druk. 6,125	Podpisano do druku w maju 2008 r.
Papier offset, kl. III 80 g. B-1	Druk ukończono w maju 2008 r.

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

IMPORTANT MESSAGE FOR THE AUTHORS

The Editoral Board during their meeting on the 18th of January 2006 authorized the Editoral 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. COVERTING 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 Quarterly 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 writting, the editorial staff may decide to relieve authors form 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 54 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 apporaches 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 knows, 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 *Ministry of Science and Higher Education* GRANT's meets of the requirements for those work.

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 otained 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

G.

M.

A.
J. S
R. S

K.

Info

CONTENTS

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2008, 54, no 1, pp. 9–28

Improved State Encoding for FSM Implementation in FPGA Structures with Embedded Memory Blocks*

GRZEGORZ BOROWIK

Warsaw University of Technology
Institute of Telecommunications, Poland
gborowik@tele.pw.edu.pl

Received 2008.01.20
Authorized 2008.03.21

Modern FPLD devices have a very complex structure. They combine PLA-like structures as well as FPGA's and even memory-based structures. However, the lack of an appropriate synthesis method does not allow the features of the modern FPLD's to be fully exploited. In this paper, an important problem of state assignment for an FSM as an extension of the previous research on ROM-based FSM implementation is presented. We pinpoint the sources of additional optimization of the functional decomposition and relate them to the state encoding conditions. The method is based on a reduction of a state assignment problem to a graph coloring problem. To this end, the so called multi-graph of incompatibility of memory T-words is applied. As a result, a new design technique for implementation of sequential circuits using embedded memory blocks of FPGA's has been developed. Preliminary experimental results are extremely encouraging.

Keywords: decomposition, state encoding, sequential circuit, finite state machine, FPGA, embedded memories, logic cell, multigraph, weighted graph

1. INTRODUCTION

In modern logic synthesis of PLD, FPGA modules as well as PLA (*Programmable Logic Array*), GA (*Gate Array*), SC (*Standard Cell*) structures, the problem of finite state machine synthesis is significant because of its practical application, but in particular internal states encoding. Encoding influences both, the structure of the realization of

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

FSM, its mutual connections between the combinational block and the memory block, as well as the complexity of the combinational block.

Any attempt to solve the above problem results in many methods of the structural synthesis of FSM. Their diversity results from different calculations, assumptions and subsequently directing the method for specific type of elements. Hence, there are separate methods of the synthesis of FSM for PLA structures [8, 10], for ROM memories [1] and PLD modules [2, 9].

The application of minimization before the state encoding process is a characteristic feature of these methods. Such a minimization is possible by the representation of the combinational part of the FSM with polyvalent symbolic variables. Unfortunately, such methods are limited to two-level structures. At the moment the research is being conducted in two directions. One concerns the realizations in multilevel gate structures, and the second covers realizations in FPGA's and FLEX's.

In the former case of such synthesis, the combinational part of the FSM is joined to inputs of the memory register, as synthesis for two-level structures, and in the latter – the other way round – combinational part is joined to outputs of such register.

Up till now, the methods of the optimization of sequential machines apply the first model of the mentioned synthesis. In these methods the optimization consisted in such a selection of state encoding, in which two-level or multilevel gate structures are applied for the reduction of the resources.

The second model was used in so-called microprogrammable control units, in which combinational part was implemented in ROM. In this implementation the memory block was a separate element – distinguished from the other parts of the circuit. The advantage of such structure made it possible to reprogram the memory of microcode. In those times it was the only way to reconfigure the circuit. Due to these advantages the capacity of the memory was no longer a critical factor. Nevertheless, it was used to make this capacity smaller. It was a typical technique of constructing special memory addressing units [1, 2], in which the method of selection of the exterior signals was applied to manufacture the address of the next microinstruction. In result the realization of the function lasted few cycles, hence it had its influence on the speed reduction of the circuit. Supposedly the speed reduction is the main factor that make such methods of the synthesis not very attractive for today technologies.

For these reasons, main research works within the scope of the optimization methods of sequential machines concern the first model. It focuses on the realization in the two-level structures and in the multilevel structures.

Unfortunately, none of the current solutions [3, 12, 13, 14, 15, 16] concerns FPGA's with embedded memory blocks.

Modern Field-Programmable Logic Devices have a very complex structure. They combine PLA-like structures as well as FPGA's and even memory-based structures. In many cases designers cannot exploit all opportunities such as complex architectures provide due to the lack of appropriate synthesis methods. Embedded memory arrays

enable an implementation of memory-like blocks such as large registers, FIFO's, RAM or ROM modules [8, 18].

These memory resources make up considerably large part of the devices, i.e., EP20K1500E devices provide 51,840 logic cells and 442 Kbit of SRAM. Taking into consideration the conversion factors of logic elements and memory bits to logic gates (12 gates/logic element and 4 gates/memory bit) it turns out that embedded memory arrays make up over 70% of all logic resources. Since not every design consists of such modules as RAM or ROM, in many cases, these resources are not utilized, though. However, such embedded memory blocks can be used for implementation of sequential machines in a way that requires fewer logic cells than the traditional flip-flop based implementation. This may be used to implement "non-vital" sequential parts of the design, saving logic cell resources for more important sections [30, 33]. Since the size of embedded memory blocks is limited, such an implementation may require more memory than is available in a device. To reduce a memory usage in ROM-based sequential machine implementations, a structure with next state logic partially implemented in the ROM and partially implemented in logic cells was proposed [5, 6].

In the considered FSM implementation the combinational logic is split into two parts. One part is implemented in embedded memory blocks which are configured as ROM, with its content determined at the time of the programming. The second part, called the address modifier, is used to reduce the number of memory address lines. The address modifier is implemented in programmable logic blocks containing LUT's. This proposal is a cross-fertilized approach between recent progress in finite-state machine synthesis and in micro-computer architectures. Similar ideas can be found in [2].

In this paper, we present an original solution to this problem. We demonstrate that the problem is intimately related to the encoding problem of FSM which is of fundamental importance in a sequential synthesis, especially the state-machine synthesis [3, 14].

Basic information is given first. Secondly, an application of a serial decomposition in the implementation of sequential machines is presented. Subsequently, experimental results are discussed. These results are obtained with a prototype tool that implements functional decomposition.

The experimental results demonstrate that decomposition is capable of constructing solutions (utilizing embedded memory blocks) of better quality than the methods implemented in commercial systems [21].

2. BASIC NOTIONS

Let $\mathcal{A} = (S, V, \delta, Y, \lambda)$ be an FSM (completely or incompletely specified), where: S – set of internal states, V – set of input symbols, δ – state transition function, Y – set of output symbols, λ – output function, and values $m \geqslant \lceil \log_2 |V| \rceil$, $p \geqslant \lceil \log_2 |S| \rceil$ denote the number of input and state variables respectively.

To describe logic dependencies in such an FSM special partition description [7] and special partition algebra [12] are applied.

Let $\mathbb{T}$ be an isomorphic function between the domain D_δ of the transition function and the set $T = 1, \ldots, t$, where $t = |D_\delta|$. Set T represents the ROM cells needed to store the next state pair $\delta(v, s)$ for each pair (v, s). Thus, the characteristic partition P_c of the FSM is defined in the following way:

Each block B_{P_c} of the characteristic partition includes these elements from the set T which correspond to these pairs (v, s) from the domain D_δ which the transition function $\delta(v, s) = s'$ maps onto the same next state s'.

Example 1. For the FSM and function $\mathbb{T}$ shown in Table 1 the characteristic partition is:

$$P_c = \{\overline{1, 8, 12, 14}; \overline{2, 7, 10, 16}; \overline{6, 9, 13}; \overline{3, 5, 11, 15}; \overline{4}\}. \quad \square$$

Table 1

FSM transition table and mapping $\mathbb{T}$

	V			
S	v_1	v_2	v_3	v_4
s_1	s_1	s_2	s_4	-
s_2	-	-	s_5	s_4
s_3	s_3	s_2	s_1	s_3
s_4	s_2	-	s_4	s_1
s_5	s_3	s_1	s_4	s_2

	V			
S	v_1	v_2	v_3	v_4
s_1	1	2	3	-
s_2	-	-	4	5
s_3	6	7	8	9
s_4	10	-	11	12
s_5	13	14	15	16

A partition P on the set T is related to a partition π on the states set S if for any inputs v_a, v_b the condition that s_i, s_j belong to one block of the partition π implies that the elements from T corresponding to pairs (v_a, s_i) and (v_b, s_j) belong to one block of the partition P.

A partition P on set T is related to a partition θ on the input symbols set V if for any state s_a, s_b the condition that v_i, v_j belong to one block of the partition implies that the elements from T corresponding to pairs (v_i, s_a) and (v_j, s_b) belong to one block of the partition P.

In particular, a partition P on set T is related to the set $\{\pi, \theta\}$ if it is related to both π and θ.

Example 2. For FSM from Table 1 partition

$$P_1 = \{\overline{1, 2, 3, 4, 5, 6, 7, 8, 9}; \overline{10, 11, 12, 13, 14, 15, 16}\}$$

is related to the partition $\pi = \{\overline{s_1, s_2, s_3}; \overline{s_4, s_5}\}$, while partition

$$P_2 = \{\overline{1, 2, 6, 7}; \overline{10, 13, 14}; \overline{3, 4, 5, 8, 9}; \overline{11, 12, 15, 16}\}$$

is related to the set $\{\pi, \theta\}$, and $\theta = \{\overline{v_1, v_2}; \overline{v_3, v_4}\}$. □

Let P_a and P_b be partitions on the set T, and $P_a \geqslant P_b$. Then a partition $P_a|P_b$, whose elements are blocks of P_b and whose blocks are those of P_a, is the quotient partition of P_a over P_b.

Example 3. For set $T = \{1,2,3,4,5,6\}$ and partitions $P_1 = \{\overline{1,2,5}; \overline{3,4,6}\}$, $P_2 = \{\overline{1,2}; \overline{3,6}; \overline{4}; \overline{5}\}$, the quotient partition is:

$$P_1|P_2 = \{\overline{(1,2)(5)}; \overline{(3,6)(4)}\}. \quad \square$$

For a partition $P_a \geqslant P_b$ let $P_a|P_b$ denote the quotient partition and let $\varepsilon(P_a|P_b)$ be the number of elements in the largest block of partition $P_a|P_b$. Let $e(P_a|P_b)$ be the smallest integer equal to or larger than $\log_2(\varepsilon(P_a|P_b))$ (i.e., $e(P_a|P_b) = \lceil \log_2(\varepsilon(P_a|P_b)) \rceil$). Then, the R-admissibility of the two-block partitions' set $\{P_1, \ldots, P_k\}$ on S in relation to the partition P on S, is defined as $R = k + e(\sigma|\rho)$, where σ is the product of $\{P_1, \ldots, P_k\}$ and ρ is the product of σ and P.

Example 4. For the FSM from the Table 1 and partitions P_2 from Example 2 and P_c from Example 1, the quotient partition $P_2|(P_2 \cdot P_c)$ is:

$$P_2|(P_2 \cdot P_c) = \{\overline{(1)(2,7)(6)}; \overline{(10)(13)(14)}; \overline{(3,5)(4)(8)(9)}; \overline{(11,15)(12)(16)}\}.$$

And hence, the R-admissibility of P_2 in relation to P_c is $R = 4$. □

3. IMPLEMENTATION OF FINITE STATE MACHINES IN FPGA'S

Any FSM, $\mathcal{A} = (S, V, \delta, Y, \lambda)$, can be implemented as in Fig. 1 using an address modifier.

If $\Pi = \{\pi_1, \ldots, \pi_p\}$ is the set of two-block partitions on S and $\Theta = \{\theta_1, \ldots, \theta_m\}$ is the set of two-block partitions on V, while P_k is a partition on the set T which is related to either π_i or θ_j, then $\mathfrak{p} = \{P_1, \ldots, P_{m+p}\}$ is the set of all partitions related to partitions $\{\pi_1, \ldots, \pi_p, \theta_1, \ldots, \theta_m\}$. Partitions in Π correspond to the state variables and partitions in Θ correspond to the input variables.

Fact 1. *To achieve unambiguous encoding of address variables and at the same time maintain the consistency relation* $\mathbb{T}$ *with the transition function, two-block partitions* $\mathfrak{P} = \{P_1, \ldots, P_w\}$ *have to be found, such that:*

$$P_1 \cdot P_2 \cdot \ldots \cdot P_w \leqslant P_c. \tag{1}$$

This is a necessary and sufficient condition for $\{P_1, \ldots, P_w\}$ to determine the address variables. This is because each memory cell is associated with a single block

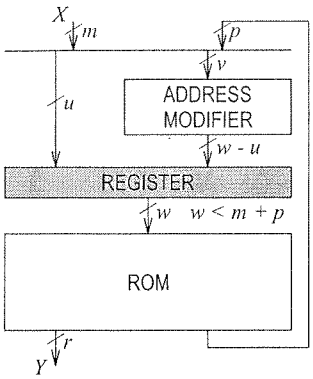

Fig. 1. Implementation of FSM using an address modifier

of P_c, i.e., with those elements from T which map the corresponding (v, s) pairs onto the same next state.

The selection of partitions $\mathfrak{P}$ from the set $\mathfrak{p}$ is made in such a way that they produce the simplest addressing unit [7]. Such a selection is possible thanks to the notion of R-admissibility.

3.1. STATES ENCODING

Assume that, the partitions $\{\pi_1, \ldots, \pi_l\}$ and $\{\theta_1, \ldots, \theta_{u-l}\}$ which were chosen for the state and input symbol partial encoding, and partitions P_i related to them, correspond to those address lines which are driven by a single variable, either a state variable q or an external variable x, i.e.,

$$a_1 = q_1, \ldots, a_l = q_l, a_{l+1} = \theta_1, \ldots, a_u = \theta_{u-l}.$$

The basic steps of a selection of such partitions P_i to the set $\mathfrak{P}$ are presented in Fig. 2. The encoding of state variables is possible thanks to the method of construction and coloring weighted graphs [4].

Corollary 1. *Inequality* (1) *can be written as:*

$$P_{i_1} \cdot P_{i_2} \cdot \ldots \cdot P_{i_u} \cdot P_{i_{u+1}} \cdot \ldots \cdot P_{i_w} \leqslant P_c, \tag{2}$$

where $P_U = P_{i_1} \cdot P_{i_2} \cdot \ldots \cdot P_{i_u}$ *is related to the partitions* $\{\pi_1, \pi_2, \ldots, \pi_l, \theta_1, \theta_2, \ldots, \theta_{u-l}\}$.

The encoding of the part of the state variables remaining after the partial encoding (input variables, in general) can be obtained from the following rules:

$$\pi_1 \cdot \pi_2 \cdot \ldots \cdot \pi_l \cdot \pi = \pi(\mathbf{0}), \tag{3}$$

$$\theta_1 \cdot \theta_2 \cdot \ldots \cdot \theta_{u-l} \cdot \theta = \theta(\mathbf{0}), \tag{4}$$

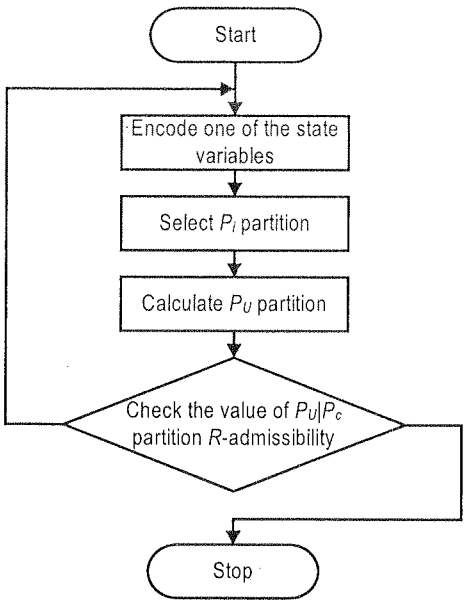

Fig. 2. Schematic representation of the selection of partitions $\mathfrak{P}$

where π and θ represent partitions corresponding to these remaining state variables. $\pi(\mathbf{0})$ as well as $\theta(\mathbf{0})$ are partitions whose blocks are equal to their elements.

Since the design process may be considered as a decomposition of the memory block into two blocks: a combinational address modifier and a smaller memory block, we need to find function G which will determine the second part of the memory address bits.

Inequality (2) can be transformed into:

$$P_U \cdot P_G \leqslant P_c. \tag{5}$$

Corollary 2. *A partition P_G has to be constructed, such that:*

$$P_G \geqslant P_V, \tag{6}$$

where $P_G = P_{i_{u+1}} \cdot \ldots \cdot P_{i_w}$ and P_V is related to the partition set $\{\pi, \theta\}$.

Let us assume that input variables are encoded.

Theorem 1. *Partition P_V can be constructed in the following way:*

$$P_V = P_S \cdot P_{V_\theta}, \tag{7}$$

where P_S is the partition related to $\pi(\mathbf{0})$ on the set of states S, and P_{V_θ} is the partition related to θ.

Proof. Let us assume that $P_V = P_{V_\pi} \cdot P_{V_\theta}$, where P_{V_π} is related to π, and P_{V_θ} is related to θ. Since P_U and P_V satisfy $P_U \cdot P_V \leqslant P_c$, we have $P_U \cdot P_{V_\pi} \cdot P_{V_\theta} \leqslant P_c$. As a result, $P_U \cdot P_S \cdot P_{V_\theta} \leqslant P_c$. □

Let $\langle V, R, E, P\rangle$ be a quadruple where: V – set of elements, R – an equivalence relation on the set V, E – set of pairs in relation P on the set V, P – two-element relation. A triple $M(V|R, E, P)$ is a multi-graph, where $V|R$ – is an equivalence class for an equivalence relation on the set V.

Since there exists an isomorphism $V|R \leftrightarrow V'$, we can construct a natural mapping from graph $M(V|R, E, P)$ to $G(V', E', P)$. Such a mapping $\psi\colon M \to G$ allows for calculation of a chromatic number $\chi(G) = \chi(M)$.

Let us apply these notions to the construction of the P_G partition. The basic steps of the algorithm are presented in Fig. 3.

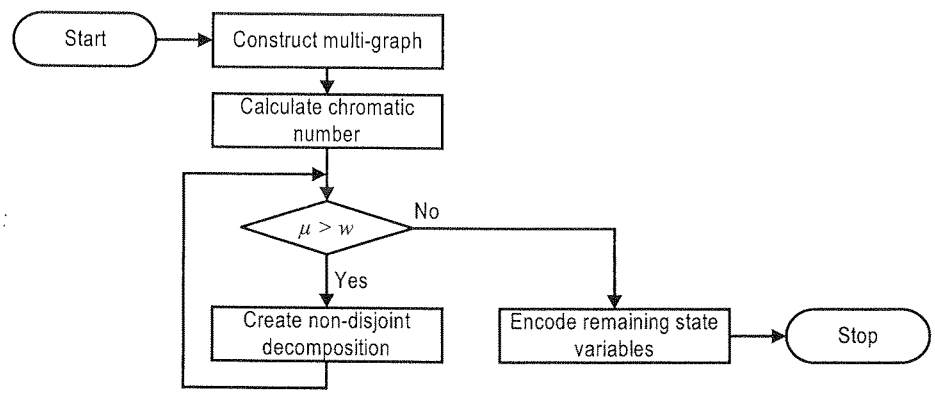

Fig. 3. Scheme of construction of the P_G partition

Inequality (5) allows us to construct a quotient partition $P_U|P_c$.

Corollary 3. *The triple* $\langle P_V, E_1, P_1\rangle$, *where:* P_V *is a partition given by equation* (7), P_1 *is a relation which represents incompatibilities in quotient partition* $P_U|P_c$ *on the set* T (relation of incompatibility in quotient partition $P_U|P_c$ is a relation among all elements in each block of the partition separately) *and* E_1 *is the set of pairs in the relation* P_1*; is a multi-graph* $M_1(P_V, E_1, P_1)$.

After mapping $\psi_1\colon M_1 \to G_1$ we calculate a chromatic number $\chi(G_1)$ which is equal to $\chi(M_1)$.

The coloring of the graph G_1 determines the P_G partition.

Example 5. Let us assume that input variables for the transition table 1 are encoded (Tab. 2).

Table 2

FSM transition table

	00	01	11	10	(x_1, x_2)
S	v_1	v_2	v_3	v_4	V
s_1	s_1	s_2	s_4	-	
s_2	-	-	s_5	s_4	
s_3	s_3	s_2	s_1	s_3	
s_4	s_2	-	s_4	s_1	
s_5	s_3	s_1	s_4	s_2	

Based on [7], we obtain set $U = \{q_1, q_2, x_2\}$, where q_1, q_2 are internal variables which generate partitions, respectively:

$$\pi_1 = \{\overline{s_1, s_2, s_4}; \overline{s_3, s_5}\},$$
$$\pi_2 = \{\overline{s_1, s_3, s_4}; \overline{s_2, s_5}\}.$$

Then

$$P_U|P_c = \{\overline{(1, 12)(10)}; \overline{(2)(3, 11)}; \overline{(5)}; \overline{(4)}; \overline{(6, 9)}; \overline{(7)(8)}; \overline{(13)(16)}; \overline{(14)(15)}\}.$$

As

$$P_{V_\theta} = P(x_1) = \{\overline{1, 2, 6, 7, 10, 13, 14}; \overline{3, 4, 5, 8, 9, 11, 12, 15, 16}\},$$

and

$$P_S = \{\overline{1, 2, 3}; \overline{4, 5}; \overline{6, 7, 8, 9}; \overline{10, 11, 12}; \overline{13, 14, 15, 16}\},$$

we obtain

$$P_V = \{\overline{1, 2}; \overline{3}; \overline{4, 5}; \overline{6, 7}; \overline{8, 9}; \overline{10}; \overline{11, 12}; \overline{13, 14}; \overline{15, 16}\}.$$

According to Corollary 3 we construct a multi-graph M_1 and its image G_1 (Fig. 4). As a result, we obtain $\chi(M_1) = 3$, and hence $\mu = 5$. □

The value of

$$\mu = |U| + \lceil \log_2(\chi(M_1)) \rceil \tag{8}$$

determines memory size required. In case of $\mu > w$, a new partition P'_V has to be constructed. Then, P_V has to be multiplied by appropriately chosen two-block partitions related to those which are generated by input variables from the set U. In that case, we obtain a non-disjoint decomposition.

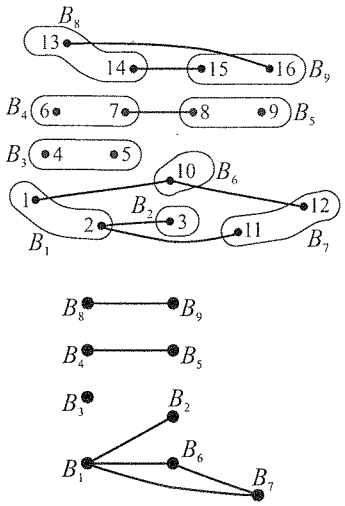

Fig. 4. Multi-graph M_1 and graph G_1

Example 6. Let $w = 4$. Selecting additional external variable to generate partition P_{V_θ}, we obtain:

$$P_{V_\theta} = P(x_1, x_2) = \{\overline{1, 6, 10, 13}; \overline{2, 7, 14}; \overline{3, 4, 8, 11, 15}; \overline{5, 9, 12, 16}\},$$

and then

$$P_V = \{\overline{1}; \overline{2}; \overline{3}; \overline{4}; \overline{5}; \overline{6}; \overline{7}; \overline{8}; \overline{9}; \overline{10}; \overline{11}; \overline{12}; \overline{13}; \overline{14}; \overline{15}; \overline{16}\},$$

and consequently a new multi-graph M_1 (Fig. 5).

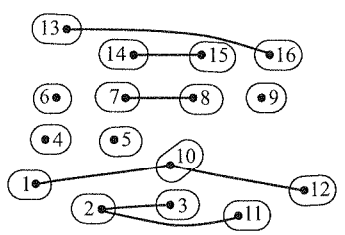

Fig. 5. Multi-graph M_1

As a result $\chi(M_1) = 2$ and $\mu = 4$. □

In the next step we calculate the remaining state variables.

Corollary 4. *The triple $\langle P_S, E_2, P_2\rangle$, where: P_S is the partition related to $\pi(\mathbf{0})$ on the states set S, P_2 is a relation which represents incompatibilities in quotient partition $P_{V_\theta}|P_G$ and E_2 is the set of pairs in the relation P_2; is a multi-graph $M_2(P_S, E_2, P_2)$.*

Similarly to the case discussed above, by coloring an image graph G_2 for the multi-graph M_2, we obtain a new partition on the set S.

Finally, we encode the new partition with the minimal binary code. $\lceil \log_2(\chi(M_2)) \rceil$ determines the number of bits needed to encode the remaining state variables. So

$$v = |V_\theta| + \lceil \log_2(\chi(M_2)) \rceil \tag{9}$$

determines the number of inputs to address modifier.

Example 7. As a result of coloring the image graph G_1 for the multi-graph M_1 presented in Fig. 5, we obtain partition

$$P_G = \{\overline{1, 2, 5, 7, 9, 12, 14, 16}; \overline{3, 4, 6, 8, 10, 11, 13, 15}\},$$

and then

$$P_{V_\theta}|P_G = \{\overline{(1)(6, 10, 13)}; \overline{(2, 7, 14)}; \overline{(3, 4, 8, 11, 15)}; \overline{(5, 9, 12, 16)}\}.$$

According to Corollary 4, we construct a multi-graph M_2 and its image G_2 (Fig. 6).

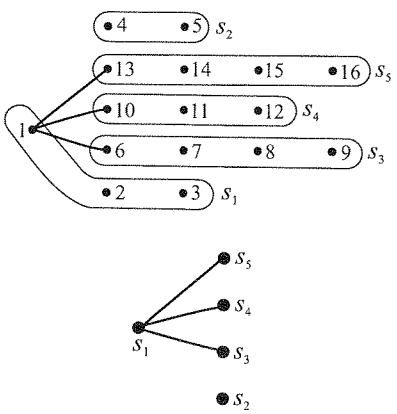

Fig. 6. Multi-graph M_2 and graph G_2

By coloring the image graph G_2, we obtain two possible partitions on the set S:

$$\pi^1 = \{\overline{s_1, s_2}; \overline{s_3, s_4, s_5}\},$$
$$\pi^2 = \{\overline{s_1}; \overline{s_2, s_3, s_4, s_5}\}.$$

One of those is chosen and encoded with natural binary code.

For example, partition π^1 can be generated by internal variable q_3. We encode partition π^1, so that:

$$\pi^1 = \{\overline{s_1, s_2}\ (0); \overline{s_3, s_4, s_5}\ (1)\};$$

thus

$$P_{V_{\pi^1}} = \{\overline{1,2,3,4,5};\overline{6,7,8,9,10,11,12,13,14,15,16}\}.$$

Consequently

$$P_V = P_{V_\theta} \cdot P_{V_{\pi^1}} = \{\overline{1};\overline{2};\overline{3,4};\overline{5};\overline{6,10,13};\overline{7,14};\overline{8,11,15};\overline{9,12,16}\}. \quad \square$$

Finally, we can construct the truth table of address modifier and the memory ROM content.

Example 8. In presented examples, we obtain a non-disjoint decomposition. Then, argument sets are:

$$U = \{q_1, q_2, x_2\},$$
$$V = \{x_1, x_2, q_3\}.$$

Assuming that the encoding of internal variables q_1 and q_2 for partition $\pi_1 \cdot \pi_2$ is:

$$\pi_1 \cdot \pi_2 = \{\overline{s_1, s_4}\ (00);\overline{s_2}\ (01);\overline{s_3}\ (10);\overline{s_5}\ (11)\},$$

and the encoding for P_G is:

$$P_G = \{\overline{1,2,5,7,9,12,14,16}\ (0);\overline{3,4,6,8,10,11,13,15}\ (1)\},$$

we have the truth table of address modifier and the ROM content, the presented in Table 3 and Table 4. $\square$

3.2. CONSTRUCTION OF PARTITION P_G

The graph G_1 can be colored in many different ways. As a consequence, many partitions P_G could be obtained. Although the number of blocks in all partitions P_G is the same, the number of address modifier inputs (equation 9) could be different. The construction of a partition on the set S is similar to that of the quotient partition $P_{V_\theta}|P_G$. It leads to a difference in the number of remaining (after partial encoding) internal variables for different partitions P_G.

