

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK
ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND
TELECOMMUNICATIONS
QUARTERLY

TOM 50 — ZESZYT 2

WARSZAWA 2004

RADA REDAKCYJNA

Przewodniczący

Prof. dr hab. inż. STEFAN HAHN
czł. rzecz. PAN

Członkowie

prof. dr hab. inż. DANIEL JÓZEF BEM — czł. koresp. PAN, prof. dr hab. inż. MICHAŁ BIAŁKO — czł. rzecz. PAN, [prof. dr hab. inż. JAN CHOJCAN], prof. dr hab. inż. MAREK DOMAŃSKI, prof. dr hab. inż. ANDRZEJ HAŁAŚ, prof. dr hab. inż. JÓZEF MODELSKI, prof. dr inż. JERZY OSIOWSKI, prof. dr hab. inż. EDWARD SĘDEK, prof. dr hab. inż. MICHAŁ TADEUSIEWICZ, prof. dr hab. inż. WIESŁAW WOLIŃSKI — czł. koresp. PAN, prof. dr inż. MARIAN ZIENTALSKI

REDAKCJA

Redaktor Naczelny

prof. dr hab. inż. WIESŁAW WOLIŃSKI

Zastępca Redaktora Naczelnego

doc. dr inż. KRYSTYN PLEWKO

ADRES REDAKCJI

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

Dyżury Redakcji: poniedziałki i piątki, godz. 14–16
tel/fax (022) 660 77 37

Telefony domowe: Redaktora Naczelnego: 812 17 65
Zast. Red. Naczelnego: 826 83 41
Sekretarza Odpowiedzialnego: 633 92 52

Ark. wyd. 14,0	Ark. druk. 11,25	Podpisano do druku w maju 2004 r.
Papier offset. kl. III 80 g. B-1		Druk ukończono w czerwcu 2004 r.

Skład, druk i oprawa: Warszawska Drukarnia Naukowa PAN
00-656 Warszawa, ul. Śniadeckich 8
Tel./fax: 628-87-77

Szanowni Autorzy

„Kwartalnik Elektroniki i Telekomunikacji” — Electronics and Telecommunications Quarterly jest kontynuatorem tradycji powstałego 50 lat temu kwartalnika p.t. „Rozprawy Elektrotechniczne”.

Kwartalnik jest czasopismem Komitetu Elektroniki i Telekomunikacji Polskiej Akademii Nauk. Wydawany jest przez Warszawską Drukarnię Naukową PAN. Kwartalnik jest czasopismem naukowym, na którego łamach są publikowane artykuły i komunikaty prezentujące wyniki oryginalnych prac teoretycznych i doświadczalnych, a także przeglądowych. Związane są one z szeroko rozumianymi dziedzinami współczesnej elektroniki, telekomunikacji, mikroelektroniki, oproelektroniki, radiotechniki i elektroniki medycznej.

Autorami publikacji są wybitni naukowcy, znani specjaliści o wieloletnim doświadczeniu, a także młodzi badacze — głównie doktoranci.

Artykuły charakteryzują się oryginalnym ujęciem zagadnienia, interesującymi wynikami badań, krytyczną oceną teorii lub metod, omówieniem aktualnego stanu, lub postępu danej gałęzi techniki oraz omówieniem perspektyw rozwojowych. Sposób pisania matematycznej części artykułów zgodny jest z wytycznymi IEC (International Electronics Commision) oraz ISO (International Organization of Standarization).

Wszystkie publikowane w Kwartalniku artykuły są recenzowane przez znanych krajowych specjalistów, co zapewnia że publikacje te są uznawane jako autorski dorobek naukowy. Opublikowane w kwartalniku wyniki prac naukowych zrealizowanych w ramach „GRANTów” Komitetu Badań Naukowych spełnia więc jeden z wymogów stawianych tym pracom.

Czasopismo dociera do wszystkich zajmujących się elektroniką i telekomunikacją krajowych ośrodków naukowych oraz technicznych, a także szeregu instytucji zagranicznych. Jest ponadto prenumerowane przez liczne grono specjalistów i biblioteki.

Każdy Autor otrzymuje bezpłatnie 20 egzemplarzy nadbitek swojego artykułu, co ułatwia przesłanie go do indywidualnych wybranych przez Autora osób i instytucji w kraju lub za granicą. Ułatwia to dodatkowo fakt, że w Kwartalniku są publikowane artykuły w języku angielskim.

Nadesłane do redakcji artykuły są publikowane w terminie około pół roku, w przypadku sprawnej współpracy Autora z Redakcją. Wytyczne dla Autorów dotycząc formy publikacji są zamieszczone w zeszytach Kwartalnika, można je także otrzymać w siedzibie Redakcji.

Artykuły można dostarczać osobiście, lub pocztą pod adresem zamieszczanym na stronie redakcyjnej w każdym zeszycie.

Redakcja

SPIS TREŚCI

S. Kozieł: Kanoniczne struktury filtrów G_m -C nieparzystego rzędu.....	143
W. Rakowski: Pakiety falkowe w kompresji sygnałów i obrazów.....	159
S. Żaba: Szeregowanie dynamiczne dla grup wiadomości w rozproszonych systemach czasu rzeczywistego.....	179
J. Kowalski, T. Kacprzak and K. Ślot: Nowy układ scalony CMOS VLSI analogowego filtru medianowego do aplikacji przetwarzania obrazu w czasie rzeczywistym	209
M. Gajer: Szeregowanie niezależnych, wyłączalnych i periodycznych zadań wieloprocessorowych dla trzech procesorów.....	231
A. Jakubiak, A. Chrustowski: Zakłócenia bierne modelowane rozkładem Suzuki	251
S. Żaba: Badania eksperymentalne magistral miejscowych dla wybranych metod szeregowania wiadomości.....	261
E. Jaszczyszyn: Analiza pełnofalowa anteny skanującej na podłożu ferroelektrycznym.....	287
Informacje dla Autorów (w jęz. polskim)	313

CONTENTS

S. Kozieł: Canonic Structures of Odd-Order Elliptic G_m -C Filters	143
W. Rakowski: Wavelet Packets in signal and image compression.....	159
S. Żaba: Dynamic message groups scheduling in distributed real times systems.....	179
J. Kowalski, T. Kacprzak and K. Ślot: New VLSI CMOS analogue median filter for real-time video applications	209
M. Gajer: The scheduling of independent, pre-emptive, periodic multiprocessor tasks for three procesors	231
A. Jakubiak, A. Chrustowski: Urban clutter modelled by Suzuki distribution.....	251
S. Żaba: Experimental testing of fieldbuses for some message scheduling methods.....	261
E. Jaszczyszyn: Full wave analysis of scan antenna on ferroelectric substrate	287
Information for the Authors (w jęz. ang.)	316

Canonic Structures of Odd-Order Elliptic G_m -C Filters

SŁAWOMIR KOZIEL

*Faculty of Electronics, Telecommunications and Informatics,
Gdańsk University of Technology, 80-952 Gdańsk, Poland
e-mail: koziel@ue.eti.pg.gda.pl.*

Otrzymano 2003.02.21

Autoryzowano 2003.05.14

In the paper, novel canonic structures of odd-order continuous-time elliptic G_m -C filters are proposed and discussed. The structures are characterized by very small number of components, especially transconductance amplifiers. This makes them suitable for low-power and low-sensitivity applications. A detailed comparison of the novel third order structure with other implementations such as leap-frog with distributed input and leap-frog with floating capacitor is carried out including sensitivity and other properties important for IC implementation. It follows that the proposed filter structures constitute a very interesting alternative to the structures known from the literature so far.

Keywords: G_m -C filters, elliptic filters, canonic structures

1. INTRODUCTION

Continuous-time (CT) filters with $j\omega$ -axis transmission zeros, specifically elliptic filters, are required in many applications because their steeper transition (better cut-off rate) between pass- and stop-band [1]–[3], [5] permits better utilization of the frequency spectrum. In active implementations, based on transconductance amplifiers and capacitors (G_m -C filters), an elliptic filter is typically realized as a topology that simulates a passive LC ladder. Since other configurations with possibly advantageous properties are conceivable, our particular aim is to find alternative structures that may lead to practical filters, and to investigate their properties. We are especially interested in so-called canonic structures which are defined as structures containing minimum number of components. This interest in canonic structures follows from the fact that small number of components implies reduction of power consumption, chip area, as well as improvement of sensitivity properties.

Recent years have seen considerable interest in G_m -C structures since they are easy to synthesize, provide suitable solutions for various signal-processing tasks [1]–[5], and offer attractive possibilities for realizing high-frequency CT filters in fully integrated form. Indeed, several integrated-circuit (IC) applications for G_m -C filters have been proposed [12]–[21]. Although many synthesis and design methods for different specific types and architectures of G_m -C filters have been reported [1]–[5], [8]–[22], no generally valid design theory appears to be available. Recent paper [8] presents a systematic procedure of generating elliptic filter structures based on a general G_m -C filter model introduced in [6] and [7]. In the present paper, we propose novel canonic structures of odd-order G_m -C filters that implement elliptic transfer function. These structures are a “mixture” of the structures presented in [8] in a sense that they contain both input signal distribution and floating capacitors.

The paper is organized as follows. In Section 2, we recall some basic facts concerning elliptic filters and give a short survey of two classes of filters developed in [8]. In Section 3 we introduce new canonic structures of elliptic G_m -C filters. Section 4 provides us with a detailed performance comparison of third-order structures. We take into consideration their sensitivity properties as well as other characteristics that are important in practical applications, especially in monolithic form. Remarks and conclusions are contained in Section 5.

2. ELLIPTIC FILTERS

In this work we are dealing with elliptic filters. Recall that elliptic approximation is one of the widely used pole-zero approximations. It is also referred to as the Cauer approximation [1]. In an elliptic filter the transfer function has zeros on the imaginary axis which are distributed across the stop band. A typical gain response of an elliptic filter is shown in Fig. 1. An elliptic filter is characterized by the minimum inband gain A_1 (corresponding inband ripple is $\delta = -20 \log A_1$), the minimum stop-band attenuation $-20 \log(A_2)$, the selectivity ω_s/ω_p (here $\omega_p = 1$ since we use normalized frequency), and the order n . A general form of the transfer function for an elliptic filter is the following [1]

$$H(s) = \frac{K \prod_{i=1}^m (s^2 + a_i)}{s^n + b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \dots + b_1s + b_0} \quad (1)$$

where K is chosen so that the peak gain equals unity (0 dB).

In papers [6] and [7] a general structure of G_m -C filter and its matrix description was introduced. Based on this model, two classes of G_m -C filters implementing odd-order elliptic transfer function (1) was introduced in [8]. A general structure of the filter of the first class is shown in Fig. 2. The order n of the filter is written as $n = 2k + 1$ (k is a natural number), which is for the convenience of notation and to emphasize

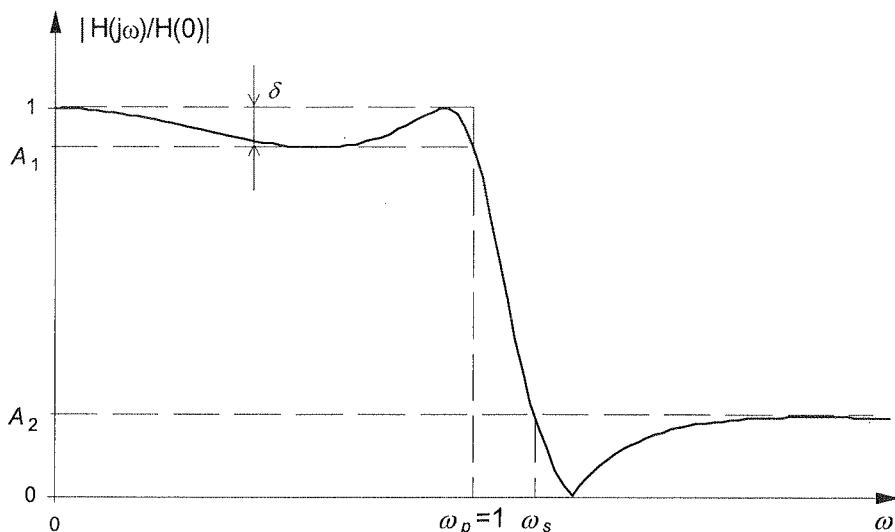


Fig. 1. Typical gain response of an elliptic filter (here: a third order one)

Rys. 1. Typowa charakterystyka częstotliwościowa filtru eliptycznego (tu: 3-go rzędu)

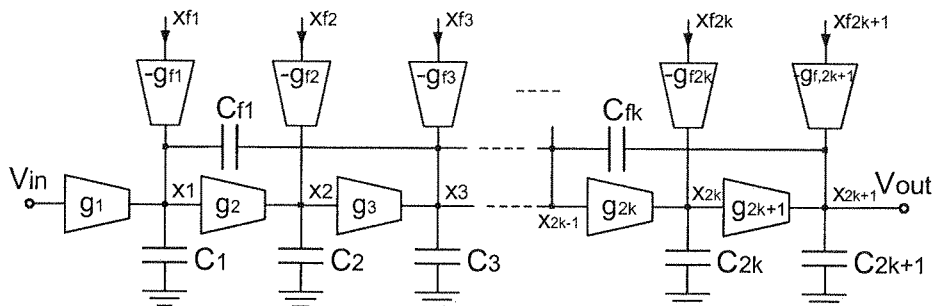


Fig. 2. General odd order elliptic filter with floating capacitors

Rys. 2. Ogólna struktura filtru eliptycznego zawierającego kondensatory nieuziemiione

the fact that n is odd. Internal nodes of the filter are denoted as x_i , $i = 1, 2, \dots, n$. The filter in Fig. 2 contains grounded capacitors C_i , $i = 1, 2, \dots, n$, floating capacitors C_{fj} , $j = 1, 2, \dots, k$, feedforward transconductors g_i , $i = 1, 2, \dots, n$, and feedback transconductors g_{fi} , $j = 1, 2, \dots, n$. The symbol x_{fi} in the input of transconductor g_{fi} means that the feedback signal at node x_i is taken from the node x_{fi} . In general, the feedback signal at node x_j can be taken from one of the nodes $x_j, x_{j+1}, \dots, x_{2k+1}$ (i.e. $x_{fj} \in x_j, x_{j+1}, \dots, x_{2k+1}$). However, it is evident that if $x_{fj} = x_j$ for some $j < 2k + 1$, then the filter becomes a cascade of sections of order lower than $2k + 1$. Otherwise, the filter is a multiple integrator loop feedback one. Two most popular structures are: leap-frog (LF) and inverse follow-the-leader (IFLF). For LF filter we have $x_{fj} = x_{j+1}$ for $1 \leq j < 2k + 1$ and $x_{f2k+1} = x_{2k+1}$, while for IFLF one: $x_{fj} = x_{2k+1}$, $1 \leq j \leq 2k + 1$.

As the feedback signal at node $x_i (i < n)$ can be taken from $n - i$ different nodes (as far as we consider multiple integrator loop feedback structures) we have in total $(n - 1)!$ different multi-loop structures (cf. [8]). With respect to the feedback scheme, LF and IFLF structures can be considered as two 'extreme' ones since in LF filter the feedback signal at node x_i is taken from the closest node, while in IFLF filter from the farthest one (see [8] for more details). Figs. 3 and 4 show diagrams of LF and IFLF structures with floating capacitors.

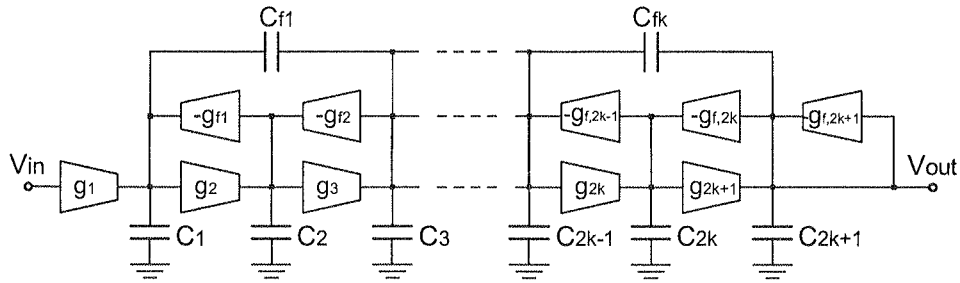


Fig. 3. Odd order LF elliptic filter with floating capacitors

Rys. 3. Filtr LF nieparzystego rzędu zawierający kondensatory nieuziemiene

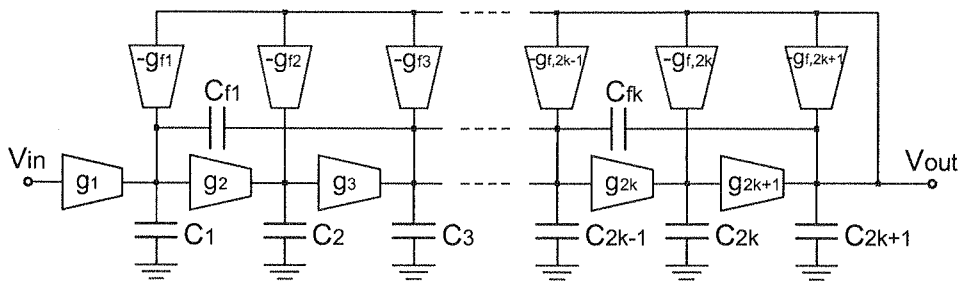


Fig. 4. Odd order IFLF elliptic filter with floating capacitors

Rys. 4. Filtr IFLF nieparzystego rzędu zawierający kondensatory nieuziemiene

The second class of elliptic filter structures can be obtained by assuming a distribution of the input signal into some of the internal nodes of the filter. A diagram of a respective general structure of odd-order elliptic filter is shown in Fig. 5. The filter in Fig. 5 contains only grounded capacitors and $k + 1$ input transconductors denoted as g_{bi} , $i = 1, 3, \dots, 2k + 1$. The feedback scheme is similar as for the filter in Fig. 2. However, not all feedback schemes are allowed. The necessary and sufficient condition for the filter in Fig. 5 to implement an elliptic transfer function is the following [8]:

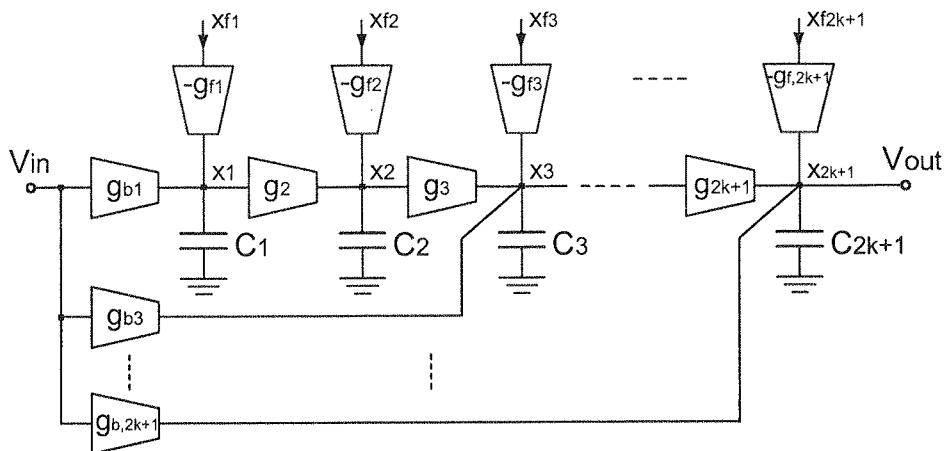


Fig. 5. General odd order elliptic filter with distributed input

Rys. 5. Ogólna struktura filtru eliptycznego z dystrybucją sygnału wejściowego

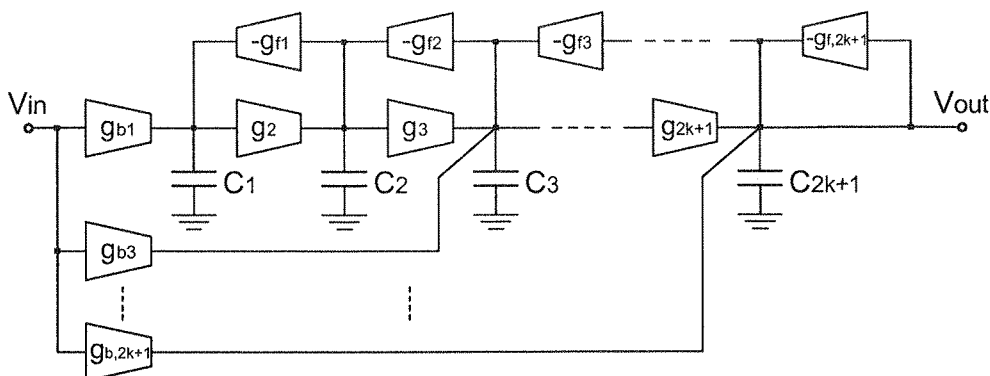


Fig. 6. Odd order LF elliptic filter with distributed input

Rys. 6. Filtr LF nieparzystego rzędu z dystrybucją sygnału wejściowego

$$\begin{aligned} \text{if } i \text{ is an odd (even) number and } x_{fi} = x_j \text{ for some } j, \\ \text{then } j \text{ is even (odd) or } j = 2k + 1 \end{aligned} \quad (2)$$