Example 9. Let us assume that as a result of coloring the image graph G_1 for multi-graph from Fig. 5, we obtain partition,

$$P_G = \{\overline{1,3,5,8,9,11,12,15,16};\overline{2,4,6,7,10,13,14}\};$$

then

$$P_{V_\theta}|P_G = \{\overline{(1)(6,10,13)};\overline{(2,7,14)};\overline{(3,8,11,15)(4)};\overline{(5,9,12,16)}\}.$$

Table 3

Truth table of address modifier

P_V	x_1	x_2	q_3	g
$\overline{1}$	0	0	0	0
$\overline{2}$	0	1	0	0
$\overline{3,4}$	1	1	0	1
$\overline{5}$	1	0	0	0
$\overline{6,10,13}$	0	0	1	1
$\overline{7,14}$	0	1	1	0
$\overline{8,11,15}$	1	1	1	1
$\overline{9,12,16}$	1	0	1	0

Table 4

Content of ROM

$P_U \cdot P_G$	q_1	x_2	q_{2_1}	g	q_1	q_{2_1}	q_3
$\overline{1,12}$	0	0	0	0	0	0	0
$\overline{2}$	0	1	0	0	0	1	0
$\overline{3,11}$	0	1	0	1	0	0	1
$\overline{4}$	0	1	1	1	1	1	1
$\overline{5}$	0	0	1	0	0	0	1
$\overline{6}$	1	0	0	1	1	0	1
$\overline{7}$	1	1	0	0	0	1	0
$\overline{8}$	1	1	0	1	0	0	0
$\overline{9}$	1	0	0	0	1	0	1
$\overline{10}$	0	0	0	1	0	1	0
$\overline{13}$	1	0	1	1	1	0	1
$\overline{14}$	1	1	1	0	0	0	0
$\overline{15}$	1	1	1	1	0	0	1
$\overline{16}$	1	0	1	0	0	1	0

By constructing the multi-graph M_2 and coloring its image graph G_2 (Fig. 7), we obtain the partition on the set S,

$$\pi = \{\overline{s_1}; \overline{s_2}; \overline{s_3, s_4, s_5}\}.$$

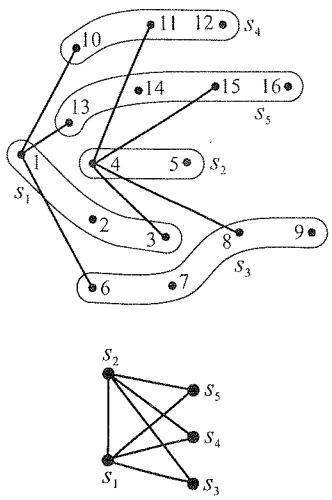

Fig. 7. Multi-graph M_2 and graph G_2

As a result, we need two additional internal variables for encoding this partition. Consequently, it is a worse solution than that used in Example 7. □

A special construction of P_G partition is proposed.

Note that the number of partition blocks on the states set S is closely related to the incompatibility relation in the quotient partition $P_{V_\theta}|P_G$. In consequence of joining P_V blocks to the P_G partition blocks, it has to be done in such a way to obtain as least incompatibilities in $P_{V_\theta}|P_G$ as it possibly can.

It is easily seen that $P_{V_\theta} \leqslant P_V$. Let us calculate the quotient partition $P_{V_\theta}|P_V$ and remove those elements which are incompatible in the quotient partition $P_U|P_c$. One can show that the remaining part of P_{V_θ} could be joined to a new partition (according to $P_U|P_c$), that the number of blocks is less or equal to $\chi(M_1)$. That new partition is so-called core of P_G. One can observe that graph G_1 has to be colored in accordance with the core of P_G. As a result, we obtain partition P_G whose number of blocks is equal to $\chi(M_1)$.

Example 10. Based on Example 6

$$P_{V_\theta}|P_V = \{\overline{(1)(6)(10)(13)}; \overline{(2)(7)(14)}; \overline{(3)(4)(8)(11)(15)}; \overline{(5)(9)(12)(16)}\}.$$

According to

$$P_U|P_c = \{\overline{(1,12)(10)}; \overline{(2)(3,11)}; \overline{(4)}; \overline{(5)}; \overline{(6,9)}; \overline{(7)(8)}; \overline{(13)(16)}; \overline{(14)(15)}\},$$

elements (1) and (10) are incompatible, and new P_{V_θ} is:

$$P'_{V_\theta} = \{\overline{2,7,14}; \overline{3,4,8,11,15}; \overline{5,9,12,16}; \overline{6,10,13}\}.$$

Thus, we obtain two solutions:

$$\text{core}_1 P_G = \{\overline{2, 6, 7, 10, 13, 14}; \overline{3, 4, 5, 8, 9, 11, 12, 15, 16}\},$$
$$\text{core}_2 P_G = \{\overline{2, 5, 7, 9, 12, 14, 16}; \overline{3, 4, 6, 8, 10, 11, 13, 15}\}.$$

As a result

$$P_{G_1} = \{\overline{1, 3, 4, 5, 8, 9, 11, 12, 15, 16}; \overline{2, 6, 7, 10, 13, 14}\},$$
$$P_{G_2} = \{\overline{1, 2, 5, 7, 9, 12, 14, 16}; \overline{3, 4, 6, 8, 10, 11, 13, 15}\}.$$

Solution P_{G_2} was used in Example 7. Note that

$$P_{V_0}|P_{G_1} = \{\overline{(1)(6, 10, 13)}; \overline{(2, 7, 14)}; \overline{(3, 4, 8, 11, 15)}; \overline{(5, 9, 12, 16)}\}$$

is a similar solution. □

4. EXPERIMENTAL RESULTS

The proposed method was applied to implement several FSM's (Table 5) from the standard benchmark set [34] in Flex10K and Stratix devices using Altera Quartus 2 (v6.0, SP1). In Table 6 and 7, a comparison of different FSM implementation techniques is presented.

In column labeled 'ROM' (Tab. 5), the number of bits required to implement a given FSM in ROM is presented (without address modifier). The Flex10K10 device is equipped with 3 EMB's only, of 2048 bits each. All of the examined FSM's cannot be implemented, because their implementation requires much more memory resources than available in this device. The 'LUT' column (Tab. 6, 7) shows the number of logic cells required to implement a given FSM in the "traditional" way using logic cells only. In this case two different state encoding methods available in Altera Quartus 2 were applied. In column 'AM + ROM', the results of implementation of a given FSM using the concept of address modifier are presented. In this approach, the address modifier can be implemented using logic cells, and ROM can be implemented with EMB blocks (Fig. 1). The number of logic cells and the number of memory bits are presented in this column. It can be easily observed that decomposition improves the quality of implementation in ROM resources, as well as the quality of implementation in logic cells only. An improvement of this method by using decomposition into the mixed structure built of LC's and EMB's with application of state encoding method presented in this article is shown in the last column 'AM + ROM (new method)' (Tab. 6, 7). The number of state encoding bits is presented in brackets.

For some examples more than one solution is possible, e.g., benchmark *cse* can be implemented in memory of size 32,768 bits (not available in small programmable devices) or with the use of 90 logic cells (Stratix), 92 logic cells (Flex10K). However,

Table 5

Benchmark Data and FSM ROM Implementation

Benchmark	Benchmark Data			ROM [bit]
	Inputs	Outputs	States	
bbsse	7	7	16	32768[1)]
cse	7	7	16	32768[1)]
ex1	9	19	20	524288[1)]
ex4	6	9	14	16384[1)]
mark1	5	16	15	16384[1)]
pma	8	8	24	131072[1)]
s1	8	6	20	131072[1)]
s208	11	2	18	524288[1)]
s386	7	7	13	32768[1)]
sse	7	7	16	16384[1)]
styr	9	10	30	262144[1)]
tbk	6	3	32	16384[1)]
tma	7	6	20	65536[1)]

[1)] not enough memory bits to implement the project using Flex10K10 device

application of the new method allows for an implementation with 21 logic cells and 1 EMB or with 9 logic cells and 2 EMB's when using a Flex10K device or 77 logic cells and 1 M512 or 8 logic cells and 1 M4K when using a Stratix device.

It can be noted that the new approach allows for much more efficient utilization of available resources. It is also possible to trade off the number of logic cells used with the number of embedded memory blocks.

5. CONCLUSIONS

The functional decomposition method with internal states encoding presented in this paper is a novel approach to FSM synthesis. With the concept of the address modifier, memory usage can be significantly reduced. In contrast to the traditional methods, it does not break the FSM design process into internal states encoding and mapping the encoded state transition table into a target architecture.

The presented results prove the efficiency of the new approach. The application of new FSM synthesis concept allows for obtaining decompositions of better quality than in any other methods. In comparison with other encoding methods [19], it significantly reduces the complexity of implementation of finite-state machines with FPGA's.

Table 6

Implementation Results Comparing for FSM in the Standard FPGA Flex10K Structure (EPF10K10LC84-3)

Benchmark	Quartus 2 Encoding (bit)	Quartus 2 LUT [LC]	AM + ROM [LC + bit]	AM + ROM (new method) [LC + EMB[2)]] (Encoding bits)
bbsse	Auto (16)	37	3 + 5632	9 + 2 (7)
	Minimal (4)	57		
cse	Auto (16)	129	2 + 5632	21 + 1 (8), 9 + 2 (7)
	Minimal (4)	92		
ex1	Auto (20)	199	145 + 3072	98 + 3 (6)
	Minimal (5)	181		
ex4	Auto (14)	29	2 + 3328	5 + 1 (7)
	Minimal (4)	32		
mark1	Auto (15)	38	2 + 5120	14 + 1 (5), 9 + 2 (6)
	Minimal (4)	40		
pma	Auto (24)	157	17 + 3328	29 + 1 (8)
	Minimal (5)	284		
s1	Auto (20)	174	96 + 5632	107 + 1 (7), 69 + 2 (7)
	Minimal (5)	162		
s208	Auto (18)	201	12 + 3584	5 + 2 (8)
	Minimal (5)	59		
s386	Auto (13)	54	9 + 5632	35 + 1 (8), 12 + 2 (8)
	Minimal (4)	55		
sse	Auto (16)	37	3 + 5632	9 + 2 (7)
	Minimal (4)	57		
styr	Auto (30)	212	151 + 7680	174 + 1 (5), 137 + 2 (9)
	Minimal (5)	210		
tbk	Auto (32)	895[1)]	333 + 4096	264 + 1 (5), 25 + 2 (5)
	Minimal (5)	1077[1)]		
tma	Auto (20)	75	4 + 2816	8 + 1 (7)
	Minimal (5)	129		

[1)] implementation not possible – not enough CLB resources

[2)] device has 3 EMB blocks with 2048 bits each

Table 7

Implementation Results Comparing for FSM in the Standard FPGA Stratix Structure (EP1S10F484C5)

Benchmark	Quartus 2 Encoding (bit)	Quartus 2 LUT [LC]	AM + ROM [LC + bit]	AM + ROM (new method) [LC + EMB[1)]] (Encoding bits)
bbsse	Auto (16)	37	3 + 5632	9 + 1 M4K (7)
	Minimal (4)	57		
cse	Auto (16)	112	2 + 5632	77 + 1 M512 (4), 8 + 1 M4K (7)
	Minimal (4)	90		
ex1	Auto (20)	181	145 + 3072	96 + 7 M512 (6)
	Minimal (5)	139		
ex4	Auto (14)	24	2 + 3328	4 + 1 M512 (7)
	Minimal (4)	33		
mark1	Auto (15)	32	2 + 5120	14 + 2 M512 (5), 8 + 1 M4K (6)
	Minimal (4)	38		
pma	Auto (24)	145	17 + 3328	18 + 1 M4K (9)
	Minimal (5)	254		
s1	Auto (20)	168	96 + 5632	129 + 1 M512 (5), 71 + 1 M4K (7)
	Minimal (5)	152		
s208	Auto (18)	154	12 + 3584	4 + 1 M4K (8)
	Minimal (5)	50		
s386	Auto (13)	44	9 + 5632	12 + 1 M4K (8)
	Minimal (4)	50		
sse	Auto (16)	37	3 + 5632	9 + 1 M4K (7)
	Minimal (4)	57		
styr	Auto (30)	189	151 + 7680	169 + 2 M512 (5), 137 + 1 M4K (9)
	Minimal (5)	216		
tbk	Auto (32)	902	333 + 4096	261 + 1 M512 (5), 21 + 1 M4K (5)
	Minimal (5)	959		
tma	Auto (20)	85	4 + 2816	10 + 2 M512 (7)
	Minimal (5)	132		

[1)] device has 920448 bits of memory in 512 bit blocks (M512) as well as in 4096 bit blocks (M4K)

The presented method can also be applied when designing sequential circuits with enhanced testability using FPGA's with embedded memory blocks [17].

6. REFERENCES

1. M. Adamski, A. Barkalov: *Architectural and Sequential Synthesis of Digital Devices*, University of Zielona Góra Press, 2006.
2. M. Adamski, A. Karatkevich, M. Wegrzyn: *Design of Embedded Control Systems*, Springer-Verlag, 2005.
3. P. Ashar, S. Devadas, A. R. Newton: *Sequential Logic Synthesis*, Kluwer Academic Publishers, Boston, MA (USA), 1992.
4. G. Borowik: *Finite State Machines Synthesis for FPGA Structures with Embedded Memory Blocks* (in Polish), PhD Dissertation, Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw 2007.
5. G. Borowik: *Optimal State Assignment for Serial Decomposition of FSM*, Proceedings of the International Workshop Control and Information Technology, pp. 229–234, VSB – Technical University of Ostrava, Ostrava (Czech Republic), September 2005.
6. G. Borowik, B. Falkowski, T. Łuba: *Cost-Efficient Synthesis for Sequential Circuits Implemented Using Embedded Memory Blocks of FPGA's*, Proc. of 10th IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems (2007), pp. 99–104, Cracow, Poland, 11-13 April 2007.
7. J. A. Brzozowski, T. Łuba: *Decomposition of Boolean Functions Specified by Cubes*, Journal of Multi-Valued Logic and Soft Computing, vol. 9, pp. 377–417, Old City Publishing Inc., Philadelphia, 2003.
8. J. Cong, K. Yan: *Synthesis for FPGAs with embedded memory blocks*, Proceedings of the 2000 ACM/SIGDA 8th International Symposium on Field Programmable Gate Arrays, ACM Press NY, pp. 75–82, Monterey (California), 2000.
9. R. Czerwiński, D. Kania: *State Assignment for PAL-based CPLDs*, Proc. of 8th Euromicro Conference on Digital Systems Design, Architectures, Methods and Tools, IEEE Computer Society, pp. 127–134, Porto, Portugal, 30 August – 3 September 2005.
10. G. De Micheli, R. K. Brayton, A. Sangiovanni-Vincentelli: *Optimal State Assignment for Finite State Machines*, IEEE Transactions on CAD, vol. CAD-4, no. 3, pp. 269–284, 1985.
11. S. Devadas, R. Newton: *Exact Algorithms for Output Encoding, State Assignment and Four-level Boolean Minimization*, IEEE Transactions on CAD, vol. 10, no. 1, pp. 13–27, 1991.
12. J. Hartmanis, R. E. Stearns: *Algebraic Structure Theory of Sequential Machines*, Prentice Hall, New York, 1966.
13. S. Hassoun, T. Sasao, R. Brayton (Eds.): *Logic Synthesis and Verification*, Kluwer Academic Publishers, New York, 2002.
14. L. Józwiak: *An Efficient Heuristic Method for State Assignment of Large Sequential Machines*, Journal of Circuits, Systems and Computers, vol. 2(1), pp. 1–26, 1991.
15. L. Jóźwiak, A. Ślusarczyk: *A new state assignment method targeting FPGA implementations*, Proc. of the 26th Euromicro Conference, vol. 1, pp. 50–59, Maastricht (The Netherlands), September 2000.
16. L. Jóźwiak, A. Ślusarczyk, A. Chojnacki: *Fast and Compact Sequential Circuits through the Information-Driven Circuit Synthesis*, Proceedings of the Euromicro Symposium on Digital Systems Design, pp. 46–53, Warsaw (Poland), September 2001.

17. A. Kraśniewski: *Cost-free Enhancement of Testability for Sequential Circuits Implemented Using Embedded Memory Blocks of FPGA's*, Proceedings of the 8th IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems, pp. 61–68, Sopron (Hungary), April 2005.
18. S. Krishnamoorthy, R. Tessier: *Technology Mapping Algorithms for Hybrid FPGAs Containing Lookup Tables and PLAs*, Proceedings of the IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 22(5), 2003.
19. H. Kubatova, M. Becvar: *FEL-Code: FSM Internal State Encoding Method*, Proc. of the 5th International Workshop on Boolean Problems, pp. 109–114, Technische Universitat Bergakademie, Freiberg, 2002.
20. T. Łuba, C. Moraga, S. Yanushkevich, M. Opoka, V. Shmerko: *Evolutionary Multi-Level Network Synthesis in Given Design Style*, Proc. of the IEEE 30th International Symposium on Multiple-Valued Logic, pp. 253-258, Portland, 2000.
21. T. Łuba, 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(3–4), pp. 289–300, 1995.
22. M. Perkowski, R. Malvi, S. Grygiel, M. Burns, A. Mishchenko: *Graph coloring algorithms for Fast Evaluation of Curtis Decompositions*, Proc. of Design Automation Conference, pp. 225–230, New Orleans, 1999.
23. M. Rawski, T. Łuba: *FSM Implementation in Embedded Memory Blocks Using Concept of Decomposition*, IFAC Workshop Programmable Devices and Systems, pp. 301-306, Gliwice, 2001.
24. M. Rawski, H. Selvaraj, T. Łuba: *An Application of Functional Decomposition in ROM-based FSM Implementation in FPGA Devices*, Journal of Systems Architecture vol. 51, pp. 424–434, Elsevier 2005.
25. C. J. Shi, J. A. Brzozowski: *An Efficient Algorithm for Constrained Encoding and Its Applications*, IEEE Trans. on CAD, vol. 12, no. 12, pp. 1813–1826, 1993.
26. C. Scholl: *Functional Decomposition with Application to FPGA Synthesis*, Kluwer Academic Publisher, Boston, 2001.
27. V. V. Solovjev: *Synthesis of Sequential Circuits on Programmable Logic Devices Based on New Models of Finite State Machines*, Proc. of Euromicro Symposium on Digital Systems Design, pp. 170–173, Warsaw, Poland, 4-6 September 2001.
28. U. Stańczyk, B. Pochopień: *Boolean Function Optimization in Binary Mathematical Morphology*, Proc. of IFAC Workshop on Programmable Devices and Systems, pp. 411–416, Cracow, 18-19 November 2004.
29. E. M. Sentovich, K. J. Singh, L. Lavagno, C. Moon, R. Murgai, A. Saldanha, H. Savoj, P. R. Stephan, R. K. Brayton, A. Sangiovanni-Vincentelli: *SIS: A System for Sequential Circuit Synthesis*, Memorandum, no. UCB/ERL M92/41, Electronics Research Laboratory, Department of Electrical Engineering and Computer Science, University of California, Berkley, 1992.
30. J. Qiao, M. Ikeda, K. Asada: *Optimum Functional Decomposition for LUT-Based FPGA Synthesis*, Proceedings of the FPL '2000 Conference, pp. 555–564, Villach, 2000.
31. T. Villa, T. Kam, R. K. Brayton, A. Sangiovanni-Vincentelli: *Synthesis of Finite State Machines: Logic Optimization*, Kluwer Academic Publishers, Boston, 1998.
32. T. Villa, A. Sangiovanni-Vincentelli: *NOVA: State Assignment for Optimal Two-level Implementations*, IEEE Transactions on CAD, vol. 9, no. 9, pp. 905–924, 1990.
33. S. J. E. Wilton: *SMAP: Heterogeneous Technology Mapping for FPGAs with Embedded Memory Arrays*, Proceedings of the ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, pp. 171–178, 1998.
34. S. Yang: *Logic Synthesis and Optimization Benchmarks User Guide*, ver. 3.0, Microelectronic Center of North Carolina, 1991.

On Application of Polynomial Algebra for Identification of Dynamic Faults in Interconnects

MICHAŁ KOPEĆ, TOMASZ GARBOLINO, KRZYSZTOF GUCWA, ANDRZEJ HŁAWICZKA

Silesian University of Technology
mkopec@onet.pl; {tgarbolino, kgucwa, ahlawiczka}@polsl.pl

Received 2008.01.03
Authorized 2008.02.28

The paper presents a new method that is an effective instrument for investigating sources of dynamic faults in interconnects (i.e. crosstalk, delay faults, etc.). It is an extension of the previous work of the authors published in the Proceedings of the European Test Symposium 2006, where fault identification was limited to static faults only. In the proposed approach an erroneous bit sequence coming from the faulty net is reconstructed on the basis of a set of signatures. This facilitates precise identification of dynamic faults.

Discussed method is applicable to interconnects between ICs mounted on the PCBs as well as interconnect networks connecting IP cores in SoCs. Moreover, it is easily scalable to any number of nets in the interconnect network and can be used with any type of test sequence and test pattern generator. There are several variant s of hardware implementation of the method. This supports finding a trade-off between area overhead and testing time.

Keywords: Interconnect test, Built-In Self Test, Interconnect BIST, Chinese Remainder Theorem, Fault localization, Fault identification

1. INTRODUCTION

In current deep-submicron technologies (DSM), dynamic failures are becoming common [11, 12, 13]. They are particularly associated with the interconnections between the IP in System-on-a-chip or with the external interconnections between the integrated circuits on a PCB. During the debug phase of a chip or a PCB, the knowledge of the erroneous signal can simplify the design modification. Even if it were possible to use an additional probe, its capacity could disturb the gigahertz digital signal.

During the chip debug phase, it would be important to know the signals at certain chip points in order to find out the reason of the failure. Such points may include the scan-chain flip-flops or certain interconnections. The problem of localization of the error-capturing scan-chain

flip-flops has been thoroughly studied in [3, 4] to mention a few. The key assumption is a small number of errors (scan-chain flip-flops that capture errors). This assumption is valid in case of scan-chains. In contrast, with the interconnections (between the IP cores, or on the PCB) the situation looks quite different. A small number of dynamic faults can generate a large number of errors across the interconnections.

Let us assume that a certain interconnection is stimulated with a sequence of length 250 that contains 25 rising edges and 25 falling edges. When the delay fault affects both edges, at least 50 errors will be observed in this interconnection. Moreover, some additional number of errors may be induced by the crosstalk. A large number of the response errors also occurs in case of a combination of static faults with dynamic ones (for example, a short and delay combination). Therefore, we investigated a method to determine an interconnection signal that could be insensible to the number of errors.

The serious drawback of the Reed-Solomon code technique [4] is the need to collect a huge number of signatures to detect a huge number of errors. This results from the requirements related to the linear codes used to correct a huge number of errors. Thus, this method is suitable for scan-chains only and not to diagnose the errors in the interconnections at all. In order to implement this method to diagnose the faults in interconnections in the above example, as high as 2t=2*50=100 (!) signatures would be required to collect and to repeat the test the same times. The probabilistic technique presented in [3] produces even worse results.

2. MOTIVATION

In [6], the authors propose a new technique of static fault identification based on a normalised signature derived from a traditional signature derived from the compaction of the test response of each interconnection. The normalised signatures of static faults (stuck-ats and shorts) have specific characteristics that allow for quick static fault identification. The most important feature of this approach is that the interconnect testing can be performed at-speed. However, the identification of dynamic faults is limited to the only conclusion that they are not static faults. No approach is indicated how to examine them. The extension of this approach characterised by the elimination of aliasing / crossaliasing (confounding) phenomenon and by a direct determination of the fault types based on the form of the signature was presented in [7].

Versatility is a key feature of the method presented in [6], since it allows to use any Test Pattern Generator (TPG) and it can be easily scaled to address the testing of any number of interconnections. Tests are performed at-speed, however the test results (the set of signatures) can be sent to the external tester at much lower clock frequency.

The paper will present a new method to determine the erroneous signal, the signature of which indicates a dynamic fault. The only additional cost is associated with a slight extension of a hardware mentioned in [6]. This approach is based on the Chinese remainder theorem. It does not limit the number of errors in any way and is independent on the structure of the interconnections. It is the best solution to handle a very large number of errors, but in case of a small number of errors in the interconnections it is worse than that described in [4], although the difference is rather small. Such situations happen hardly ever. As it was mentioned in the

introduction, the number of errors in an interconnection resulting from a single fault is usually large. The Chinese remainder theorem (CRT) has many applications (for example [5]). However in the area of the test technology, the CRT was only used to design a modular compactor of better diagnostic properties [9]. The paper [1] proposes an on-line detector, which detects on-line signal violations in interconnects. However it does not provide any additional information allowing to identify faults.

The organisation of the paper is as follows. Section 3 presents the concept related to the reconstruction of an erroneous signal based on the Chinese remainder theorem. Section 4 describes a hardware implementation of the proposed solution, whilst the scenario of the dynamic fault identification is illustrated in Section 5. The conclusions are summarized in Section 6.

3. THE CONCEPT

A polynomial representation of the sequences and binary vectors will be used in the next part of the paper. In the polynomial, the highest power of the variable x corresponds to the most significant bit of sequence or vector (the leftmost bit). For example, to the binary sequence 101101 corresponds the polynomial notation $a(x) = x^5 + x^3 + x^2 + 1$. We will also indicate the sequences and binary vectors and their corresponding polynomials using the hexadecimal notation. For example, to the vector 111101 and representing it polynomial notation $b(x) = x^5 + x^4 + x^3 + x^2 + 1$ corresponds the hexadecimal notation $b(x) = 3D$. We moreover assume that the binary sequences are applied at the interconnects starting from the LSB.

Let us assume, that a t-bit long sequence $w(x)$, the value of which we want to determine, is applied at the input 0 of n-bit register MISR illustrated in Figure 1. Hence, the following questions arise.

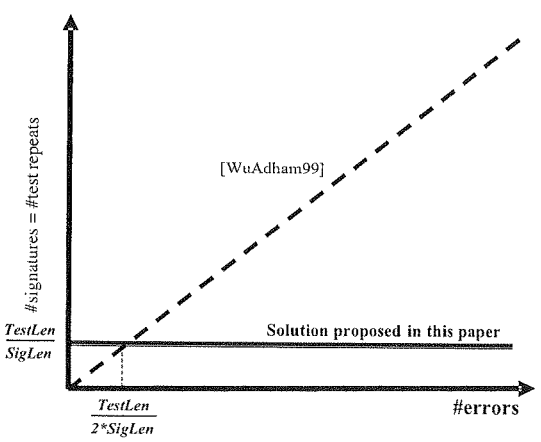

Fig. 1. A comparison between the technique [4] used to diagnose the errors in an interconnection and that presented in this paper

Problem 1

- Is it possible to determine unambiguously the sequence $w(x)$ knowing the set of its remainders under various modules $p_i(x)$ (i.e. at various feedbacks of the MISR) and how to do it?
- How many reminders are required to be known in order to determine the value of $w(x)$ unambiguously?

The Chinese remainder theorem [5] provides the affirmative answer and indicates how the sequence $w(x)$ can be determined. Let all symbols in the theorem and in its proof be the polynomials.

Theorem 1 (The Chinese remainder theorem)

There is given k-element set of the reminders $r_i(x)$ calculated according to various modules $p_i(x)$:

$$\{w(x) = r_i(x) \bmod p_i(x)\} \ (i=1...k) \tag{1}$$

We assume, that every two polynomials $p_i(x)$ and $p_j(x)$ $(i, j=1...k, i\neq j)$ are relatively prime (that is they do not have common factors in their prime factorization). There exists a common solution $w(x)$ of these congruencies and every two solutions are congruent modulo $P(x)$, where:

$$P(x) = \prod_{j=1}^{k} p_j(x) \tag{2}$$

The proof.

The complete proof is provided in [5]. Below, there is described only the manner of determination of the polynomial $w(x)$.