The number of different structures depending on x_{fi} is smaller than $(n-1)!$ (i.e. $(2k)!$) because the condition (2) significantly restricts the “feasible” feedback schemes. The total number of different structures in this case is $(k+1)k!$. Thus, there are many more elliptic filter structures with floating capacitors than with only grounded capacitors. For example, if we consider filters of seventh order, i.e. $k = 3$, there are $(2k)! = 720$ structures with floating capacitors but just $(k+1)k! = 144$ containing only grounded

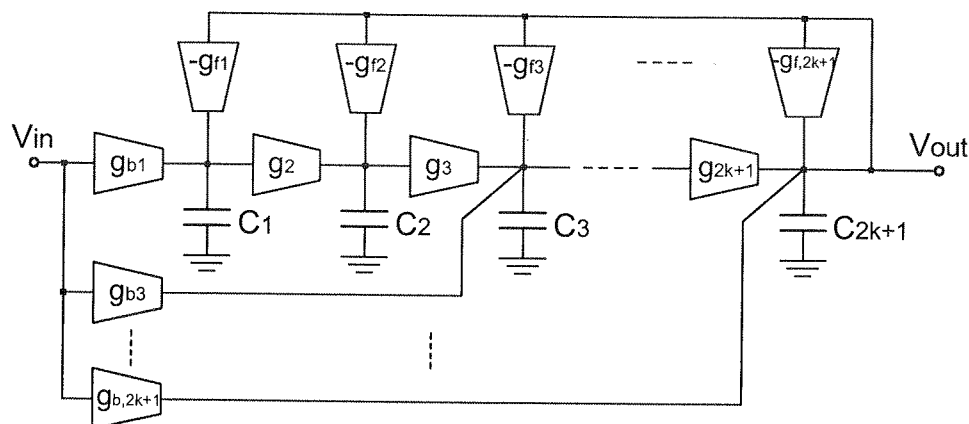


Fig. 7. Odd order IFLF elliptic filter with distributed input

Rys. 7. Filtr IFLF nieparzystego rzędu z dystrybucją sygnału wejściowego

capacitors. One can easily check that the LF and IFLF structures, shown in Figs. 6 and 7, satisfy (2).

3. NOVEL STRUCTURE OF AN ELLIPTIC G_m -C FILTER

In this section we introduce a novel structure of G_m -C filter that implements an odd-order elliptic transfer function (1). A diagram of this structure is shown in Fig. 8.

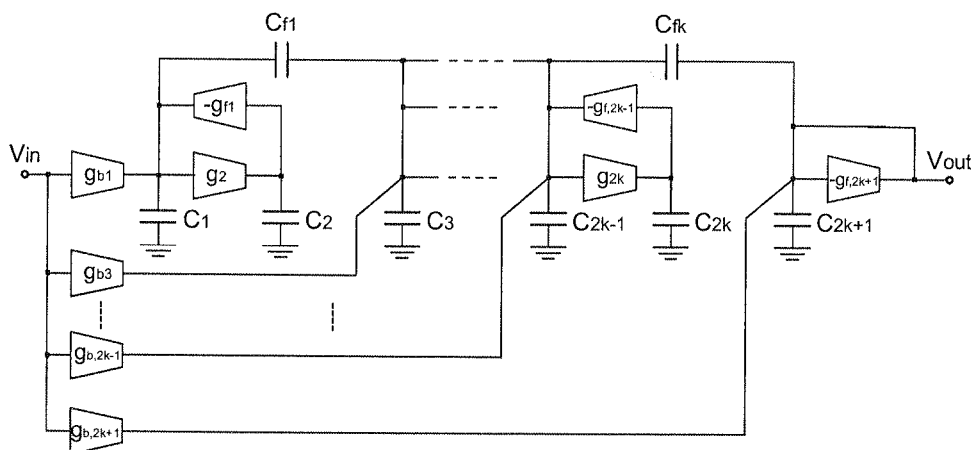


Fig. 8. Odd-order canonic elliptic filter with distributed input and floating capacitors — general structure

Rys. 8. Filtr eliptyczny nieparzystego rzędu z dystrybucją sygnału wejściowego i kondensatorami nieziemionymi — struktura ogólna

Note that the structure in Fig. 8 contains both grounded and floating capacitors but also $k + 1$ input transconductors. Thus, it can be considered as a kind of combination of the structures in Figs. 2 and 5. It should be also emphasized that the filter in Fig. 8 contains very small number of transconductors and the total number of elements is smaller than the number of elements of filters in Figs. 2 and 5. This means that we can call it a canonic structure. Table 1 shows the number of elements of filters in Figs. 2, 5 and 8 as a function of filter order n .

Table 1
Number of elements for the filter structures in Figs. 2, 5 and 8 as function of transfer function order n

Liczba elementów struktur z Rys. 2, 5 i 8 w funkcji rzędu funkcji przenoszenia

Number of elements	Structure		
	Fig. 2 (floating capacitors)	Fig. 5 (distributed input)	Fig. 8 (distributed input & floating capacitors)
Transconductors	$2n$	$(5n - 1)/2$	$(3n + 1)/2$
Capacitors	$(3n - 1)/2$	n	$(3n - 1)/2$
Total	$(7n - 1)/2$	$(7n - 1)/2$	$3n$

It is possible to derive simple recursive formulas that allow us to calculate the transfer function of the filter structure in Fig. 8. Let us first introduce the following notation (recall that $n = 2k + 1$, where n is the order of the transfer function):

$$\begin{aligned}
 D_0(s) &= 1 \\
 D_1(s) &= s\hat{C}_1 \\
 D_j(s) &= \begin{cases} sC_j D_{j-1}(s) + g_j g_{f,j-1} D_{j-2}(s) & j = 2, 4, \dots, 2k \\ s\hat{C}_j D_{j-1}(s) - s^3 C_{f,(j-1)/2}^2 C_{j-1} D_{j-3}(s) & j = 3, 5, \dots, 2k - 1 \end{cases}
 \end{aligned} \quad (3)$$

where

$$\hat{C}_{2l+1} = \begin{cases} C_{2l+1} + C_{f,l+1} & l = 0 \\ C_{2l+1} + C_{fl} + C_{f,l+1} & l = 1, \dots, k - 1 \\ C_{2l+1} + C_{fl} & l = k \end{cases} \quad (4)$$

and

$$N_i(s) = D_{i-1}(s) s^{2k+1-i} C_{fi} C_{f,i+2} \dots C_{f,2k-1} C_{i+1} C_{i+3} \dots C_{2k}, \quad i = 1, 3, \dots, 2k + 1 \quad (5)$$

Then the transfer function of the filter takes the form

$$H(s) = \frac{N(s)}{D(s)} \quad (6)$$

where

$$N(s) = \sum_{i=0}^k g_{b,2i+1} N_{2i+1}(s) \quad (7)$$

and

$$D(s) = (s\hat{C}_{2k+1} + g_{f,2k+1})D_{2k} - s^3 C_{f,2k-1}^2 C_{2k} D_{2k-2} \quad (8)$$

If we specialize to the case $n = 3$, then, using (3)–(8) we get the following formulas:

$$\begin{aligned} D_0 &= 1 \\ D_1 &= s(C_1 + C_{f1}) \\ D_2 &= s^2 C_2 (C_1 + C_{f1}) + g_2 g_{f1} \end{aligned} \quad (9)$$

and

$$N(s) = s^2 [g_{b1} C_2 C_{f1} + g_{b3} C_2 (C_1 + C_{f1})] + g_{b3} g_2 g_{f1} \quad (10)$$

$$\begin{aligned} D(s) &= s^3 C_2 (C_1 C_3 + C_1 C_{f1} + C_3 C_{f3}) + s^2 C_2 (C_1 + C_{f1}) g_{f3} + \\ &+ s (C_3 + C_{f1}) g_2 g_{f1} + g_2 g_{f1} g_{f3} \end{aligned} \quad (11)$$

Then, we obtain the following transfer function formula:

$$\begin{aligned} H(s) &= \frac{s^2 [g_{b1} C_2 C_{f1} + g_{b3} C_2 (C_1 + C_{f1})] +}{s^3 C_2 (C_1 C_3 + C_1 C_{f1} + C_3 C_{f3}) + s^2 C_2 (C_1 + C_{f1}) g_{f3} +} \\ &\quad \frac{+ g_{b3} g_2 g_{f1}}{+ s (C_3 + C_{f1}) g_2 g_{f1} + g_2 g_{f1} g_{f3}} \end{aligned} \quad (12)$$

Fig. 9 shows a diagram of the corresponding filter structure. If we assume that all

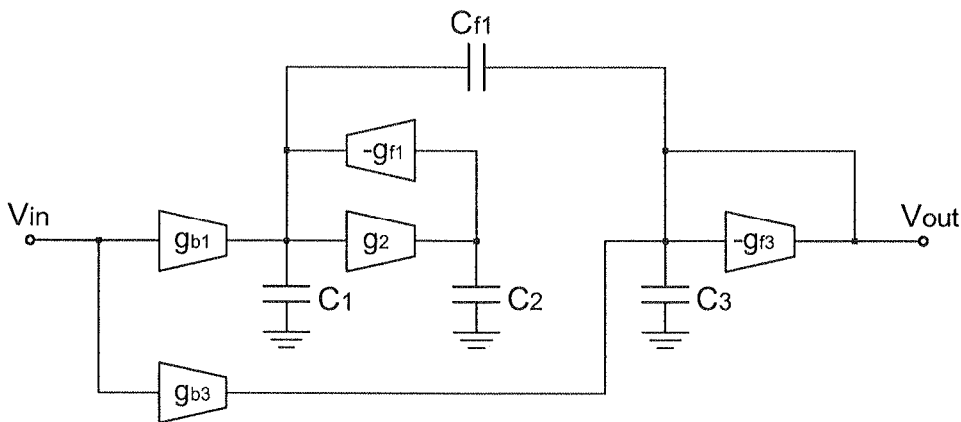


Fig. 9. Canonic elliptic filter with distributed input and floating capacitor — third order structure

Rys. 9. Filtr eliptyczny z dystrybucją sygnału wejściowego i kondensatorami nieziemionymi
— struktura trzeciego rzędu

transconductances in the filter in Fig. 9 are the same (up to the sign) i.e. $g_2 = g_{f1} = g_{f3} = -g_{b1} = g_{b3} = g$, then (12) simplifies to the form:

$$H(s) = \frac{\frac{C_1 C_2}{g^2} s^2 + 1}{\frac{C_2 (C_1 C_3 + C_1 C_{f1} + C_3 C_{f3})}{g^3} s^3 + \frac{C_2 (C_1 + C_{f1})}{g^2} s^2 + \frac{C_3 + C_{f1}}{g} s + 1} \quad (13)$$

Suppose that one wants to implement third order elliptic transfer function of the form

$$H(s) = \frac{b_2 s^2 + 1}{a_3 s^3 + a_2 s^2 + a_1 s + 1} \quad (14)$$

Coefficients a_{1-3} and b_2 depend on the filter parameters A_1 and ω_s defined in Fig. 1. By comparing (13) and (14) we can derive explicit design formulas, which are:

$$\begin{aligned} C_1 &= \frac{b_2 (a_1 a_2 - a_3)}{(a_2 - b_2)^2} g, & C_2 &= \frac{(a_2 - b_2)^2}{a_1 a_2 - a_3} g \\ C_3 &= \frac{a_3 - a_1 b_2}{a_2 - b_2} g, & C_{f1} &= \frac{a_1 a_2 - a_3}{a_2 - b_2} g \end{aligned} \quad (15)$$

It follows from (15) that the values of capacitors are positive (which is a necessary condition for the filter to be realizable) if and only if the following inequalities hold:

$$a_1 a_2 > a_3 > a_1 b_2 \quad (16)$$

According to (16), the filter in Fig. 9 can not be realizable for some particular transfer function coefficients a_{1-3} and b_2 that violate (16). However, arbitrary third order elliptic transfer function can be still implemented with the filter structure in Fig. 9 if we give up the assumption of equal transconductances.

4. PERFORMANCE COMPARISON OF THIRD ORDER STRUCTURES

In this section we present the results of the performance comparison of our novel canonic elliptic filter structure (denoted in the sequel as F_c) shown in Fig. 10 and two representatives of filter classes described in Section 2. In particular we shall consider LF filter with distributed input shown in Fig. 11 (denoted just as LF) LF structure with a floating capacitor shown in Fig. 12 (we shall denote it as LF_f). Table 2 shows explicit design formulas for the considered filters, that is element values as functions of the parameters a_{1-3} and b_2 of the transfer function (14).

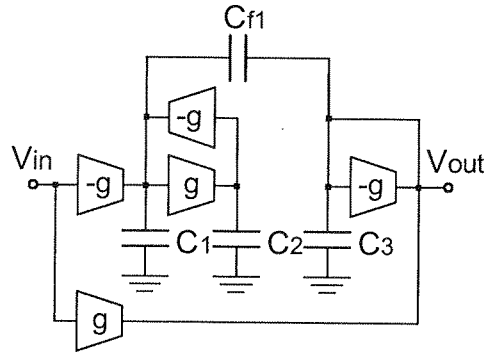
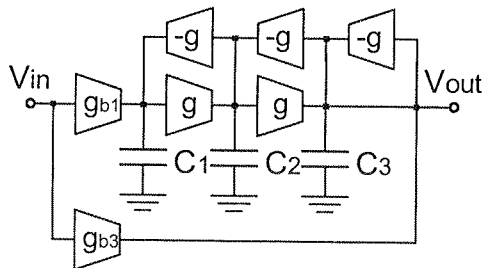
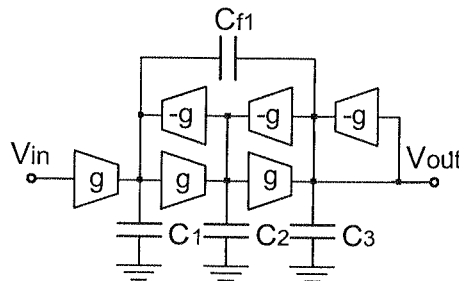
Fig. 10. Third order LF filter with distributed input and floating capacitor (F_c)Rys. 10. Filtr LF 3-go rzędu z dystrybucją sygnału wejściowego i kondensatorem nieziemionym (F_c)

Fig. 11. Third order LF filter with distributed input (LF)

Rys. 11. Filtr LF 3-go rzędu z dystrybucją sygnału wejściowego (LF)

Fig. 12. Third order LF filter with floating capacitor (LF_f)Rys. 12. Filtr LF 3-go rzędu z kondensatorem nieziemionym (LF_f)

For practical reasons we would like all the filter transconductances to have the same value. This is important from the point of view of automatic tuning on chip and the range of tunability of the filter. However, we cannot assume in general that $g_{b1}=g_{b3}=g$ for structure LF (cf. Table 2). We shall compare the above fil-

Table 2

Element values for the filter structures in Figs. 9, 10 and 11 as functions of the parameters a_1 , a_2 , a_3 and b_2 of the transfer function (14)

Wartości elementów dla struktur z Rys. 9, 10 i 11 w funkcji parametrów a_1 , a_2 , a_3 oraz b_2 funkcji przenoszenia (14)

Structure	Element values and realizability conditions
LF	$C_1 = \frac{a_1 a_2 - a_3}{a_2} g$, $C_2 = \frac{a_2^2}{a_1 a_2 - a_3} g$, $C_3 = \frac{a_3}{a_2} g$, $g_{b1} = \frac{a_2 - b_2}{a_2} g$, $g_{b3} = \frac{b_2}{a_2} g$ $a_2 > b_2$ and $a_1 a_2 > a_3$
LF _f	$C_1 = \frac{a_1 a_2 - a_3}{a_2 - b_2} g$, $C_2 = \frac{(a_2 - b_2)^2}{a_1 a_2 - a_3} g$, $C_3 = \frac{a_3 - a_1 b_2}{a_2 - b_2} g$, $C_4 = \frac{b_2(a_1 a_2 - a_3)}{(a_2 - b_2)^2} g$ $a_1 a_2 > a_3 > a_1 b_2$
F _c	$C_1 = \frac{b_2(a_1 a_2 - a_3)}{(a_2 - b_2)^2} g$, $C_2 = \frac{(a_2 - b_2)^2}{a_1 a_2 - a_3} g$, $C_3 = \frac{a_3 - a_1 b_2}{a_2 - b_2} g$, $C_4 = \frac{a_1 a_2 - a_3}{a_2 - b_2} g$, $a_1 a_2 > a_3 > a_1 b_2$

ter structures with respect to their sensitivity performance and other characteristics that are important for practical applicability, especially in monolithic implementation. We shall consider sensitivity integral $I_{[0, \omega_s]}^S$, that is the total sensitivity measure S (we use Schoeffler's measure [23] defined as $S(j\omega) = \sum_i |S_{z_i}^{H(j\omega)}|^2$, where i indexes filter elements — capacitors and transconductors, while $S_{z_i}^{H(j\omega)} = \partial \ln H(j\omega) / \partial \ln z_i$) integrated over the frequency interval $[0, \omega_s]$, the c_R coefficient (i.e. the maximum to minimum capacitance ratio) and g_R coefficient (i.e. the maximum to minimum transconductance ratio). Obviously, the optimal value of c_R and g_R is 1. Another important feature of the filter structure is its complexity, i.e. the number of capacitors and transconductors. Moreover, it turns out that on the assumption of equal transconductances, it is not possible to implement arbitrary third-order elliptic transfer functions with F_c (see Section 3) and LF_f (see [8]) structures, which is also some guideline for choosing the filter structure to particular application.

As stated in Section 3, the values of a_1 , a_2 , a_3 , and b_2 of (14) depend on the filter parameters A_1 and ω_s defined in Fig. 1 (A_2 is implied; it is a function of $n = 3$, A_1 , and ω_s) [1]. Table 3 indicates whether the realizability conditions are satisfied for different values of A_1 and ω_s in the considered third-order filter structures and also shows the values of c_R and g_R . The values of the total capacitance of the filter as well as sensitivity integrals over the passband, $I_{[0,1]}^S$, and the transition band, $I_{[1, \omega_s]}^S$, are presented in Table 3. Of course, $I_{[0, \omega_s]}^S = I_{[0,1]}^S + I_{[1, \omega_s]}^S$ will give the sensitivity integral over both bands. All results have been obtained for the normalized transconductance value $g = 1$. Fig. 13 shows total sensitivity S (Schoeffler's measure) versus normalized

Table 3

Realizability and the values of c_R and g_R for realizations of third-order elliptic transfer functions; the best values in each row are shaded

Realizowalność oraz wartości parametrów c_R i g_R dla filtrów eliptycznych 3-go rzędu; wartości najlepsze zakreskowano

Transfer function parameters		Realizability (Y — possible, N — not possible)			g_R			c_R		
A_1	ω_s	LF	LF _f	F _c	LF	LF _f	F _c	LF	LF _f	F _c
0.99	1.10	Y	N	N	7.73	-	-	34.28	-	-
0.99	1.20	Y	N	N	3.48	-	-	9.42	-	-
0.99	1.40	Y	Y	Y	2.10	1.00	1.00	3.34	21.32	21.32
0.95	1.10	Y	Y	Y	3.31	1.00	1.00	10.60	2419.0	2419.0
0.95	1.20	Y	Y	Y	2.06	1.00	1.00	4.23	5.07	5.07
0.95	1.40	Y	Y	Y	3.12	1.00	1.00	2.16	2.67	2.67
0.90	1.10	Y	Y	Y	2.65	1.00	1.00	6.96	6.44	6.44
0.90	1.20	Y	Y	Y	2.24	1.00	1.00	3.17	2.47	2.47
0.90	1.40	Y	Y	Y	3.63	1.00	1.00	1.79	2.63	2.63
0.80	1.10	Y	Y	Y	2.26	1.00	1.00	4.39	3.32	3.32
0.80	1.20	Y	Y	Y	2.53	1.00	1.00	2.21	2.04	2.04
0.80	1.40	Y	Y	Y	4.14	1.00	1.00	1.35	3.14	3.14

frequency for the considered third-order filter structures (here for the transfer function parameters: $A_1 = 0.90$, $A_2 = 0.16$, $\omega_s = 1.20$, $\delta = 0.92$ dB).

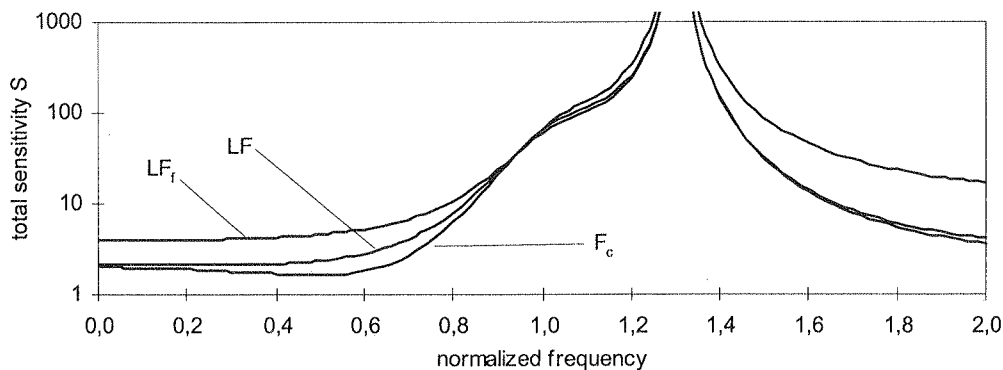


Fig. 13. Total sensitivity S (Schoeffler's measure) vs. normalized frequency for the considered third-order filter structures. Transfer function parameters: $A_1 = 0.90$, $A_2 = 0.16$, $\omega_s = 1.20$, $\delta = 0.92$ dB

Rys. 13. Miara wrażliwości Schoeffler'a S w funkcji znormalizowanej częstotliwości dla rozważanych struktur 3-go rzędu. Parametry funkcji przenoszenia: $A_1 = 0.90$, $A_2 = 0.16$, $\omega_s = 1.20$, $\delta = 0.92$ dB

Having in mind the results presented in Tables 1, 3 and 4 we can draw some conclusions concerning the performance of the novel filter structure F_c , especially in comparison to the structures LF and LF_f . First of all the structure F_c contains the smallest number of transconductors (i.e. only 5, while the structures LF_f and LF

Table 4

Total capacitance and sensitivity integrals for the considered filter structures;
the best values in each row are shaded

Całkowita pojemność oraz całki wrażliwości dla rozważanych struktur filtrów;
wartości najlepsze zakreskowano

Transfer function parameters		C_{total}			$I_{[0,1]}^S$			$I_{[1,\omega_s]}^S$		
A_1	ω_s	LF	LF _f	F _c	LF	LF _f	F _c	LF	LF _f	F _c
0.99	1.10	Y	N	N	7.73	-	-	34.28	-	-
0.99	1.10	5.91	-	-	2.60	-	-	8.90	-	-
0.99	1.20	3.55	-	-	2.68	-	-	13.84	-	-
0.99	1.40	2.72	2.38	2.38	3.16	5.88	2.42	21.01	20.33	21.61
0.95	1.10	4.37	3.74	3.74	5.51	10.80	4.85	20.49	20.53	20.14
0.95	1.20	3.49	2.95	2.95	5.34	8.18	4.49	22.86	21.31	24.85
0.95	1.40	3.24	2.88	2.88	5.45	6.86	4.38	26.51	23.63	37.92
0.90	1.10	4.22	3.56	3.56	8.50	12.08	7.83	25.10	23.48	26.72
0.90	1.20	3.68	3.20	3.20	7.57	9.45	6.90	25.44	23.20	30.44
0.90	1.40	3.56	3.26	3.26	7.12	8.00	6.43	27.66	24.33	46.52
0.80	1.10	4.27	3.90	3.90	13.72	15.56	13.86	28.51	26.24	32.61
0.80	1.20	3.99	3.73	3.73	11.24	12.12	11.45	27.03	24.48	35.48
0.80	1.40	4.01	3.84	3.84	9.81	10.14	10.08	28.17	24.54	55.87

contain 6 and 7 transconductors, respectively; according to Table 1 the reduction of transconductors number for high order filters is as much as 40% over LF structure and 25% over LF_f one; e.g. for 9th order filter the figure is 13, 18 and 22 transconductors, for F_c, LF_f and LF structures, respectively). This means the F_c structure is particularly suited for low-power filter implementations. Furthermore, g_R factor for F_c filter is always 1 which is not the case for LF filter. Thus F_c structure has advantage over LF one while implementing filters with automatic on-chip tuning. The value of c_R and C_{total} for F_c is comparable or better than for LF filter (except two high- A_1 and low- ω_s cases). Note also that g_R , c_R and C_{total} are the same for F_c and LF_f. Eventually, F_c structure is characterized by very low pass-band sensitivity. Its stop-band sensitivity is comparable to other structures (except the case $\omega_s = 1.40$).

5. CONCLUSIONS

In the paper a novel class of odd-order elliptic G_m-C filter structures has been introduced. The structure contains both grounded and floating capacitors and is characterized by a very small number of transconductors which makes it suitable for low-power filter implementations. We have presented recursive transfer function formulas for the considered general filter structure and derived explicit design formulas for the

third order case. A detailed performance comparison of the novel structure and two representatives of other classes of elliptic filter structures (leap-frog with distributed input and leap-frog with floating capacitor) has been carried out for the third order case. It follows that the novel elliptic filter structure can be implemented with all transconductors equal, has very good maximum to minimum capacitance ratio and low total capacitance as well as low sensitivity, especially in the pass-band. All these features indicate that the novel elliptic filter structure may be very interesting and competitive in many practical applications.

6. REFERENCES

1. R. Schaumann, M.S. Ghausi, K.R. Laker: *Design of Analog Filters, Passive, Active RC, and Switched Capacitor*. Englewood Cliff, NJ: Prentice-Hall, 1990.
2. M.S. Ghausi, K.R. Laker: *Modern Filter Design: Active RC and Switched Capacitor*. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1981.
3. T. Deliyannis, Y. Sun, J.K. Fidler: *Continuous-time active filter design*. USA, CRC Press, 1999.
4. C. Toumazou, J. Lidgey, D. Haigh, Eds.: *Analogue IC Design - The Current-Mode Approach*. London, U.K.: IEE, Apr. 1990.
5. R. Schaumann, M.A. Van Valkenburg: *Design of Active Filters*. Oxford University Press, New York, Jan. 2001.
6. S. Koziel, S. Szczepanski, R. Schaumann: *General Approach to Continuous-Time G_m -C Filters Based on Matrix Description*. Proc. Int. Symp. Circuits, Syst., ISCAS, vol.IV, Phoenix, AZ, USA, 2002, pp. 647-650.
7. S. Koziel, S. Szczepanski, R. Schaumann: *General Approach to Continuous-Time G_m -C Filters*. Int. J. Circuit Theory and Applications, 2003 (to be published).
8. S. Koziel, S. Szczepanski: *Structure Generation and Performance Comparison of Canonical Elliptic G_m -C Filters*. Proc. Int. Conf. Electron. Circuits, Syst., ICECS, vol. I, Dubrovnik, Croatia, 2002, pp. 157-160.
9. R. L. Geiger, E. Sánchez-Sinencio: *Active filter design using operational transconductance amplifiers: A tutorial*. IEEE Circuit and Devices Mag., vol. 1, Mar. 1985, pp. 20-32.
10. E. Sánchez-Sinencio, R.L. Geiger, H. Nevarez-Lozano: *Generation of Continuous-Time Two Integrator Loop OTA Filter Structures*. IEEE Trans. Circuits Syst., vol. 35, Aug. 1988, pp. 936-946.
11. M.A. Tan, R. Schaumann: *Design of a General Biquadratic Filter Section with Only Transconductances and Grounded Capacitors*. IEEE Trans. Circuits Syst.-II, vol. 35, Apr. 1988, pp. 478-480.
12. B. Nauta: *Analog CMOS filters for very high frequencies*, Kluwer Academic Publishers, 1993.
13. J. Silva-Martinez, M.S.J. Steyaert, W. Sansen: *A 10.7MHz 68-dB CMOS continuous-time filter with on-chip automatic tuning*. IEEE J. Solid-State Circuits, vol. 27, No.12, 1992, pp. 1843-1853.
14. R. Allini, A. Baschiroto, R. Castello: *Tuneable BiCMOS continuous-time filter for high-frequency applications*. IEEE J. Solid-State Circuits, 27, (12), 1992, pp. 1905-1915.
15. G.A. De Veirman, R.G. Yamasaki: *Design of a bilopar 10 MHz programmable continuous-time 0.05 ° equiripple linear phase filter*. IEEE J. Solid-State Circuits, vol. 27, Mar. 1992, pp. 324-331.
16. H. Khorramabadi, M.J. Tarsie, N.S. Woo: *Baseband filters for IS-95 CDMA receiver applications featuring digital automatic frequency tuning*. in int. Solid State Circuits Conf., San Francisco, 1996, pp. 172-173.

17. W. Dehaene, M.S.J. Steyaert, W. Sansen: *A 50-MHz Standard CMOS Pulse Equalizer for Hard Disk Read Channels*. IEEE J. Solid-State Circuits, vol. 32, July 1997, pp. 977-988.
18. C.A. Laber, P.R. Gray: *A 20-MHz Sixth-Order BiCMOS Parasitic-Insensitive Continuous-Time Filter and Second-Order Equalizer Optimized for Disk-Drive Read Channels*. IEEE J. Solid-State Circuits, vol. 28, Mar. 1993, pp. 462-469.
19. P.K.D. Pai, A.A. Abidi: *A 40-mW 55 Mb/s CMOS Equalizer for Use in Magnetic Storage Read Channels*, IEEE J. Solid-State Circuits, vol. 29, Mar. 1994, pp. 489-499.
20. Y.P. Tsividis: *Integrated continuous-time filter design - An overview*. IEEE J. Solid-State Circuits, vol. 29, Mar. 1994, pp. 166-176.
21. J. Mahattanakul, C. Toumazou: *Current-Mode Versus Voltage-Mode G_m -C Biquad Filters: What the Theory Says*. IEEE Trans. Circuits Syst.-II, vol. 45, Feb. 1998, pp. 173-186.
22. R. Mackay, A.S. Sedra: *Generation of low-sensitivity state-space active filters*. IEEE Trans. Circuit Syst., vol. CAS-27, Oct. 1980, pp. 863-870.
23. J.D. Schoeffler: *The synthesis of minimum sensitivity networks*. IEEE Trans. Circuit Theory, vol. 11, 1964, pp. 271-276.

S. KOZIEL

KANONICZNE STRUKTURY FILTRÓW G_m -C NIEPARZYSTEGO RZĘDU

Streszczenie

W pracy przedstawiono nowe struktury kanonicznych filtrów eliptycznych G_m -C czasu ciągłego. Zaproponowane struktury charakteryzują się bardzo małą liczbą elementów, w szczególności wzmacniaczy transkonduktancyjnych. Powoduje to, że nadają się one do realizacji filtrów o małym poborze mocy i niewielkiej wrażliwości. Przedstawiono ogólne formuły rekurencyjne umożliwiające wyznaczenie funkcji przenoszenia filtru dowolnego rzędu i pozwalające na otrzymywanie wzorów projektowych dla rozważanych struktur. Praca zawiera również szczegółowe wyniki porównania proponowanych struktur z przedstawicielami innych klas filtrów eliptycznych opisanych w literaturze pod względem parametrów istotnych dla realizacji monolitycznych. Porównania dokonano dla przypadku filtrów trzeciego rzędu. Okazuje się, że nowa struktura jest konkurencyjna w porównaniu do istniejących zarówno pod względem wrażliwości (szczególnie w paśmie przepustowym) jak też stosunku maksymalnej do minimalnej pojemności oraz całkowitej pojemności w układzie. Ponadto, może ona być realizowana przy zachowaniu równości wszystkich transkonduktancji w układzie, co jest istotne z punktu widzenia układów automatycznego dostrajania.

Słowa kluczowe: filtry G_m -C, filtry eliptyczne, struktury kanoniczne

w
n
d
z
P
l
v
c
i
n
f
c

Pakiety falkowe w kompresji sygnałów i obrazów

WALDEMAR RAKOWSKI

*Politechnika Białostocka, Wydział Informatyki
ul. Wiejska 45A, 15-351 Białystok
e-mail: waldi@ii.pb.bialystok.pl*

*Otrzymano 2003.06.13
Autoryzowano 2003.08.30*

W pracy przedstawiono podstawy matematyczne reprezentacji sygnałów za pomocą pakietów falkowych. Pokazano bazy do aproksymacji sygnałów oparte na pakietach falkowych Walsha. Podano algorytm szukania najlepszej bazy i jego implementację w języku ANSI C. Zamieszczono informację o złożoności obliczeniowej algorytmów reprezentacji sygnałów za pomocą pakietów falkowych. Omówiono wykorzystanie pakietów falkowych w kompresji dźwięków i obrazów.

Słowa kluczowe: falka, filtr falkowy, transformata falkowa, pakiet falkowy, najlepsza baza, kompresja dźwięku, kompresja obrazu

1. WPROWADZENIE

W połowie lat siedemdziesiątych odkryto, że kompresja dźwięku w postaci cyfrowej polegająca na kwantyzacji próbek i kodowaniu entropijnym skwantowanych próbek, może być polepszona poprzez wprowadzenie dodatkowej fazy kompresji, a mianowicie dekompozycję pasmową sygnału, poprzedzającą fazę kwantyzacji. Ideę takiej dekompozycji w układzie dwóch filtrów pasmowych: dolnoprzepustowego i górnoprzepustowego podali jako pierwsi A. Croisier, D. Esteban i C. Galand [5] (rys. 1). W drugiej połowie lat osiemdziesiątych S. Mallat odkrył związek filtracji pasmowej sygnału z *analizą wielorozdzielczą* przestrzeni funkcji o skończonej energii [6, 7]. Pod koniec lat osiemdziesiątych I. Daubechies wynalazła *ortogonalne filtry cyfrowe o skończonej odpowiedzi impulsowej* [18, 19], spełniające *warunki doskonałej rekonstrukcji sygnału* sformułowane dla układu filtrów z rysunku 1 przez M. Vetterliego [32, 24]. Wymienionym wyżej filtrom Daubechies towarzyszą funkcje o zwartym nośniku (falki) generujące bazy ortonormalne przestrzeni $L^2(R)$. Falki Daubechies oprócz zwartego nośnika są również

„zlokalizowane” w dziedzinie częstotliwości i stanowią „atomy”, których energia skoncentrowana jest w kostkach czas-częstotliwość pokazanych na rysunku 2. Na podstawie falki matki $\psi(t)$ generowana jest rodzina falek

$$\psi^{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k), \quad j \in \mathbb{Z}, k \in \mathbb{Z}, \quad (1)$$

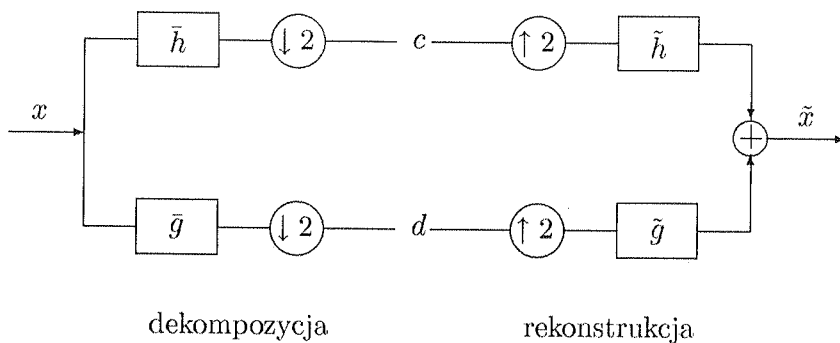
umożliwiająca przedstawienie funkcji o skończonej energii w postaci szeregu [17, 23]

$$f(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} d_j[k] \psi_{j,k}(t). \quad (2)$$

Falka $\psi_{j,k}(t)$ ma najwięcej energii skoncentrowanej w kostce Heisenberga

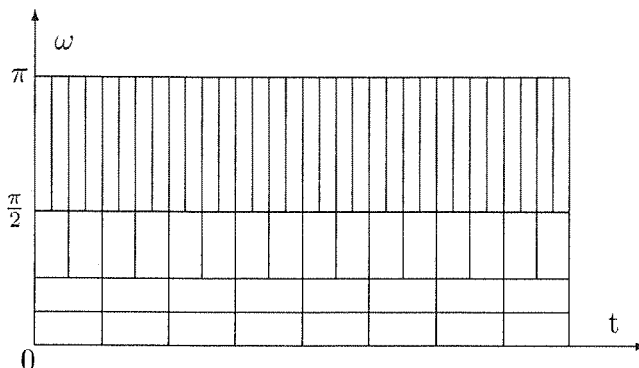
$$\left[2^j k, 2^j (k+1)\right] \times \left[\pi/2^j, 2\pi/2^j\right]. \quad (3)$$

Podział diadyczny osi częstotliwości, charakterystyczny dla reprezentacji falkowej sygnału, może nie odpowiadać własnościom czasowo-częstotliwościowym analizowanego sygnału. W górnej połowie pasma częstotliwości sygnał jest reprezentowany przez jedną falkę i nie ma możliwości podziału tego pasma na podpasma. Znacznie więcej możliwości wybrania do aproksymacji sygnału atomów o energii skoncentrowanej w węższych pasmach częstotliwości, pokrywających się lepiej z obszarami skupień energii sygnału na płaszczyźnie czas-częstotliwość, daje reprezentacja sygnału za pomocą *pakietów falkowych* (ang. *wavelet packets*) wprowadzonych przez R. Coifmana, Y. Meyera i V. Wickerhausera [3, 1, 2] na początku lat dziewięćdziesiątych.



Rys. 1. Bank filtrów dekompozycji sygnału (z lewej) i bank filtrów rekonstrukcji sygnału (z prawej). W przypadku filtrów ortogonalnych $\bar{h} = h$, $\bar{g} = g$, a $\bar{h}$ oznacza ciąg rewersyjny względem ciągu h , tj. $\bar{h}[n] = h[-n]$; analogiczna uwaga dotyczy $\bar{g}$ i g . Symbol $(\downarrow 2)$ oznacza decymator (usuwa co drugi element ciągu), a $(\uparrow 2)$ oznacza ekspander (wstawia pomiędzy dwa elementy ciągu dodatkowy element równy zero)

Fig. 1. Analysis/synthesis filter banks. For orthonormal filters $\bar{h} = h$, $\bar{g} = g$, where $\bar{h}$ is in a reverse order to h , i.e. $\bar{h}[n] = h[-n]$; $(\downarrow 2)$ — decimator, $(\uparrow 2)$ — expander



Rys. 2. Podział płaszczyzny czas-częstotliwość odpowiadający trzypoziomowej dekompozycji falkowej sygnału. Prostokąty tworzące rozłączne pokrycie dziedziny czasu i częstotliwości sygnału ilustrują obszary najsilniejszego skupienia energii odpowiadających im funkcji bazowych rozkładu (falek)

Fig. 2. Time-frequency plane decomposition for 3-level wavelet transform. Rectangles, covering time-frequency plane correspond to regions in which the energy of base functions is the most concentrated

2. PAKIETY FALKOWE

Punktem wyjścia w analizie sygnału za pomocą pakietów falkowych jest, podobnie jak w analizie falkowej, filtr cyfrowy h , o którym zakłada się że ma skończoną odpowiedź impulsową spełniającą warunki:

$$\sum_k h[n] = \sqrt{2}, \quad (4)$$

$$\sum_k h[n]h[n-2k] = \delta[k]. \quad (5)$$

Filtrowi takiemu odpowiada *funkcja skalująca* $\phi(x)$ będąca rozwiązaniem *równania dylatacyjnego*

$$\phi(t) = \sqrt{2} \sum_k h[k]\phi(2t-k). \quad (6)$$

Rozwiązanie powyższego równania może być znalezione za pomocą *algorytmu kaskadowego* podanego przez Ingrid Daubechies [7]. Skończona odpowiedź impulsowa filtru h zapewnia zwarty nośnik funkcji skalującej $\phi(t)$.