In order to determine the polynomial $w(x)$, we first have to calculate the quantities:

$$P_i(x) = \prod_{j=1, i\neq j}^{k} p_j(x) \tag{3}$$

For $i = 1...k$, the polynomials $p_i(x)$ and $P_i(x)$ are relatively prime (they have no common divisor other than unity):

$$GCD(p_i(x), P_i(x)) = 1 \tag{4}$$

Thus, there exist such values $P_i(x)$ and $Q_i(x)$ that satisfy the equation:

$$P_i(x)Q_i(x) = 1 \bmod p_i(x) \tag{5}$$

If so, always the value $Q_i(x)$ can be determined (e.g. using the Euclid's algorithm [10]), $(i = 1...k)$ such, that

$$Q_i(x) = (P_i(x))^{-1} \bmod p_i(x) \tag{6}$$

Finally, the searched sequence $w(x)$ is given as:

$$w(x) = \sum_{i=1}^{k} r_i(x)P_i(x)Q_i(x) \bmod P(x) \tag{7}$$

Corollary 1

Notice that the values $P_i(x)$, $Q_i(x)$ and $P(x)$ do not depend on the sequence $w(x)$ thereby so it will be sufficient to determine them only once. The sequence $w(x)$ is the sum modulo $P(x)$ of the reminders $r_i(x)$ (the signatures) and its determination involves no complex calculations.

Corollary 2

If $\deg(w(x))<\deg(P(x))$, this set of solutions is a single-element one. In other words, there exists only one value $w(x)$ satisfying the system of equations (1).

Corollary 3

In order to determine unambiguously the value of a t-long sequence $w(x)$ it is necessary to determine $k = \lceil t/n \rceil$ remainders $r_i(x)$ at k various modules $p_i(x)$, where $i = 1,2,\ldots, k$. The polynomials $p_i(x)$ have to satisfy the necessary conditions of Theorem 1.

This indeed supports our original intuition, that the knowledge of k of n-bit signatures permits to reproduce unambiguously the compacted sequence $w(x)$ if is not longer than $t = nk$ bits.

Corollary 4

The reliability of the method can be improved by reading the additional $k + 1$ signature. If the sequences determined on the basis of all k-element subsets of a $k + 1$ element set of signatures are identical, this means that the process of determination of the sequence $w(x)$ is error-free.

The following examples illustrate some practical aspects of the use of this Chinese remainder theorem based approach.

Example 1

In order to reproduce unambiguously 1000-bit sequence it is necessary to collect $\lceil 1000/32 \rceil = 32$ 32-bit signatures.

Example 2

The searched binary sequence was applied four times at the input No. 0 of a linear register having its feedback described successively by the following primitive polynomials:

$p1(x) = x^{16}+x^5+x^3+x^2+1 = 1002D$

$p2(x) = x^{16}+x^5+x^4+x^3+1 = 10039$

$p3(x) = x^{16}+x^5+x^4+x^3+x^2+x+1 = 1003F$

$p4(x) = x^{16}+x^6+x^4+x+1 = 10053$

The following set of signatures was obtained: $\{S_i(x)\}=\{A3AC,44A7,953F,3B41\}$. Based on this set of signatures, there was determined a sequence having its hexadecimal value of EF0123456789ABCD, which corresponds to a binary sequence:

1110111100000001001000110100010101100111100010011010101111001101

Example 3

The sequence S illustrated in Table 1 below was applied by the TPG to the faulty interconnect. Based on the sequence R obtained at the output, the right number of signatures was determined. The value of the sequence R was determined with the use of the algorithm proposed in Theorem 1.

It can be seen from the comparison between these two sequences S and R that the delay of the rising edge is t_r=1, the delay of the falling edge is t_f=2, whilst the short impulse 0 could be a result of a crosstalk. Further interpretation of the results will depend on the structure and other specific properties of the network under test (NUT).

Table 1

Test sequence S and response sequence R

S:	00000000111111110000000011111111000000001111111
R:	00000000111111111000000011111111100000001110111

In the section above there was presented the method to determine the sequence $w(x)$ applied at the input number j=0 of an n bit MISR. Now, the next question is.

Problem 2

If and how the values of the sequence $w(x)$ applied at one of the remaining inputs number j=1 ... n-1 of the MISR illustrated in Fig. 1 can be determined using Theorem 1.

The solution of this problem represents the Property 1.

Property 1

Let us denote by $r_{i,j}(x)$ the reminder, when the signal $w(x)$ is applied at the input j-th of the MISR illustrated in Fig.2. The following is true: $r_{i,j}(x)=x^j r_{i,0}(x) \bmod p_i(x)$ (see [6]). Obviously, $r_{i,0}(x)=r_i(x)$.

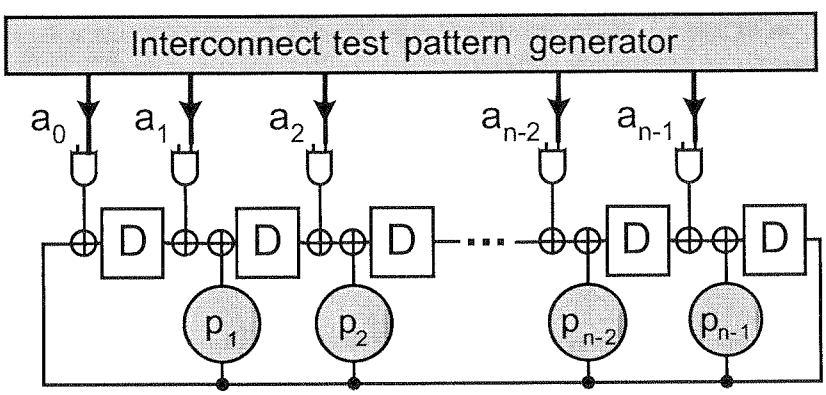

Fig. 2. A generic n bit MISR with a programmable feedback. The MISR performs compaction of the test responses of an interconnection network

Let us denote by $r_{i,j}(x)$ the reminder, when the signal $w(x)$ is applied at the input j-th. Obviously $r_{i,0}(x)=r_i(x)$.

This draws ahead $r_{i,j}(x)=x^j r_{i,0}(x) \bmod p_i(x)$ (see [6]) and $r_{i,0}(x)=x^{-j} r_{i,j}(x) \bmod p_i(x)$

Example 4

Using a MISR with its polynomials $\{p_i(x)\}$={1002D,10039,1003F,10053} and applying the signal $w(x)$ at the input j = 1, the set of signatures $\{r_{i,1}(x)\}$={51D6,A24F,CA80,9D89}

was obtained. The problem is to determine the signatures (the reminders) that would have been obtained if the same signal $w(x)$ had been applied at the input j=0. Using Property 1, we receive $\{r_i(x)\}$={A3AC,44A7,953F,3B41}

The possible practical applications of the proposed here method of the determination of the sequence on the basis of the corresponding set of the reminders (signatures) are provided in the next chapter.

4. THE HARDWARE

As it was already mentioned in section 2, the authors decided to implement the proposed method and use the enhanced hardware [6] to detect, localize and identify static faults in the interconnections. Fig. 3. is a block diagram of the hardware, whilst Fig. 4. is a schematic diagram of a single COMP module. The main unit of the COMP is a 32-bit MISR that is connected at its inputs with the lines of the NUT via the AND gates controlled by the CEN (Compaction ENable) signal. Also, the outputs of a 32-bit CSR (Control Shift Register) are connected to the

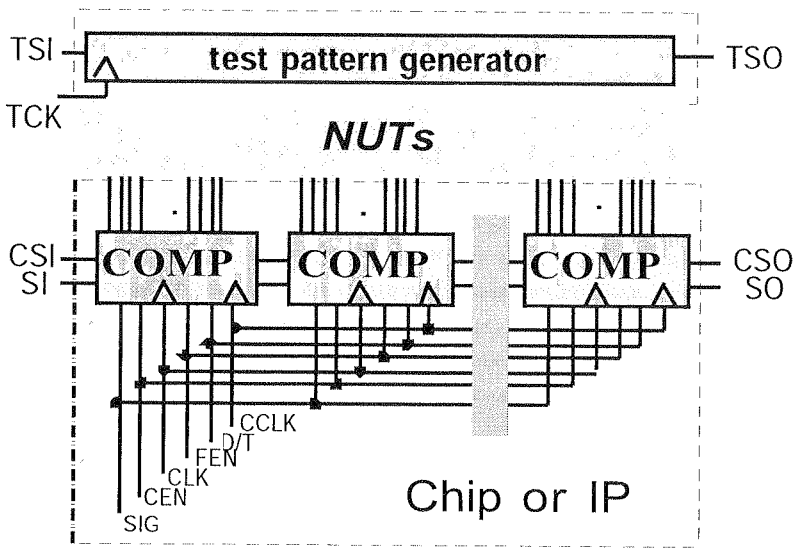

Fig. 3. Block diagram of the testing hardware [6]

inputs of these AND gates. The MISR is composed of T flip-flops that can function as either D flip-flops (D/T = 0) or T flip-flops (D/T = 1), depending on the value of the signal D/T. The same register has also the serial input SI gated by the signal SIG and the output SO. The feedback can be disabled by setting the signal FEN to 0. Table 2 provides the basic modes of operation of the module COMP and the corresponding values of the control signals. The „Scan mode" serves to set the initial state of the MISR and to output the signature. The „Parallel compaction" mode runs a parallel compaction of the test responses of all connections, whilst the resulting

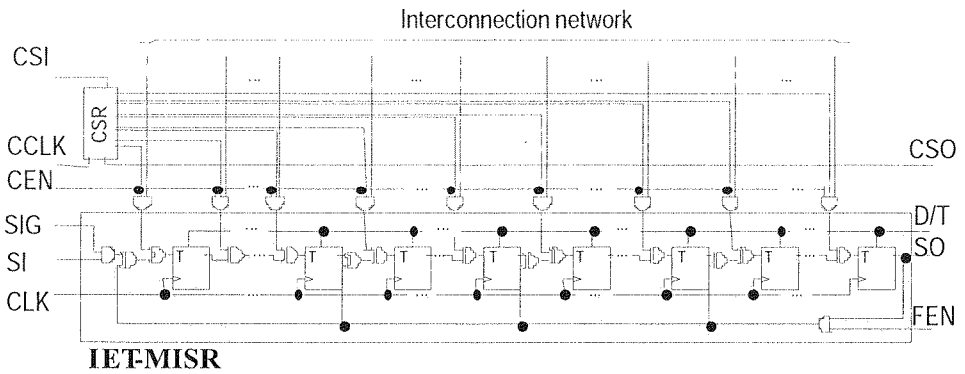

Fig. 4. The structure of a single COMP module [6]

Table 2

Modes of operation of the COMP module

Mode	Function	FEN	D/T	CEN	SIG	CSR
Scan mode	Initialisation / signature reading	0	0	0	1	xx...xxx...xx
Parallel compaction in a chain of COMPs	MISR	1	1	1	1	11...111...11
Serial compaction in a single COMP	SISR	1	1	1	0	00...010...00

signature makes possible to detect the faults in the interconnections. In this mode, the signal D/T = 1 and all MISR flip-flops work as T flip-flops. The authors decided to use in this mode a MISR composed of T flip-flops, because they exhibit better antialiasing properties than the classic D flip-flop - based MISR in case of the testing sequences commonly used in testing of the interconnections [8]. The „Serial compaction" mode is used during the localization and identification of faults in the interconnections. The MISR is used to compact the set of responses associated with a single interconnection only, indicated by a mask written into the CSR. Based on the received signature, the interconnection is considered either faulty or fault-free. If the interconnection is faulty, the external tester determines the type of fault based upon the signature value and the fault simulation results.

Fig. 5 illustrates the module PCOMP that is an enhanced version of the module COMP. The added components are marked in grey colour. These are a 31 -bit PREG shift register, thirty one XOR gates and thirty one 2-input AND gates. In the „Scan mode", a PMISR feedback programming vector (PMISR is a MISR with Programmable feedback) is fed serially to the PREG. Fig. 6 a, b and c illustrate three block diagrams of a compactor that employs PCOMP modules.

As it can be seen in Fig. 6a, only the module No. m is a PCOMP-type one, whilst the other modules 1 to m-1 are of a COMP–type. This circuit serves to analyse the set of the test responses coming only from one chosen connection. The MISRs in the COMP modules should be configured in the manner to create a single MIShR (Multi Input Shift Register) having the length of $(m\text{-}1)n = 32(m\text{-}1)$ bits. To do this, it is necessary, for example, to add the input MEN (MIShR

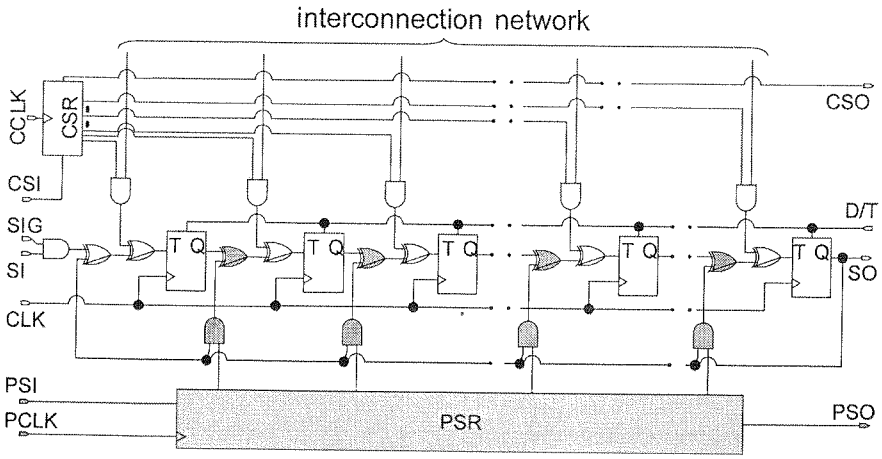

Fig. 5. The structure of a single PCOMP module

ENable)which controls the input SIG of the COMP1 and to gate the input of the MIShR. If the modules COMP are required to work as a single long MIShR, then MEN = 0. The PCOMP works as a 32-bit PMISR having its serial input SI connected with the serial output SO of the MIShR itself. The values of masks introduced to the CSRs in all COMPs and in the PCOMP must assure that the sequence related to the chosen interconnection is applied, via an AND, at one of the parallel inputs of the corresponding COMP/PCOMP. In case, when the chosen interconnection is brought to the j-th input in the i-th COMP, the MIShR will function as an intermediate w-bit shift register – where $w=[(m-1)-i]n+(n-j)=(m-i)n-j$ and $i=1..m-1$, $j=0,..,n-1$, $n=32$ to bring the set of responses to the input SI of the PMISR in the PCOMP. In order to assure correct operation of the MIShR it is necessary to initially reset the MISRs in all COMPs. The PMISR in the PCOMP can also serve to compact the sequences coming from the interconnections brought directly to its $n = 32$ parallel inputs.

In order to reproduce the set of responses, the interconnect network requires that a t-clock pulse long testing sequence is applied k times, and that k signatures, where $k = \lceil t/n \rceil$ and $n = 32$, are determined in the PCOMP. It is possible to reduce the number of the necessary signatures, thereby reducing the testing time. Let us assume that the given set of test responses is applied at the j-th input of the i-th COMP. Then, it is sufficient to compact only first t-w bits of the set of responses in the MISR of the PCOMP (the formula for the value of w is given hereinabove). The remaining w bits of the sequence is to be merely introduced into the MIShR, from which they could be directly output once into an external tester (ATE) in „Scan mode", together with the first signature. It is sufficient to determine the $k' = \lceil (t-w)/n \rceil$ signatures, which requires only that the testing sequence is applied k' times to the interconnections. Notice, that $k' < k$. In total, $b'=w+k'n$ bits have to be sent to the external ATE (w-bit portion of the test response and k' n-bit signatures), where $t+n>b' \geq t$. In the optimal case, when the only faulty interconnection is that brought to the input $j = 0$ of the COMP1, then $k' = \lceil [t-(m-1)n]/n \rceil = \lceil t/n-(m-1) \rceil$ and $b'=(m-1)n+\lceil t/n-m-1 \rceil n= [(m-1)+\lceil t/n-(m-1) \rceil]n = \lceil t/n \rceil n$. The above mentioned reduction of testing time is not possible only in the case of having $n = 32$ interconnections brought directly to the parallel inputs of the PMISR

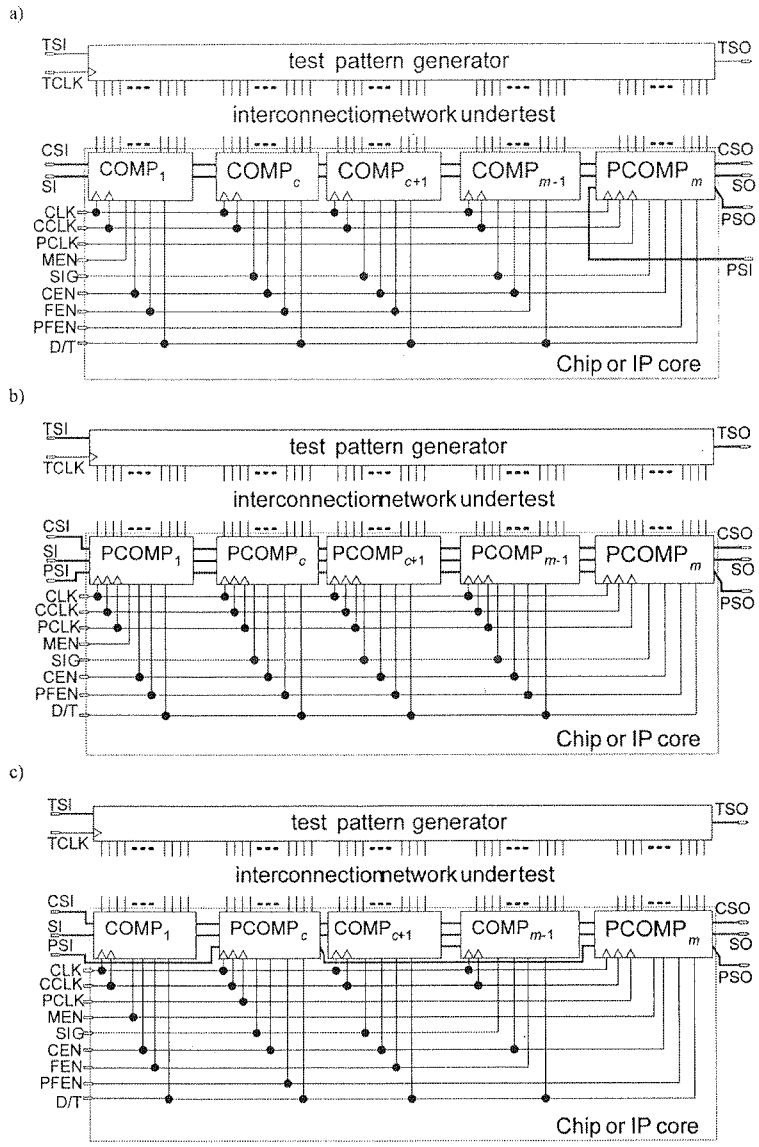

Fig. 6. The block diagram of the proposed testing hardware: (a) containing a single PCOMP module (low speed & overhead), (b) all modules are PCOMP type (high speed & overhead), (c) containing two PCOMP modules (medium speed & overhead)

in the PCOMP. Here, the interconnect network requires that a t-bit sequence is applied k times, and that $b= kn$ bits (i.e. k n-bit signatures), where $t+n>b\geq t$ are output to the external ATE.

In the solution depicted in Fig. 6b, all modules are PCOMP-type. This allows to compact simultaneously the sequences of test responses coming from maximum m interconnections, provided that each interconnection is brought to one of the parallel inputs of another PCOMP.

In the worst case, when $n = 32$ of faulty interconnections are brought to all parallel inputs of the first PCOMP ($PCOMP_1$), the interconnect network requires that the testing sequence is applied nk times, where $k = n\lceil t/n \rceil$. Then, kmn bits must be sent to the external ATE. In the optimal case, when the only faulty interconnection is that brought to the input of the $PCOMP_m$. Here, the interconnect network requires that a t-bit sequence is applied k times, and that $b = kn$ bits (i.e. k n-bit signatures), where $t+n>b\geq t$ are output to the external ATE.

The circuit illustrated in Fig. 6c employs two modules: $PCOMP_c$ and $PCOMP_m$. A set of $COMP_e$ modules, where $e = 1,2,...,(c-1)$, along with the $PCOMP_c$ provides the solution analogous to that illustrated in Fig . 6a. A similar situation is when considering the group of the $COMP_f$ modules, where $f=(c + 1), (c + 2),...,(m-1)$, along with the $PCOMP_m$. Notice only, that now the signals SIG of the COMP1 and COMP(c + 1) are connected to the global signal MEN. If we assume $c=\lceil m/2 \rceil$, then we can in the best case obtain approx. twofold reduction of the testing time of that compactor structure compared to the circuit presented in Fig. 6a.

Depending on the requirements relating to the admissible hardware overhead and the duration of the testing time, the proposed solution allows to reach a compromise between the above mentioned two factors. It is possible to use a single PCOMP only to ensure a small overhead at the cost of long testing time. This latter might be reduced by increasing the number of PCOMPs what, however, will imply a larger hardware overhead. When using solely PCOMPs, the testing time minimizes whilst the hardware overhead increases.

In the next section there is provided a detailed description of the testing scenario for the circuit illustrated in Fig 6a.

5. THE TESTING SCENARIO

This section deals with the testing scenario to be used for the circuit shown in Fig. 6a. The description for the circuits illustrated in Fig 6b and 6c are similar. The fragments relating to the verification of the functioning of the hardware designed for interconnection testing, fault detection and localization of faulty interconnections and the identification of the type of static faults were described in details in [6]. Focus is instead given to the fragment relating to the reproduction of the set of test responses of a faulty interconnection, what is necessary to identify dynamic faults.

- Verification of correct operation of the hardware used to test the interconnections [6]. If the hardware is faulty, FINISH
- Detection of faults in the interconnections [6]. If the final signature is correct - that is no fault was detected - FINISH
- Localization of faulty connections and identification of the type of static faults [6]. If only static faults appear, FINISH
- In case of each interconnection manifesting the dynamic fault for which the value of the set of test responses should be determined:
  - load the set of masks into the CSR in the COMP and PCOMP
  - load a primitive polynomial $p_f(x)$ into the PREG in the PCOMP

– reset all COMPs and the PCOMP by setting SIG=MEN=CEN=D/T=FEN=PFEN=0 and applying 32 CLK cycles

– set SIG=PFEN=CEN=1, FEN=MEN=D/T=0

– start the test lasting t cycles,

– set SIG=1, MEN=FEN=PFEN=CEN=0. Output serially the reminder (signature) $r_1(x)$ from the PCOMP or alternatively the fragment of the set of responses from the COMP (see Section 4),

– reset all COMPs and the PCOMP by setting MEN=SIG=CEN=D/T=FEN=PFEN=0 and applying 32 CLK cycles

– for i=2 to k'

◊ load the next primitive polynomial $p_i(x)$p into the PREG

◊ set SIG=PFEN=CEN=1, MEN=FEN=D/T=0

◊ start the test lasting t cycles,

◊ set SIG=MEN=PFEN=FEN=CEN=D/T=0 and output serially the reminder (signature) $r_1(x)$ from the PCOMP.

– based on the obtained set of the k' signatures, determine the sequence of test responses of the interconnection under test.

THE END

6. CONCLUSIONS

In the paper, we proposed a new tool to investigate the causes of dynamic faults of interconnections (crosstalks, delays and other) during the silicon debug phase or on a PCB. This is an extension of the method proposed in [6] in which the identification was limited to the static faults only. By reproducing the faulty signal on the basis of the set of signatures it is possible to identify the dynamic faults. The method can be applied to any number of interconnections between the chips or IP and allows using any test generator. A compromise is possible between the hardware overhead and the testing time.

Compared to the approach proposed in [4], this new tool has the following advantages:

- it employs a simple equation instead of the necessity of solving complex equations over GF(2^n)
- it employs a classic IE-MISR instead of a PMISR functioning in GF(2^n)
- it does not have any limitations on the number of faults.

In addition, exact location and identification might be necessary to the reconfiguration.

Because of small delays introduced by linear registers, this method can assure a high diagnostic resolution. Testing is carried out at-speed, whilst the set of signatures can be send to the external tester at much lower clock frequency. The drawback is associated with the necessity of the repetition of the test, but this procedure is to be executed only once, during the chip prototyping phase.

Further advantage of the proposed approach is that we analyse the whole sequence of the chosen interconnection and determine its signature. Thus, no need exists to introduce fault-sensitive signal to control memorizing of interconnect response in the shift register flip-flops.

Let us imagine a delay at the falling edge of that control signal. There will appear "phantom" errors in the reconstructed interconnect sequence. Similar effect can be produced by glitch faults in the control signal line. It is therefore recommended to avoid introducing such additional control signal.

The hardware overhead of this approach is similar as that in [2] (approx. 2FF / the input connection). It guaranties full identification of both static and dynamic faults. It can be increased to 3FF for chosen connections, thereby obtaining shorter testing times. The hardware components (COMP and PCOMP) can be easily integrated with the P1500 wrapper's serial structure.

7. REFERENCES

1. A. Attarha, M. Nourani: "Testing Interconnects for Noise and Skew in Gigahertz SoC", Proc. of Int. Test Conf., 2001, pp. 305-314.
2. A. Jutman: "At-Speed On-Chip Diagnosis of Board-Level Interconnect Faults", Proc. of ETS'04, pp. 2-7.
3. J. Rajski, J. Tyszer: "Diagnosis of Scan Cells in BIST Environment", IEEE Trans. Comp., Vol. 48, No. 7, 1999, pp. 724-731.
4. W. Yuejian, Saman M. I. Adham: Scan-Based BIST Fault Diagnosis, IEEE Trans. on CAD, Vol. 18, No. 2, 1999, pp. 203-211.
5. N. Koblitz: „A Course in Number Theory and Cryptography", Springer-Verlag, New York, 1994.
6. M. Kopeć, T. Garbolino, K. Gucwa, A. Hławiczka: „Test-per-clock detection, localization and identification of interconnect faults", Proc. ETS 2006, pp. 233-238.
7. T. Garbolino, M. Kopeć, K. Gucwa, A. Hławiczka: "Detection, Localisation and Identification of Interconnection Fault Using MISR Compactor", Proc. DDECS 2006, pp. 230-231.
8. A. Hławiczka, K. Gucwa, T. Garbolino, M. Kopeć: „Can a D flip-flop based MISR compactor reliably detect interconnect faults?" Proc. DDECS 2005, pp. 2-8.
9. W. Rajski, J. Rajski: „Modular Compactor of Test Responses," pp. 242-251, 24th IEEE VLSI Test Symposium, 2006.
10. V. Shoup: NTL: A Library for doing Number Theory, version 5.4. http://shoup.net/ntl/.
11. M. Cuviello, S. Dey, X. Bai, Y. Zhao: Fault Modeling and Simulation for Crosstalk in System-On-Chip Interconnects. Proceedings of the International Conference on Computer-Aided Design (ICCAD), pp. 297-303, 1999.
12. X. Aragones, J.L. González, F. Moll, A. Rubio: "Noise Generation and Coupling Mechanisms in Deep-Submicron ICs," IEEE Design & Test Computers, vol.19, no.5, pp.27-35, September-October 2002.
13. K. Shu-Min Li, Chung Len Lee, Chauchin Su, Jwu E Chen: „A Unified Approach to Detecting Crosstalk Faults of Interconnects in Deep Sub-Micron VLSI," pp. 145-150, 13th Asian Test Symposium (ATS'04), 2004.

com
is hi
at-sp
❑ T
PT c
❑ T
dept
❑ T

betw
regis
well

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2008, 54, no 1, pp. 43–51

On Design of High Speed Test Pattern Generators Based on Ring LFSRs

ANDRZEJ HŁAWICZKA, TOMASZ GARBOLINO

Silesian University of Technology
Institute of Electronics
ul. Akademicka 16, 44-100 Gliwice, Poland
ahlawiczka@polsl.pl, tgarbolino@polsl.pl

Received 2008.02.14
Authorized 2008.03.20

The paper describes a method of designing ring registers that have short feedback connections and contain cells functioning as D or T flip-flops. The proposed approach enables obtaining a ring register operating with a maximum frequency for a wide range of polynomials for which existing methods are unable to provide optimal results. The paper also contains a number of examples illustrating different techniques of designing ring registers.

Keywords: linear ring register, linear feedback shift register, operating frequency

1. INTRODUCTION

Linear registers are commonly used as random test pattern generators (TPG) as well as main components of deterministic and mixed-mode TPGs [8, 1, 7]. One of the desired features of each TPG is high operating frequency, what is particularly important if a TPG is to be used as a component of an at-speed testing BIST. The following factors influence an operating frequency of a linear register:

- The length of nets connecting cells of a linear register what has impact on propagation time PT of the nets,
- The number XL of layers of XOR gates in feedback paths of the register (the feedback logic depth),
- The fan-out FO at the outputs of cells of the register.

In order to obtain a linear register working with maximum frequency the connections between cells of the register should be kept as short as possible. Moreover, the structure of the register needs to provide the minimum number of layers of XOR gates in its feedback paths as well as the minimum fan-out at the outputs of its cells, i.e. $XL \leq 1$ and $FO \leq 2$.

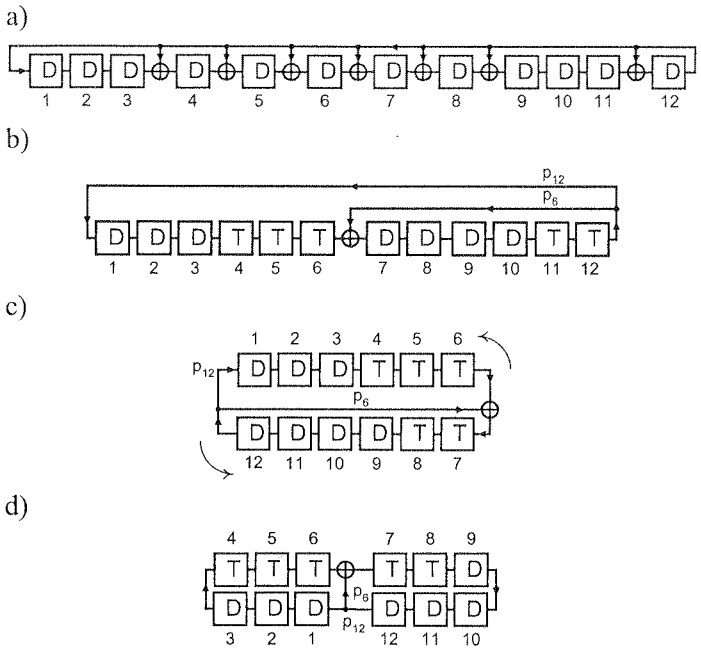

Fig. 1. Various structures of linear registers implementing polynomial $p_1(x)=1+x^3+x^4+x^5+x^6+x^7+x^8+x^{11}+x^{12}$: (a) classic LFSR structure, (b) LFSR containing D- and T-type flip-flops, (c) ring register before rotation

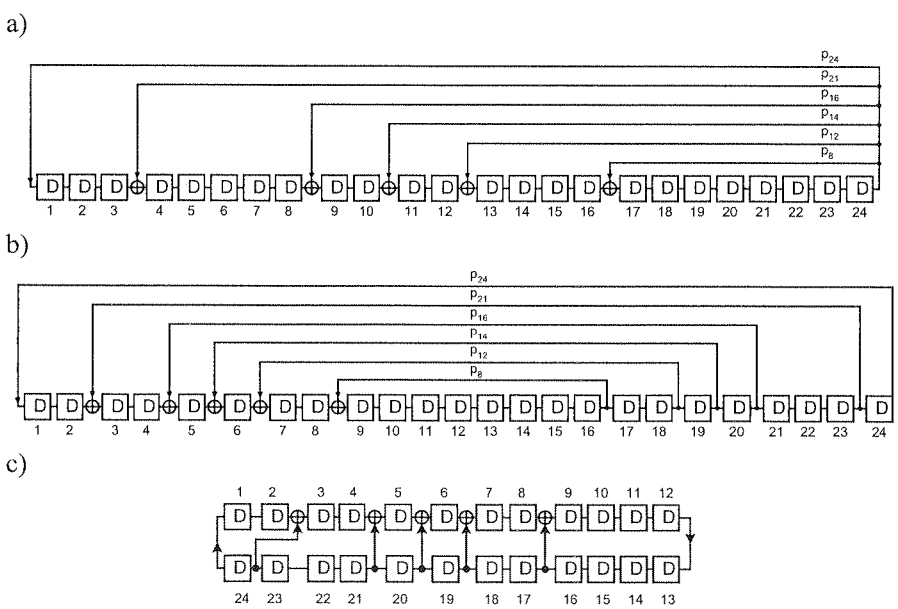

Fig. 2. Various structures of linear registers implementing polynomial $p_2(x)=1+x^3+x^8+x^{10}+x^{12}+x^{16}+x^{24}$ [6]: (a) classic LFSR structure, (b) LFSR with shifted feedback loop connections, (c) ring register structure

Recently the authors of [3, 4] and [5, 6] have described a specific kind of linear registers called ring registers. A ring register is produced by segmenting a long chain of flip-flops of a classical LFSR into two short series arranged one under another (see Fig. 1d and Fig. 2c). In [3, 4], the authors explain how to optimize designing process for such a ring structure by transforming all long connection loops into short links, thereby minimizing the fan-out FO and the propagation time PT. Their approach differs from the technique proposed in [5, 6]. The authors of [3, 4] used a transformation of the characteristic polynomial p(x) into its identical ring-like structure and methods that served to rotate rings, whilst the papers [5, 6] exploited an approach of left or right shifting of each long connection loop across a series of flip-flops.

Thus, due to the use of two different methods, each approach is dedicated to handle a separate group of characteristic polynomials (i.e. usually, the given characteristic polynomial can be successfully handled by one of the above-mentioned approaches but not by both).

The first method (Method 1) is dedicated exclusively for the use with the polynomials p(x) that are transformable into one of the following ring forms of the structural polynomials $p_{wr}(x)$ (see [3] – Table 5.1 on page 231):

$$1+x^b(1+x)^f;\ 1+x^b[1+x^c];\ 1+x^b(1+x)^f[1+x^c];$$
$$1+(1+x)^f[1+x^c];\ 1+x^b(1+x)^f[1+x^c(1+x)^g];\ 1+x^b[1+x^c(1+x)^g];\ 1+(1+x)^f[1+x^c(1+x)^g]; \quad (1)$$

where every variable x corresponds to a D-type flip-flop of a linear register [2, 3, 4] whilst every binomial (1+x) corresponds to a T-type flip-flop [2, 3, 4] or a (⊕D) cell functioning as a T-type flip-flop [2, 3, 4]. Each of these forms reflects ring-like structure that is similar to the one illustrated in Fig. 1d and has only a single short connection between the bottom row of flip-flops and the XOR gate located in the top row of flip-flops. The main drawback of Method 1 is that it cannot be applied to characteristic polynomials other than these specified by generic formulas (1).

The second method (Method 2) is designed to handle polynomials having different odd numbers of non-zero coefficients that are not associated with the consecutive powers of variable x, like for example ... $x^i + x^{i+1}+x^{i+2}+x^{i+3}$... [6]. For majority of polynomials having different odd number of non-zero coefficients but associated with consecutive powers of variable x, like for example ... $x^i+x^{i+1}+x^{i+2}+x^{i+3}$ ..., Method 1 is useful. As it was found in [6], with such polynomials, it is impossible to apply Method 2 to obtain a ring register with a minimum fan-out FO ≤ 2.

Each of methods presents a different philosophy and each of them is applicable to a different set of characteristic polynomials. What is important, the sum of these two sets does not contain all possible polynomials. Therefore, the main objective of this paper is to propose a new method that is applicable to polynomials that cannot be successfully handled by either of existing approaches. In this context the term "successfully handled" corresponds to obtaining a ring register implementing the given polynomial and operating with the maximum frequency.

Further sections of the paper deal with the following topics: Section 2 describes procedures that are exploited in two above mentioned methods of designing ring registers; Section 3 proposes a novel method of designing ring registers; the final Section 4 summarizes obtained results.

2. EXISTING METHODS OF DESIGNING RING REGISTERS

Generally, structures of the linear registers associated with the assumed (e.g. primitive) characteristic polynomial p(x) are given as respective tables, containing the information about

these cells of the register that should be connected to XOR gates to obtain a linear register corresponding to its (primitive) characteristic polynomial given in a table. In [2, 3, 4] there has been developed a different approach to designing linear registers with different feedback logic, including also ring registers: the authors propose to replace cumbersome and hardly available tables with a simple and versatile technique. Such technique requires that the source polynomial p(x) is transformed to bear its structural form $p_w(x) = p(x)$, from which the feedback structure of the linear register is taken directly.

2.1. METHOD 1

Fig.1a shows a classic LFSR implementing polynomial $p_1(x)=1+x^3+ x^4+ x^5+ x^6+ x^7+ x^8+ x^{11}+ x^{12}$. The polynomials p(x) can be transformed to the equivalent structural form $p_{wr1}(x) = 1+x^3(1+x)^3[1+x^4(1+x)^2] = p_1(x)$, to which the following simplified symbolic representation corresponds: $D_1D_2D_3T_4T_5T_6 \oplus D_7D_8D_9D_{10}T_{11}T_{12}$ or even a simpler one: $DDDTTT \oplus DDDDTT = D^3T^3 \oplus D^4T^2$. It illustrates the distribution of cells containing D or T flip-flops of the linear register illustrated in Fig.1b which has a classic serial structure containing two long loops p_{12} and p_6 (the lower index denotes the number of flip-flops within a loop). If T flip-flops are replaced with cells containing a D flip-flop with a XOR gate on its input, and denoted with (⊕D), then the simplified symbolic representation will take the form:

$$DDD(\oplus D)(\oplus D)(\oplus D) \oplus DDDD(\oplus D)(\oplus D).$$

In order to eliminate the loop p_{12}, a ring structure for such a register can be used (Fig.1c). To do this, the series of flip-flops from Fig.1b is broken in two halves and the resulting first six flip-flops are placed in the top row whilst the next six flip-flops are placed in the bottom row respectively. We assume that such quantitative partition of the flip-flops between the upper and lower rows will be consequently followed in further examples. Here, the long loop p_{12} has been transformed into a short, local connection. Unfortunately, the second loop p_6 has persisted that is as long as six flip-flops (see Fig.1c). In order to transform the p_6 into a short, local connection, the whole ring needs to be rotated in steps, for example to the left, by every flip-flop until the moment when the signal source of the loop p_{12} (the output of the flip-flop D_{12}) is positioned in the bottom row just opposite to the XOR gate in the top row. Rotation of the whole exemplary ring in steps by every flip-flop is accomplished by shifting a flip-flop from the least significant position within the top row to the most significant position of the bottom row and simultaneously shifting a flip-flop from the least significant position within the bottom row to the most significant position within the top row. Such process is similar to that used in a bicycle to rotate the chain every gear to the left. The resulting ring structure is illustrated in Fig.1d.

Observation 1: During a step left (right) rotation of the ring, the length of the loop p_6 is reduced. By shifting the half flip-flops situated before the XOR gate and placing them from the top row to the bottom row and making the same operation with the identical number of flip-flops situated after the XOR gate, i.e. shifting the bottom row flip-flops to the top row, we will transform the loop p_6 in the shortest local connection.

As it can be seen from the above example, the given classic LFSR can be transformed by the Method 1 into the corresponding ring register structure with a maximum operating frequency

provided that its characteristic polynomial can be transformed into one of the expressions described in (1). On the other hand, all classic LFSRs described by characteristic polynomials that cannot be transformed into any form included in (1) cannot be implemented in the form of the ring register working with maximum frequency. A good example of a polynomial belonging to this group is a primitive characteristic polynomial $p(x) = 1+x^3+x^8+x^{10}+x^{12}+x^{16}+x^{24}$ with seven non-zero coefficients [6]. For such cases, the Method 2 should be applied.

2.1. METHOD 2

Fig.2a represents a classic form of the linear register with the feedback logic associated with the polynomial $p_2(x) = 1+x^3+ x^8+ x^{10}+ x^{12} +x^{16}+ x^{24}$. The output of the utmost flip-flop D_{24} is the source of signal for six loop-like long connections p_{24}, p_{21}, p_{16}, p_{14}, p_{12} and p_8. The value of the fan-out at the output of this flip-flop is FO = 6. The fan-out can be reduced by employing solution proposed in [5, 6]. The authors of [5, 6] have proven, that the shifting of any loop to the left (right) to the output of any flip-flop staying before the flip-flop D_{24} will not change the characteristic polynomial associated with the structure of the linear feedback so transformed. Hence, by shifting every loop to the left and locating the loops more or less symmetrically in relation to the centre (the output of the flip-flop D_{12}) of the row of twenty four D-type flip-flops we will obtain the scheme (illustrated in Fig. 2b) of the linear register associated with the polynomial $p_2(x)$. After this row of flip-flops has been broken down into two halves, we obtained the ring register as illustrated in Fig. 2c.

Observation 2: Locating the shifted loops more or less symmetrically in relation to the centre of the linear register guarantees obtaining a ring register with the shortest possible length of the local connections (Fig. 3) after the row of flip-flops is broken down into two halves.

Note, that this method does not employ the rotation of the ring from Method 1. It is also easy to see from the example, that the signal source supplying each of four loops has been assigned to another flip-flop, thereby reducing the fan-out from FO = 6 down to FO ≤ 2. Similarly, due to

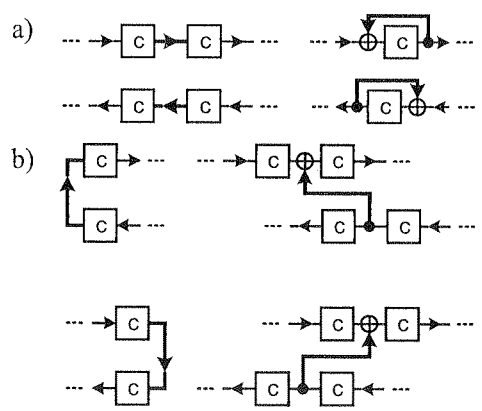

Fig. 3. Examples of short local connections in ring register with cell c∈{D,T}: (a) in top or bottom cell row, (b) between top and bottom cell rows

the location of each feedback XOR gate between different pairs of D-type flip-flops resulted in the minimum number of layers of the XOR gates in the feedback paths XL ≤ 1. Unfortunately, it is obvious that such optimal shifting of the loops is impossible in the case of polynomials with non-zero coefficients, which are associated with consecutive powers of variable x, like for example ... $x^i + x^{i+1}+x^{i+2}+x^{i+3}$ ... [6]. Then, the value of FO ≤ 2 is obtained at the cost of increased coefficient XL higher than unity or the value of XL ≤ 1 is obtained at the cost of the increased FO (higher than 2).

The above analysis of pitfalls and opportunities of Method 1 and Method 2 has become the basis for the development of a novel, versatile and uniform approach that is free of these drawbacks and can be applied to a much wider range of characteristic polynomials p(x) than the above-mentioned methods.

3. A NEW METHOD OF DESIGNING RING REGISTERS

The proposed approach is in fact the result of combining techniques that are involved in both methods discussed in previous section. Owing to this feature it is applicable to a large group of the polynomials that cannot be successfully handled by existing methods.

Polynomial $p_3(x) = 1+x^4+x^{11}+x^{13}+x^{14}+x^{15}+x^{16}+x^{17}+x^{18}$ is a good example of a characteristic polynomial that, if handled separately by either method presented in Section 2, results in the ring register with the fan-out FO>2 and/or the number of layers of XOR gates in some feedback paths XL > 1.

The classic form of the linear register implementing the above-mentioned polynomial $p_3(x)$ is shown in Fig. 4. The value of the fan-out at the output of the utmost flip-flop D_{18} is FO = 8.

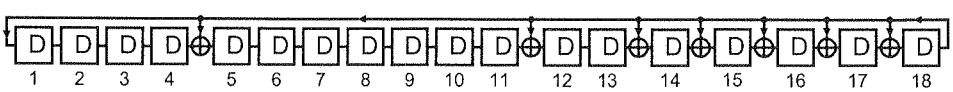

Fig. 4. Classic LFSR structure implementing polynomial $p_3(x) = 1+x^4+x^{11}+x^{13}+x^{14}+x^{15}+x^{16}+x^{17}+x^{18}$

The polynomial $p_3(x)$ can be represented in the equivalent structural form $p_{wr3}(x) = 1+(1+x)^7x^4[1+x[1+x^6]]$. Note that the polynomial $p_{wr3}(x)$ does not correspond to any generic form described by (1). Fig. 5 presents a structure of the ring register resulting from application of the Method 1 to the considered polynomial $p_3(x)$. The fan-out at the output of the flip-flop D_{18} in this ring register is FO = 3.

If, in turn, the Method 2 is solely used to transform the classic LFSR from Fig. 4, one can either reduce the fan-out to FO≤2 at the cost of increasing the number of layers of XOR gates

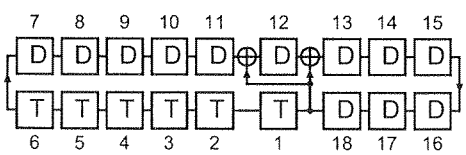

Fig. 5. Ring register structure resulting from application of Method 1 to polynomial $p_3(x)$

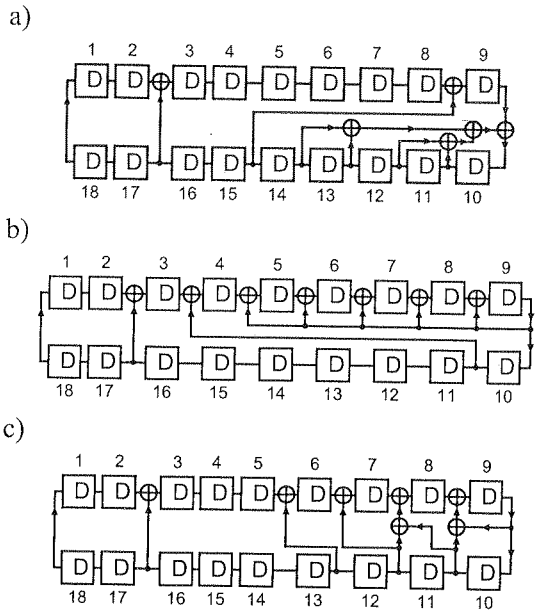

Fig. 6. Various ring register structures resulting from application of Method 2 to polynomial $p_3(x)$: (a) ring register with fan-out FO≤2 reduced at the cost of feedback logic depth XL=3, (b) ring register with feedback logic depth XL≤1 reduced at the cost of fan-out FO=5, (c) ring register with minimum feedback logic depth XL≤2 and fan-out FO≤3

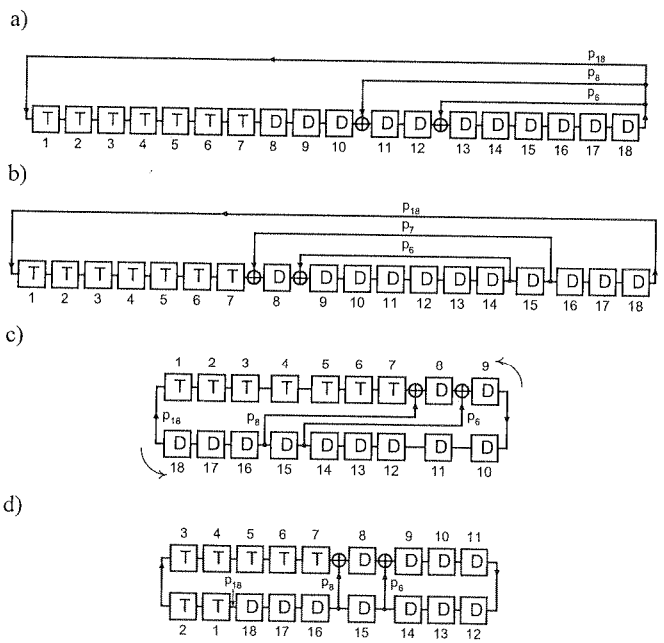

Fig. 7. Stages of designing a ring LFSR according to a new method: (a) LFSR structure corresponding to structural polynomial $p_{wr3}(x)$, (b) LFSR with shifted loops, (c) ring register before rotation, (d) ring register after rotation

in one feedback path to XL=3 (see Fig. 6a) or reduce the number of layers of XOR gates to XL≤1 by increasing the fan-out at the output of the flip-flop D_9 to FO = 5 (see Fig. 6b). Some trade-off between the above extreme solutions is the ring register with FO≤3 and XL≤2 (see Fig. 6c).

Note that none of the approaches presented in Section 2, if used separately from another one, has resulted in the optimal solution. However, if the two methods are combined and applied together achieving the ring register with FO ≤ 2 and XL ≤ 1 becomes possible. Fig. 7a presents the LFSR structure corresponding to the structural polynomial $p_{wr3}(x) = 1+(1+x)^7x^4[1+x[1+x^6]]$ that is a result of one of transformations involved in Method 1. In order to reduce the fan-out at the output of the flip-flop D_{16} loops p_6 and p_7 have to be shifted to the left four times, as it is shown in Fig. 7b. The both loops cannot be shifted further to the left in order to locate them in the middle of the register because this will increase the number of layers of XOR gates in these two feedback paths to XL=2 (since a T-type flip-flop should be considered as a D-type flip-flop with a XOR gate at the input). Another reason is that shifting a feedback loop across T-type flip-flops alters the polynomial implemented by the register [6]. The ring LFSR that is equivalent to the structure from Fig. 7b is shown in Fig. 7c. The fan-out in that register is FO ≤ 2 and the maximum number of XOR layers in the feedback paths is XL=1. However, the length of the feedback loops p_6 and p_8 can further be reduced by rotating the ring two times anticlockwise. The result of the rotation is shown in Fig. 7d.

4. CONCLUSIONS

The paper describes a new approach to designing linear ring registers. It is based on alternately applying techniques involved in Method 1 and Method 2, which have been presented in Section 2. The proposed algorithm includes the following steps:

1) The given polynomial p(x) is transformed to the equivalent structural form $p_{wr}(x)$ that directly reflects a structure of the LFSR implementing polynomial p(x). In general case the resulting LFSR is composed of D- and T-type flip-flops as well as XOR gates in the feedback loops.

2) The feedback loops in the LFSR are shifted to the left until the number of layers of XOR gates in each feedback path is XL ≤ 1 and maximum fan-out in the register is FO ≤ 2. At the same time the centre of each loop should be placed as close of the centre of the register as possible. However, shifting loops across T-type flip-flops and crossing the loops has to be avoided, since this may influence polynomial implemented by the register.

3) The LFSR is then transformed into ring register.

4) The clockwise or anticlockwise rotation of the ring register is performed until the length of its feedback loops achieves minimum value.

Note that if step 2) is skipped in the above algorithm, the latter becomes Method 1. Similarly omitting steps 1) and 4) alters the algorithm into Method 2. Therefore, the proposed new uniform approach to designing linear ring registers is much more versatile than methods used so far. It makes it possible to design a ring register operating with maximum frequency for much wider range of polynomials than Method 1 or Method 2.

Some disadvantage of the proposed method is necessity of avoidance of shifting loops across T-type flip-flops and crossing the loops. In the case of some polynomials this may make it difficult or even impossible to find an optimum structure of the ring register. The future author's research will be devoted to the above issue.

5. REFERENCES

1. M. Bellos, D. Kagaris, D. Nikolos: "Test Set Embedding Based on Phase Shifters". Proc. of Fourth European Dependable Computing Conference – EDCC-4, Toulouse, France, October 23-25, 2002, pp. 90-101.
2. A. Hławiczka: "D or T Flip-Flop Based Linear Registers". Archives of Control Sciences (former Archiwum Automatyki i Telemechaniki), vol. 1 (XXXVII), 1992, no. 3-4, pp. 249-268.
3. A. Hławiczka: "Linear Registers-Analysis, Synthesis and Applications in Digital Circuits Testing" (Rejestry liniowe – analiza, synteza i zastosowania w testowaniu układów cyfrowych). Skrypt Politechniki Śląskiej nr 1370, seria Elektronika z. 9, 1997.
4. T. Garbolino, A. Hławiczka: "A New LFSR with D and T Flip-Flops as an Effective Test Pattern Generator for VLSI Circuits", Proc. of Third European Dependable Computing Conference – EDCC-3, Prague, Czech Republic, September 15-17, 1999, Lecture Notes in Computer Science, Springer Verlag Press, pp. 321-338.
5. G. Mrugalski, N. Mukherjee, J. Rajski, J. Tyszer: "Dense Ring Generators of Pseudo-Random Test Patterns", Proc. of the 6th IEEE Inter. Workshop on Design and Diagnostics of Electronic Circuits and Systems (DDECS'03), Poznań, April 14-16, 2003, pp. 65-72.
6. G. Mrugalski, J. Rajski, J. Tyszer: "High Speed Ring Generators and Compactors of Test Data", Proc. of the 21st IEEE VLSI Test Symposium (VTS'03) pp. 57-62.
7. P. Rosinger, B. M. Al-Hashimi, N. Nicolici: "Dual multiple-polynomial LFSR for low-power mixed-mode BIST", IEE Proc. Computers and Digital Techniques, Vol. 150, Issue 4, 18 July 2003, pp. 209-217.
8. C. E. Stroud: "A Designer's Guide to Built-In Self-Test", Kluwer Academic Publishers, 2002.

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2008, 54, no 1, pp. 53–66

New Methods for Location Service in the WCDMA System

JACEK STEFAŃSKI

Gdańsk University of Technology
Department of Radiocommunictaion Systems and Networks
ul. Narutowicza 11/12, 80-952 Gdańsk
jstef@eti.pg.gda.pl

Received 2008.01.10
Authorized 2008.03.12

New methods for a location service (LCS) in the Wideband Code Division Multiple Access (WCDMA) system are elaborated. These methods base on the observed time difference of arrival (OTDOA) algorithm. The classic OTDOA in the WCDMA system, to calculation of the geographical position of a mobile station (MS), needs a knowledge of relative time differences (RTDs) between base stations. In this reason, the LCS implementation is difficult and rather expensive. The proposed methods do not require RTDs. The elimination of the RTD parameters significantly simplifies the localization process in the 3G (3rd Generation) real-life cellular networks. The usefulness of the methods for location service implementation is analysed. The simulation model is outlined and simulation results of the proposed methods are presented. These results show that in the extremely bad conditions, i.e. in urban environment, the proposed algorithms fulfil the requirements for emergency calls from cellular phones in relation to accuracy of MS position calculation.

Keywords: location service, LCS, location based services, LBS, radio navigation, OTDOA

1. INTRODUCTION

The interest in mobile station positioning for cellular networks is rapidly growing. Many services are based on position information LBS (Location Based Services), for instance access to the databases including restaurant addresses, hotels and offices in the vicinity. For a network operator the location information can be used to select channels or for management of handover processes. One of the most important applications of location services is to automatically locate people who make emergency calls from their cellular phone.

To fulfil the requirements for emergency call service in the European Union or in the USA, the location service in cellular phone networks should be based on the observed time difference of arrival (OTDOA) between a MS and a minimum of three (one serving and two auxiliary) base

stations (BSs). Implementation of a location service in the WCDMA system, in the frequency division duplex (FDD) mode, is difficult and rather expensive due to asynchronous operation of the base stations in this system. As depicted in Fig. 1, the Universal Mobile Telecommunications System (UMTS) requires dedicated components to control positioning [1]. Base stations are equipped with an *associated* location measurement unit (LMU) that provides for timing synchronization with neighbouring base station. To prevent accurate measurements in a bad radio conditions, operators can alternatively establish *stand-alone* LMUs. In this context, the LCS in the WCDMA/FDD is based on the actual RTDs between base stations. Each RTD slowly changes with time. The rate of change depends on the frequency differences and jitters between BSs. Therefore, the measurements of this relatively slow rate of drift are needed using the LMUs. The measurement results are then stored in special databases which are later used for the calculation of MS position. The LMUs complicate positioning architecture in WCDMA/FDD system. When the RTD measurements to estimate the position of the MS are not needed, the location service is cheaper.