Najprostszym przykładem omawianego wyżej filtru jest filtr Haara o współczynnikach

$$h[0] = h[1] = \frac{1}{\sqrt{2}},$$

któremu odpowiada funkcja skalująca

$$\phi(t) = \mathbf{1}(t),$$

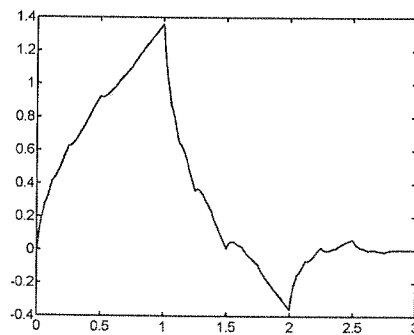
gdzie

$$\mathbf{1}(t) = \begin{cases} 1 & \text{jeżeli } 0 \leq t < 1 \\ 0 & \text{dla pozostałych } t. \end{cases}$$

Innym przykładem jest filtr Daubechies o 4 współczynnikach:

$$h[0] = \frac{1 + \sqrt{3}}{4\sqrt{2}}, \quad h[1] = \frac{3 + \sqrt{3}}{4\sqrt{2}}, \quad h[2] = \frac{3 - \sqrt{3}}{4\sqrt{2}}, \quad h[3] = \frac{1 - \sqrt{3}}{4\sqrt{2}} \quad (7)$$

któremu odpowiada funkcja skalująca pokazana na rysunku 3.



Rys. 3. Funkcja skalująca odpowiadająca filtrowi Daubechies o 4 współczynnikach

Fig. 3. Scaling function for Daubechies' 4-tap filter

Niech analizowany sygnał $f(t)$ będzie funkcją o skończonej energii, tj. $f(t) \in L^2(R)$ i niech dane będą współczynniki rozwinięcia sygnału $f(t)$ w przestrzeni V_0 , której bazą jest zbiór funkcji $\{\phi(t-k)\}_{k \in \mathbb{Z}}$. Filtr h spełniający warunek (5) umożliwia rozbić przestrzeni V_0 na dwie ortogonalne przestrzenie V_1 i W_1 , takie że

$$V_0 = V_1 \oplus W_1, \quad (8)$$

gdzie $\oplus$ oznacza sumę ortogonalną. Bazę przestrzeni V_1 tworzą funkcje

$$\{\phi_1(t-2k)\}_{k \in \mathbb{Z}}, \quad (9)$$

gdzie

$$\phi_1(t) = \frac{1}{\sqrt{2}} \sum_k h[k] \phi(t/2 - k); \quad (10)$$

bazę przestrzeni W_1 , tworzą funkcje

$$\{\psi_1(t-2k)\}_{k \in \mathbb{Z}}, \quad (11)$$

gdzie

$$\psi_1(t) = \frac{1}{\sqrt{2}} \sum_k g[k] \phi(t/2 - k), \quad (12)$$

a g oznacza filtr *lustrzany* względem filtru h , zdefiniowany następująco:

$$g[n] = (-1)^{1-n} h[1-n]. \quad (13)$$

Współczynniki rozwinięcia sygnału $f(t)$ względem bazy przestrzeni V_1 i względem bazy przestrzeni W_1 znajduje się obliczając *dyskretną transformatę falkową* [17, 23] ciągu współczynników rozwinięcia sygnału względem bazy V_0 . Do tego miejsca analiza za pomocą pakietów falkowych jest identyczna z analizą falkową. Różnica pojawia się przy dalszym „rozszczipianiu” przestrzeni. W analizie falkowej rozszczipieniu podlega tylko przestrzeń V_1 , a w analizie za pomocą pakietów falkowych zarówno V_1 jak i W_1 . W celu ujednoliconego przedstawienia całego procesu analizy za pomocą pakietów falkowych wprowadza się następujące oznaczenia. Przestrzenie pakietów falkowych oznaczane będą przez w_j^p , gdzie indeks j odpowiada *skali rozwinięcia* równej 2^j , a indeks p oznacza numer pakietu falkowego będącego funkcją bazową p -tej przestrzeni w skali 2^j ; pakiety falkowe będą odpowiednio oznaczane przez $w_j^p(t)$. Przy założeniu, że sygnał $f(t)$ określony jest w przedziale $[0, 2^l]$, $l > 0$, indeks j może przyjmować wartości od 0 do l , a indeks p od 0 do $2^j - 1$. Zgodnie z powyższym funkcja skalująca $\phi(t)$ jest pakietem falkowym $w_0^0(t)$, a przestrzeń V_0 przestrzenią W_0^0 .

Na rysunku 4 pokazano trzypoziomową dekompozycję przestrzeni wyjściowej V_0 (W_0^0) na przestrzenie pakietów falkowych. Z każdą przestrzenią W_j^p , $j = 0, 1, \dots, l$;

W_0^0							
W_1^0				W_1^1			
W_2^0		W_2^1		W_2^2		W_2^3	
W_3^0	W_3^1	W_3^2	W_3^3	W_3^4	W_3^5	W_3^6	W_3^7

Rys. 4. Podział wyjściowej przestrzeni aproksymacji sygnału na podprzestrzenie za pomocą pakietów falkowych

Fig. 4. Decompositions of approximation space by wavelet packets

$p = 0, 1, \dots, 2^j - 1$ związany jest pakiet falkowy o takich samych indeksach $w_j^p(t)$, którego przesunięte wersje

$$\{w_j^p(x - 2^j k)\}_{k \in \mathbb{Z}} \quad (14)$$

tworzą bazę przestrzeni.

Pakiet falkowy $w_0^0(t) = \phi(t)$, a pozostałe pakiety zdefiniowane są rekurencyjnie

$$w_{j+1}^{2p}(t) = \sum_k h[k] w_j^p(t - 2^j k), \quad (15)$$

$$w_{j+1}^{2p+1}(t) = \sum_k g[k] w_j^p(t - 2^j k), \quad (16)$$

2.1. PAKIETY FALKOWE WALSHA

Pakiety falkowe Walsha generowane są przez filtry:

$$h[0] = \frac{1}{\sqrt{2}}, h[1] = \frac{1}{\sqrt{2}} \quad (17)$$

$$g[0] = \frac{1}{\sqrt{2}}, g[1] = \frac{1}{\sqrt{2}} \quad (18)$$

zgodnie poniższymi wzorami

$$w_{j+1}^p(t) = \frac{1}{\sqrt{2}} (w_j^p(t) + w_j^p(t - 2^j)), \quad (19)$$

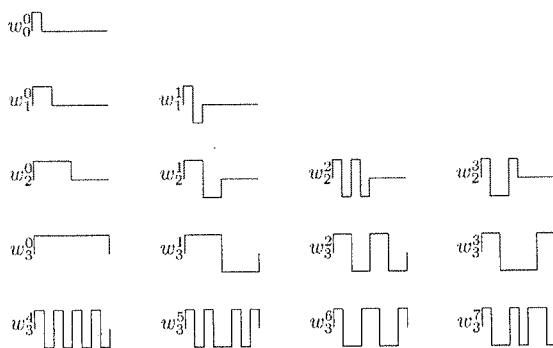
$$w_{j+1}^{p+1}(t) = \frac{1}{\sqrt{2}} (w_j^p(t) + w_j^p(t - 2^j)). \quad (20)$$

Nazwa pakietów pochodzi od funkcji Walsha¹, które mają taki sam kształt jak pakiety falkowe Walsha $w_j^p(t)$, dla sygnału określonego na przedziale $[0, 2^j]$. Na rysunku 5 pokazano zestaw pakietów falkowych Walsha do analizy sygnałów określonych na przedziale $[0, 8]$. W tabeli 1 wyszczególniono zbiory przesuniętych wersji pakietów falkowych tworzących bazy poszczególnych podprzestrzeni. Pakiety falkowe o wyraźnie zlokalizowanym nośniku w dziedzinie czasu dla $j = 0$, dla największej możliwej wartości j mają nośnik równy całemu przedziałowi, na którym określony jest sygnał. Ze wzrostem j , kosztem utraty lokalizacji w dziedzinie czasu, polepsza się lokalizacja w dziedzinie częstotliwości² pakietu.

¹ Funkcje Walsha są zdefiniowane w przestrzeni $L^2(0, 1)$ [30], ale po przeskalowaniu przedziału określoności $[0, 1]$ do przedziału $[0, 2^j]$ pokrywają się z pakietami $w_j^p(t)$

² Jest to ogólna własność sygnałów, patrz [30]

l^2 (Z
 w_j^p
 elen
 zbic
 wsp
 ana
 dys
 jako
 ozn
 o 2



Rys. 5. Pakiety falkowe Walsha na przedziale $[0, 8]$. Pakiety w_0^j przyjmują wartości od 0 do $(\frac{1}{\sqrt{2}})^j$; pakiety $w_j^p, p > 0$ przyjmują wartości od $-(\frac{1}{\sqrt{2}})^j$ do $(\frac{1}{\sqrt{2}})^j$

Fig. 5. Walsh wavelet packets on interval $[0, 8]$; w_0^j ranges from 0 to $(\frac{1}{\sqrt{2}})^j$; $w_j^p, p > 0$ ranges from $-(\frac{1}{\sqrt{2}})^j$ to $(\frac{1}{\sqrt{2}})^j$

Tabela 1

Bazy przestrzeni pakietów falkowych Walsha na odcinku $[0, 8]$

Walsh wavelet packet bases on interval $[0, 8]$

Przestrzeń	$\{p\}$	Baza
W_0^p	$\{0\}$	$\{w_0^p(t - k)\}_{k=0, \dots, 7}$
W_1^p	$\{0, 1\}$	$\{w_1^p(t - 2k)\}_{k=0, \dots, 3}$
W_2^p	$\{0, 1, 2, 3\}$	$\{w_2^p(t - 4k)\}_{k=0, 1}$
W_3^p	$\{0, 1, 2, 3, 4, 5, 6, 7\}$	$\{w_3^p(t)\}$

2.1.1. DYSKRETNE PAKIETY FALKOWE WALSHA

Dyskretne pakiety falkowe są elementami przestrzeni ciągów o skończonej energii $l^2(\mathbb{Z})$ i będą oznaczane przez $w_j^p[n], n \in \mathbb{Z}$, gdzie $w_j^p[n]$ oznacza n -ty element ciągu w_j^p lub cały ciąg, w zależności od kontekstu. Taki sposób oznaczania ciągu i jego elementu jest zgodny ze sposobem oznaczania funkcji f będącej odwzorowaniem np. zbioru liczb rzeczywistych R w zbiór liczb rzeczywistych R . $f(t)$ oznacza zarówno wspomniane odwzorowanie $f: R \rightarrow R$ jak i wartość funkcji f dla argumentu t . Jeżeli analizowany sygnał cyfrowy jest ciągiem skończonym o długości $N = 2^l, l > 0$, to dyskretne pakiety falkowe służące do aproksymacji tego sygnału można traktować jako elementy przestrzeni R^N . Zgodnie z powyższym $w_j^p[n], 0 \leq j < l, 0 \leq p < 2^j$ oznaczać będzie dyskretny pakiet falkowy, a $w_j^p[n - 2^j k]$ pakiet przesunięty w prawo o $2^j k$ elementów.

Dyskretny pakiet falkowy Walsha przestrzeni W_0^0 ma elementy równe:

$$w_0^0[n] = \delta[n], 0 \leq n < N. \quad (21)$$

Pozostałe pakiety spełniają następującą zależność rekurencyjną:

$$w_{j+1}^{2p}[n] = \frac{1}{\sqrt{2}} (w_j^p[n] + w_j^p[(n - 2^j)]), \quad (22)$$

$$w_{j+1}^{2p+1}[n] = \frac{1}{\sqrt{2}} (w_j^p[n] + w_j^p[(n - 2^j)]), \quad (23)$$

Obliczone według powyższych zależności dyskretny pakiety Walsha dla $N = 8$ są następujące:

$$\begin{aligned} w_0^0 &= [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \\ w_1^0 &= \frac{1}{\sqrt{2}} [1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \\ w_1^1 &= \frac{1}{\sqrt{2}} [1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \\ w_2^0 &= \frac{1}{2} [1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0] \\ w_2^1 &= \frac{1}{2} [1 \ 1 \ -1 \ -1 \ 0 \ 0 \ 0 \ 0] \\ w_2^2 &= \frac{1}{2} [1 \ -1 \ 1 \ -1 \ 0 \ 0 \ 0 \ 0] \\ w_2^3 &= \frac{1}{2} [1 \ -1 \ -1 \ 1 \ 0 \ 0 \ 0 \ 0] \\ w_3^0 &= \frac{1}{2\sqrt{2}} [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1] \\ w_3^1 &= \frac{1}{2\sqrt{2}} [1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0] \\ w_3^2 &= \frac{1}{2\sqrt{2}} [1 \ 1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1] \\ w_3^3 &= \frac{1}{2\sqrt{2}} [1 \ 1 \ -1 \ -1 \ -1 \ -1 \ 1 \ 1] \\ w_3^4 &= \frac{1}{2\sqrt{2}} [1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1] \\ w_3^5 &= \frac{1}{2\sqrt{2}} [1 \ -1 \ 1 \ -1 \ -1 \ 1 \ -1 \ 1] \\ w_3^6 &= \frac{1}{2\sqrt{2}} [1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 1] \end{aligned}$$

$$(21) \quad w_3^7 = \frac{1}{2\sqrt{2}} [1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \quad 1 \quad -1]$$

(22) W tabeli 2 wyszczególniono zbiory przesuniętych wersji dyskretnych pakietów falkowych tworzących bazy poszczególnych podprzestrzeni. I tak np. bazą przestrzeni W_2^1

Tabela 2

(23) Bazy dyskretnych przestrzeni Walsha dla $N = 8$

Bases of Walsh discrete spaces for $N = 8$

Przestrzeń	$\{p\}$	Baza
W_0^p	$\{0\}$	$\{w_0^p[n-k]\}_{k=0,\dots,7}$
W_1^p	$\{0, 1\}$	$\{w_1^p[n-2k]\}_{k=0,\dots,3}$
W_2^p	$\{0, 1, 2, 3\}$	$\{w_2^p[n-4k]\}_{k=0,1}$
W_3^p	$\{0, 1, 2, 3, 4, 5, 6, 7\}$	$\{w_3^p[n]\}$

jest zbiór pakietów $\{w_2^1[n], w_2^1[n-4]\}$, gdzie

$$w_2^1[n-4] = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix}$$

Pakiety falkowe można zestawić w macierz, która staje się operatorem dekompozycji w określonej bazie. Np. operatorem dekompozycji przestrzeni W_0^0 w przestrzenie W_1^0 i W_1^1 jest następująca macierz:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Współczynniki rozkładu sygnału w przestrzeniach W_1^0 i W_1^1 otrzymuje się mnożąc powyższą macierz przez wektor współczynników rozkładu sygnału względem bazy przestrzeni W_0^0 . W praktyce opisaną wyżej dekompozycję realizuje się za pomocą transformacji falkowej, której złożoność obliczeniowa jest $O(N)$ [17, 23]. Pakiety falkowe Walsha dobrze ilustrują własności, w szczególności w dziedzinie czasu, pakietów

falkowych wywodzących się z falek o zwartym nośniku. Jednakże ich zastosowanie jest ograniczone, ponieważ brakuje im "gładkości". Znacznie częściej wykorzystywane są pakiety falkowe pochodzące od filtrów Daubechies, które przy odpowiednio długich filtrach, są ciągłe i różniczkowalne. Są one jednak gorzej zlokalizowane w dziedzinie czasu, gdyż nośnik pakietu w_j^p może mieć szerokość równą $2^j + (L - 2)(2^j - 1)$, gdzie L —liczba współczynników filtru h [34].

2.2. WŁASNOŚCI CZASOWO-CZĘSTOTLIWOŚCIOWE PAKIETÓW FALKOWYCH

Analizując rysunek 5 można zauważyć, że dla ustalonej skali, wzrost numeru pakietu (indeks p) nie idzie w parze ze wzrostem "głównej" częstotliwości pakietu. Powyższe stwierdzenie można sformalizować określając ile razy pakiet zmienia znak. W tabeli 3 podano liczbę zmian znaku dla $j = 3$.

Tabela 3

Liczba zmian znaku pakietów falkowych Walsha $W_3^p, p = 0, \dots, 7$

Number of sign changes of Walsh wavelet packets $W_3^p, p = 0, \dots, 7$

Numer pakietu	0	1	2	3	4	5	6	7
Liczba zmian znaku	0	1	3	2	7	6	4	5

Z tabeli 3 widać, że pakiety $W_3^p(t)$ uporządkowane w kolejności zwiększającej się liczby zmian znaków tworzą następujący ciąg:

$$w_3^0(t), w_3^1(t), w_3^3(t), w_3^2(t), w_3^6(t), w_3^7(t), w_3^5(t), w_3^4(t) \quad (24)$$

Dowodzi się [33] że, powyższe uporządkowanie można otrzymać przedstawiając wartości górnego indeksu ciągu

$$w_3^0(t), w_3^1(t), w_3^2(t), w_3^3(t), w_3^4(t), w_3^5(t), w_3^6(t), w_3^7(t) \quad (25)$$

w kodzie binarnym Graya

$$w_3^{000}(t), w_3^{100}(t), w_3^{110}(t), w_3^{010}(t), w_3^{011}(t), w_3^{111}(t), w_3^{101}(t), w_3^{001}(t) \quad (26)$$

a następnie przyjmując, że rozwinięcia binarne reprezentują liczby w naturalnym kodzie binarnym. Powyższe stwierdzenie jest ogólne i dotyczy odwzorowania indeksu p pakietu falkowego $w_3^p(t)$ generowanego według wzorów (15) i (16) na podpasmo częstotliwości, w którym jest najsilniej skoncentrowana energia pakietu. Energia pakietu $w_j^p(t - 2^j k)$ jest najsilniej skoncentrowana w kostce Heisenberga

$$\left[2^j k, 2^j(k+1)\right] \times \left[\tilde{p}\pi/2^j, (\tilde{p}+1)\pi/2^j\right], 1 \leq j \leq l, 0 \leq \tilde{p} \leq 2^j, k \geq 0, \quad (27)$$

nie jest
wane są
długich
edzinie
, gdzie

gdzie $\tilde{p}$ i p związane są zależnością określoną następująco. Niech $p_1 p_2 \dots$ będą cyframi binarnymi rozwinięcia liczby p w naturalnym kodzie binarnym. Wówczas rozwinięcie binarne $\tilde{p}_1 \tilde{p}_2 \dots$ liczby $\tilde{p}$ w naturalnym kodzie binarnym ma cyfry równe

$$\tilde{p}_i = \left(\sum_{k \geq i} p_k \right) \bmod 2. \quad (28)$$

3. DYSKRETNA TRANSFORMACJA PAKIETOWA

numeru
pakietu.
a znak.