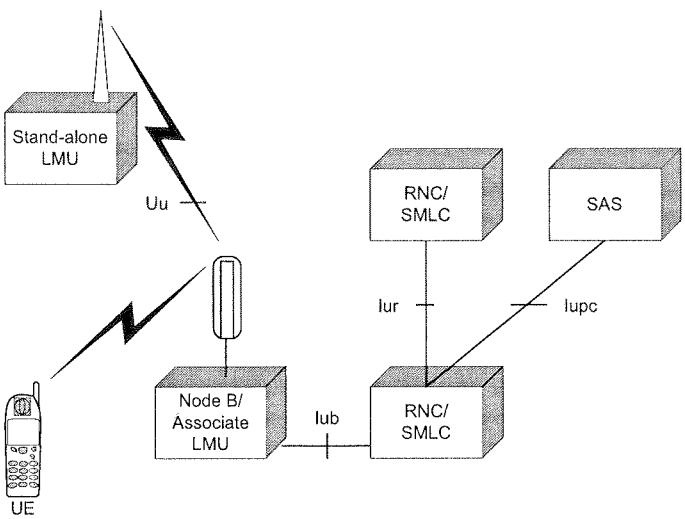

Fig. 1. UMTS positioning architecture. (LMU – location measurement unit; RNC – Radio Network Controller; SMLC – Serving Mobile Location Centre; SAS – Stand-alone SMLC; UE – User Equipment (mobile station); *Uu* – radio interface; *Iub*, *Iur* and *Iupc* – telecommunications interfaces)

In this paper two methods, which enable the calculation of the geographical position of a mobile station without knowledge of relative time differences, are described. There are following methods:

- OTDOA+RTT-LMU, which is supported by round trip time (RTT) [2],
- OTDOA+IPDL-LMU, which is based on idle period downlink (IPDL) mode [3].

The new methods were tested in a finite microcellular environment with outdoor and indoor users – similar to Manhattan structure. The simulation investigations show that the estimation accuracy of the mobile station using the new methods is close to the classic one.

2. THE OTDOA+RTT-LMU METHOD

2.1. PROPOSED SOLUTION

The classic OTDOA method, in a two-dimensional plane, is reduced to a solution of the following set of non-linear equations [4]:

$$\begin{cases} c\cdot[t_1(n)-RTD_2-t_2(n)]=\sqrt{[x_1-x(n)]^2+[y_1-y(n)]^2}+ \\ \qquad -\sqrt{[x_2-x(n)]^2+[y_2-y(n)]^2} \\ c\cdot[t_1(n)-RTD_3-t_3(n)]=\sqrt{[x_1-x(n)]^2+[y_1-y(n)]^2}+ \\ \qquad -\sqrt{[x_3-x(n)]^2+[y_3-y(n)]^2} \end{cases} \tag{1}$$

where $t_1(n)$, $t_2(n)$ and $t_3(n)$ denote the measured signal of transfer times from the BS_1, BS_2 and BS_3 to the MS at the same discrete time n, (x_1,y_1), (x_2,y_2) and (x_3,y_3) represent the co-ordinates of base stations, $((x(n),y(n))$ are the co-ordinates of a mobile station at the discrete time n, c is a speed of light and RTD_2, RTD_3 describe the relative time differences between serving BS_1 and auxiliaries BS_2 or BS_3. However, the solution of this set of equations (1) is possible only when we know RTD_2 and RTD_3 (in the UMTS standard the RTD_2 and RTD_3 are measured in LMUs.). In connection with this, the second set of non-linear equations at the discrete time $n+1$ is proposed (2):

$$\begin{cases} c\cdot[t_1(n+1)-RTD_2-t_2(n+1)]= \\ \quad =\sqrt{[x_1-x(n+1)]^2+[y_1-y(n+1)]^2}+ \\ \quad -\sqrt{[x_2-x(n+1)]^2+[y_2-y(n+1)]^2} \\ c\cdot[t_1(n+1)-RTD_3-t_3(n+1)]= \\ \quad =\sqrt{[x_1-x(n+1)]^2+[y_1-y(n+1)]^2}+ \\ \quad -\sqrt{[x_3-x(n+1)]^2+[y_3-y(n+1)]^2} \end{cases} \tag{2}$$

The relative time differences RTD_2 and RTD_3 are the constants at the discrete time n and $n+1$. By transforming the non-linear equations (1) and (2) respectively, we can get the new non-linear set of equations with six variables (3a):

$$\begin{cases} \sqrt{[x_1-x(n)]^2+[y_1-y(n)]^2}-c\cdot t_1(n)=0 \\ \sqrt{[x_2-x(n)]^2+[y_2-y(n)]^2}-c\cdot[t_2(n)+RTD_2]=0 \\ \sqrt{[x_3-x(n)]^2+[y_3-y(n)]^2}-c\cdot[t_3(n)+RTD_3]=0 \\ \sqrt{[x_1-x(n+1)]^2+[y_1-y(n+1)]^2}-c\cdot t_1(n+1)=0 \\ \sqrt{[x_2-x(n+1)]^2+[y_2-y(n+1)]^2}-c\cdot[t_2(n+1)+RTD_2]=0 \\ \sqrt{[x_3-x(n+1)]^2+[y_3-y(n+1)]^2}-c\cdot[t_3(n+1)+RTD_3]=0 \end{cases} \tag{3a}$$

or with four variables (3b):

$$\begin{cases} \sqrt{[x_1 - x(n)]^2 + [y_1 - y(n)]^2} - c \cdot t_1(n) = 0 \\ \sqrt{[x_1 - x(n+1)]^2 + [y_1 - y(n+1)]^2} - c \cdot t_1(n+1) = 0 \\ \sqrt{[x_2 - x(n+1)]^2 + [y_2 - y(n+1)]^2} + \\ \quad - \sqrt{[x_2 - x(n)]^2 + [y_2 - y(n)]^2} - c \cdot [t_2(n+1) - t_2(n)] = 0 \\ \sqrt{[x_3 - x(n+1)]^2 + [y_3 - y(n+1)]^2} + \\ \quad - \sqrt{[x_3 - x(n)]^2 + [y_3 - y(n)]^2} - c \cdot [t_3(n+1) - t_3(n)] = 0 \end{cases} \tag{3b}$$

These results can be used to calculate the geographical position of a mobile station without the knowledge of relative time differences. In the UMTS, the variables of $t_1(n)$ and $t_1(n+1)$ can be calculated by measuring the round trip times (RTTs), and the elements $t_i(n)+RTD_i$ and $t_i(n+1)+RTD_i$ (for $i=2,3$) by measuring observed time differences in the system frame number SFN [5].

The solution of the above problem is shown in Fig. 2. As we can see, it is necessary to find the coordinates of a mobile station ($(x(n), y(n))$ and $(x(n+1), y(n+1))$). However, we must assume that we know:

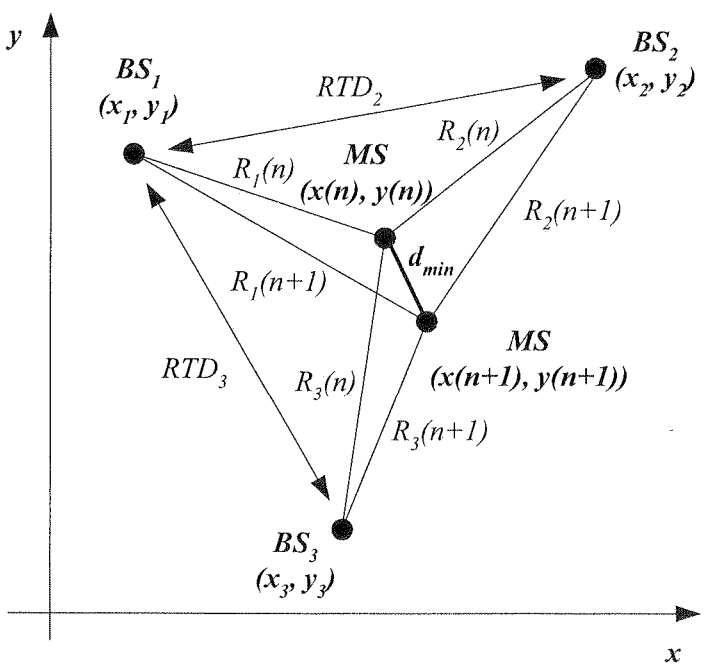

Fig. 2. The graphical illustration of the problem's solution. (MS – mobile station, BS – base station, RTD – relative time difference)

- the coordinates of BS_1, BS_2 and BS_3,
- the distances between BS_1 and the mobile station ($R_1(n)$ and $R_1(n+1)$),
- the distance differences between BS_2, BS_3 and mobile station ($R_2(n+1) - R_2(n)$ *and* $R_3(n+1) - R_3(n)$) at the discrete time n and $n+1$.

The distance, d_{min}, between two distinct positions of a mobile station at the discrete time n and $n+1$, depends on the resolution of the measured times of transition signals from BSs to MS. It is obvious that the smaller the d_{min} distance, the more useful is the new algorithm for the position estimation of a very slowly moving mobile station (uncertainty of 10 nanoseconds contributes to the 3 meter error in the position estimation).

2.2. CALCULATION OF A POSITION

The resulting relative time differences cause inaccuracies in the calculation of signal transit times and distances. As a result, the distances between BS_2, BS_3 and MS are measured, which are known as pseudo distances (PSDs). In this way, the pseudo distances at the discrete time n and $n+1$ can be expressed as:

$$\begin{aligned} PSD_i(n) &= R_i(n) + c \cdot RTD_i \\ R_i(n) &= \sqrt{[x_i - x(n)]^2 + [y_i - y(n)]^2} \\ PSD_i(n+1) &= R_i(n+1) + c \cdot RTD_i \\ R_i(n+1) &= \sqrt{[x_i - x(n+1)]^2 + [y_i - y(n+1)]^2} \end{aligned} \tag{4}$$

where R_i (i = 1, 2, 3) is the real distance between BS_i and a mobile station, RTD_i equals zero. The solution of the non-linear equation (3) is difficult. One of the possible ways is the linearization of the equation on the basis of the Taylor-series expansion and then solving it iteratively [6]. In order to solve the set of equations (3), only the first part of the Taylor model is being exploited. After transforming all six equations, we can solve them according to the rules of a linear algebra:

$$\mathbf{g} = \mathbf{A}^{-1} \cdot \mathbf{d} \tag{5}$$

where

$$\mathbf{g} = \begin{bmatrix} \Delta x(n) \\ \Delta y(n) \\ \Delta x(n+1) \\ \Delta y(n+1) \\ RTD_2 \\ RTD_3 \end{bmatrix} \tag{6}$$

$$\mathbf{d} = \begin{bmatrix} PSD_1(n) - R_1(n) \\ PSD_2(n) - R_2(n) \\ PSD_3(n) - R_3(n) \\ PSD_1(n+1) - R_1(n+1) \\ PSD_2(n+1) - R_2(n+1) \\ PSD_3(n+1) - R_3(n+1) \end{bmatrix} \tag{7}$$

$$\mathbf{A} = \begin{bmatrix} \frac{x(n)-x_1}{R_1(n)} & \frac{y(n)-y_1}{R_1(n)} & 0 & 0 & 0 & 0 \\ \frac{x(n)-x_2}{R_2(n)} & \frac{y(n)-y_2}{R_2(n)} & 0 & 0 & c & 0 \\ \frac{x(n)-x_3}{R_3(n)} & \frac{y(n)-y_3}{R_3(n)} & 0 & 0 & 0 & c \\ 0 & 0 & \frac{x(n+1)-x_1}{R_1(n+1)} & \frac{y(n+1)-y_1}{R_1(n+1)} & 0 & 0 \\ 0 & 0 & \frac{x(n+1)-x_2}{R_2(n+1)} & \frac{y(n+1)-y_2}{R_2(n+1)} & c & 0 \\ 0 & 0 & \frac{x(n+1)-x_3}{R_3(n+1)} & \frac{y(n+1)-y_3}{R_3(n+1)} & 0 & c \end{bmatrix} \tag{8}$$

The unknown variables $\Delta x(n)$, $\Delta y(n)$, $\Delta x(n+1)$, $\Delta y(n+1)$ represent the estimation error. The values obtained from the solution of $\Delta x(n)$, $\Delta y(n)$, $\Delta x(n+1)$, $\Delta y(n+1)$ are used to recalculate the estimated position of $x(n)$, $y(n)$, $x(n+1)$ and $y(n+1)$ in accordance to the equations (9):

$$\begin{aligned} x_{new}(n) &= x_{old}(n) + \Delta x(n) \\ y_{new}(n) &= y_{old}(n) + \Delta y(n) \\ x_{new}(n+1) &= x_{old}(n+1) + \Delta x(n+1) \\ y_{new}(n+1) &= y_{old}(n+1) + \Delta y(n+1) \end{aligned} \tag{9}$$

The estimated values of $x_{new}(n)$, $y_{new}(n)$, $x_{new}(n+1)$, and $y_{new}(n+1)$ can now be substituted into the set of equations (5) by applying the normal iterative process, until the error of components $\Delta x(n)$, $\Delta y(n)$, $\Delta x(n+1)$, $\Delta y(n+1)$ is smaller than the assumed one. As a result of simulation, depending on the initial estimation of parameters, three to five iterative calculations are generally required to obtain an error of the unknowns of less than 1 cm. The calculated values of RTD_2 and RTD_3 using the equation (5), correspond to the synchronization error of base station and they can be used to reduce the algorithm of a location service in the next steps of position calculations.

3. THE OTDOA+IPDL-LMU METHOD

3.1. PROPOSED SOLUTION

As we know, in the WCDMA/FDD system a mobile station measures observed time differences in the system frame number (SFN) for downlink. When the base stations work asynchronously and taken into account the expression (1), the MS measures (in this case, to simplify the notation a symbol of the discrete time, n, was omitted):

$$\begin{cases} t_{SFN1} = t_1 - t_2 - RTD_2 \Rightarrow t_1 - t_2 = t_{SFN1} + RTD_2 \\ t_{SFN2} = t_1 - t_3 - RTD_3 \Rightarrow t_1 - t_3 = t_{SFN2} + RTD_3 \end{cases} \tag{10}$$

Normally, in order to facilitate the OTDOA location measurements and to avoid near-far problems, the WCDMA/FDD standard includes idle periods in downlink, during which transmission of all channels from a base station is temporarily seized. A typical frequency of idle periods is 1 slot (about 667 μs) every 100 ms, i.e. 0.7 % of the time [7]. In this respect the mobile station is able to receive the pilot signal of the neighbour cells even if the best pilot signal on the same frequency is very strong. For elimination of the relative time differences (RTDs), the idle period downlink mode is suggested. A graphical illustration of the proposed method is shown in Fig. 3. This method is based on finding time differences of the radio signal arrival between the

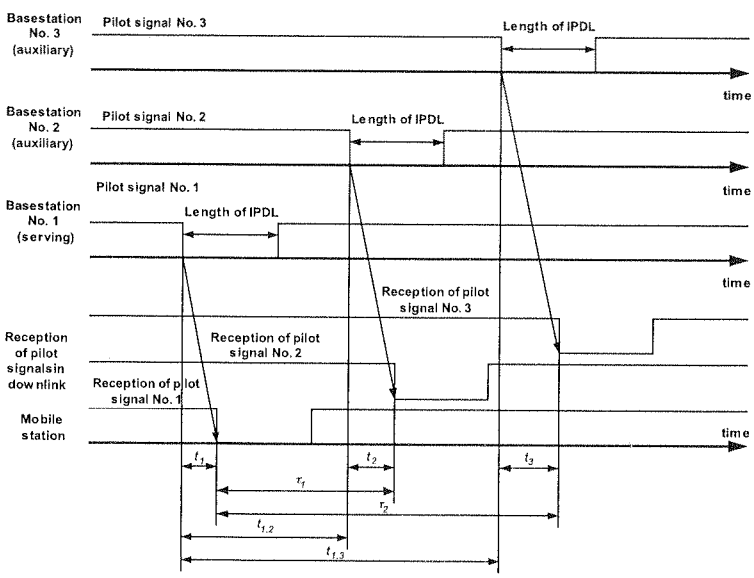

Fig. 3. The graphical illustration of the proposed method. (t_1 – measured signal of transfer times from BS_1 to MS; t_2 – measured signal of transfer times from BS_2 to MS; t_3 – measured signal of transfer times from BS_3 to MS; τ_1 – time difference between moments of switch off of BS_1 transmission and BS_2 detected by mobile station; τ_2 – time difference between moments of switch off of BS_1 transmission and BS_3 detected by mobile station; $t_{1,2}$ – real time difference between moments of switch off of BS_1 transmission and BS_2; $t_{1,3}$ – real time difference between moments of switch off of BS_1 transmission and BS_3; IPDL – idle period downlink; BS – base station; MS – mobile station)

mobile station and base station No. 1 and No. 2, and between the mobile station and base station No. 1 and No. 3. For this purpose, the moments of switch off of pilot channel transmission can be used. In connection with this, according to Fig. 3, we can get the following expressions (11):

$$\begin{cases} t_1 - t_2 = t_{1,2} - \tau_1 \\ t_1 - t_3 = t_{1,3} - \tau_2 \end{cases} \tag{11}$$

where $t_{1,2}$ and $t_{1,3}$ represent the real time differences between moments of switch off BS_1 transmission and BS_2 or BS_3, τ_1 and τ_2 denote the time difference between moments of switch off BS_1 transmission and BS_2 or BS_3 detected by a mobile station. In the WCDMA/FDD system all base stations work asynchronously, therefore the variables $t_{1,2}$ and $t_{1,3}$ can be expressed as:

$$\begin{cases} t_{1,2} = t_{per1} - RTD_2 \\ t_{1,3} = t_{per2} - RTD_3 \end{cases} \tag{12}$$

where t_{per1} and t_{per2} represent the theoretical time differences between moments of switch off of BS_1 transmission and BS_2 or BS_3, when RTD_2 and RTD_3 equal zero. These variables can be calculated on the basis of the parameters sent to the mobile station via higher layers [7]. By transforming the equations (1) and (10) ÷ (12), we can get the new non-linear set of equations with two variables (13):

$$\begin{cases} c \cdot \left(\dfrac{t_{SFN1} + t_{per1} - \tau_1}{2} \right) = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} + \\ \qquad - \sqrt{(x_2 - x)^2 + (y_2 - y)^2} \\ c \cdot \left(\dfrac{t_{SFN2} + t_{per2} - \tau_2}{2} \right) = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} + \\ \qquad - \sqrt{(x_3 - x)^2 + (y_3 - y)^2} \end{cases} \tag{13}$$

This method can be used to calculate the geographical position of a mobile station without knowledge of RTDs. The solution of the non-linear equation (13) is commonly known [4]. To sum up, the new method consists of:

- measurement of observed time differences of arrival between a mobile station and a minimum of three base stations (BS_1, BS_2 and BS_3) in the system frame number for downlink (classic method which is proposed to be implemented in WCDMA/FDD system) – measurement of t_{SFN1} and t_{SFN2},
- measurement of time differences between moments of switch off of pilot channel transmission for serving base station and BS_2 and BS_3 detected by mobile station during IPDL mode – measurement of τ_1 and τ_2,
- calculation of t_{per1} and t_{per2} variables using parameters which are sent to the mobile station via higher layers,
- the geographical position calculation of the mobile station using equations (13).

3.2. DETECTION OF SWITCH OFF OF BASE STATION TRANSMISSION

The detection of switch off of base station transmission is based on a matched filter correlator (MFC) [8], which is matched to the common pilot channel (CPICH). In the WCDMA, each radio frame consists of 38400 chips (10 ms) and is divided into 15 slots, i.e. 2560 chips per slot. The CPICH is scrambled by the primary downlink scrambling code of the cell. There are 10 pilot symbols within each time slot. The position in the radio frame and the length of idle period downlink mode is expressed in the number of CPICH symbols [7]. Therefore, according to Fig. 4, where the MFC output is presented, the moment of switch off of base station transmission can be measured.

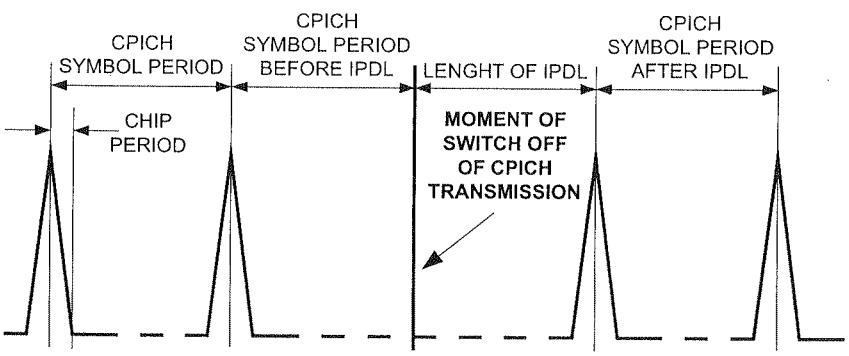

Fig. 4. The MFC output. (IPDL – idle period downlink; CPICH – common pilot channel)

To simplify our analysis, we can consider the single multipath situation where the receiver signal $r(t)$ is given by [9]

$$r(t) = s(t) + as(t - D_0) + n(t), \quad 0 \le t \le T \tag{14}$$

where $s(t)$ and $n(t)$ are uncorrelated Gaussian random processes, a is an attenuation coefficient, and D_0 is a multipath time delay. The autocorrelation function of $r(t)$, $R_r(D)$, is

$$R_r(D) = \frac{1}{T}\int_0^T r(t)r(t - D)dt \tag{15}$$

Note that $R_r(D)$ has a peak at $D=0$ and peaks at D close to $\pm D_0$. The latter peaks may not be exactly at $\pm D_0$ due to slight biases caused by the superposition of the various correlograms. We can now determine the error performance (variance) of the multipath time delay estimate via autocorrelation, i.e. we estimate the location of the peak near D_0 by finding where $R_r(D)/dD$ is equal to zero in the vicinity of D_0; we call this point $\hat{D}$. A useful approximate result for $\mathrm{var}[\hat{D}]$ can be found if we assume that $s(t)$ and $n(t)$ have uniform, low pass spectra in the band $|f| \le B/2$. For highly resolvable multipath ($BD_0 >> 1$), large observation time with respect to multipath delay ($T/D_0 >> 1$), and $BT >> 1$ we find [9]:

$$\mathrm{var}[\hat{D}] \cong \frac{1}{BT}\frac{3}{\pi^2 B^2}\frac{[(1+a^2)^2+a^2]S^2+2(1+a^2)SN+N^2}{a^2S^2} \tag{16}$$

where S and N are the spectrum levels of $s(t)$ and $n(t)$.

In Fig. 5, we show estimation variance (normalized by T^2) versus SNR (Signal to Noise Ratio) for the autocorrelator. The CPICH has a fixed rate (30 kbps, SF=256 – *Spreading Factor*), hence T=33.3μs and B=5MHz. Additionally, we assumed that D_0 is equal to 2.3μs – the mean value for typical urban environment, and a=1. The above analyses can be used to the error of switch off of pilot channel transmission detected by the mobile station.

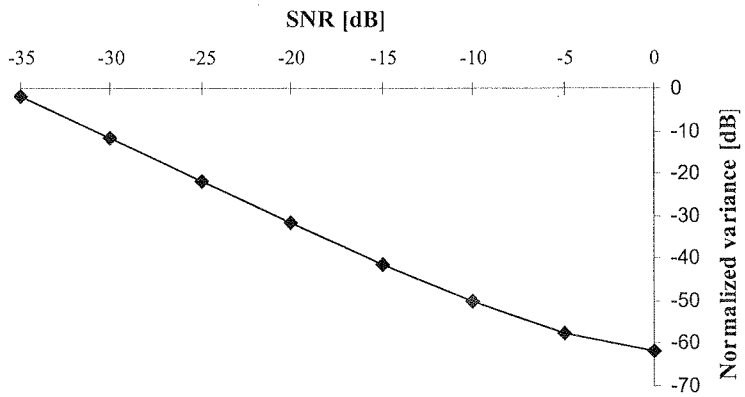

Fig. 5. Normalized variance versus SNR. (T=33.3μs; D_0=2.3μs; B=5MHz; BT=167; BD_0=11.3; T/D_0=14; a=1)

4. SIMULATION INVESTIGATIONS

4.1. SIMULATION MODEL

The experiments were carried out using the simulation model required for the UMTS [10]. This is a typical bad urban environmental model (Manhattan model). The area consisted of 12 by 11 blocks with a total number of 72 base stations. The street width was 30 m and the distance between two street corners was 230 m (see Fig. 6). Base station antennas were placed 10 m above the mobile users but below rooftops. The mobile station was located outdoor (25% of simulation time) and indoor (75% of simulation time). In our implementation, we covered the simulation area with a regular grid with resolution of 10 m. In the simulation model, the effect of a multipath propagation was implemented. The time of radio signal arrival between the mobile station and base stations under the multipath environment was modelled by the sum of true value τ_0 and non-line of sight (NLOS) error τ_m [11]

$$\tau = \tau_0 + \tau_m \tag{17}$$

The variable τ_m is defined as mean excess delay and essentially correlated with the root mean squared delay spread τ_{rms} [12]

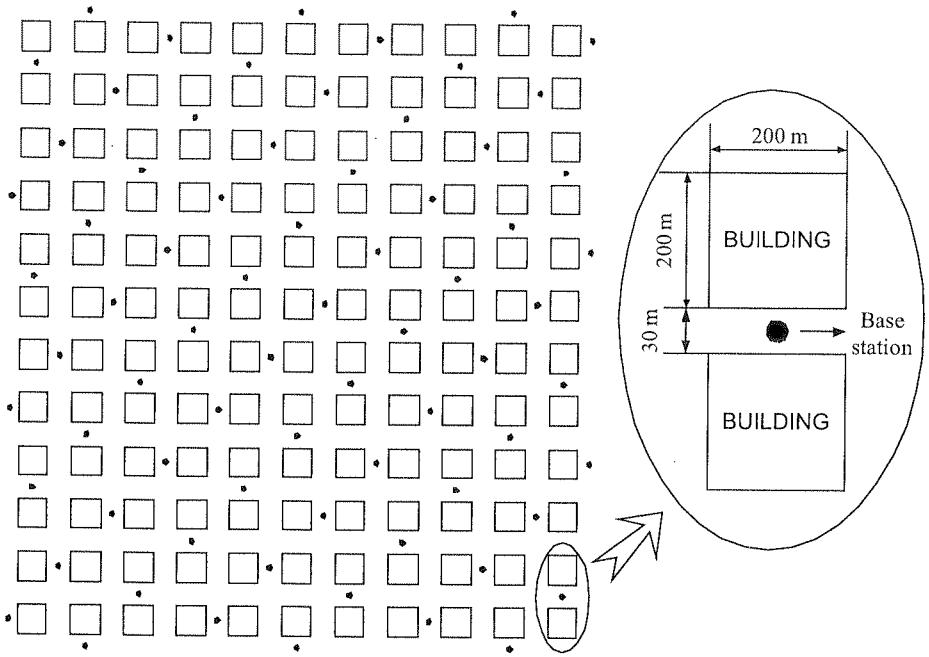

Fig. 6. Simulation model of the bad urban environment (Manhattan model)

$$\tau_m \approx k \cdot \tau_{rms} = k \cdot T_1 \cdot d^{\varepsilon} \cdot y \tag{18}$$

where k is a constant proportional coefficient ($k=1$ in urban region), T_1 is the median value of τ_{rms} at d=1km (for urban environment T_1=0.7μs), d is the distance between the mobile station and base station, ε is an exponent (for urban environment ε=0.5) and y is a lognormal variate. Specifically, Y=10log$y$ is a Gaussian random variable over the terrain at the distance d, having zero mean and a standard deviation σ_y (for urban environment σ_y=4dB).

4.2. SIMULATION RESULTS

The cumulative probability distribution functions (CDF) of the absolute position error were obtained from the simulation investigations. The absolute position error is defined as

$$\Delta d = \sqrt{(x - x_0)^2 + (y - y_0)^2} \tag{19}$$

where x and y represent the real coordinates of a mobile station and x_0 and y_0 denote the estimated coordinates of a mobile station. All timing values have been assumed to be accurate within ± ¼ of a WCDMA chip (the uniform random time error corresponds to a maximum distance error of about ± 19 m). Additionally for the second method, the error of switch off of pilot channel transmission detected by the mobile station was determined from (16) for two SNR values: -15dB and -20dB.

The simulation results are presented in Fig. 7. The proposed methods (OTDOA+RTT-LMU and OTDOA+IPDL-LMU) were compared with the classical OTDOA method. Our investigation show that all methods in the analysed environment (bad urban) fulfil the US E911 phase II requirements of positioned emergency calls within 125m in 67% of the time [13]. In the classic OTDOA method calls from cellular phones were located within 125m in 75.7% of the time. In the OTDOA+RTT-LMU method cellular phones were located within 125m in 67.7% of the time. In the last method OTDOA+IPDL-LMU, when the SNR decreases, the mobile station was located within 125m in 69.5% (SNR=-15dB) and 41.9% (SNR=-20dB) of the time. When the moment of switch off of pilot channel transmission detected by the mobile station reaches the ideal, for high SNR, the CDF of the proposed OTDOA+IPDL-LMU method converges to the classical one.

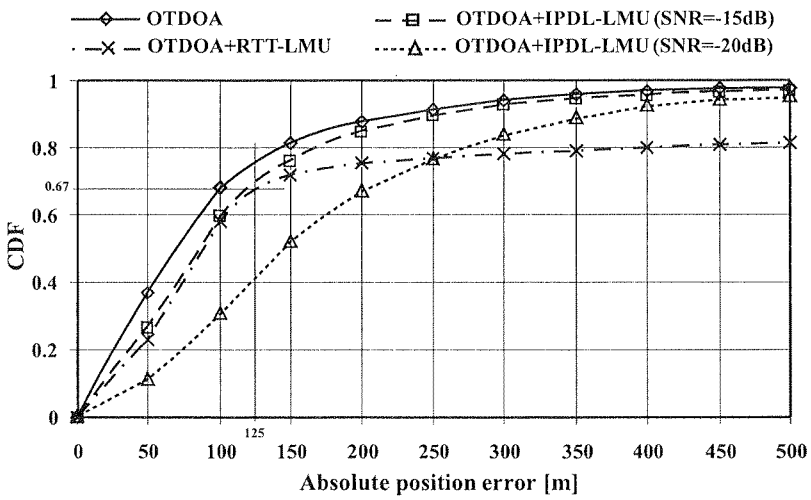

Fig. 7. The CDFs of absolute position error for the classical OTDOA and proposed methods.
(CDF – cumulative probability distribution function; OTDOA – observed time difference of arrival; OTDOA+RTT-LMU and OTDOA+IPDL-LMU – OTDOA methods without using location measurement unit; SNR – Signal to Noise Ratio)

All simulation investigation was carried out in the bad urban environment, where a multipath propagation effect is very intensive. In other environments, i.e. urban area or rural area, the US E911 phase II requirements for emergency calls in relation to MS position will be fulfilled in excess.

Even though the proposed methods are worse than the conventional one with respect to the location estimation of a MS, they significantly simplify the localization process in the WCDMA/FDD system.

5. CONCLUSIONS

A new approach for the location service in the WCDMA/FDD system are proposed. As a result, the new techniques of calculation of the geographical position of a MS without

the knowledge of relative time differences are proposed. The complexity of localization processes of these methods is comparable. New methods fulfil the requirements of E911 phase II in the bad urban environment in relation to accuracy of MS geographical position calculation. The OTDOA+RTT-LMU method requires active connection with the cellular network to determine round trip time parameter. The accuracy of this method depends on initial estimation of parameters: co-ordinates of a mobile station at the discrete time n and $n+1$. The OTDOA+IPDL-LMU method is based on the moment of switch off of pilot channel transmission during IPDL mode detected by a mobile station. Therefore, this method needs extra information from higher layer to calculate the theoretical time differences between moments of switch off of base station transmission, when the relative time differences equal zero. The practical usefulness of this method depends on the "hearing" of distant base stations and detection accuracy moment of switch off of pilot channel transmission. Several authors have proved, for example in [14], that the same solutions allow for a good measurement precision of the time of radio signal arrival for distant base stations without using switch off of all channel transmission in a serving BS. Therefore the detection of switch off pilot channel transmission is possible. Moreover, using a modified IPDL method suggested in [15], which improves the system capacity, the proposed method OTDOA+IPDL-LMU can be available in 95% of the area network. In this case, the described method should be slightly modified. Instead of detecting the switch off of pilot channel transmission, we have to detect the moment of switch off of other channels.

Implementation of the OTDOA method for LCS in contemporary cellular networks is very expensive and brought by operators into effect rather reluctantly. Taking into account the obtained results, the proposed new methods for the geographical position of a mobile station exemplifies the low cost alternative to the classic one.

6. REFERENCES

1. A. Küpper: *Location-based Services, Fundamentals and Operation*, Wiley & Sons, 2005.
2. J. Stefański: *Simplified Algorithm for Location Service for the UMTS*, Proc. of IEEE 62nd Vehicular Technology Conference, vol. 4, 2005, pp. 2741-2744.
3. J. Stefański: *Method of Location of a Mobile Station in the WCDMA System without Knowledge of Relative Time Differences*, Proc. of IEEE 65th Vehicular Technology Conference, April 2007, pp. 674-678.
4. Y.T. Chan, K.C. Ho: *A Simple and Efficient Estimator for Hyperbolic Location*, IEEE Transactions on Signal Processing, vol. 42, no. 8, August 1994, pp. 1905-1915.
5. 3GPP Technical Specification: *Technical Specification Group Radio Access Network; Physical layer – Measurements (FDD)*, 3G TS 25.215.
6. W.H. Foy: *Position-location solutions by Taylor-series estimation*, IEEE Trans. Aerosp. Electron. Syst., vol. AES-12, March 1976, pp. 187-194.
7. 3GPP Technical Specification: *Technical Specification Group Radio Access Network; Physical layer procedures (FDD)*, 3G TS 25.214.
8. W.H. Sheen, J.K. Tzeng, C.K. Tzou: *Effects of Cell Correlations in a Matched-Filter PN Code Acquisition for Direct-Sequence Spread-Spectrum Systems*, IEEE Tran. on Vehicular Tech., vol. 48, no. 3, 1999, pp. 724-732.
9. J.P. Ianniello: *Large and Small Error Performance Limits for Multipath Time Delay Estimation*, IEEE Tran. on Acoust., Speech, and Sig. Proc., vol. ASSP-34, No. 2, 1986, pp. 245-251.

10. ETSI: *Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS*, TR 101 112, ver. 3.2.0, Apr. 1998.
11. Y. Jeong, H. You, C. Lee: *Calibration of NLOS error for positioning systems*, Proc. of IEEE 53rd Vehicular Technology Conference, vol. 4, 2001, pp. 2605-2608.
12. L.J. Greenstain, V. Erceg, Y.S. Yeh, M.V. Clark: *A new path-gain/delay-spread propagation model for digital cellular channels*, IEEE Trans. on Vehicular Tech., vol. 46, 1997, pp. 477-485.
13. D.N. Hatfield: *A Report on Technical and Operational Issues Impacting the Provision of Wireless Enhanced E911 Services*, Federal Communications Commission, September 2002.
14. E. Grosicki, K. Abed-Meraim, P. Loubaton, J.M. Chaufray: *Comparison of Downlink Mobile Positioning Methods for the UMTS FDD Mode without Using IPDL Periods*, Proc. of 7th International Symposium on Signal Processing and its Applications, vol. 2, 2003, pp. 347-350.
15. S. Moshavi, D. Yelli, J.S. Sadowsky, Y. Perets, K. Pick: *Pilot Interference Cancellation Technology for CDMA Cellular Networks*, IEEE Trans. on Vehicular Tech., vol. 54, no. 5, 2005, pp. 1781-1792.

Optimization of time between the attempts to gain access to a service system (a generalized problem of an impatient customer)*

ROMAN SZOSTEK

Rzeszów University of Technology
Department of Quantitative Methods in Economics
ul. Wincentego Pola 2, 35-959 Rzeszów
rszostek@prz.rzeszow.pl

Received 2008.01.07
Authorized 2008.03.05

The purpose of the paper is to solve an optimization problem consisting in finding the optimal time between the attempts to access a service system. To obtain a solution the method of queueing systems analysis with the aid of Markov chain was used [1][4]. It allows to solve a problem which is presented by non Markovian model.

The discussed task is called a generalized problem of an impatient customer. The basic problem of an impatient customer is presented in [5]. The author formulates and solves the problem of an impatient customer in this paper.

Keywords: Probabilistic methods, Markov process, Non- Markovian process, included Markov chain, system parameters, objective (cost) function, optimization

1. INTRODUCTION

The method of the included Markov chains allows to analyse non-Markovian systems in which, one process is non-exponentially distributed, but all the others are exponentially distributed. The method allows to determine a probability of a system being in the particular states in the time instants when, the distinguished non-exponentially distributed process generated an arrival. We can consider the instants just before and just after an arrival is generated.

The method of the included Markov chains enables determination of many system parameters e.g. S mean service time of arrival and R mean sojourn time in the system.

* Artykuł finansowany ze środków Sieci Naukowej Aeronautica Integra z pozycji promocja wyników i zastosowań praktycznych badań naukowych prowadzonych w ramach sieci.

2. PROBLEM FORMULATION

The given configuration of three satellites is shown in Fig 1. The task of satellite A is to transmit data packet to satellite B. Because the direct communication between the satellites is not possible, so the satellite C has to mediate in data transmission. It charges a fee for the time during which its resources are used. All the time satellite C is idle and ready to cooperate, while satellite B is idle or busy interchangeably. When it is busy doing other tasks, it cannot receive data from satellite A.

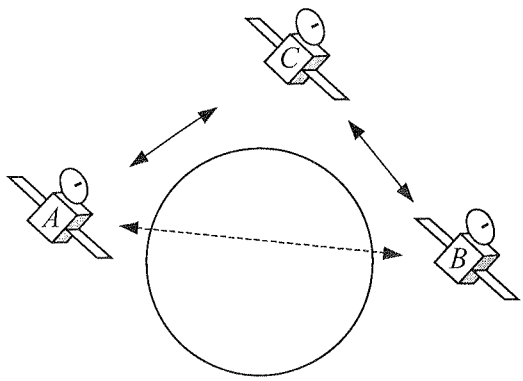

Fig. 1. Problem illustration

The task of satellite A is to gain access to the resources of satellite B, when they are accessible. However, satellite A has no information about the state of satellite B until it undertakes an attempt of data transmission. If an attempt fails (satellite B is busy), after time T satellite A will undertake another attempt of data transmission. Consecutive attempts are undertaken at the regular intervals T, till satellite B will be ready to obtain data packet.

Every attempt of data transmission by use of satellite C, even it fails, involves a cost of c_1. We assume that the waiting time of satellite A for accessible resources of satellite B also involves a unit cost of c_2. A cost function of obtaining connection by satellite A is given by

$$F_C(T) = c_1 N_{sr}(T) + c_2 T_{sr}(T) \tag{1}$$

where: $N_{sr}(T)$ – the mean number of attempts to obtain connection by satellite A,
$T_{sr}(T)$ – the mean waiting time of obtaining connection by satellite A,
$c_1 \in \boldsymbol{R}_+$ – cost of one attempt to obtain connection,
$c_2 \in \boldsymbol{R}_+$ – cost of (connection) waiting time unit.

The $Cox_2/M/1/-/1$ Markov system is assumed to be a model describing a state, in which satellite B is. Therefore, the time when the satellite is idle, is subjected to the Cox distribution of the 2 nd order, but the time when the satellite is busy, is exponentially distributed. It is not possible to queue to the resources of that satellite. Graph of states of satellite B is shown in Fig. 2. In states W_1 and W_2 satellite B is idle. In state Z satellite B is busy.

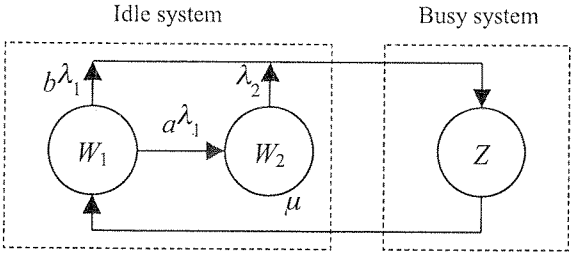

Fig. 2. Graph of state for the $Cox_2/M/1/-/1$ system

Attempts of obtaining connection by satellite A can be regarded as an additional determined stream of arrivals to the resources of satellite B. The system in Fig. 2 together with the additional determined stream of arrivals is called a developed system, which is non Markovian. Graph of states of the developed system is presented in Fig. 3.

Additional absorbing state K shows a case, when satellite A obtained access to the resources of satellite B. Transition to the state K is possible only from state W_1 and W_2 (satellite B was idle). Additional transition from state Z to the state Z represents a case when an attempt to connect by satellite A has failed.

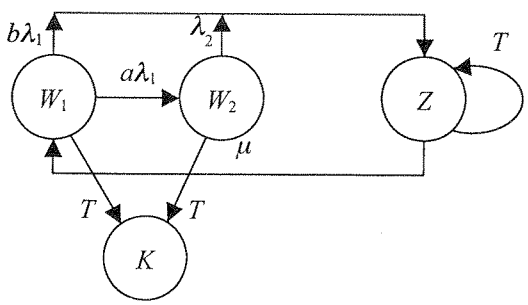

Fig. 3. Graph of states for the $Cox_2/M/1/-/1$ developed system by an additional determined stream of arrivals that refers to a satellite

Further, the optimal time T between consecutive attempts to obtain connection by satellite A with satellite B, that is, the time which minimizes a cost function (1) will be found.

3. ANALYSIS OF TRANSIENT PROCESSES IN THE $COX_2/M/1/-/1$ SYSTEM

The aim of author's consideration in this chapter is to determine characteristics describing satellite B, i.e. the following functions: $p_{1w}(T), p_{2w}(T), p_z(T)$ with the initial conditions: $p_{1w}(0)=0$, $p_{2w}(0)=0$, $p_z(0)=1$. For that purpose an analysis of transient processes of the system in Fig. 2 is necessary.

Denote by:

A_{1i} – event that, a connection to satellite B in i-th attempt was obtained, when satellite was in state W_1 ($i=0, 1, 2, ...$),

A_{2i} – event that, a connection to satellite B in i-th attempt was obtained, when satellite was in state W_2 (i =0, 1, 2, ...),

B_i – event that, a connection to satellite B was not obtained in i-th attempt (i = 0, 1, 2, ...).

In time intervals T the consecutive attempts to obtain connection are undertaken. An attempt i (i=0,1,2,...) is undertaken at time iT.

At the zero-test instant satellite B is in the stationary state. Therefore, we have

$$\begin{cases} \lambda_1 p(A_{10}) = \mu p(B_0) \\ \lambda_2 p(A_{20}) = a\lambda_1 p(A_{10}) \\ p(A_{10}) + p(A_{20}) + p(B_0) = 1 \end{cases} \tag{2}$$

that is

$$\begin{cases} p(A_{10}) = \dfrac{\lambda_2 \mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \\ p(A_{20}) = \dfrac{a\lambda_1 \mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \\ p(B_0) = \dfrac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \end{cases} \tag{3}$$

If the first attempt of obtaining connection fails, after time T satellite A tries again. We assume that the system in Fig. 2 is Markovian (memoryless), so we have:

$$\begin{cases} p_{1w}(T) := p(A_{1(i+1)} / B_i) = p(A_{11} / B_0), & i = 0,1,2,... \\ p_{2w}(T) := p(A_{2(i+1)} / B_i) = p(A_{21} / B_0), & i = 0,1,2,... \\ p_z(T) := p(B_{i+1} / B_i) = p(B_1 / B_0), & i = 0,1,2,... \end{cases} \tag{4}$$

The following notation is given:

$p_{1w}(T)$ – probability, that the system at a time instant T is idle in the state W_1, assuming that at the time instant T=0 it was busy, that is $p_{1w}(0)$=0,

$p_{2w}(T)$ – probability, that the system at a time instant T is idle in the state W_2, assuming that at the time instant T=0 it was busy, that is $p_{2w}(0)$=0,

$p_z(T)$ – probability, that the system at a time instant T is busy, assuming that at the time instant T=0 it was busy, that is $p_z(0)$=1,

To determine these probabilities it is necessary to solve a system of differential equations, which describes transient processes in the system shown in Fig. 2.

$$\begin{cases} \dot{p}_{1w}(T) = -\lambda_1 p_{1w}(T) + \mu p_z(T) \\ \dot{p}_{2w}(T) = a\lambda_1 p_{1w}(T) - \lambda_2 p_{2w}(T) \\ \dot{p}_z(T) = b\lambda_1 p_{1w}(T) + \lambda_2 p_{2w}(T) - \mu p_z(T) \end{cases} \tag{5}$$

A solution of equations (5) takes the following complex form

$$\begin{cases} p_{1w}(T) = A + S \cdot e^{s_1 T} + T \cdot e^{s_2 T} \\ p_{2w}(T) = D + U \cdot e^{s_1 T} + W \cdot e^{s_2 T} \\ p_z(T) \;\; = G + X \cdot e^{s_1 T} + Y \cdot e^{s_2 T} \end{cases} \tag{6}$$

where:

$$s_1 = \frac{-(\lambda_1 + \lambda_2 + \mu) + \sqrt{[\lambda_2 - (\lambda_1 + \mu)]^2 - 4a\lambda_1\mu}}{2} = -K + \frac{\sqrt{\Delta}}{2} \tag{7}$$

$$s_2 = \frac{-(\lambda_1 + \lambda_2 + \mu) - \sqrt{[\lambda_2 - (\lambda_1 + \mu)]^2 - 4a\lambda_1\mu}}{2} = -K - \frac{\sqrt{\Delta}}{2} \tag{8}$$

$$A = \frac{\lambda_2\mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \tag{9}$$

$$D = \frac{a\lambda_1\mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \tag{10}$$

$$G = \frac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \tag{11}$$

$$S = \frac{-\lambda_2\mu}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} + \frac{\lambda_2\mu(\lambda_1 - \lambda_2 + \mu) + 2a\lambda_1\mu^2}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)\sqrt{\Delta}} \tag{12}$$

$$T = \frac{-\lambda_2\mu}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} - \frac{\lambda_2\mu(\lambda_1 - \lambda_2 + \mu) + 2a\lambda_1\mu^2}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)\sqrt{\Delta}} \tag{13}$$

$$U = \frac{-a\lambda_1\mu}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} - \frac{a\lambda_1\mu(\lambda_1 + \lambda_2 + \mu)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)\sqrt{\Delta}} \tag{14}$$

$$W = \frac{-a\lambda_1\mu}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} + \frac{a\lambda_1\mu(\lambda_1 + \lambda_2 + \mu)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)\sqrt{\Delta}} \tag{15}$$

$$X = \frac{\mu(\lambda_2 + a\lambda_1)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} + \frac{\lambda_2\mu(-\lambda_1 + \lambda_2 - \mu) + a\lambda_1\mu(\lambda_1 + \lambda_2 - \mu)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)\sqrt{\Delta}} \tag{16}$$

$$Y = \frac{\mu(\lambda_2 + a\lambda_1)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} - \frac{\lambda_2\mu(-\lambda_1 + \lambda_2 - \mu) + a\lambda_1\mu(\lambda_1 + \lambda_2 - \mu)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)\sqrt{\Delta}} \tag{17}$$

$$K = \frac{\lambda_1 + \lambda_2 + \mu}{2} > 0 \tag{18}$$

$$\Delta = [\lambda_2 - (\lambda_1 + \mu)]^2 - 4a\lambda_1\mu \tag{19}$$

To present a solution of equations (5) in a real form, it is necessary to distinguish three cases in dependence of the sign which the expression $[\lambda_2 - (\lambda_1 + \mu)]^2 - 4a\lambda_1\mu$ takes.

I. First variant

$$[\lambda_2 - (\lambda_1 + \mu)]^2 - 4a\lambda_1\mu > 0 \tag{20}$$

In this case s_1 and s_2 are real negative values. Characteristics $p_{1w}(T)$, $p_{2w}(T)$, $p_z(T)$, we were seeking, are in the form (6) with parameters described by (7) – (17).

Graphs of functions $p_{1w}(T)$, $p_{2w}(T)$, $p_z(T)$ for two different systems are displayed in Fig. 4 and Fig. 5, respectively. These systems have a significant qualitative difference. For the first of them the function $p_z(T)$ takes, in a certain period of time, the values smaller than the set value and it has its mimimum. For the second system the values of function $p_z(T)$ are always bigger than the set value and the function is decreasing in all domain.

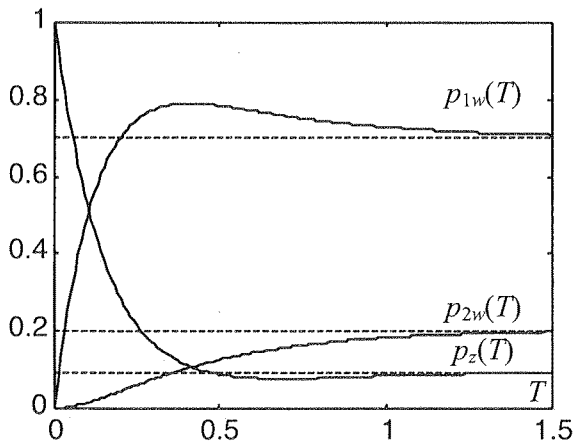

Fig. 4. Graphs of $p_{1w}(T)$, $p_{2w}(T)$, $p_z(T)$ functions for λ_1=1, λ_2=2, μ=7.5, a=0.58. Δ= 24.85. Overcontrolled system

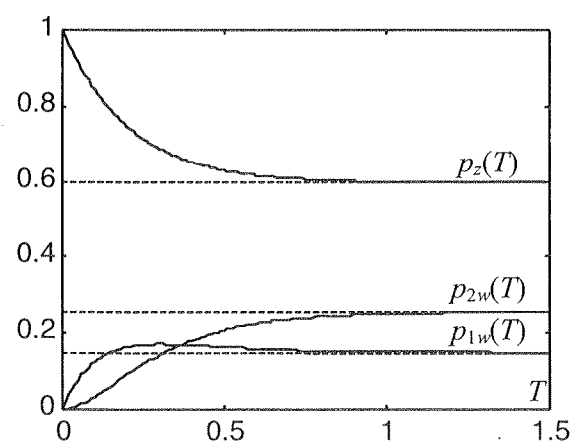

Fig. 5. Graphs of $p_{1w}(T)$, $p_{2w}(T)$, $p_z(T)$ functions for λ_1=8, λ_2=3, μ=2, a=0.64. Δ=8.04. Not overcontrolled system

The systems, for which a function $p_z(T)$ has its minimum, are called overcontrolled systems, otherwise they are called not overcontrolled systems. It can be proved that the system in the first variant is overcontrolled when

$$\frac{-s_1 Y}{s_2 X} > 1 \tag{21}$$

II. Second variant

$$[\lambda_2 - (\lambda_1 + \mu)]^2 - 4a\lambda_1\mu < 0 \tag{22}$$

In that case s_1 and s_2 take complex values. Characteristics $p_{1w}(T)$, $p_{2w}(T)$, $p_z(T)$, we were seeking, are in the form (23) with parameters displayed by (24) – (30).

$$\begin{cases} p_{1w}(T) = A + B \cdot e^{-KT} \cos(\frac{1}{2}\sqrt{-\Delta} \cdot T) + \frac{2P}{\sqrt{-\Delta}} \cdot e^{-KT} \sin(\frac{1}{2}\sqrt{-\Delta} \cdot T) \\ p_{2w}(T) = D + E \cdot e^{-KT} \cos(\frac{1}{2}\sqrt{-\Delta} \cdot T) + \frac{2Q}{\sqrt{-\Delta}} \cdot e^{-KT} \sin(\frac{1}{2}\sqrt{-\Delta} \cdot T) \\ p_z(T) = G + H \cdot e^{-KT} \cos(\frac{1}{2}\sqrt{-\Delta} \cdot T) + \frac{2R}{\sqrt{-\Delta}} \cdot e^{-KT} \sin(\frac{1}{2}\sqrt{-\Delta} \cdot T) \end{cases} \tag{23}$$

where:

$$A = -B = \frac{\lambda_2\mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \tag{24}$$

$$D = -E = \frac{a\lambda_1\mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \tag{25}$$

$$G = \frac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \tag{26}$$

$$H = \frac{\mu(\lambda_2 + a\lambda_1)}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \tag{27}$$

$$P = \frac{\lambda_2\mu(\lambda_1 - \lambda_2 + \mu) + 2a\lambda_1\mu^2}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} \tag{28}$$

$$Q = -\frac{a\lambda_1\mu(\lambda_1 + \lambda_2 + \mu)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} \tag{29}$$

$$R = \frac{\lambda_2\mu(-\lambda_1 + \lambda_2 - \mu) + a\lambda_1\mu(\lambda_1 + \lambda_2 - \mu)}{2(\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu)} \tag{30}$$

Diagrams of functions $p_{1w}(T)$, $p_{2w}(T)$, $p_z(T)$ are shown in Fig. 6.

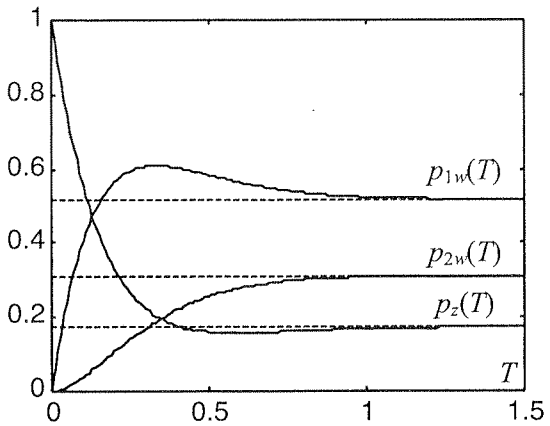

Fig. 6. Graphs of $p_{1w}(T), p_{2w}(T), p_z(T)$ functions for λ_1=2, λ_2=3, μ=6, a=0.9. Δ= -18.2

In the second variant, the $p_z(T)$ function oscillated round its horizontal asympthote. In Fig. 6 oscillations of the function cannot be observed because they are of small amplitudes.

It can be proved that all the systems in the second variant are overcontrolled, i.e. for each of them a function $p_z(T)$ has its minimum for $t \in (0,\infty)$.

III. Third variant

$$[\lambda_2 - (\lambda_1 + \mu)]^2 - 4a\lambda_1\mu = 0 \tag{31}$$

In that case s_1=s_2 and it takes a real value. Characteristics $p_{1w}(T)$, $p_{2w}(T)$, $p_z(T)$, we were seeking, are in the form (32) with parameters described by the same relationships as in the second variant.

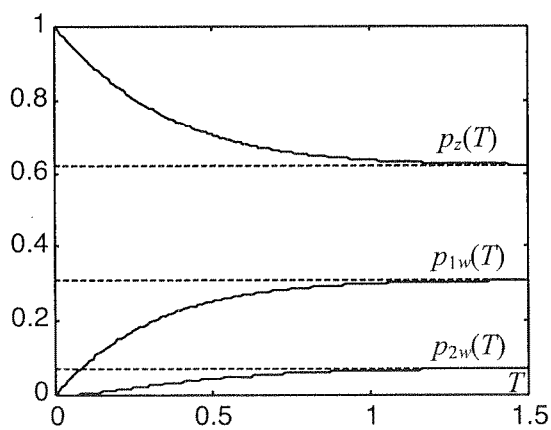

Fig. 7. Graphs of $p_{1w}(T), p_{2w}(T), p_z(T)$ functions for λ_1=2, λ_2=5.191, μ=1, a=0.6. Δ = 0

$$\begin{cases} p_{1w}(T) = A + B\cdot e^{-KT} + P\cdot t\cdot e^{-KT} \\ p_{2w}(T) = D + E\cdot e^{-KT} + Q\cdot t\cdot e^{-KT} \\ p_z(T) = G + H\cdot e^{-KT} + R\cdot t\cdot e^{-KT} \end{cases} \tag{32}$$

It can be proved that all the systems in the third variant are not overcontrolled. Diagrams of functions $p_{1w}(T), p_{2w}(T), p_z(T)$ are displayed in Fig. 7.