Zakłada się, że dany jest ciąg współczynników $d_0^0[n] \in l^2(Z)$, będących współczynnikami rozwinięcia analizowanego sygnału $f(t) \in L^2(R)$ względem bazy przestrzeni W_0^0 , tzn.

$$f(t) \approx \sum_k d_0^0[k] \phi(t-k) \quad (29)$$

Tabela 3

cej się

Dalej przyjmuje się, że $0 \leq k < N$, co oznacza, że $d_0^0[n] \in R^N$, przy czym zakłada się, że N jest potęgą dwójki. Transformacji pakietowej poddawany jest ciąg $d_0^0[n]$ w banku filtrów dekompozycji pokazanym na rysunku 1. Wynikiem transformacji są dwa ciągi: $d_1^0[k]$ i $d_1^1[k]$, $0 \leq k < N/2$, będące produktem splotu i decymacji ciągu $d_0^0[n]$ odpowiednio z filtrami dekompozycji $\tilde{h}$ i $\tilde{g}$. Każdy z ciągów $d_1^p[k]$, $p = 0, 1$ może być poddany analogicznej dekompozycji jak ciąg wyjściowy $d_0^0[n]$. Procedurę transformacji można powtarzać, otrzymując l -poziomową, $0 < l \leq \log_2 N$, reprezentację pakietową ciągu $d_0^0[n]$. Złożoność obliczeniowa dekompozycji ciągu w banku filtrów z rysunku 1 jest liniowa, zatem złożoność obliczeniowa pełnej dekompozycji pakietowej ciągu $d_0^0[n]$ jest $O(N \log_2 N)$. l -poziomową transformatę pakietową zapisuje się w tablicy o wymiarach $(l+1) \times N$. W wierszu o numerze 0 jest zapisany ciąg wyjściowy $d_0^0[n]$. W wierszu o numerze $0 < j \leq l$ zapisane są podciągi $d_j^p[k]$, $p = 0, 1, \dots, 2^j - 1$, $k = 0, 1, \dots, N/2^j$. Każdemu podciągowi $d_j^p[k]$ odpowiada wektor bazowy $w_j^p \in R^N$, którego przesunięte wersje w liczbie $N/2^j$ tworzą bazę podprzestrzeni W_j^p . Wykorzystując np. wiersz j -ty transformaty pakietowej i wektory bazowe w_j^p , n -ty współczynnik ($0 \leq n < N$) ciągu wyjściowego $d_0^0[n]$ można wyrazić następująco³:

(24)

ąc war-

(25)

(26)

ym ko-
eksu p
no czę-
pakietu

$$d_0^0 = \sum_{p=0}^{2^j-1} \sum_{k=0}^{N/2^j} d_j^p[k] w_j^p[n - 2^j k]. \quad (30)$$

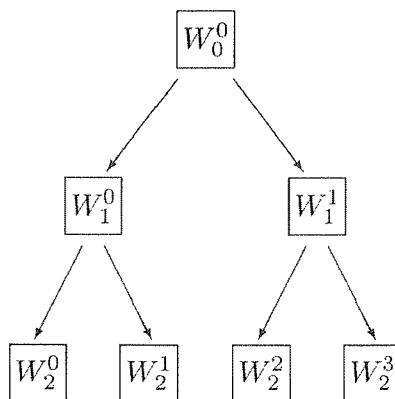
Powyższy wzór pokazuje $l(1 \leq j \leq l)$ możliwych reprezentacji ciągu $d_0^0[n]$ za pomocą współczynników transformaty pakietowej. W rzeczywistości możliwych reprezentacji jest znacznie więcej, co przedstawiono w następnym punkcie.

³ W praktyce, rekonstrukcję ciągu $d_0^0[n]$ ze współczynników transformaty pakietowej wykonuje się w banku filtrów rekonstrukcji pokazanym na rysunku 1

(27)

4. BAZY PAKIETÓW FALKOWYCH

Dowolne pokrycie szerokości prostokąta symbolizującego przestrzeń W_0^0 na rysunku 2 rozłącznymi prostokątami symbolizującymi podprzestrzenie W_j^p dla $j > 0$ pokazuje sposób utworzenia bazy przestrzeni W_0^0 . Zbiór wektorów bazowych podprzestrzeni stanowiących wspomniane pokrycie jest bazą przestrzeni W_0^0 . Np. zbiór wektorów bazowych podprzestrzeni $W_2^0, W_3^2, W_3^3, W_3^4, W_3^5, W_2^3$ jest bazą przestrzeni W_0^0 . Wygodnym sposobem ilustracji dekompozycji przestrzeni wyjściowej W_0^0 na podprzestrzenie pakietów falkowych jest drzewo binarne (rysunek 2), którego korzeń reprezentuje przestrzeń wyjściową W_0^0 , a pozostałe węzły podprzestrzenie W_j^p dla $j > 0$. Na rysunku 6 pokazano pełne trójpoziomowe drzewo binarne. Jeżeli sygnały przestrzeni W_0^0 są wektorami należącymi do R^N i $N = 2^l, l > 0$, to wysokość pełnego drzewa jest równa $l + 1$. "Przyciętym" drzewem jest drzewo, w którym każdy węzeł ma dwóch



Rys. 6. Trójpoziomowe drzewo binarne dekompozycji pakietowej przestrzeni W_0^0

Fig. 6. 3-level binary tree of wavelet packet spaces W_0^0

potomków, albo nie ma żadnego potomka. Zgodnie z tą definicją pełne drzewo jest szczególnym przypadkiem drzewa "przyciętego". Węzeł, który nie ma potomka nazywa się liściem. Zbiór wszystkich liści "przyciętego" drzewa odpowiada bazie przestrzeni W_0^0 . Liczba różnych baz przestrzeni W_0^0 jest co najmniej równa liczbie różnych drzew "przyciętych". Drzewo jednopoziomowe ($l = 0$) ma jedną bazę. Niech K^l oznacza liczbę wszystkich "przyciętych" drzew drzewa o $l + 1$ poziomach. Ze zbioru liczącego K^l różnych drzew można wybierać pary drzew od siebie niezależnych i tworzyć z nich lewe i prawe poddrzewo nowego drzewa (dokładając korzeń). Liczba różnych par ze zbioru K^l elementowego wynosi $K^l \cdot K^l$. Ponieważ drzewo złożone z samego korzenia jest również nowym drzewem, więc liczba drzew o $l + 2$ poziomach wynosi $K_l^2 + 1$, zatem

$$K_{l+1} = 1 + K_l^2, \quad (31)$$

przy czym $K_0 = 1$. W tabeli 4 pokazano liczbę możliwych baz przestrzeni W_0^0 dla kilku wartości l . Ze wzrostem lK_l mocno rośnie i np. dla pięciopoziomowej pakietowej dekompozycji obrazu (2D) liczba różnych baz wynosi $5,6 \cdot 10^{78}$. Mnogość baz stwarza szansę wyboru takiej bazy, w której koszt zakodowania współczynników rozkładu analizowanego sygnału jest mniejszy niż w innych bazach. W następnym punkcie scharakteryzowano możliwe funkcje kosztu i algorytm wyboru *najlepszej bazy*.

Tabela 4

Liczba baz dopuszczalnych

Number of admissible bases

Liczba dekompozycji l	1	2	3	4	5
Liczba baz dopuszczalnych	2	5	26	677	458330

4.1. WYBÓR NAJLEPSZEJ BAZY PAKIETÓW FALKOWYCH

Coifman i Wickerhauser w pracy [4] podali prosty algorytm szukania najlepszej bazy pakietów falkowych. Jeżeli dla każdej podprzestrzeni W_j^p w pełnym drzewie binarnym określony jest "koszt" użycia do reprezentacji analizowanego sygnału współczynników rozwinięcia sygnału w danej podprzestrzeni, to można wybrać zbiór podprzestrzeni $\{W_j^p\}$ minimalizujący całkowity koszt reprezentacji w następujący sposób. Na początku oznacza się liście pełnego drzewa binarnego, tj. $\{W_l^p\}_{0 \leq p < 2^l}$, jako „mogące tworzyć najlepszą bazę”. Następnie dla $j = l - 1, l - 2, \dots, 1, 0; 0 \leq p, 2^j$ sprawdza się, czy koszt przypisany podprzestrzeni W_j^p jest mniejszy od sumy kosztów podprzestrzeni potomnych W_{j+1}^{2p} i W_{j+1}^{2p+1} . Jeżeli tak, to podprzestrzeń W_j^p zaznacza się jako „mogącą tworzyć najlepszą bazę”. Jeżeli nie, to koszt przypisany podprzestrzeni W_j^p zastępuje się sumą kosztów podprzestrzeni W_{j+1}^{2p} i W_{j+1}^{2p+1} . Proces sprawdzania powtarza się dla następnej podprzestrzeni, aż do W_0^0 włącznie. Na zakończenie przegląda się drzewo zaczynając od jego korzenia (W_0^0) i odcina gałęzie prowadzące do węzłów potomnych węzła oznaczonego jako „mogący tworzyć najlepszą bazę”. Liście tak przyciętego drzewa odpowiadają najlepszej bazie. Liczba operacji niezbędnych do wykonania algorytmu jest proporcjonalna do liczby węzłów w pełnym drzewie binarnym, a ta nie przekracza $2N - 1$, co świadczy o złożoności obliczeniowej $O(N)^4$. Poniżej podano prostą implementację opisanego wyżej algorytmu w języku ANSI C. W kodzie programu „zaszyto” dane testowe — tablica dwuwymiarowa $c[l][p]$ zainicjowana jest przykładowymi „kosztami”. Element tablicy $s[j][p]$ przechowuje informację o tym, czy podprzestrzeń W_j^p „może tworzyć najlepszą bazę” czy nie.

⁴ Oszacowanie to nie uwzględnia operacji niezbędnych do obliczenia funkcji kosztu. Łącznie, obliczenie funkcji kosztu na podstawie współczynników pakietowych i znalezienie najlepszej bazy wymaga $2N \log 2N$ operacji

```

/* Szukanie najlepszej bazy (algorytm Coifmana-Wickerhausera) */
#include <stdio.h>
#define l 3 /* liczba poziomow transformacji pakietowych */
#define L 8 /* L=2^l */
void main(void) {
    int c[l+1][L]={50},
        {20,22},
        {11,12,13,14},
        { 1, 2, 3, 4, 5, 6, 7, 8}},
        s[l+1][L],
        j, p, pmax, px, x, xmax, y;

    for (j=0; j<l; j++)
        for (p=0; p<L; p++)
            s[j][p]=0;
    /*zaznaczenie podprzestrzeni (l,p) jako "najtanszych" */
    for (p=0; p<L; p++)
        s[l][p]=1;
    /* zaznaczenie "tanszych" podprzestrzeni */
    for (j=l-1; j>=0; j--) {
        pmax=1;
        for (p=1; p<=j; p++)
            pmax*=2; /* pmax=2^j */
        for (p=0; p<pmax; p++)
            if (c[j][p]<c[j+1][2*p]+c[j+1][2*p+1])
                s[j][p]=1; /* zaznaczenie "tanszej" podprzestrzeni */
            else
                c[j][p]=c[j+1][2*p]+c[j+1][2*p+1];
    }
    /* usuwanie zaznaczen "nienajtanszych" podprzestrzeni */
    pmax=1;
    for (j=0; j<l; j++) {
        pmax*=2;
        for (p=0; p<pmax; p++)
            if (s[j][p]==1) {
                px=p; xmax=1;
                for (y=j+1; y<=l; y++) {
                    px*=2; xmax*=2;
                    for (x=0; x<xmax; x++)
                        s[y][px+x]=0;
                }
            }
    }
}

```

}
 pri
 for
 f
 }
Przyk
 strzen
 Wicke
 pozyc

która
 powyż

V
 klucz
 preze
 od wa
 otrzy
 dzie
 Autor
 zana
 kodow
 Ramo
 jąc te
 nia (r
 dla d


```

}
printf("\nNajlepsza baza tworza bazy podprzestrzeni (j,p):\n");
for (j=0; j<=1; j++)
    for (p=0; p<L; p++)
        if (s[j][p]==1) printf("(d,d) ",j,p);
}

```

Przykład 1: Na rysunku 7 podano przykładowe koszty wyboru do reprezentacji przestrzeni W_0^0 poszczególnych podprzestrzeni. Wynik wykonania algorytmu Coifmana-Wickerhausera wskazuje na bazę przestrzeni W_0^0 odpowiadającą następującej dekompozycji pakietowej:

$$W_0^0 = W_3^0 \oplus W_3^1 \oplus W_3^2 \oplus W_3^3 \oplus W_1^1,$$

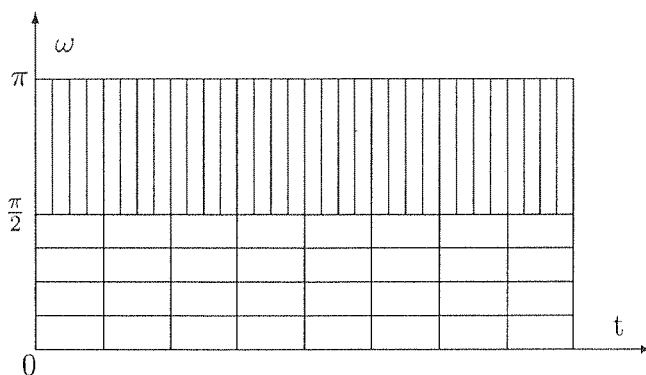
która minimalizuje całkowity koszt do 32. Na rysunku 8 pokazano odpowiadający powyższej dekompozycji podział płaszczyzny czas-częstotliwość.

W_0^0 50							
W_1^0 21				W_1^1 22			
W_2^0 11		W_2^1 12		W_2^2 13		W_2^3 14	
W_3^0 1	W_3^1 2	W_3^2 3	W_3^3 4	W_3^4 5	W_3^5 6	W_3^6 7	W_3^7 8

Rys. 7. Przykładowe "koszty" przypisane podprzestrzeniom

Fig. 7. Example of "costs" assigned to different spaces

W zastosowaniach algorytmu Coifmana-Wickerhausera szukania najlepszej bazy kluczową sprawą jest określenie sposobu ustalania kosztu związanego z wyborem do reprezentacji analizowanego sygnału bazy danej podprzestrzeni. Koszt powinien zależeć od wartości współczynników rozwinięcia sygnału w podprzestrzeni. Algorytm pozwala otrzymać bazę minimalizującą koszt, jeżeli funkcja kosztu M będzie addytywna, tj. będzie określona zarówno dla skalarów jak i ciągów i $M(0) = 0$, a $M(\{x_k\}) = \sum_k M(x_k)$. Autorzy algorytmu zaproponowali kilka różnych funkcji kosztu. Jedna z nich jest związana z entropią ciągu i nie jest addytywna. Wprawdzie entropia ma związek z kosztem kodowania symboli, jednakże przydatność praktyczna takiej miary jest ograniczona. Ramchandran i Vetterli w pracy [25] pokazali jak szukać najlepszej bazy uwzględniając teorię Shannona określającą zależność zniekształceń od średniej bitowej kodowania (rate-distortion theory) i algorytm Shohama-Gersha efektywnego przydziału bitów dla dowolnego zbioru kwantyzatorów [28]. Funkcja kosztu, którą wykorzystano nie



Rys. 8. Podział płaszczyzny czas-częstotliwość odpowiadający dekompozycji pakietowej przestrzeni

$$W_0^0 = W_3^0 \oplus W_3^1 \oplus W_3^2 \oplus W_3^3 \oplus W_1^1$$

Fig. 8. Time-frequency plane partition corresponding to wavelet packet decomposition

$$W_0^0 = W_3^0 \oplus W_3^1 \oplus W_3^2 \oplus W_3^3 \oplus W_1^1$$

była addytywna. Pewne odmiany podejścia do omawianego zagadnienia przedstawiono w pracach [26, 36], jednakże wszystkie one wymagają bardzo dużych nakładów obliczeniowych związanych z rozwiązywaniem nieliniowych zagadnień optymalizacyjnych, co ogranicza ich przydatność. Algorytm poszukiwania „prawie” najlepszych baz przy zastosowaniu nieaddytywnych funkcji kosztu podano w pracy [31]. Autorzy pracy [20] zrezygnowali ze stosowania teorii „średnia bitowa-zniekształcenia”. Wybrali funkcję kosztu, która ma dwa składniki: jeden związany z kosztem kodowania znaku i amplitudy współczynnika pakietowego po kwantyzacji, drugi związany z kosztem kodowania położenia w tablicy współczynnika znaczącego (różnego od zera po kwantyzacji). W każdej podprzestrzeni (podpasmie częstotliwości) rozkład współczynników aproksymowano rozkładem Laplace’a. Do kwantyzacji użyto skalarne kwantyzatora jednorodnego z symetryczną „strefą martwą”. Współczynniki w podpasmie przeglądano w kolejności wynikającej z krzywej Hilberta, zaczynając od podpasma o najniższej częstotliwości. Położenie współczynników znaczących kodowano entropijnie z uwzględnieniem kontekstu. Zamieszczone w pracy wyniki eksperymentów pokazują, że wizualnie, ale również i w sensie kryterium PSNR, metoda jest konkurencyjna w stosunku do metody falkowej wykorzystującej algorytm SPIHT [22]. Wyższość omawianej metody w sensie percepcyjnym dotyczy zwłaszcza tekstur. Informacje o rozwoju algorytmów pakietowych i ich zastosowań można znaleźć w pracy Wickerhausera [35], jednego z twórców koncepcji pakietów falkowych.

5. WYBRANE ZASTOSOWANIA PAKIETÓW FALKOWYCH

Pierwsze zastosowania znalazły pakiety falkowe w kompresji dźwięku [34]. W cytowanej pracy zamieszczono wyniki eksperymentu kompresji mowy. Zarejestrowano

zдание „Joe brought a young girl”, próbując je z częstotliwością 22kHz i przeznaczając 8 bitów na próbkę. Otrzymano blisko 2^{15} próbek i przyjęto długość sygnału równą 32768. Cały sygnał podzielono na osiem równych segmentów po $4096 = 2^{12}$ próbek. Każdy segment kompresowano niezależnie. Do transformacji pakietowej używano filtrów o dwóch, dziesięciu i osiemnastu współczynnikach. Współczynniki pakietów falkowych o modułach mniejszych niż z góry zadany próg ε zastępowano zerami. „Najlepszą” bazę wyszukiwano spośród dwunastu, wybierając do aproksymacji współczynniki pakietów falkowych z tego samego poziomu transformacji, tj. z tym samym indeksem skali j . Poprzez takie ograniczenie dopuszczalnych baz, zmniejszono liczbę bitów niezbędnych do kodowania położenia współczynnika pakietowego w transformacji, gdyż indeks j mógł być zakodowany tylko raz dla wszystkich współczynników. Kryterium wyboru stanowiła liczba zerowych współczynników pakietowych (po progowaniu). Ten wiersz transformaty pakietowej, który miał najwięcej zer odpowiadał najlepszej bazie. Eksperyment wykonano dla głosu kobiety i dla głosu mężczyzny. Osiągnięto współczynnik kompresji od 5 do 10, przy zniekształceniach rzędu kilku procent. Stwierdzono, że zależność pomiędzy stopniem kompresji i zniekształceniami była liniowa, i że dłuższe filtry dawały mniejsze zniekształcenia. Zniekształcenia wyższe niż kilka procent objawiały się mową zaszumioną. Znacznie lepsze efekty kompresji dźwięku osiągnięto stosując bardziej wyrafinowaną kwantyzację współczynników pakietowych oraz wykorzystując do kodowania algorytmu oparte na koncepcji drzew zer współczynników falkowych [22]. Dalszą poprawę efektywności kompresji osiągnięto uwzględniając własności psychoakustyczne narządu słuchu, w szczególności tzw. efekt maskowania, polegający na tym, że słaby sygnał z danego podpasma nie będzie słyszalny, jeżeli w sąsiednim podpasmie jest silny sygnał. Dzięki temu można zerować współczynniki pakietowe takich słabych sygnałów nie wprowadzając zniekształceń [12]. Tego typu system kompresji, pozwalający zachować wysoką jakość dźwięku próbkowanego z częstotliwością 44.1 kHz przy transmisji z szybkością 45 kbitów/sek i posiadający własność progresywnej transmisji, przedstawiono w pracy [29].