4. CONSTRUCTION OF THE INCLUDED MARKOV CHAIN

The included Markov chain for the system presented in Fig. 3 has two states. At a time instant just after an attempt of gaining access to the resources of satellite B, a system can be in state K (satellite A obtained connection) or Z (satellite A did not obtained connection). Graph of states of included Markov chain is displayed in Fig. 8.

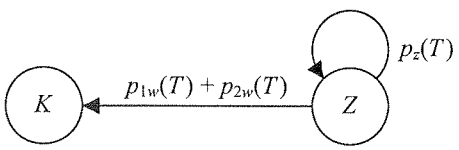

Fig. 8. Graph of states of included Markov chain

Denote by (i=0,1,2,...is the attempt number):
$q_K(i)$ – probability that the Markov chain is in state K,
$q_Z(i)$ – probability that the Markov chain is in state Z.
Equations describing the chain are given by:

$$[q_K(i+1), q_Z(i+1)] = [q_K(i), q_Z(i)]\begin{bmatrix} 1 & 0 \\ p_{1w}(T)+p_{2w}(T) & p_z(2T) \end{bmatrix} \tag{33}$$

where: i=0,1,2,...

The following intial conditions are also fulfilled (on the basis of (3))

$$[q_K(0), q_Z(0)] = [\frac{\lambda_2\mu + a\lambda_1\mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu}, \frac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu}] \tag{34}$$

Equations (33) and (34) define the Markov chain, which describes transient processes in the examined system.

Let us determine the probabilities p_i, that satellite A obtains a connection exactly in the attempt i (i=0,1,2,...). It can be done in two ways

$$\begin{cases} p_0 = p(A_{10} \cup A_{20}) = p(A_{10}) + p(A_{20}) \\ p_i = p([A_{1i} \cup A_{2i}] \cap B_{i-1} \cap \ldots \cap B_0) = \\ \quad = p([A_{1i} \cup A_{2i}] / B_{i-1}) p(B_{i-1} \cap \ldots \cap B_0) = \ldots = \\ \quad = [p(A_{1i} / B_{i-1}) + p(A_{2i} / B_{i-1})] p(B_{i-1} / B_{i-2}) \ldots p(B_1 / B_0) p(B_0), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1,2,3,\ldots \end{cases} \tag{35}$$

or exploiting from (33)

$$\begin{cases} p_0 = q_K(0) \\ p_i = q_K(i) - q_K(i-1) \qquad i = 1,2,3,\ldots \end{cases} \tag{36}$$

In both the first and the second cases we get (after exploiting (4))

$$\begin{cases} p_0 = \dfrac{\lambda_2\mu + a\lambda_1\mu}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \\ p_i = \dfrac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} p_z^{i-1}(T)[p_{1w}(T) + p_{2w}(T)], \qquad i = 1,2,3,\ldots \end{cases} \tag{37}$$

The values of p_i describe the distribution of satellite A's waiting time for the connection to satellite B. The following condition is fulfilled

$$\sum_{i=0}^{\infty} p_i = 1 \tag{38}$$

5. DETERMINATION OF SYSTEM PARAMETERS $N_{SR}(T)$ AND $T_{SR}(T)$

Mean number of attempts to obtain connection is given by

$$N_{sr}(T) = 1p_0 + 2p_1 + 3p_2 + \ldots = 1 + \sum_{i=1}^{\infty} i p_i \tag{39}$$

Mean waiting time to obtain connection is given by

$$T_{sr}(T) = 0T\,p_0 + 1T\,p_1 + 2T\,p_2 + \ldots = T\sum_{i=1}^{\infty} i\,p_i \tag{40}$$

Substituting (37) we have

$$N_{sr}(T) = 1 + \frac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \cdot \frac{1}{1 - p_z(T)} \tag{41}$$

$$T_{sr}(T) = T\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \cdot \frac{1}{1 - p_z(T)} \tag{42}$$

Diagram of function $N_{sr}(T)$ is displayed in Fig. 9 and Fig. 10. The first case concerns the system overcontrolled, while the second one concerns the system not overcontrolled. For the overcontrolled systems the function $N_{sr}(T)$ has its minimum. For the not overcontrolled systems the function $N_{sr}(T)$ has not minimum.

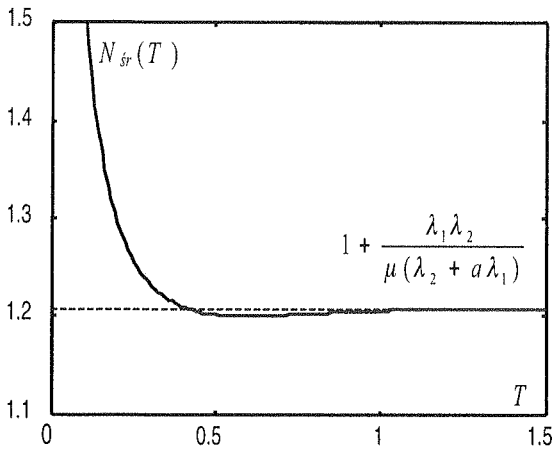

Fig. 9. Graph of $N_{sr}(T)$ function for λ_1=2, λ_2=3, μ=6, a=0.9. Δ = -18.2. Overcontrolled system

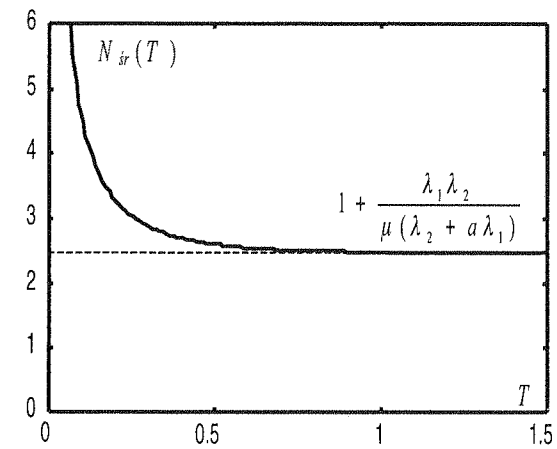

Fig. 10. Graph of $N_{sr}(T)$ function for λ_1=8, λ_2=3, μ=2, a=0.64. Δ = 8.04. Not overcontrolled system

Graph of the function $T_{sr}(T)$ is displayed in Fig. 11. Function $T_{sr}(T)$ can approach to its skew asymptote in different ways. For the systems of the second variant the function oscillates round the asymptote. For the systems of the first variant it aims to the asymptote from above (systems not overcontrolled) or from below (systems overcontrolled).

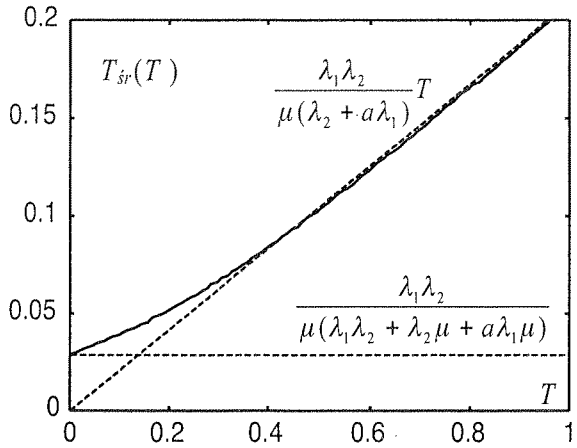

Fig. 11. Graph of $T_{sr}(T)$ function for λ_1=2, λ_2=3, μ=6, a=0.9. Δ = -18.2. Overcontrolled system

6. DETERMINATION OF THE OPTIMAL TIME T

Cost function $F(T)$ given by (1) after substituting (41) and (42) can be rewritten as

$$F_C(T) = \frac{\lambda_1\lambda_2}{\lambda_1\lambda_2 + \lambda_2\mu + a\lambda_1\mu} \cdot \frac{c_2T + c_1}{1 - p_z(T)} + c_1 \tag{43}$$

Diagram of function $F_C(T)$ is displayed in Fig. 12.

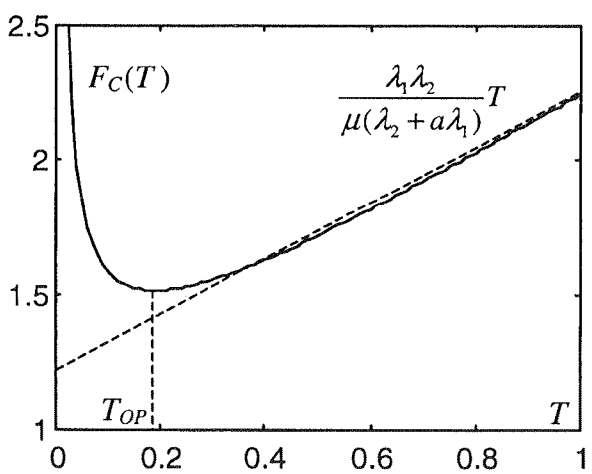

Fig. 12. Graph of $F_C(T)$ function for λ_1=2, λ_2=3, μ=6, a=0.9, c_1=1, c_2=5. Δ = -18.2. Overcontrolled system

Value T_{OP} can be calculated numerically. It is such $T \in (0,\infty)$, for which function $F_C(T)$ described by (43) has its minimum.

After calculating T_{OP} from (41) we obtain mean number of attempts which have to be made by satellite A to obtain connection to satellite B, but from (42) we obtain mean time of making these attempts in the situation when cost function $F_C(T)$ is minimized. From the equation (43) we can calculate the minimal mean cost $F_C(T_{OP})$.

7. CONCLUSION

In this work a problem of optimal time determination between attempts of gaining access to a service system in the generalized problem of impatient customer is solved.

The method of included Markov chain can be effectively exploited for analysis and optimization of non Markovian queueing systems. For the presented problem the method was sufficient to obtain precise results.

In general, in the problems solved by included Markov chain an analysis of stationary states is sufficient. To solve the stated task the analysis of transient processes in the tested system is necessary (eqs. (33) and (34)).

The methodology of solving a problem, can be easily applied in case when the model of service system is more complex than the model used in the work. The only condition of its exploitation is a possibility of transient processes determination in the service system (satellite B), i.e. characteristics of the service system (chapter 3). They can be obtained also by numerical methods. For the assumed model of satellite B, it was possible to solve the task by analytical methods.

8. REFERENCES

1. G. T. Artamonow, O. M. Briechow: Analiticzeskije wierojatnostnyje modieli funkcionirowanija EWM. Moskwa, Wyd. Energija, 1978.
2. T. Czachórski: Modele kolejkowe systemów komputerowych. Gliwice, Politechnika Śląska, skrypty uczelniane nr 1844, 1994.
3. B. Filipowicz: Modelowanie i optymalizacja systemów kolejkowych cz.I. Systemy markowskie. Kraków, Wyd. T. Rudkowski, 1995.
4. B. Filipowicz: Modelowanie i optymalizacja systemów kolejkowych cz.II. Systemy niemarkowskie. Kraków, Wyd. FHU Poldex, 2000.
5. R. Szostek: Systemy kolejkowe z niewykładniczym węzłem obsługi. Kraków, Wydawnictwa AGH, Elektrotechnika i Elektronika, 2000, Tom 19/1, str. 1-8.
6. F. Zitek: Stracony czas. Elementy teorii obsługi masowej. Poznań, PWN, 1977.

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY, 2008, 54, no 1, pp. 81–96

Radio Wave Propagation Conditions for Terrestrial Radiocommunications in the EHF Band

KRZYSZTOF BRONK[1], RYSZARD J. KATULSKI[2], ADAM LIPKA[1]

[1] The National Institute of Telecommunications
Jaskowa Dolina 15, 80-252 Gdańsk
e-mail: <K.Bronk, A.Lipka>@itl.waw.pl
[2] The Gdansk University of Technology
Department of Radiocommunictaion Systems and Networks
ul. Narutowicza 11/12, 80-952 Gdańsk
e-mail: rjkat@eti.pg.gda.pl

Received 2008.02.14
Authorized 2008.03.20

Designing an effective radiocommunication system which operates at EHF band is not an easy task. To achieve this goal it is important to take the meteorological effects into account. It is noteworthy that a combined impact of many weather factors should be considered when transmission frequency exceeds 10 GHz. The understanding of these mechanisms as well as their mitigation is a key to design and build efficient wireless communication systems operating at EHF. Therefore, in this study an influence of different weather conditions on effectiveness of the transmission is discussed. In addition, some possible solutions of reducing these meteorological effects are mentioned. Finally, the simulation results which confirm the usefulness of the EHF band are presented. It is important to note here that the main concern for this study is the fixed and mobile short-range terrestrial radiocommunication and all considerations included in this article are focused mainly on this part of the wireless communication.

Keywords: EHF band, radio wave propagation, attenuation, link availability, millimetre wave communication systems

1. INTRODUCTION

The utilization of extremely high frequencies is a very promising option in modern radiocommunications whose major merit is the possibility of implementing very wide channels and reaching high data rates with relatively simple modulation schemes. On the other hand, shifting the transmission into the band of 30 – 300 GHz is hardly a straightforward task and requires a substantially different approach comparing to the situation when lower frequencies

are being used. Another obstacle is a fact that propagation in the EHF band has not been fully researched yet and in some areas further work is still necessary. Nevertheless, the available analyses enable to indicate several most important factors that influence this kind of propagation, which also constitute the basis for designing radiolinks operating in the EHF Band. Among those factors that adversely affect the discussed propagation are: non-line-of-sight (NLOS) propagation and lack of the first Fresnel Zone clearance, diffraction due to uneven terrain, frequency selective fading due to multipath and attenuation and scattering due to vegetation.

It is easy to realize that all the above aspects are relevant to the transmission in lower frequency bands as well as in very high; when analyzing the EHF band transmission, the two most important factors are those that are virtually negligible at low frequencies, namely attenuation due to atmospheric gases and attenuation and depolarization due to precipitation (most notably rain). The first one should be treated as a constant, inevitable factor, depending on a given climatic zone and system parameters. Despite that, the current research show that the communication is realizable due to several attenuation windows that can be identified throughout the EHF Band, which is a similar situation to the optical transmission through fibers. The latter factor is variable and its negative effects are diminished by the assumed built-in link margin, which influences link availability. Rain and other kind of precipitations generally increase link attenuation, which results in the fact the radio link budget might not be balanced. To prevent from this situation, every real link is designed with a particular link margin which should eliminate a negative influence of the precipitation and determine the percentage of time in which radiolink is available. For practical reasons, existing links are not designed with the assumption of full availability (100 percent of time), because the required link margin would have to be too large.

Generally, due to increasing demand of broadband wireless communication services, in addition to the frequency congestion and management problems at lower frequency bands, design engineers have directed their attention towards the use of the EHF band for communication. This is caused by the recent demand for multimedia communication and high data rate capabilities.

2. MAJOR SOURCES OF THE RADIO WAVES DEGRADATION IN THE EHF BAND

Radio wave propagation negative influences at EHF bands include rain attenuation, rain and ice depolarization, gaseous absorption, cloud attenuation, melting layer attenuation, troposphere scintillation, and antenna wetting. These effects are related to the troposphere and weather conditions. Rain attenuation, gaseous absorption, cloud attenuation and melting layer attenuation are absorptive effects producing both signal attenuation and increase in antenna thermal noise. Troposphere scintillation is not absorptive but can produce attenuation as well as enhancement.

Terrestrial millimetre wave communication systems are mainly exploited in clear line-of-sight (LOS) environments to avoid high penetration losses due to buildings and foliage, multipath and diffraction effects as well. For these communication systems, in order to facilitate planning, installation and high-quality fiber-like system performance, it is vital to quantitatively and qualitatively characterize the terrestrial mm-wave propagation channel conditions.

In general, distortions of the radio waves at frequencies higher than 10 GHz are caused mainly by:

– Non-line-of-sight (NLOS) propagation and lack of the first Fresnel Zone clearance;
– Diffraction due to uneven terrain;
– Frequency selective fading due to multipath;
– Attenuation and scattering due to vegetation;
– Attenuation due to atmospheric gases;
– Attenuation and depolarization due to precipitation (rain, snow, fog, clouds etc).

These propagation effects are discussed below.

2.1. NLOS, DIFFRACTION AND FRESNEL ZONES

Diffraction allows radio signals to propagate around the curved surface of the Earth, beyond the horizon, and to propagate behind obstructions. The phenomenon of diffraction can be explained by Huygen's principle, which states that all points on a wavefront can be regarded as point sources and give rise to secondary wavelets. Combination of all the wavelets produces a new wavefront in the direction of propagation.

Fresnel Zones represent successive regions (figure 1) where secondary waves have a path length from the transmitter to receiver which is $n\lambda/2$ greater than the total path length of a line-of-sight path. Fresnel zones are elliptical in shape with the transmitting and receiving antenna at their foci. The successive Fresnel zones alternatively provide constructive and destructive interference to the total received signal. The radius of the nth Fresnel zone circle, denoted by r_n, can be expressed by:

$$r_n = \sqrt{\frac{n \cdot \lambda \cdot d_1 \cdot d_2}{d_1 + d_2}}, \tag{1}$$

where: r_n – nth Fresnel zone radius [m],
d_1 – distance to the transmitter [m],
d_2 – distance to the receiver [m],
λ – wavelength of the transmitted signal [m].

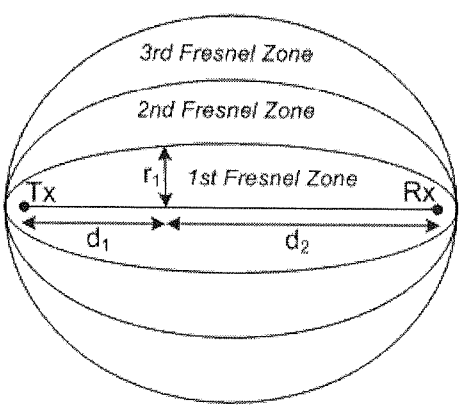

Fig. 1. The Fresnel Zones representation

Diffraction losses occur when secondary waves are blocked such that only a part of the energy is diffracted around the corner. The total received signal energy is given by the vector summation of energy contributions from all unobstructed Fresnel zones. The diffraction losses caused by obstacles can be calculated and taken into account as shown in [1]. In general, however, the attenuation can be approximated, as:

$$L_{diff} = \frac{-20 \cdot h}{r_1} + 10, \tag{2}$$

where h [m] denotes a difference between the highest terrain obstacle and the Line-Of-Sight propagation path (this value is negative when the obstacle stands out above the LOS line), r_1 [m] represents a radius length of the first Fresnel zone (see eq. 1) calculated in the place of the obstacle occurrence.

The given above expression shows that the most important, when considering radio wave propagation, is the first Fresnel zone in which most of the electromagnetic field energy is propagated.

It is important to note that, according to ITU-R recommendation [2], for millimetre radio systems the NLOS propagation is not valid due to large diffraction losses experienced when obstacles cause the propagation path to become NLOS. For these situations, multipath reflections and scattering will be the most likely signal propagation methods.

2.2. MULTIPATH PROPAGATION

In wireless telecommunications, multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. Causes of multipath include atmospheric ducting, ionospheric reflection and refraction, and reflection from terrestrial objects, such as mountains and buildings. It is important to note that many of these effects are highly unlikely in the case of the short-range terrestrial millimetre systems.

The effects of multipath include constructive and destructive interference, and phase shifting of the signal. This causes Rayleigh fading. The standard statistical model of it gives a distribution known as the Rayleigh distribution. Rayleigh fading with a strong line of sight content is said to have a Rician distribution.

In the case of millimetre radio systems, however, this effect is not as destructive as others described in this study. It is so because for EHF band mechanisms related with absorbing and scattering are much more dangerous than multipath propagation. Additionally, during millimetre transmission the LOS conditions are essential as well as narrow antennas beams are exploited in most cases, which leads to conclusion that multipath propagation is unlikely. In most cases this effect would only reduce a quality of a wireless connection but would not decrease its availability. It is, however, possible to take this effect into account because there are ITU-R recommendations [3, 9] which present this issue. Unfortunately, it is done only to a certain extent which is caused by the lack of the measurement data for frequencies higher than 15 GHz.

2.3. EFFECTS OF VEGETATION

When a radio wave reaches a rough surface, the reflected energy is spread in all directions due to scattering. At millimetre-wave frequencies, the dimensions of tree leaves and tree branches are large as compared to the wavelength. The tree leaves and branches also usually contain water and hence result in absorption and scattering of electromagnetic waves as they propagate through vegetation. Foliage can not only introduce attenuation and broadening of the beam but also depolarization of the electromagnetic wave. The transmission losses through vegetation are affected by various parameters such as the dielectric constant, density, physical size and shape. ITU-R provides an appropriate recommendation [4] and the effects of vegetation can be taken into account.

Attenuation in vegetation can be important in the case of terrestrial millimetre wave propagation. However, the wide range of conditions and types of foliage makes it difficult to develop a generalized prediction procedure. There is also a lack of suitable experimental data. According to ITU-R recommendation, however, some solutions are possible. It is important to note that they are valid only for frequencies lower than 60 GHz. There is no data available for higher bands.

In order to estimate the total field loss, the diffracted, ground reflected and through-vegetation scattering components are first calculated and then combined. The diffracted components consist of those bent over the top of the vegetation and those bent around the sides of the vegetation. These components and the ground reflected component are calculated using ITU-R Recommendations. The through or scattered component is calculated using a model based upon the theory of radiative energy transfer (RET). Detailed information can be found in the given above recommendation. The combined loss caused by all of these factors can be calculated as follows:

$$L_{veg} = -10 \log_{10} \left\{ 10^{\left(\frac{-L_{sidea}}{10}\right)} + 10^{\left(\frac{-L_{sideb}}{10}\right)} + 10^{\left(\frac{-L_{top}}{10}\right)} + 10^{\left(\frac{-L_{ground}}{10}\right)} + 10^{\left(\frac{-L_{scat}}{10}\right)} \right\}, \tag{3}$$

where: L_{veg} – total loss due to vegetation,

L_{sidea} – loss due to diffraction around side a of vegetation,

L_{sideb} – loss due to diffraction around side b of vegetation,

L_{top} – loss due to diffraction over top of vegetation,

L_{ground} – loss due to ground reflected component,

L_{scat} – loss due to scattered component.

Example charts given by the recommendation are shown in the figure 2. Unfortunately, these results are only for the case of 40 GHz (5 GHz and 10 GHz – not the EHF band). For different frequencies, however, calculations can be done as well with exploiting equations given in the above recommendation.

Another issue, considering the vegetation influence on millimetre wave propagation, is depolarization effect. Measurements at 38 GHz suggest that depolarization through vegetation may well be large, i.e. the cross-polar signal may be of a similar order to the co-polar signal when both transmitted through the vegetation. However, for the larger vegetation depths, required for this to occur, the attenuation would be so high that both the co-polar and cross-polar components would be below the dynamic range of the receiver.

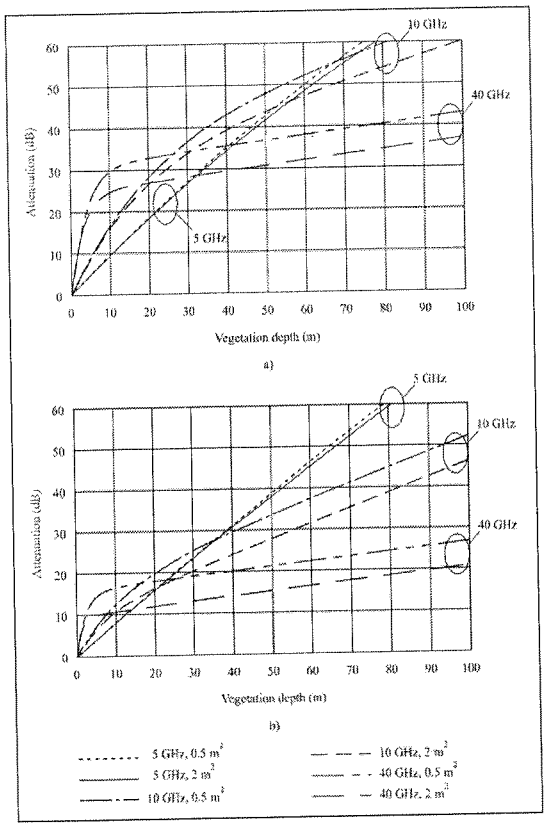

Fig. 2. Total loss due to vegetation for 0,5 m^2 and 2 m^2 illumination area, a) in leaf, b) out of leaf [4]

According to the recommendation it is also important to take into account some dynamic effects of the vegetation. It has been observed that where a link passes through foliage the received signal amplitude varies rapidly when the vegetation moves which is caused mainly by wind. Measurements at 38 GHz and 42 GHz have demonstrated that there is a strong correlation between the amplitude fluctuation rate and the wind speed.

When considering the effects of vegetation it is clear that the environment will not remain static. A receiver site may have one or more trees along the signal path that do not give a sufficient mean attenuation to take the received signal level below the system margin. However, it has been found that as the trees move, the signal level varies dynamically over a large range making the provision of a service unlikely. Several measurements of the attenuation levels during transmission through trees, as a function of time, have been made [4] and show an average reduction of the signal level of about 20 dB per tree. Considerable signal variability was found, with frequent drop-outs of up to 50 dB attenuation lasting for around 10 ms.

The ITU-R recommendation provides some data related to the dynamic effects of wind. Modelled time series and the standard deviations of signal amplitude for wind speeds, ranging from 0 to 20 m/s, are presented in figure 3 in comparison with measured data.

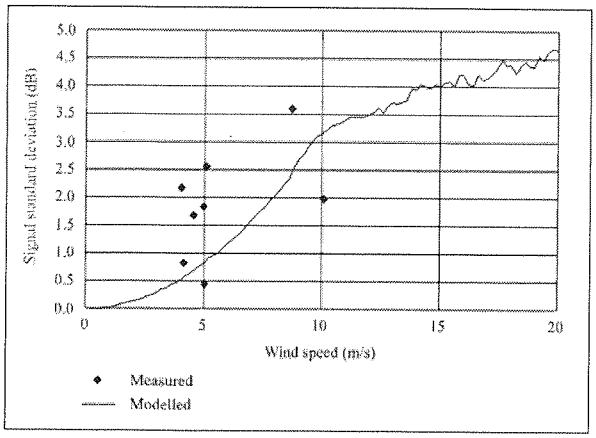

Fig. 3. Standard deviation of measured and modeled 40 GHz time series as a function of wind speed [4]

It should be noted that despite the fact that this type of model shows an inherent frequency dependence, the path length differences through trees are small and the fading across a typical 40 MHz bandwidth will appear flat. Rapid fading is due to the time variability of the medium.

Unfortunately, results and solutions given by the ITU-R recommendation are sufficient only in the particular cases but not in general. It is so due to the lack of the measured data for frequencies higher than 40 GHz. The situation is not hopeless, though, because it is often, when considering EHF systems in the terrestrial environment, that degradation of the radio waves is caused mainly by different, more destructive factors, and the influence of the vegetation is marginal. Therefore, in most cases a coarse estimation, shown in the ITU-R recommendation, can be sufficient.

2.4. GASEOUS ABSORPTION

Compared to other absorptive factors, gaseous absorption from oxygen and water in the atmosphere is small. However, gaseous absorption must be considered in link budget analysis since it is always present and its effect increases for higher frequencies, humidity and temperature. Specific attenuation due to gaseous influence varies as a function of frequency. ITU-R recommendation on how to predict absorption due to atmospheric gases is available [5].

In general, the clear cloudless atmosphere also absorbs radio signals, but below 30 GHz absorption losses are small except around the 22 GHz (resonance frequencies of water vapour) and they vary appreciably with water vapour content (figure 4).

Above 30 GHz (EHF band), however, the gaseous absorption issue becomes more destructive. Another resonance frequencies of water vapour can be observed at around 183 GHz. Additionally, at 60 and 119 GHz the resonance frequencies of oxygen are also present (figure 4). Between these marked thresholds in the EHF band atmosphere attenuation windows exist, as shown in table 1.

There are many atmospheric gases and pollutants that have absorption lines in the millimetre bands (such as SO_2, NO_2, O_2, H_2O, and N_2O), however, the absorption loss is present mainly due to water vapour and oxygen.

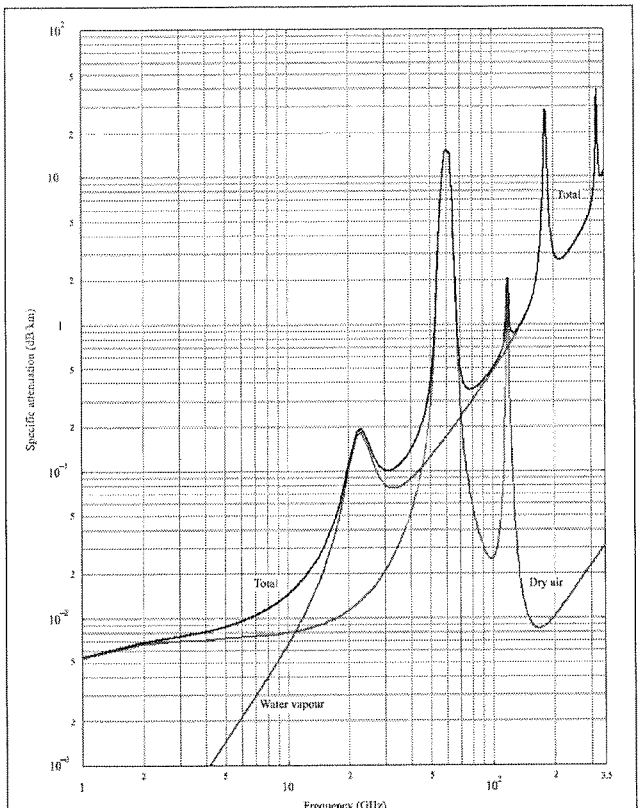

Fig. 4. Specific attenuation due to atmospheric gaseous absorption [5]

Table 1.

Attenuation windows of the atmosphere for the EHF band

Window ID	**Frequency range [GHz]**	**Clear air attenuation (Sea level, 15°C, RH* = 50%) [dB/km]**
W1	25 to 50	0,1 to 0,5
W2	70 to 115	0,5 to 2
W3	125 to 160	2 to 5
W4	200 to 250	5 to 10
*) RH – relative humidity [%]		

The ITU-R recommendation shows how to take the gaseous absorption effect into account. For a horizontal path or for slightly inclined paths close to the ground the attenuation, L_{abs} [dB], may be written as:

$$L_{abs} = \gamma \cdot d = (\gamma_o + \gamma_w) \cdot d, \tag{4}$$

where: d – path length [km],

γ_o – dry air attenuation [dB/km] (figure 4),

γ_w – water vapour attenuation [dB/km] (figure 4).

The mentioned above recommendation includes also some equations which allow to approximate the attenuation in the range of 1 to 350 GHz for both horizontal paths as well as inclined ones. This estimate is sufficient in the case of the terrestrial millimetre wave propagation.

It is important to note that the gaseous absorption effect is not only related with humidity (water vapour density) and frequency but also with a pressure and temperature of the air. The recommendation enables to exploit the knowledge of these factors and calculate the proper attenuation. The results in the case of standard values of these parameters and for the horizontal paths are presented in the figure 4. The *Total* attenuation depicted in this figure represents the one calculated according to the expression (4).

2.5. RAIN ATTENUATION

Rain attenuation is the most degrading effects of the atmosphere. Rain on a radio path causes fading, or "rain attenuation". Rain attenuation is a function of frequency, elevation angle, polarization angle, rain intensity, raindrop size distribution and raindrop temperature. This is the interference between raindrops and electromagnetic signals travelling through the atmosphere, which causes the radio waves distortions. When this phenomenon occurs, the transmission is weakened by absorption and scattering of the radio wave. Raindrops absorb and scatter radio wave energy, resulting in amplitude fluctuation and phase randomness in the received signal which degrades the reliability and performance of the communication link.

Any raindrop in the path of the signal, which is comparable to the half of the wavelength in diameter, will cause attenuation. Rain also depolarizes signals, converting energy from one polarization to another. The severity of rain attenuation and depolarization depends on how hard it is raining (described by the "rain rate" in millimetres of accumulation per hour), not on the total rain accumulation. Thus, areas subject to intense thunderstorms experience more severe propagation problems from rain than areas with a high average rainfall but few thunderstorms.

The ITU-R recommendation [6] provides a "semi-empirical" model for the rain attenuation calculations. The parameter, L_{rain} [dB], is obtained from the rain rate R [mm/h] using the power-law relationship:

$$L_{rain} = d_{eff} \cdot k \cdot R^{\alpha}. \tag{5}$$

Values for the coefficients k and α are determined as functions of frequency, f [GHz], in the range from 1 to 1 000 GHz. These coefficients can be rewritten either as k_H and α_H in the case of the horizontal polarization or as k_V and α_V in the case of the vertical polarization. The

ITU-R recommendation provides expressions which enable to calculate specific rain attenuation for different types of the polarization. Numerical values for the coefficients k and α at given frequencies as well as equations for calculations are given in the recommendation. The effective radio link length d_{eff}, which denotes a statistically calculated distance within rain, is a value which depends on the actual radio link length and the rainfall rate. Appropriate equations can be found in the recommendation.

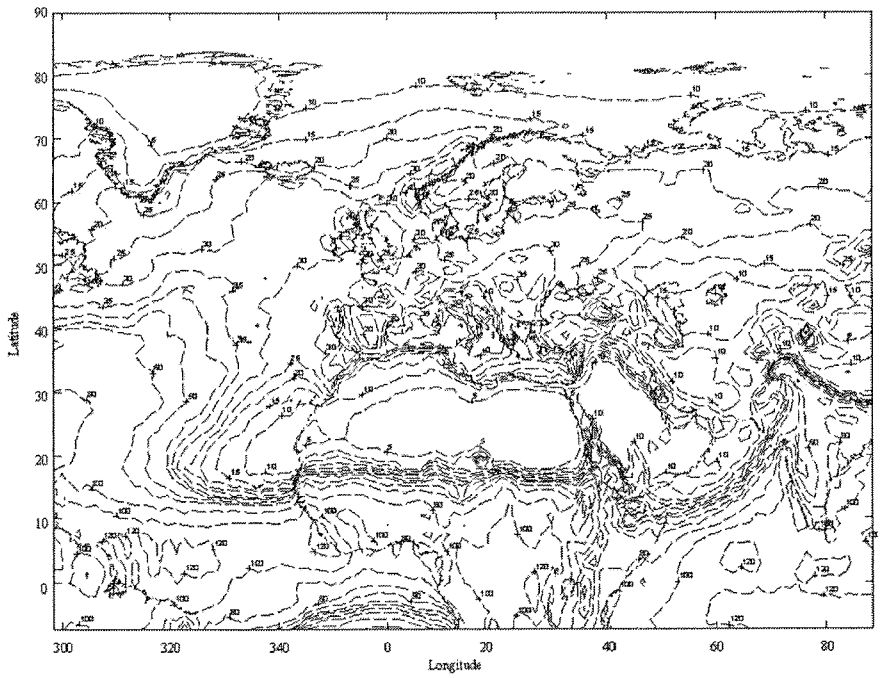

Fig. 5. Rain rate [mm/h] exceeded for 0,01% of the average year [7]

In order to compute the value of the specific rain attenuation, the rain rate must be obtained. It can be done with exploiting another ITU-R recommendation [7] which gives a solution for this issue. It presents how to derive the rainfall rate, exceeded for $p\%$ of the average year, for the desired latitude and longitude. Figure 5 presents the rainfall rates for latitudes and longitudes which are within the scope of this article. It is easy to note that in the case of Poland this value, exceeded for 0,01% of the average year, numbers about 25 mm/h. For Europe rainfall rate changes in the range of 15 (northern Europe) to 40 (southern Europe) mm/h.

2.6. CLOUD AND FOG ATTENUATION

Despite the fact that the effect of cloud attenuation is the main concern for the satellite communication it is described here because fog attenuation can still influence the terrestrial millimetre propagation to some degree. Radio waves are depolarized and attenuated by clouds

and fog but this mechanism is difficult to predict or model because it is not described by anything analogous to rain rate that can be measured on the ground. The small size of cloud/fog particles relative to wavelength makes cloud attenuation a function of cloud/fog temperature and water content along the propagation path. The ITU-R provides the recommendation which shows a method for predicting cloud and fog attenuation [8].

The propagation loss due to clouds and fog may be a factor of importance especially for microwave systems well above 10 GHz. For clouds or fog consisting entirely of small droplets, generally less than 0.01 cm, the Rayleigh approximation is valid for frequencies below 200 GHz and it is possible to express the attenuation in terms of the total water content per unit volume. Thus the cloud/fog attenuation can be written as:

$$L_{fog} = K_l \cdot M \cdot d \,, \tag{6}$$

where: L_{fog} – attenuation [dB] caused by the cloud/fog,

K_l – specific attenuation coefficient [(dB/km)/(g/m^3)],

M – liquid water density in the cloud or fog [g/m^3],

d – propagation path length within the cloud/fog [m].

At frequencies of the order of 100 GHz and above, attenuation due to fog may be significant. The liquid water density in fog is typically about 0.05 g/m^3 for medium fog (visibility of the order of 300 m) and 0.5 g/m^3 for thick fog (visibility of the order of 50 m).

A mathematical model based on Rayleigh scattering, which uses a double-Debye model for the dielectric permittivity of water, can be used to calculate the value of the specific attenuation coefficient for frequencies up to 1000 GHz. This solution can be also found in the above mentioned recommendation. For the purpose of this study, however, figure 6, which shows the values of K_l at frequencies from 5 to 200 GHz and temperatures between –8° C and 20° C, is presented. It is important to note that for cloud attenuation, the curve corresponding to 0°C should be used.

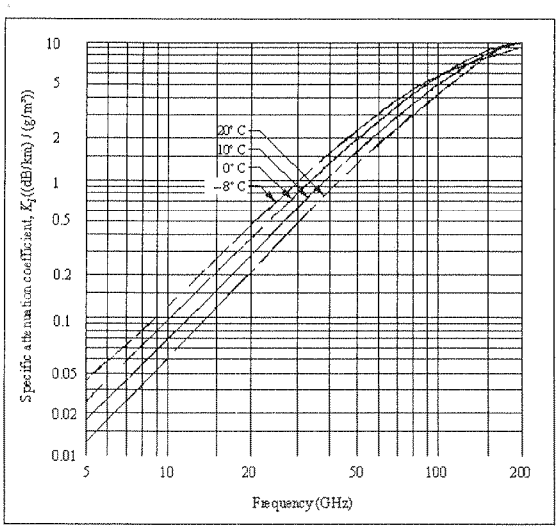

Fig. 6. Specific attenuation coefficient at various temperatures as a function of frequency [8]

2.7. COMBINED EFFECT OF PROPAGATION INFLUENCES

The factors described above are the individual propagation impairments that affect radio wave propagation. These effects have been studied individually for lower frequencies. Little data are available for frequencies above 30 GHz. However, with a few assumptions, which can be done due to specific of the millimetre and LOS propagation in the short-range terrestrial environment, a solution can be found. The propagation effects are related and can occur simultaneously. Hence, a model which describes the combined effect of propagation impairment is needed. There is a recommendation provided by ITU-R [9], which solves this issue.

In general, for frequencies above 10 GHz the total transmission loss for a radio system is given as:

$$L\,[dB] = 10 \cdot \log\left(\frac{4 \cdot \pi \cdot d}{\lambda}\right)^{\alpha} + L_{add}\,[dB]\,, \tag{7}$$

$$L_{add}\,[dB] = L_{diff} + L_{abs} + L_{rain} + L_{veg} + L_{fog}\,, \tag{8}$$

where: L_{add} – total excess attenuation [dB],
L – total attenuation [dB],
d – propagation path length [m],
λ – wavelength [m],
α – free space exponent.

The free space exponent in the case of the real free space equals 2. The ITU-R recommends, however, to use 2.2 as the value of this parameter in order to estimate the basic propagation loss for frequencies higher than 10 GHz when considering the short range propagation in urban areas. Due to the higher value of the free space exponent, an additional link margin is obtained, which can be useful in avoiding multipath propagation effects.

Other mechanisms such as antenna-wetting, depolarization due to rain or ice, dispersion, and troposphere scintillation influence also the propagation of radio waves. These effects, however, will not be discussed here because their importance as a destructive factors, comparing with the previously presented, is small.

3. COMPENSATION OF THE RADIO WAVES DISTORTIONS IN THE EHF BAND

To diminish distortions of the radio signals, system designers must incorporate compensation techniques in their designs. The main task of it is to mitigate the negative propagation influences caused by the effects described before. At higher frequencies compensation is an essential process for maintaining a proper quality of the service. It is important to add that most of the mechanisms inflecting the radio waves changes dynamically, therefore the compensation methods should be adaptive and even preventive.

The most common compensation techniques exploited in modern radiocommunication systems are as follows:

– **Built-in Link Margin**

In general, a link or system margin is allocated to a communication system in order to provide specific link availability. Since link margins are fixed, it is not possible to achieve an increase in capacity or availability using this technique.

– **Diversity**

At frequencies above 10 GHz, diversity techniques can be used e.g. to mitigate rain fade. Although diversity can be used to mitigate fading, the requirement of complex system design and separate antenna system (such as MIMO) may not justify the benefit.

– **Uplink Power Control (UPC)**

Uplink power control is a technique where the uplink EIRP is adjusted on an adaptive basis based on continuous measurement of the downlink carrier power. The amount of uplink power is estimated from the measured downlink fade.

– **Adaptive Modulation and Coding**

The AMC technique is used in wireless communications to match the modulation and coding parameters to the conditions of the radio link. Adaptive modulation systems require, however, some channel information at the transmitter in order to improve rate of transmission, and/or bit error rate.

4. THE PRACTICAL ASPECTS OF THE PROPAGATION CONDITIONS FOR FREQUENCIES HIGHER THAN 5 GHz

In this part of the article some simulation results are presented in order to show a propagation conditions for frequencies above 5 GHz in a more practical way. The given above frequency band can be divided into two different ranges. The former one is called *Super High Frequencies* (SHF) and is comprised of frequencies in the range of 3 to 30 GHz, which in general, are beyond the scope of this article but are presented for the comparison purposes. The latter one is called *Extremely High Frequencies* (EHF) and consists of frequencies above the SHF band up to 300 GHz.

The ITU-R recommendation [2], however, provides a considerably different division. Two separate propagation models, for the short range terrestrial radio systems at the standard SHF band, are available. In the case of frequencies up to 15 GHz a LOS propagation within street canyons can be considered. Proper equations for calculating the transmission loss are given in the recommendation. It is important to add that this model allows to estimate the attenuation in some range only (upper and lower bound of attenuation can be computed) and results are dependent on some inputs: distance, base station antenna height, mobile station antenna height and a level of traffic. The latter parameter is characterized by an effective height of the road (details can be found in the recommendation). Some standard values of these parameters were chosen and proper calculations were made. Results are presented in table 2. These results clearly show that the main destructive influence on propagation has the heavy road traffic which can cause additional attenuation of even 20 dB comparing with the light one. Additionally, a general rule can be created with reference to antennas height. Higher base station antennas do not reduce the propagation loss as clearly as higher mobile ones. It is important to add, at this point, that the presented here model do not take any meteorological phenomena into account in contrast to the next one.

Table 2.

Propagation attenuation for LOS situations within street canyons

ID	Conditions	Frequency [GHz]					
		5	7	8	11	13	15
		Attenuation [dB] (for the path length equals 1 km)					
1.	hm* = 1.8 m, hb** = 10 m, light traffic (hs*** = 0,5 m)	102-122	103-123	104-123	107-125	108-127	109-127
2.	hm = 1.8 m, hb = 20 m, light traffic (hs = 0,5 m)	100-119	103-121	104-122	107-124	108-125	109-126
3.	hm = 1.8 m, hb = 30 m, light traffic (hs = 0,5 m)	100-118	103-120	104-121	107-123	108-124	109-125
4.	hm = 1 m, hb = 20 m, light traffic (hs = 0,5 m)	104-124	104-124	104-124	107-126	109-127	110-128
5.	hm = 3 m, hb = 20 m, light traffic (hs = 0,5 m)	100-117	103-120	104-120	107-123	109-124	110-125
6.	heavy traffic (hs > hm)	117-137	120-140	121-141	124-144	125-145	127-147

*) hm – mobile terminal antenna height [m];
**) hb – base station antenna height [m];
***) hs – effective height of the road [m].

There is another propagation model provided by the mentioned before ITU-R recommendation. It can be exploited in the case of designing the short range terrestrial millimetre radio systems and also for those terrestrial solutions which operates in LOS conditions and at frequencies higher than 10 GHz. In general, it is described by the equation (7) and the whole model components were shown in the paragraph 2 of this article. Calculations results for those frequencies which are in the interest of this article and fit in the range of 10 GHz and more are presented in table 3.

The given table clearly shows that the atmospheric gases cause noticeable overall attenuation increment only at the band 52-55 GHz. The additional 4 dB may turn out to be an important component of the total transmission loss. It is noteworthy, however, that atmospheric gases are not the main concern when considering the given above frequencies.

As expected, rain appears to be the most destructive factor inflecting badly the transmission, even at the SHF band. Additionally, this kind of meteorological phenomenon is difficult to predict. Moreover, its influence on the radio wave propagation varies in the wide range which is a result of different rain rates that can occur. There is only one proper solution for this issue. Designer of the wireless SHF/EHF systems needs meeds to remember about this problem during their work. They should consider exploiting a proper link margin or different methods for mitigating the destructive influence of this factor.

The fog influence on the propagation is marginal at most frequencies. Its impact begins to be a problem only at the EHF band but even though the loss of 2 dB is much smaller than 11 dB in the case of the rain. Moreover, that level of attenuation is achievable only for a very thick fog.

The above results present as well the difference between standard free space transmission loss ($\alpha = 2$) and the one which utilizes coefficient $\alpha = 2.2$. More than 10 dB higher attenuation

Table 3.
Propagation attenuation for frequencies above 10 GHz and for terrestrial LOS propagation scenarios

Conditions	Frequency [GHz]																
	11	13	15	18	23	26	28	38	52-55 GHz band			71-76 GHz band			81-86 GHz band		
									52	53.5	55	71	73.5	76	81	83.5	86
	Attenuation [dB] (for the effective path length equals 1 km)																
Free space (α=2)	113.3	114.7	116	117.5	119.7	120.7	121.4	124	126.8	127	127.2	129.5	129.8	130.1	130.6	130.9	131.1
Free space (α=2.2)	124.6	126.2	127.6	129.3	131.6	132.8	133.5	136.4	139.4	139.7	140	142.4	142.7	142.9	143.7	144	144.2
Attenuation by atmospheric gases (ρ = 1013 hPa, T = -5°C, humidity = 5 g/m³)	0.016	0.02	0.026	0.048	0.138	0.1	0.087	0.112	0.739	1.65	4.578	0.424	0.34	0.308	0.294	0.298	0.305
Attenuation by atmospheric gases (ρ = 1013 hPa, T = 0°C, humidity = 5 g/m³)	0.015	0.019	0.025	0.046	0.137	0.098	0.084	0.107	0.709	1.601	4.455	0.404	0.324	0.293	0.28	0.283	0.29
Attenuation by atmospheric gases (ρ = 1013 hPa, T = 7.5°C, humidity = 7.5 g/m³)	0.018	0.023	0.032	0.063	0.196	0.137	0.114	0.133	0.725	1.595	4.345	0.478	0.409	0.387	0.39	0.401	0.416
Attenuation by atmospheric gases (ρ = 1013 hPa, T = 15°C, humidity = 12.5 g/m³)	0.024	0.033	0.048	0.098	0.315	0.219	0.18	0.198	0.808	1.663	4.32	0.667	0.616	0.611	0.647	0.675	0.708
Attenuation by atmospheric gases (ρ = 1013 hPa, T = 20°C, humidity = 12.5 g/m³)	0.023	0.032	0.046	0.096	0.314	0.214	0.175	0.189	0.778	1.62	4.213	0.636	0.586	0.581	0.614	0.64	0.67
Attenuation by rain (rain rate = 15 mm/h)	0.474	0.701	0.939	1.325	2.045	2.506	2.82	4.355	6.105	6.257	6.403	7.596	7.732	7.858	8.079	8.177	8.267
Attenuation by rain (rain rate = 25 mm/h)	0.882	1.267	1.666	2.303	3.445	4.153	4.624	6.832	9.181	9.377	9.563	11.04	11.2	11.35	11.61	11.72	11.82
Attenuation by rain (rain rate = 40 mm/h)	1.561	2.184	2.825	3.829	5.568	6.608	7.287	10.34	13.36	13.6	13.83	15.57	15.75	15.92	16.2	16.33	16.44
Attenuation by fog (T = 0°C, visibility = 300 m)	0.006	0.008	0.01	0.015	0.024	0.03	0.034	0.06	0.102	0.107	0.111	0.164	0.172	0.18	0.196	0.204	0.212
Attenuation by fog (T = 0°C, visibility = 50 m)	0.056	0.078	0.103	0.147	0.235	0.297	0.342	0.597	1.018	1.066	1.114	1.637	1.718	1.799	1.96	2.039	2.118
Attenuation by fog (T = 15°C, visibility = 300 m)	0.004	0.005	0.007	0.01	0.016	0.02	0.023	0.041	0.074	0.078	0.082	0.128	0.136	0.143	0.159	0.167	0.175
Attenuation by fog (T = 15°C, visibility = 50 m)	0.036	0.051	0.067	0.097	0.156	0.199	0.23	0.413	0.74	0.779	0.819	1.279	1.355	1.433	1.59	1.67	1.75

shows that additional factors like multipath propagation and diffraction have quite a huge impact on the transmission in this case of the wireless SHF/EHF terrestrial communication. Values of the free space loss (α = 2.2), for frequencies lower than 15 GHz, can be easily compared with results given in the table 2. The conclusion – attenuations are comparable. Both models seem to be compliant with each other which can, to a certain degree, confirm their correctness.

5. CONCLUSION

The recent increase in demand for multimedia wireless communication services have caused designers to develop more and more highly efficient systems operating in the EHF band. The personal communications network, the mobile broadband systems, local multipoint distribution system, fixed radio links, next generation Internet, and wireless local loop telephony, are often systems meant to operate in the EHF band. The availability of tremendous amount of bandwidth, the potential for high frequency reuse, and the reduced size of transmitting and receiving antenna elements and electronic components, make EHF band uniquely suited for these services. Implementation of these systems requires, however, a detailed knowledge of all factors influencing the radio wave propagation. Hence, there is a need for estimating not only the individual propagation effects but also the combined propagation influence for systems operating at EHF band. In general, the transmitted signals can be distorted by the multipath propagation as well as the weather conditions. It is important, therefore, to consider not only the influence of the typical propagation phenomena such as: reflection, diffraction and scattering but also an interaction of the radio waves with meteorological phenomena and atmosphere itself as well.

6. REFERENCES

1. Recommendation ITU-R – P.526-9, *Propagation by Diffraction.*
2. Recommendation ITU-R – P.1411-3, *Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz.*
3. Recommendation ITU-R – P.1407-2, *Multipath propagation and parameterization of its characteristics.*
4. Recommendation ITU-R – P.833-5, *Attenuation in vegetation.*
5. Recommendation ITU-R – P.676-6, *Attenuation by atmospheric gases.*
6. Recommendation ITU-R – P.838-3, *Specific attenuation model for rain for use in prediction methods.*
7. Recommendation ITU-R – P.837-4, *Characteristics of precipitation for propagation modelling.*
8. Recommendation ITU-R – P.840-3, *Attenuation due to clouds and fog.*
9. Recommendation ITU-R – P.530-11, *Propagation data and prediction methods required for the design of terrestrial line-of-sight systems.*

INFORMATION FOR AUTHORS OF E.T.Q.

An article published in other magazines can not be submitted for publishing in E.T.Q. TThe size of an article can not exceed 30 pages, 1800 character each, including figures and tables.

Basic requirements

The article should be submitted to the editorial stuff as a one side, clear, black and white computer printout in two copies. The article should be prepared in English. Floppy disc with an electronic version of the article should be enclosed. Preferred wordprocessors: WORD 6 or 8.

Layout of the article.

- Title.
- Author (first name and surname of author/authors).
- Workplace (institution, adress and e-mail).
- Concise summary in a language article is prepared in (with keywords).
- Main text with following layout:
  - Introduction
  - Theory (if applicable)
  - Numerical results (if applicable)
  - Paragraph 1
  - Paragraph 2
  - .....
  - .....
  - Conclusions
  - Acknowledgements (if applicable)
  - References
- Summary in additional language:
  - Author (firs name initials and surname)
  - Title (in Polish, if article was prepared in English)
  - Extensive summary, hawever not exceeching 3600 characters (along with keywords) in Polish, if artide was prepared in English). The summary should be prepared in a way allowing a reader to obtaoin essential information contained in the artide. For that reason in the summary author can place numbers of essential formulas, figures and tables from the article.

Pages should have continues numbering.

Main text

Main text cannot contain formatting such as spacing, underlining, words written in capital letters (except words that are commonly written in capital letters). Author can mark suggested formatting with pencil on the margin of the article using commonly accepted adjusting marks.

Text should be written with double line spacing with 35 mro left and right margin. Titles and subtitles should be written with small letters. Titles and subtitles should be numberd using no mor than 3 levels (i.e. 4.1.1.).

Tables

Tables with their titles should be places on separate page at the end of the article. Titles of rows and columns should be written in small letters with double line spacing. Annotations concerning tables should be placed directly below the table. Tables should be numbered with Arabic numbers on the top of each table. Table can contain algorithm and program listings. In such cases original layout of the table will be preserved. Table should be cited in the text.

Mathematical formulas

Characters, numbers, letters and spacing of the formula should be adequate to layout of main text. Indexes should be properly lowered or raised above the basic line and clearly written. Special characters such as lines, arrows, dots

should be placed exactly over symbols which they ar attributed to. Formulas should be numbered with Arabic numbers place in brackets on the right side of the page. Units of measure, letter and graphic symbols should be printed according to requirements of IEE (International Electronical Commission) and ISO (International Organisation of Standardisation).

References

References should be placed at the and of the main text with th subtitle „References". Refrences should be numbered (without brackets) adequately to references placed in the text. Examples of periodical [@], non-periodical [2] and book [3] references:

1. F. Valdoni: A new milimetre wave satelite. E.T.T. 1990, vol. 2, no 5, pp. 141–148
2. K. Anderson: A resource allocation framework. XVI International Symposium (Sweden). May 1991. paper A 2.4
3. Y.P. Tvidis: Operation and modeling of the MOS transistors. Nw York. McGraw-Hill. 1987. p. 553

Figures

Figures should be clearly drawn on plain or milimetre paper in the format not smaller than 9×12 cm. Figures can be also printed (preferred editor – CorelDRAW). Photos or diapositives will be accepted in black and whit format not gratd than 10×15 cm. On th margin of each drawing and on th back side of each ohoto authorp name and abbreviation of the title of article should be placed. Figure's captures should be given in two languages (first in the language the article is writtes in and then in additional language). Figure's captures should be also listed on separate page. Figures should be cited in the text.

Additional information

On the separate page following information should be placed:

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

Author is entiled to free of charge 20 copies of article. Additional copies or the whole magazine can be ordered at publisher at th one's expense.

Author is obliged to perform the author's correction, which should be accomplished within 3 days starting from the date of receiving the text from the editorial stuff. Corrected text should be returned to the editorial stuff personally or by mail. Correction marks should be placed on th margin of copies received from the editorial stuff or if needed on separate pages. In the case when the correction is not returned within said time limit, correction will be performed by technical editorial stuff of the publisher.

In case of changing of workplace or home address Authors are asked to inform the editorial stuff.