Jeden z wcześniejszych systemów kompresji obrazów wykorzystujący pakiety falkowe przedstawiono w pracy [25]. Otrzymane wyniki porównano z wynikami standardu kompresji JPEG. W pracy [36] przedstawiono metodę kompresji obrazów, w której połączono algorytm Shapira kodowania drzew zer [22] z zależnością zniekształceń od średniej bitowej kodowania. Uzyskane tą metodą wyniki kompresji odcisków palców porównano z wynikami systemu kompresji FBI WSQ, otrzymując wyższy PSNR. Wyczerpujący opis efektywnego systemu kompresji obrazów wykorzystującego pakiety falkowe wraz z obrazami umożliwiającymi porównanie w kryteriach percepcyjnych omawianej metody z metodą falkową wykorzystującą SPIHT zawiera praca [20].

Koncepcję nowego koder (CPSO), przewyższającego jakością (dla określonych klas obrazów) koder oparty na algorytmie SPIHT, przedstawiono w pracy [10]. Kompresję elektrokardiogramów za pomocą pakietów falkowych opisano w pracy [11]. Oprócz wymienionych tu zastosowań w kompresji, pakiety falkowe znajdują zasto-

sowania w ekstrakcji cech [27], w algorytmach klasyfikacji [14, 15], w aplikacjach telekomunikacyjnych [16], w metodach numerycznych [21].

6. UWAGI KOŃCOWE

Pakiety falkowe dostarczają baz do aproksymacji sygnałów, które mogą lepiej „pasować” do analizowanego sygnału niż bazy falkowe. Dotyczy to w szczególności sygnałów zawierających oscylacje oraz sygnałów niestacjonarnych. Dobrze znaną klasą obrazów, które efektywniej aproksymuje się pakietami falkowymi niż falkami są odciski palców. Ostatnie badania pokazały, że tekstury są bardziej „przyjazne” pakietom falkowym niż falkom. Reprezentacja sygnałów i obrazów za pomocą pakietów falkowych ma również swoje wady w porównaniu do reprezentacji falkowej. Najistotniejszą jest znacznie większa złożoność obliczeniowa algorytmów pakietowych niż falkowych. Nieustanny rozwój sprzętu i nowe implementacje równoległe algorytmów pakietowych na procesory wektorowe SIMD i systemy wieloprocessorowe MIMD [8, 9, 13] obniżają bariery obliczeniowe algorytmów pakietowych, dzięki czemu metody pakietowe stają się użytecznym uzupełnieniem metod falkowych analizy sygnałów i obrazów, jak również metod opartych na okienkowej transformacji Fouriera.

7. BIBLIOGRAFIA

1. R. Coifman, Y. Meyer, V. Wickerhauser: *Size Properties of Wavelet-Packets*. M.B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, L. Raphael, editors, Wavelets and Their Applications, Boston, Jones and Bartlett Publishers, MA, 1992, pp. 453-470.
2. R. Coifman, Y. Meyer, V. Wickerhauser: *Wavelet Analysis and Signal Processing*. M.B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, L. Raphael, editors, Wavelets and Their Applications, Boston, Jones and Bartlett Publishers, MA, 1992, pp. 153-178.
3. R.R. Coifman, Y. Meyer, S.R. Quake, M.V. Wickerhauser: *Signal Processing and Compression with Wavelet Packets*. J.S. Byrnes, J.L. Byrnes, K.A. Hargreaves, K. Berry, editors, Wavelets and Their Applications, volume 442 of NATO ASI Series C: Mathematical and Physical Sciences, Kluwer Academic Publishers, Dordrecht/Boston/London, 1994. Proceedings of the NATO Advanced Study Institute at Il Ciocco, Barga, Italy, August, 1992, pp. 363-379.
4. R.R. Coifman, M.V. Wickerhauser: *Entropy-Based Algorithms for Best Basis Selection*. IEEE Transactions on Information Theory, 38(2), 1992, pp. 713-718.
5. A. Croisier, D. Esteban, C. Galand: *Perfect channel splitting by use of interpolation/decimation techniques*. Internat. Conf. On Information Sciences and Systems, Patras, Greece, August 1976, pp. 443-446.
6. I. Daubechies: *Orthonormal bases of compactly supported wavelets*. Communications on Pure and Applied Mathematics, 41, 1988, pp. 909-996.
7. I. Daubechies: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
8. M. Feil, A. Uhl: *Wavelet Packet Decomposition and Best Basis Selection on Massively Parallel SIMD Arrays*. Proceedings of the International Conference Wavelets and Multiscale Methods (IWC'98), INRIA, Rocquencourt, April 1998, <http://citeseer.nj.nec.com/article/feil98wavelet.html>.

9. M. Feil, A. Uhl: *Algorithms and Programming Paradigms for 2-D Wavelet Packet Decomposition on Multicomputers and Multiprocessors*. ACPC, 1999, pp. 367-376, <http://citeseer.nj.nec.com/feil99algorithm.html>.
10. S.-Y. Han, S.-Y. Cho: *Coefficient Partitioning Scanning Order Wavelet Packet Algorithm for Satellite Images*. CAIP 2003, 10th International Conference on Computer Analysis of Images and Patterns, Groningen, The Netherlands, Aug., 2003, pp. 25-27.
11. M. Hilton: *Wavelet and wavelet packet compression of electrocardiograms*. IEEE Trans. Biomed. Eng., 44(2), May 1997, pp. 394-40, <http://citeseer.nj.nec.com/hilton97wavelet.html>.
12. J.D. Johnston: *Transform Coding of Audio Signals Using Perceptual Noise Criteria*. IEEE Transactions on Selected Areas in Communications, 6, 1988, pp. 314-323.
13. R. Kutil, A. Uhl: *Parallel adaptive wavelet analysis*. Future Generation Computer Systems, 18(1), 2001, pp. 97-106, <http://citeseer.nj.nec.com/kutil01parallel.html>.
14. A. Laine, J. Fan: *Texture classification by wavelet packet signatures*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(15), 1993, pp. 1186-1191.
15. R.E. Learned, A.S. Willsky: *A wavelet packet approach to transient signal classification*. Appl. Comput. Harmonic Anal., 2, 1995, pp. 265-278.
16. A.R. Lindsey, J.C. Dill: *Digital transceiver implementation for wavelet packet modulation*. H.H. Szu, editor, Wavelet Applications V, SPIE Proceedings, volume 3391, 1998, pp. 255-264.
17. S. Mallat: *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.
18. S.G. Mallat: *Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$* . Transactions of the American Mathematical Society, 315(1), September 1989, pp. 69-87.
19. S.G. Mallat: *A theory for multiresolution signal decomposition: The wavelet representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7), July 1989, pp. 674-693.
20. F.G. Meyer, A.Z. Averbuch, J.-O. Strömberg: *Fast Adaptive Wavelet Packet Image Compression*. IEEE Transactions on Image Processing, 9(5), May 2000.
21. L. Bacchelli Montefusco: *Semi-orthogonal Wavelet Packet Bases for Parallel Least-squares Approximation*. J. Comput. Appl. Math., 73, 1996, pp. 191-208.
22. W. Rakowski: *Kodowanie transformat falkowych dla obrazów*. VI Symposium Nowości w Technice Audio - Technika Audio a Multimedia, Warszawa, 22-23 października 1999, ss. 101-111.
23. W. Rakowski: *Metody falkowe kompresji obrazów z obszarami zainteresowania*. Rozprawy Naukowe Nr 99. Wydawnictwo Politechniki Białostockiej, Białystok, 2002.
24. W. Rakowski: *A Proof of the Necessary Condition for Perfect Reconstruction of Signals Using the Two-Channel Wavelet Filter Bank*. Bulletin of the Polish Academy of Sciences, Technical Sciences, 51(1), 2003, pp. 13-23.
25. K. Ramchandran, M. Vetterli: *Best Wavelet Packet Bases in a Rate-Distortion Sense*. IEEE Transactions on Image Processing, 2(2), 1993, pp. 160-175.
26. K. Ramchandran, Z. Xiong, K. Asai, M. Vetterli: *Adaptive Transforms for Image Coding Using Spatially-varying Wavelet Packets*. IEEE Transactions on Image Processing, 5, 1996, pp. 1197-1204.
27. N. Saito, R.R. Coifman: *Local Discriminant Bases*. A.F. Laine, M.A. Unser, editors, Wavelet Applications in Signal and Image Processing II, SPIE Proceedings, vol. 2303, 1995, pp. 2-14.
28. Y. Shoham, A. Gersho: *Efficient Bit Allocation for an Arbitrary Set of Quantizers*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 36(9) September 1988, pp. 1445-1453.
29. P. Srinivasan, L. Jamienson: *High Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modelling*. IEEE Transactions on Signal Processing, 46(4), 1998, pp. 1085-1093.
30. J. Szabat: *Podstawy teorii sygnałów*, Warszawa, WKŁ, 2000.
31. C. Taswell: *Satisficing Search Algorithms for Selecting Near-Best Bases in Adaptive Tree-Structured Wavelet Transforms*. IEEE Transactions on Signal Processing, 44(10), 1996, pp. 2423-2438.

32. M. Vetterli: *Filter Banks Allowing Perfect Reconstruction*. Signal Processing, 10(3), April 1986, pp. 219-244.
33. M. V. Wickerhauser: *INRIA Lectures on Wavelet Packet Algorithms*. Pierre-Louis Lions, editor, Problèmes Non-Linéaires Appliqués, Ondelettes et Paquets D'Ondes., INRIA, Roquencourt, France, 17-21 June 1991, pp. 31-99 <http://www.math.wustl.edu/victor/papers/lwpa.pdf>.
34. M. V. Wickerhauser: *Acoustic signal compression with wavelet packets*.+ Ch.K. Chui, editor, Wavelets - A Tutorial in Theory and Applications, Academic Press, Boston, 1992, pp. 679-700, <http://www.math.wustl.edu/victor/papers/acoustic.ps.gz>.
35. M. V. Wickerhauser: *Progress in Wavelet Algorithms and Applications*. H.H. Szu, editor, Wavelet and Independent Component Analysis Applications IX, vol. 4738 of SPIE Proceedings, Orlando, Florida, 3-5 April 2002, , pp. 1-12 <http://www.math.wustl.edu/victor/papers/pwaa.pdf>.
36. Z. Xiong, K. Ramchandran, M.T. Orchard: *Wavelet Packet Image Coding Using Space-frequency Quantization*. IEEE Transactions on Image Processing, 7, 1998, pp. 160-174.

W. RAKOWSKI

WAVELET PACKETS IN SIGNAL AND IMAGE COMPRESSION

S u m m a r y

In the paper mathematical basics of signal representation by wavelet packets were presented. Walsh wavelet packet bases for signal approximation were described. An algorithm to search the best basis and the implementation of the algorithm in ANSI C programming language were shown. Computational complexity of the wavelet packet representation was presented. Application of wavelet packets in both signal and image compression were discussed.

Keywords: wavelet, wavelet filter, wavelet transform, wavelet packets, best basis, sound compression, image compression

C
prze
ków
ków
liczb
Jedn
rozpr
mów
system

Szeregowanie dynamiczne dla grup wiadomości w rozproszonych systemach czasu rzeczywistego

ŚLAWOMIR ŻABA

*Instytut Elektrotechniki Teoretycznej i Automatyki, Politechnika Krakowska
ul. Warszawska 24, 31-155 Kraków
e-mail: szaba@usk.pk.edu.pl*

*Otrzymano 2003.02.19
Autoryzowano 2003.06.18*

W artykule przedstawiono nową metodę szeregowania wiadomości. W zaproponowanej metodzie (Dynamic Priority Group Scheduling) szeregowanie wiadomości odbywa się w oparciu o dynamicznie przydzielany priorytet dla grup wiadomości. Opracowany został także wariant omawianej metody umożliwiający szeregowanie wiadomości z mieszanym przydziałem priorytetu (Mixed Priority Group Scheduling). Przedstawiono sposób sprawdzania dotrzymania warunków RT dla metod DPGS i MPGS. Dla trzech wybranych magistral miejscowych (PROFIBUS FMS, CAN, InterBus-S) przeprowadzono badania, mające na celu porównanie właściwości opracowanych metod z metodami GRMS i EDF.

Słowa kluczowe: rozproszony system komputerowy, system czasu rzeczywistego, magistrala miejscowa, szeregowanie wiadomości, grupowanie wiadomości, metoda GRMS, EDF

1. WSTĘP

Obszary zastosowania sieciowych systemów komunikacyjnych we współczesnym przedsiębiorstwie obejmują wszystkie poziomy — od zarządzania do poziomu czujników i zadajników sygnałów. Komunikację na poziomie sterowników PLC oraz czujników i elementów wykonawczych, w której wykorzystuje się cykliczną transmisję małej liczby danych w czasie rzeczywistym, zapewniają magistrale miejscowe (*fieldbus*) [3]. Jednakże zastosowanie systemów sieciowych sprawia, że analiza i projektowanie takich rozproszonych systemów sterujących staje się bardziej złożona niż w przypadku systemów scentralizowanych. W szczególności niezwykle trudna jest analiza i projektowanie systemów o ostrych ograniczeniach czasowych odnośnie nienaruszalności ograniczeń

czasowych dla każdej przesyłanej wiadomości. Spełnienie ograniczeń czasowych w dużym stopniu zależy od metody szeregowania wiadomości [7, 8, 9]. Dla systemów czasu rzeczywistego z globalnym programem szeregującym zadania, rozwiniętych zostało wiele analitycznych metod szeregowania i sprawdzania dotrzymania ograniczeń czasu rzeczywistego. Istotnym zagadnieniem staje się opracowanie nowych analitycznych metod szeregowania i sprawdzania spełnienia warunków czasu rzeczywistego lub adaptacja metod szeregowania zadań opracowanych dla środowiska scentralizowanego systemów czasu rzeczywistego do wymagań środowiska rozproszonego.

2. UOGÓLNIONY MODEL DOSTĘPU DO WARSTWY FIZYCZNEJ MAGISTRAL MIEJSCOWYCH

W celu zaprezentowania opracowanych metod wprowadzone zostaną dodatkowe pojęcia i oznaczenia. Oznaczmy czas obiegu znacznika (dla systemu stosującego metodę z przekazywaniem znacznika) lub czas cyklu odpytania węzłów slave przez układ master (dla metody odpytań) jako docelowy czas obiegu znacznika *TTRT* (*target token rotation time*).

Aby w systemie dotrzymane zostały ograniczenia czasowe wartość *TTRT* musi być dobrana w ten sposób, że:

$$TTRT \leq d_{min} \quad (1)$$

gdzie: d_{min} jest najkrótszym ograniczeniem czasowym zadania w systemie.

Dla $d_i = t_i$ zachodzi $TTRT \leq t_{min}$, gdzie t_{min} jest najkrótszym okresem występowania zadania w systemie.

Jeżeli warunek (1) nie będzie dotrzymany, to nie będą spełnione ograniczenia dla danych z okresem występowania t_{min} , a tym samym ograniczeniem d_{min} .

Oznaczamy w_T jako sumę czasów przekazywania znacznika pomiędzy poszczególnymi węzłami lub sumę czasów odpytań lub sumę czasów trwania danych synchronizujących (w zależności od zastosowanej metody dostępu). Czas w_T jest czasem potrzebnym do organizacji transmisji. Dlatego pozostały czas, tzn. $TTRT - w_T$ może być efektywnie wykorzystany do transmisji danych. Czas, w którym węzeł i wyłącznie dysponuje magistralą oznaczony jest przez h_i i obliczany jest wg. Wzoru [4]:

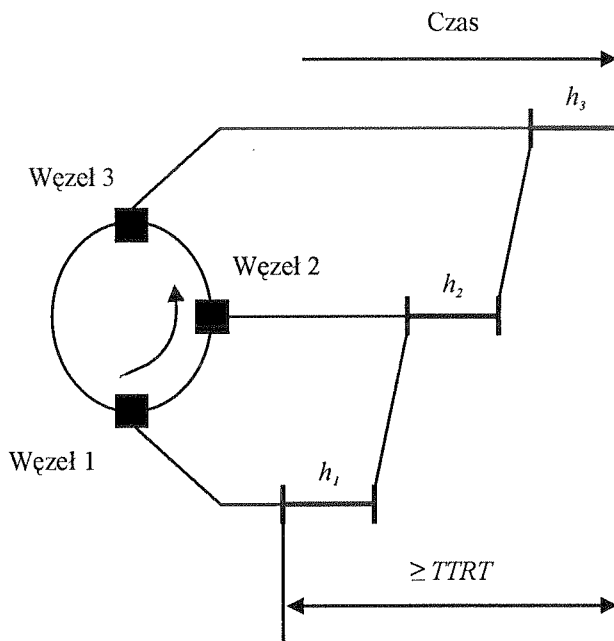
$$h_i = \frac{u_i}{u} (TTRT - w_T) \quad (2)$$

gdzie: $u = u_1 + \dots + u_n$, a u_i jest stopniem wykorzystania sieci przez stację i

$$u_i = \sum_{j \in P_j} \frac{c_j}{t_j} \quad (3)$$

gdzie: zbiór $P_i = \{P_{i1}, P_{i2}, \dots, P_{im}\}$ oznacza pakiety danych należące do węzła i .

Na rys. 1 przedstawiono omawianą sytuację.

Rys. 1. Ilustracja czasu obiegu znacznika $TTRT$ Fig. 1. Presentation of target token rotation time $TTRT$

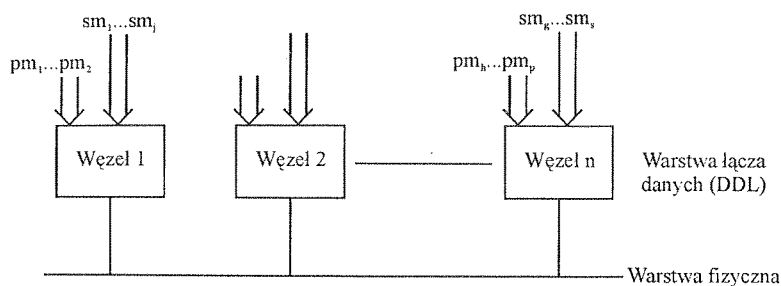
Każdy węzeł może wysyłać dane przez czas nie większy od wartości h_i (postulat ten będzie obowiązujący w pozostałej części artykułu!). W ten sposób nie zostanie przekroczona wartość $TTRT$.

3. MODEL STRUMIENIA WIADOMOŚCI

Przy opracowywaniu metod szeregowania wiadomości dla magistrali należy zdefiniować model wiadomości. Przyjęty model (rys. 2) obejmuje wiadomości okresowe oraz nieokresowe.

Elementy modelu [6]:

- PM (periodic message) oznacza zbiór wiadomości okresowych, SM (aperiodic message) oznacza zbiór wiadomości nieokresowych, czyli: $PM = \{pm_1, pm_2, \dots, pm_i, \dots, pm_p\}$, $SM = \{sm_1, sm_2, \dots, sm_j, \dots, sm_s\}$
- PM_i — oznacza zbiór wiadomości okresowych należących do węzła i , SM_i oznacza zbiór wiadomości nieokresowych należących do węzła i , czyli: $PM_i = \{pm_e, pm_f, \dots, pm_l\}$, $SM_i = \{sm_c, sm_d, \dots, sm_g\}$,
- tp_i jest okresem występowania wiadomości okresowej pm_i , $i = 1, 2, \dots, p$,
- ts_j oznacza minimalny czas pomiędzy kolejnymi pojawieniami się wiadomości nieokresowej sm_j , $j = 1, 2 \dots s$,



Rys. 2. Model strumieni wiadomości i protokołu sieciowego

Fig. 2. Model of message stream and fieldbus protocol

- $dpi(ds_j)$ oznaczają graniczne przedziały czasowe dla zrealizowania przesłania wiadomości $p_{mi}(sm_j)$. Jeżeli wiadomość $p_{mi}(sm_j)$ została przekazana w chwili t do wysłania, to ograniczenie czasowe wynosi $t + dpi$ dla wiadomości p_{mi} i $t + ds_j$ dla wiadomości sm_j . Zakłada się spełnienie warunku $dpi \leq tpi$ dla wiadomości okresowych oraz $ds_j \leq ts_j$ dla wiadomości nieokresowych.
- $cpi(cs_j)$ oznaczają przedział czasu, jaki potrzebny jest do nadawania wiadomości $p_{mi}(sm_j)$. Wartość $cpi(cs_j)$ możemy wyznaczyć z zależności:

$$cpi = \frac{1}{\alpha}(bp_i + \beta)$$

$$cs_j = \frac{1}{\alpha}(bs_j + \beta)$$
(4)

gdzie: α — prędkość transmisji w [Kb/s],
 β — dane nadmiarowe, czyli te elementy, które są dołączane do transmitowanej wiadomości przez warstwę łącza danych i warstwę fizyczną, np.: pole adresowe, pole kontrolne, itp., [b]
 bp_i — długość wiadomości p_{mi} [b],
 bs_j — długość wiadomości sm_j [b].

Oczywiście przedział czasu nadawania wiadomości powinien być mniejszy od jej ograniczenia czasowego, co zapisujemy $cpi \leq dpi$ ($cs_j \leq ds_j$). W zapisie jednostek b oznacza bity.

- wiadomości są niezależne od siebie,

Przedstawiony model wiadomości posłuży do opracowania metod szeregowania wiadomości na magistrali.

4. POTRZEBY OPRACOWANIA METOD SZEREGOWANIA WIADOMOŚCI

Stosowanie algorytmów szeregowania zadań do szeregowania wiadomości posiada kilka mankamentów. Podstawową wadą jest fakt, że metody szeregowania zadań nie

uwzględniających
 nych w
 czas. I
 nych w

Przyk

C
 nego w
 sowych
 Za
 = 5 [m
 param

Wiad

A
 zostały
 ich og
 znacz
 wiado
 obieg
 obieg
 zosta
 czasow
 system

uwzględniają specyfiki działania systemów rozproszonych. W systemach rozproszonych węzeł może wysyłać wiadomości tylko w określonych chwilach i przez określony czas. Dlatego należy dostosować parametry czasowe (ograniczenia czasowe) wysyłanych wiadomości do parametrów czasowych systemu rozproszonego.

Przykład 1

Celem przykładu jest pokazanie, że poprawne zaprojektowanie systemu rozproszonego wymaga dostosowania parametrów czasowych wiadomości do parametrów czasowych systemu rozproszonego (w szczególności do czasu TTRT).

Zostanie rozważony system rozproszony, w którym czas obiegu znacznika TTRT = 5 [ms], a czas do wyłącznej dyspozycji jednego z węzłów wynosi 1 [ms]. Przyjęte parametry czasowe wiadomości przedstawia tabela 1.

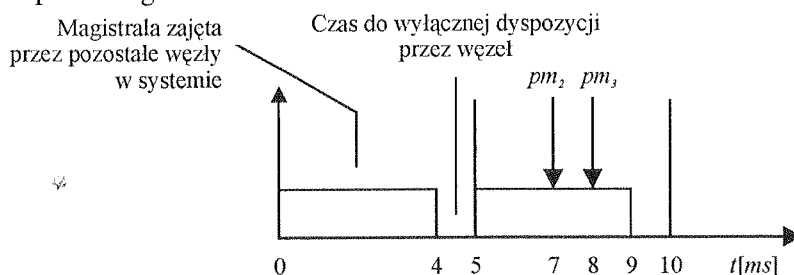
Tabela 1

Parametry czasowe wiadomości

Time properties of messages

Wiadomość okresowa	Okres występowania tp_i	Czas przesłania cp_i	Ograniczenie czasowe dp_i
pm_1	5	0.2	5
pm_2	7	0.4	7
pm_3	8	0.3	8

Aby ograniczenia czasowe dla najgorszego przypadku w rozważanym systemie zostały dotrzymane, wszystkie wiadomości z tab. 1 powinny być traktowane, jak gdyby ich ograniczenia czasowe wynosiły 5 [ms]. Wynika to z faktu, że czas jednego obiegu znacznika wynosi 5 [ms], a dwóch obiegu wynosi 10 [ms]. Ponieważ ograniczenia wiadomości pm_2 i pm_3 są większe od jednego, a mniejsze od czasu trwania dwóch obiegu znacznika, to ograniczenia te należy zmniejszyć do czasu trwania jednego obiegu znacznika. Przyjęcie ograniczenia równego dwóm obiegom znacznika może zostać nie dotrzymane, jak to ilustruje rys. 3. Należy więc dostosować parametry czasowe (ograniczenia czasowe) wysyłanych wiadomości do parametrów czasowych systemu rozproszonego.



Rys. 3. Przebiegi czasowe dla wybranego węzła w przykładowym systemie rozproszonym

Fig. 3. Time behaviour for a node in the given distributed system

5. METODA SZEREGOWANIA DYNAMICZNEGO DLA GRUP WIADOMOŚCI

5.1. TRANSFORMACJA OGRANICZEŃ CZASOWYCH WIADOMOŚCI

W systemach rozproszonych węzeł może wysyłać wiadomości tylko w określonych chwilach i przez określony czas. Należy więc dostosować ograniczenia czasowe wysyłanych wiadomości do właściwości czasowych systemu rozproszonego. Dalej opisano propozycję metody szeregowania dynamicznego dla grup wiadomości. Czas $TTRT$, który jest równy lub mniejszy od najkrótszego ograniczenia wiadomości w systemie będzie ograniczeniem bazowym, a pozostałe ograniczenia wiadomości należy zdefiniować jako wielokrotności czasu $TTRT$ [7]. Ograniczenie dm_{ij} wiadomości m_i z węzła j po transformacji wyznacza się ze wzoru:

$$dm^*_{ij} = \left\lfloor \frac{dm_{ij}}{TTRT} \right\rfloor TTRT \quad (5)$$

Wiadomości o tym samym ograniczeniu dm^*_{ij} będą tworzyć tzw. grupy GR_i wiadomości. Do grupy GR_1 (w węźle k) należeć będą wiadomości, dla których dm^*_{ik} wynosi 1, do grupy GR_2 wiadomości, dla których dm^*_{ik} wynosi 2, itd. Przyjmijmy następującą notację dla każdego węzła:

- n_{ij} — liczba wiadomości w grupie GR_i w węźle j ,
- dm^*_{ij} — ograniczenie czasowe dla grupy GR_i ,
- $k_i = \frac{dm^*_{ij}}{TTRT}$ — wskaźnik wielokrotności ograniczenia dla grupy GR_i
- lg_j — liczba grup w węźle j ,

5.2. PRZYDZIELANIE PRIORYTETÓW

Przydział priorytetów odbywa się w każdym węźle dynamicznie. Algorytm przyporządkowuje wyższy priorytet grupie z najkrótszym przedziałem przeterminowania. Obliczanie priorytetów odbywa się na podstawie długości przedziału przeterminowania:

$$NrPrzedz[i] = k_i - Modulo \frac{NrCyklu - 1}{k_i} \quad (6)$$

gdzie: $1 \leq i \leq m_j$,

$NrCyklu$ — liczba, oznaczająca kolejny numer cyklu obiegu znacznika ($NrCyklu \leq 1$).

Wartość zmiennej $NrPrzedz[i]$ obliczana jest w każdym cyklu obiegu znacznika, po otrzymaniu przez węzeł prawa do wysyłania wiadomości. Wartość ta reprezentuje liczbę obiegów znacznika, po której ograniczenia czasowe w danej grupie zostaną przekroczone. Grupie GR_i , dla której zmienna $NrPrzedz[i]$ osiąga najmniejszą wartość,

przypię
posiad
wybra
szego
wiadom
wydaj
grupy,
o wyż

N
cza w
wym v
łów pr
z najm
i tym
N

pracę
każde
grup v
domoś
wiadom
czasow
przed
sprow
przek
od gru
jest on
niższy
pie z
do gru
Priori

6. AL

W
gorytr
miejs
znaczn
spraw
wykor
w me
ną. Sp
DPGS

IOŚCI

ślonych
e wysy-
ppisano
TTRT,
systemie
definio-
z węzła

(5)

wiado-
wynosi
ępującą

n przy-
owania.
owania:

(6)

nika,po
je licz-
ą prze-
wartość,

przypisuje się najwyższy priorytet. Jeżeli grupy w danym cyklu obiegu znacznika posiadają ten sam priorytet (ta sama wartość zmiennej *Nr Przedz[i]*), to teoretycznie wybrać można dowolną grupę, z której wiadomości będą wysyłane. Nie ma to większego znaczenia, jeżeli w systemie dotrzymywane są ograniczenia czasowe dla wszystkich wiadomości. Jednakże, jeżeli w systemie wystąpią chwilowe przeciążenia, to logicznym wydaje się wybór grupy z mniejszą wartością indeksu. Im mniejszy numer indeksu grupy, tym krótsze ograniczenia czasowe wiadomości (względem wiadomości z grup o wyższym indeksie), a tym samym większe znaczenie z pkt. widzenia wymagań RT.

Należy zaznaczyć, że dla metod EDF, MLF i MUF program szeregujący oblicza wartość przedziału przeterminowania dla *wszystkich* wiadomości po *każdorazowym* wysłaniu wiadomości i aktywacji wiadomości. Po obliczeniu wartości przedziałów przeterminowania program szeregujący wybiera wartość najmniejszą. Wiadomość z najmniejszą wartością przedziału przeterminowania otrzymuje najwyższy priorytet i tym samym zostaje przekazana do wysłania.

Natomiast dużą zaletą utworzenia grup wiadomości jest mały narzut czasowy na pracę programu szeregującego. Zmienna *NrPrzedz[i]* obliczana jest tylko raz w ciągu każdego obiegu cyklu znacznika (w każdym węźle). Liczba obliczeń zależy od liczby grup wiadomości — w najgorszym przypadku (każda grupa zawiera tylko jedną wiadomość) równa jest ona liczbie wiadomości, w każdym innym jest mniejsza od liczby wiadomości w systemie. Jeżeli więc dane są wszystkie wiadomości i ich parametry czasowe przed uruchomieniem systemu, to podział na grupy odbywa się jeden raz przed uruchomieniem systemu. Rola programu szeregującego (podczas pracy systemu) sprowadzać się będzie do przyporządkowania priorytetu dla grupy, a następnie na przekazywaniu do wysłania wiadomości w określonej wcześniej kolejności poczynawszy od grupy z najwyższym priorytetem. Jeżeli wiadomość nie jest gotowa do wysłania, to jest ona pomijana i wysyłana jest kolejna wiadomość. Przejście do grupy z priorytetem niższym możliwe jest wówczas, gdy nie ma gotowych wiadomości do wysłania w grupie z wyższym priorytetem. Opisany powyżej sposób przydziału priorytetów należy do grupy metod z dynamicznym przydziałem priorytetów i nazwany zostanie *Dynamic Priority Group Scheduling (DPGS)*.

6. ANALIZA CZASOWA SYSTEMU ROZPROSZONEGO DLA METODY DPGS

W dokonywanych analizach założono, że w węzłach systemu stosowany jest algorytm szeregowania wiadomości oparty o metodę *DPGS*. Przyjęto, że magistrala miejscowa może w warstwie MAC stosować jedną z metod: odpytań, przekazywanie znacznika, TDMA. Proponowana metoda analizy dla *DPGS* składa się z dwóch części sprawdzania dotrzymywania ograniczeń RT: *DPGS_I* i *DPGS_II*. Metoda *DPGS_I* musi być wykonana zawsze, ponieważ wartości wyznaczane przy jej użyciu są wykorzystywane w metodzie *DPGS_II*. Poza tym, metoda *DPGS_I* jest metodą najmniej czasochłonną. Sprawdzanie dotrzymywania ograniczeń czasowych w systemie przy użyciu metody *DPGS_II* jest wykonywane wtedy, gdy metoda *DPGS_I* nie da pożądaných efektów.

Metoda DPGS_I (podstawowa) obejmuje etapy opisane poniżej:

- **Etap I. Wyznaczenie czasu potrzebnego do nadawania wiadomości**
Czas przesłania wiadomości wyznaczany jest ze wzoru (4).
- **Etap II. Transformacja ograniczeń czasowych wiadomości**
Transformacja i podział na grupy wykonywane jest jak opisano w rozdz. 5.1.
- **Etap III. Przydział czasu h_{jwyst} do wyłącznej dyspozycji przez węzeł**
Metoda wyznaczania opisana została w oddzielnym punkcie poniżej.
- **Etap IV. Sprawdzenie warunku czasu rzeczywistego przydziału czasów h_{jwyst} dla całej sieci**

Jeżeli dla obliczonych wartości h_{jwyst} zostanie spełniona zależność:

$$\sum_{j=1}^m h_{jwyst} \leq TTRT - w_T \quad (7)$$

m — liczba węzłów w systemie,

to w systemie zostaną dotrzymane ograniczenia czasowe dla wszystkich wiadomości. **Wyznaczenie przydziału czasu h_{jwyst} do wyłącznej dyspozycji przez węzeł (etap III metody DPGS_I)** Należy zaznaczyć, że wyznaczany poniżej parametr h_{jmin} nie będzie używany w dalszej części pracy. Wyznaczenie wartości h_{jmin} jest etapem pośrednim dla wyznaczenia h_{jwyst} . Minimalny czas, w którym węzeł j wyłącznie dysponuje magistralą wyznaczony jest z następującej zależności:

$$h_{jmin} = \frac{1}{\alpha} \left(\left(\sum_{i=1}^{n_{1j}} bm_{1j} + n_{1j} * \beta \right) + \sum_{i=2}^{lg_j} \left[\frac{\sum_{s=1}^{n_{ij}} bm_{is} + n_{ij} * \beta}{k_i} \right] \right) \quad (8)$$

gdzie: bm_{ij} - długość wiadomości okresowej lub nieokresowej m_j z grupy GR_i

W pierwszej części wzoru (8), tj:

$$\sum_{i=1}^{n_{1j}} bm_{1i} + n_{1j} * \beta$$

obliczana jest suma bajtów danych oraz narzutów protokołu dla wszystkich wiadomości z grupy 1, bowiem muszą one być wysyłane w każdym cyklu obiegu znacznika (ograniczenie czasowe wiadomości z grupy 1 równe jest czasowi $TTRT$). W drugiej części wzoru, tj:

$$\sum_{i=2}^{lg_j} \left[\frac{\sum_{s=1}^{n_{ij}} bm_{is} + n_{ij} * \beta}{k_i} \right]$$

za pomocą wskaźnika i wybierana jest kolejna grupa (od 2 do lg_j), a wskaźnik s dotyczy sumowania wiadomości w wybranej przez wskaźnik i grupie. Wprowadzenie funkcji

[$\cdot$] jest podyktowane koniecznością uniknięcia powstania wartości ułamkowych przy obliczaniu sumy długości wiadomości w grupie. Obliczona wartość h_{jmin} wystarczy, aby w węźle j zostały dotrzymane ograniczenia czasowe dla wszystkich wiadomości, jeżeli w węźle nie wystąpią żadne straty czasowe, tzn. cały przydzielony czas h_{jmin} będzie zawsze wykorzystany na wysyłanie wiadomości. Taka sytuacja ma miejsce, gdy wiadomości w grupie (lub całym systemie) mają tę samą długość. Jednakże, jeżeli wiadomości w danej grupie mają różne długości, to z reguły wystąpi sytuacja, kiedy węzeł nie wykorzysta przyznanego mu czasu ze względu na konieczność wcześniejszego oddania prawa do nadawania dla następnego węzła. Omawianą sytuację ilustruje przykład 2.

Przykład 2

Celem przykładu jest pokazanie, że przydzielona węzłowi wartość h_{jmin} nie zawsze jest wartością wystarczającą, aby w systemie dotrzymane zostały ograniczenia RT, gdy wiadomości szeregowane są przy użyciu metody DPGS.

Zostanie rozważony system, w którym czas $TTRT = 1.5$ [ms]. Przyjęte parametry czasowe wiadomości w węźle 1 po transformacji okresów występowania przedstawia tabela 2.

Tabela 2

Parametry czasowe wiadomości

Message time parameters

Wiadomość okresowa	Okres występowania dp_i [ms]	Czas przesłania cp_i [ms]	Wskaźnik wielokrotności k_i
pm_1	1.5	0.2	1
pm_2	3	0.3	2
pm_3	3	0.4	2
pm_4	3	0.3	2

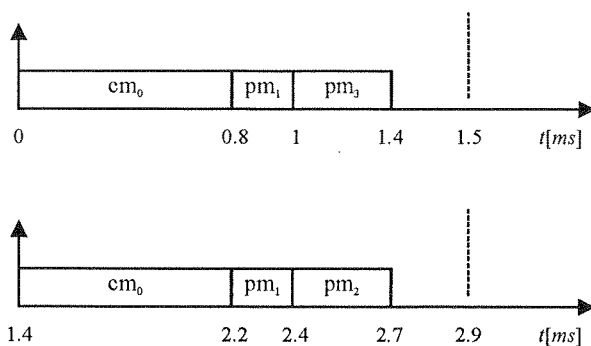
Wartość czasu h_{1min} wynosi:

$$h_{1min} = cp_1 + \frac{cp_2 + cp_3 + cp_4}{2} = 0.7 [\text{ms}]$$

Czas do dyspozycji pozostałych węzłów cm_0 wyniesie

$$cm_0 = TTRT - h_{1min} = 0.8 [\text{ms}]$$

Przebiegi czasowe w węźle przedstawia rys. 4. Analiza przeprowadzona została dla najgorszego przypadku (w węźle aktywowane zostały wszystkie wiadomości, a zarazem utracił on prawo do wysyłania wiadomości). Wiadomość pm_4 nie została wysłana, bowiem czas przesłania tej wiadomości wynosi 0.3 [ms], a czas do wyłącznej dyspozycji przez węzeł wynosi 0.2 [ms] ($2.9 - 2.7 = 0.2$ [ms]). Tak więc, dla wiadomości pm_4 nie zostaną dotrzymane ograniczenia czasowe. Można zauważyć, że w powyższym przykładzie dla przyjętych parametrów systemu węzeł nie wykorzystał całego przyznanego



Rys. 4. Przebiegi czasowe w węźle 1

Fig. 4. Time behaviour for a node 1

czasu w związku z koniecznością wcześniejszego przekazania prawa do nadawania dla kolejnego węzła. Dlatego też należy wprowadzić pojęcie czasu wystarczającego dla każdego węzła h_{jwyst} , który to czas zapewni, że wszystkie wiadomości z węzła j będą wysłane przed swoimi ograniczeniami czasowymi. Wartość h_{jwyst} zostanie obliczona ze wzoru:

$$h_{jwyst} = \frac{1}{\alpha} \left(\left(\max \left\{ \left(\sum_{i=1}^{n_{ij}} bm_{1i} + n_{1j} * \beta \right), bw_{1max} \right\} \right) + \sum_{i=2}^{lg_j} \max \left\{ \left(\frac{\sum_{s=1}^{n_{ij}} bm_{is} + n_{ij} * \beta}{k_i} \right), bw_{imax} \right\} \right) \quad (9)$$

gdzie:

$$bw_{imax} = \sum_{l=1}^{\left\lceil \frac{n_{ij}}{k_i} \right\rceil} (bw_{ilmax} + \beta)$$

bm_{ilmax} – długość kolejnej wiadomości m_l z grupy GR_i o największej liczbie bajtów

Różnica w porównaniu ze wzorem (8) polega na tym, że dla każdej grupy wiadomości przydzielana jest większa wartość czasu spośród dwu wartości: wartości h_{jmin} wyznaczonej we wzorze (8) i wartości bw_{imax} . Wartość bw_{imax} reprezentuje wymagany czas, jaki potrzebny jest do dyspozycji danej grupy w jednym cyklu obiegu znacznika.

Dla wybranej grupy GR_i w jednym cyklu obiegu znacznika musi być wysłanych $\left\lceil \frac{n_{ij}}{k_j} \right\rceil$ wiadomości, aby po k_i cyklach wysłanych zostało n_{ij} wiadomości, czyli wszystkie z danej grupy. Jeżeli w jednym cyklu obiegu znacznika zostanie zarezerwowany dla grupy GR_i przedział czasu o długości bw_{imax} , to przed upływem czasu $k_i * TTRT$ zostaną wysłane wszystkie wiadomości z grupy GR_i . Tak więc, jeżeli w cyklu obiegu znacznika dla węzła zarezerwowany zostanie czas h_{jwyst} , to wiadomości ze wszystkich

grup wysłane zostaną przed przekroczeniem ograniczeń. Dla wyznaczonych wartości h_{jwyst} należy sprawdzić spełnienie warunku (7). Jeżeli warunek (7) jest spełniony (każdemu węzłowi można przydzielić w cyklu obiegu znacznika czas h_{jwyst}), to w systemie dotrzymane zostaną ograniczenia czasowe. Jeżeli warunek (7) nie zostanie dotrzymany tzn.

$$\sum_{j=1}^m h_{jwyst} > TTRT - w_T$$

to należy dokonać dalszych sprawdzeń w oparciu o metodę DPGS_II.

Metoda DPGS_II

Każdemu węzłowi należy przydzielić dopuszczalny czas, w którym może on wyłącznie dysponować magistralą. Czas ten, dla węzła j oznaczony jest przez h_{jdop} i obliczany jest wg. wzoru:

$$h_{jdop} = \frac{u_j}{u} (TTRT - w_T) \quad (10)$$

gdzie: $u = u_1 + \dots + u_n$, n jest liczbą węzłów w systemie.

Wielkość u_j jest stopniem wykorzystania sieci przez węzeł j :

$$u_j = \sum_{i=1}^{lg_j} \frac{\left(\sum_{s=1}^{n_{ij}} b m_{is} + n_{ij} * \beta \right)}{k_i} \quad (11)$$

System będzie spełniał wymagania RT, jeżeli każdy węzeł będzie je spełniał. Dlatego analiza systemu zostanie przeprowadzona dla każdego węzła osobno. Aby sprawdzić, czy obliczona wartość h_{jdop} zapewnia spełnienie ograniczeń czasowych dla wszystkich wiadomości i dla każdego dowolnego układu wiadomości w węźle, należy dokonać sprawdzenia dla najgorszego przypadku:

1. Należy założyć, że węzeł utracił prawo do nadawania, a jednocześnie zostały mu przekazane wszystkie wiadomości do wysłania.
2. Należy dokonać sprawdzenia, czy dla takiego układu wiadomości, dotrzymane zostaną ograniczenia czasowe w czasie t_{dyn_wyst} . Wartość t_{dyn_wyst} wyznaczana jest ze wzoru:

$$t_{\{dyn_wyst\}} = \left\lceil \frac{NWW}{TTRT} \right\rceil TTRT \quad (12)$$

gdzie: NWW jest najmniejszą wspólną wielokrotnością okresów występowania wiadomości w węźle.

Dla metody GRMS sprawdzenie spełnienia ograniczeń czasowych sprowadzało się do określenia, czy wiadomość dotrzyma pierwszego ograniczenia w sytuacji, gdy aktywowane zostały wszystkie wiadomości o wyższych priorytetach. Jest to słuszne dla metod szeregowania ze statycznym przydziałem priorytetu — przy każdej aktywacji wiadomość uzyskuje ten sam priorytet. Nie można natomiast zastosować tego warunku w przypadku metod z dynamicznym przydziałem priorytetu. W zależności bowiem od

bieżącego czasu (numeru cyklu obiegu znacznika) grupy wiadomości uzyskują różny priorytet. Należy więc dokonać sprawdzenia dotrzymania ograniczeń dla czasu $t_{dyn_{wyst}}$. Dla tego okresu czasu wystąpią bowiem wszystkie możliwe przypadki przydzielenia różnych priorytetów dla grup wiadomości. Sprawdzenie dotrzymania ograniczeń odbywa się w następujący sposób:

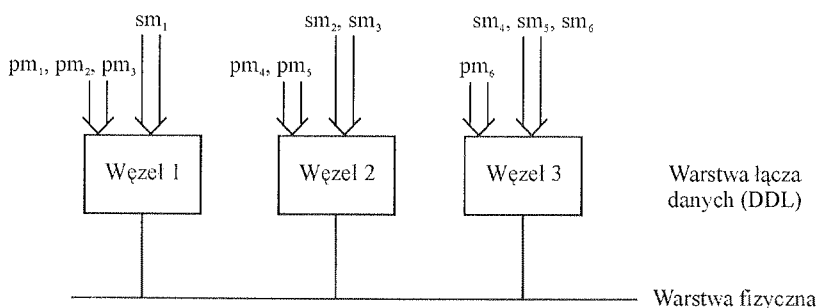
- dla każdej grupy wiadomości wyznaczamy jej priorytet dynamiczny zgodnie z zależnością (6),
- jeżeli grupy uzyskają ten sam priorytet dynamiczny, to wybierana jest grupa z mniejszą wartością indeksu,
- w obrębie wybranej grupy następuje zwiększenie czasu o czas przesłania wiadomości,
- wiadomości z grupyo niższym priorytecie mogą być wysyłane wówczas, gdy nie ma gotowych wiadomości do wysłania w grupie z wyższym priorytetem.
- po każdorazowym wysłaniu wiadomości program szeregujący sprawdza, czy są gotowe wiadomości do wysłania począwszy od grupy z najwyższym priorytetem (wiadomości mają różne okresy występowania),
- zakładając, że węzeł dysponuje czasem h_{jdp} , należy sprawdzić, czy wysyłane wiadomości dotrzymają swoich ograniczeń czasowych.

Sprawdzenie dotrzymania wg powyższego sposobu przeprowadzane jest dla najgorszego przypadku. Jeżeli powyższe ograniczenia zostaną dotrzymane, to zostaną one dotrzymane dla każdego dowolnego przypadku wystąpienia wiadomości w węźle.

Przykład 3

Celem przykładu jest przedstawienie sposobu analizy czasowej dla systemu rozproszonego z magistralą miejscową, w węzłach którego wiadomości szeregowane są przy użyciu metody DPGS.

Rozważony zostanie system sterujący z magistralą PROFIBUS FMS. Dla zadanego modelu strumienia wiadomości przyjęto system złożony z trzech węzłów. Sposób przydziału wiadomości okresowych PM i nieokresowych SM ilustruje rys. 5. Przyjęto następujące parametry wysyłanych wiadomości:



Rys. 5. Przykładowy model wiadomości

Fig. 5. A message stream model

• ok
• og
• ma
4)
• og
• dlu

Protok
go zna
ponize

Ramka
SYN

Ramka
SYN

Ramka
SYN

Poszcz
SYN
SD1
SD2
LE
LEr
DA
SA
FC
DATA
FCS
ED

Każd
Pr
R
C
ników
SD5):
 α
 β
R
ramki

- okres występowania wiadomości periodycznych $TP = \{1.6, 3.2, 6.4, 1.6, 3.2, 3.2\}$ [ms],
- ograniczenia czasowe wiadomości periodycznych $DP = TP$,
- maksymalny okres występowania wiadomości sporadycznych $TS = \{4, 4, 4, 4, 4, 4\}$ [ms],
- ograniczenia czasowe wiadomości sporadycznych $DS = TS$,
- długość wiadomości: $BP = \{3, 5, 8, 4, 2, 4\}$ [bajt],
 $BS = \{9, 5, 3, 5, 5, 8\}$ [bajt].

Protokół dostępu do medium fizycznego PROFIBUS FMS stosuje metodę krążącego znacznika. Postacie ramek znacznika, wiadomości i potwierdzenia przedstawiono poniżej.

Ramka znacznika SD4

SYN	SD4	DA	S.A.
-----	-----	----	------

Ramka wiadomości SD2

SYN	SD2	LE	Ler	DA	SA	FC	DATA_UNIT	FCS	ED
-----	-----	----	-----	----	----	----	-----------	-----	----

Ramka potwierdzenia SD5

SYN	S.C.
-----	------

Poszczególne pola mają następujące znaczenie:

SYN	bity synchronizacji, przynajmniej 33 bity
SD1	znacznik początku, kod = 10H, ramka pytania stacji master
SD2	znacznik początku, kod = 68H, ramka przesyłanej zmiennej
LE	długość, 4 do 249 bajtów
Ler	długość zwrotna
DA	adres przeznaczenia
SA	adres źródłowy
FC	znak sterujący
DATA_UNIT	jednostka danych, długość 1 do 246 bajtów
FCS	suma kontrolna
ED	znacznik końca

Każdy znak składa się z 11 bitów: 8 bitów danych i 3 bitów technicznych

Przyjęto szybkość transmisji (dla przejrzystości obliczeń) równą 1 [Mb/s].

Rozwiązanie

Czas przesyłania wiadomości będzie obliczany ze wzoru (4). Wartości współczynników α i β wynoszą odpowiednio (współczynnik β uwzględnia ramkę potwierdzenia SD5):

$$\alpha = 1000[Kb/sek] = 1000[b/ms]$$

$$\beta = 2 * SYN + SD2 + LE + Ler + DA + SA + FC + FCS + ED + SC = 165[b]$$

Ramka znacznika będzie przekazywana jako wiadomość typu SD4. Czas trwania ramki znacznika c_{TOKEN} wyznaczany jest ze wzoru:

$$c_{TOKEN} = \frac{SYN + (SD1 + DA + SA) * 11}{1000} = \frac{33 + 3 * 11}{1000} = 0,066 [ms]$$

Straty czasowe na przekazywanie znacznika przez trzy stacje w jednym cyklu wyniosą $w_T = 3 * 0.66 = 0.198$ [ms]. Ramka znacznika nie zawiera użytecznej informacji, ale jest niezbędna do prawidłowego funkcjonowania systemu.

Transformacja ograniczeń czasowych wiadomości

Pierwszym krokiem jest wyznaczenie czasu TTRT, który powinien spełniać zależność $TTRT \leq dm_{min}$. Minimalny ograniczenie czasowe wiadomości w systemie $dm_{min} = 1.6$ [ms]. W przykładzie arbitralnie przyjęto, że $TTRT = 1.6$ [ms].

Dokonując transformacji ograniczeń czasowych wiadomości otrzymujemy 3 grupy wiadomości (tab. 3):

- GR_1 , ograniczenie czasowe grupy GR_1 $dm_1^* = TTRT = 1.6$ [ms], współczynnik wielokrotności $k_1 = 1$
- GR_2 , ograniczenie czasowe grupy GR_2 $dm_2^* = 2 * TTRT = 3.2$ [ms], współczynnik wielokrotności $k_2 = 2$,
- GR_4 , ograniczenie czasowe grupy GR_4 $dm_4^* = 4 * TTRT = 6.4$ [ms], współczynnik wielokrotności $k_4 = 4$.

Tabela 3

Parametry grup wiadomości po transformacji okresów występowania

Properties of group messages after arrival period transformation

Numer węzła	Liczba wiadomości w węźle j	Nazwa grupy w węźle i	Liczba wiadomości w grupie ni _j	Nazwy wiadomości ftm	Nazwy wiadomości w grupie Gr _i
1	4	GR ₁	1	pm ₁	pm ₁₁
		GR ₂	2	pm ₂ , sm ₁	pm ₂₁ , sm ₂₂
		GR ₄	1	pm ₃	pm ₄₁
2	4	GR ₁	1	pm ₄	pm ₁₁
		GR ₂	3	pm ₅ , sm ₂	pm ₂₁ , sm ₂₂
		GR ₄	0	sm ₃	sm ₂₃
3	4	GR ₁	0	pm ₆	pm ₁₁
		GR ₂	4	sm ₄ , sm ₅	pm ₂₁ , sm ₂₃
		GR ₄	0	sm ₆	sm ₂₄

Przydział czasu do wyłącznej dyspozycji przez węzeł

Przydzielenie czasu h_{jwyst} do wyłącznej dyspozycji przez węzeł odbywa się wg wzoru (9). W poniższych wzorach długości wiadomości przedstawione zostały w bitach przy uwzględnieniu faktu, że każdy znak składa się z 11 bitów: 8 bitów danych i 3 bitów technicznych. Symbole bp_{ij}^* i bs_{ij}^* oznaczają odpowiednio:

$$bp_{ij}^* = bp_{ij} * 11[b]$$

$$bs_{ij}^* = bs_{ij} * 11[b]$$

klę wy-
formacji,

iać za-
systemie

3 grupy

czynnik

czynnik

czynnik

Tabela 3

wy-
mości
ie Gr_i

11
sm₂₂
41

11
sm₂₂
23

11
sm₂₃
24

się wg.
y bitach
ych i 3

$$h_{1wyst} = \frac{1}{\alpha} \left[(bp_{11}^* + 1^* \beta) + \max \left\{ \left\lceil \frac{(bp_{21}^* + bs_{22}^*) + 2^* \beta}{2} \right\rceil, bs_{21}^* + \beta \right\} + \max \left\{ \left\lceil \frac{bs_{41}^* + 1^* \beta}{4} \right\rceil, bp_{41}^* + \beta \right\} \right]$$

$$\begin{aligned} h_{1wyst} &= \frac{1}{1000} [(33 + 165) + \max \{242, 264\} + \max \{64, 253\}] \\ &= \frac{1}{1000} [198 + 264 + 253] = 0,715 [ms] \end{aligned}$$

$$h_{2wyst} = \frac{1}{\alpha} \left[(bp_{11}^* + 1^* \beta) + \max \left\{ \left\lceil \frac{(bp_{21}^* + bs_{22}^* + bs_{23}^*) + 3^* \beta}{2} \right\rceil, (bs_{22}^* + bp_{11}^*) + 2\beta \right\} \right]$$

$$h_{2wyst} = \frac{1}{1000} [(44 + 165) + \max \{303, 418\}] = \frac{1}{1000} [209 + 418] = 0,627 [ms]$$

$$h_{3wyst} = \frac{1}{\alpha} \left[\max \left\{ \left\lceil \frac{(bp_{21}^* + bs_{22}^* + bs_{23}^* + bs_{24}^*) + 4^* \beta}{2} \right\rceil, (bs_{24}^* + bp_{23}^*) + 2\beta \right\} \right]$$

$$h_{3wyst} = \frac{1}{1000} [\max \{451, 473\}] = \frac{1}{1000} [473] = 0,473 [ms]$$

Warunek czasu rzeczywistego przydziału czasów h_{jwyst} dla całej sieci

Sprawdzamy, czy spełniony jest warunek (7).

$$\begin{aligned} \sum_{j=1}^3 h_{jwyst} &= 0.715 + 0.627 + 0.473 = 1.815 [ms] > TTRT - w_T \\ &= 1.6 - 0.198 = 1.402 [ms] \end{aligned}$$

Ponieważ warunek (7) nie jest dotrzymany, to stosowana jest *Metoda DPGS-II*. Zgodnie z *Metoda DPGS-II* każdemu węzłowi należy przydzielić dopuszczalny czas, w którym może on wyłącznie dysponować magistralą. Czas ten, dla węzła j oznaczony jest przez h_{jdop} i obliczany jest wg. wzoru (10). Aby wyznaczyć czasy h_{jdop} należy wyznaczyć wartości u_j , czyli stopień wykorzystania sieci przez stację j , ze wzoru (11). Dla zadanego przydziału wiadomości do węzłów, stopień wykorzystania magistrali przez poszczególne węzły wynosi odpowiednio:

$$u_1 = (bp_{11}^* + 1^* \beta) + \frac{(bp_{21}^* + bs_{22}^*) + 2^* \beta}{2} + \frac{bp_{41}^* + 1^* \beta}{4} = 504$$

$$u_2 = bp_{11}^* + 1 * \beta + \frac{(bp_{21}^* + bs_{22}^* + bs_{23}^*) + 3 * \beta}{2} = 512$$

$$u_3 = \frac{(bp_{21}^* + bs_{22}^* + bs_{23}^* + bs_{24}^*) + 4 * \beta}{2} = 451$$

$$h_{1dop} = \frac{u_1}{u_1 + u_2 + u_3} (TTRT - w_T) = 0.481 [ms]$$

$$h_{2dop} = 0.489 [ms]$$

$$h_{3dop} = 0.432 [ms]$$

Do zbioru zadań w węzłach 1, 2, 3 należy dołączyć wiadomości m_{10} , m_{20} , m_{30} reprezentujące obieg znacznika. Parametry czasowe tych wiadomości przedstawia tabela 4.

Tabela 4

Parametry czasowe wiadomości m_{i0} dla poszczególnych węzłów

Time properties of m_{i0} messages for each node

Numer węzła i	Czas przesłania wiadomości m_{i0}	Okres występowania wiadomości t_{i0}
1	$cm_{10} = TTRT - h_{1dop} = 1.6 - 0.481 = 1.119ms$	$t_{10} = TTRT = 1.6 ms$
2	$cm_{20} = TTRT - h_{2dop} = 1.6 - 0.489 = 1.111ms$	$t_{20} = TTRT = 1.6 ms$
3	$cm_{30} = TTRT - h_{3dop} = 1.6 - 0.432 = 1.168ms$	$t_{30} = TTRT = 1.6 ms$

W przypadku magistrali PROFIBUS FMS wiadomość priorytetowa może być wysłana, pomimo, że spowoduje to przekroczenie limitu czasu jaki został przyznany węzłowi. W rozważanym przykładzie zastosowano rozwiązanie, w którym program szeregujący na bieżąco oblicza zależności czasowe w systemie i w przypadku stwierdzenia, że wysłanie wiadomości spowodowałoby przekroczenie limitu czasowego węzła, wiadomość nie zostanie przekazana do warstwy aplikacji magistrali miejscowej. Tym samym żaden węzeł nie przekroczy przydzielonego mu czasu h_{jmax} . Podobnie jak w przypadku omawiania metody GRMS, wartości czasu t_{i0} dla zadań m_{i0} reprezentują wartości maksymalne. System będzie spełniał wymagania RT, jeżeli każdy węzeł będzie je spełniał. Dlatego analiza systemu zostanie przeprowadzona dla każdego węzła osobno. Należy jeszcze wyznaczyć czas przesyłania wiadomości ze wzoru (4). Czasy przesyłania wyznaczane są więc ze wzorów:

$$cp_i = \frac{2 * SYN + (bp_i + SD2 + LE + LER + DA + SA + FC + FCS + ED + SC) * 11}{1000} [ms]$$

$$cs_i = \frac{2 * SYN + (bs_i + SD2 + LE + LER + DA + SA + FC + FCS + ED + SC) * 11}{1000} [ms]$$

$$cp_1 = \frac{66 + 12 * 11}{1000} = 0,198[ms] \quad cs_1 = 0.264[ms]$$

$$cp_2 = 0.220[ms] \quad cs_2 = 0.220 [ms]$$

$$cp_3 = 0.253 [ms] \quad cs_3 = 0.198 [ms]$$

$$cp_4 = 0.209 [ms] \quad cs_4 = 0.220 [ms]$$

$$cp_5 = 0.187 [ms] \quad cs_5 = 0.220 [ms]$$

$$cp_6 = 0.209 [ms] \quad cs_6 = 0.253 [ms]$$

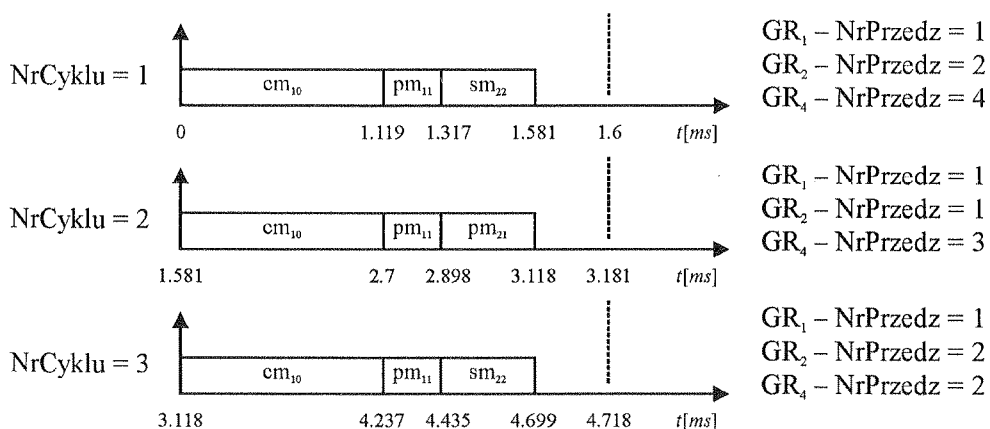
Analiza dla węzła 1

Analiza zostanie przeprowadzona przy następujących założeniach:

- węzeł utracił prawo do nadawania, a jednocześnie zostały mu przekazane wszystkie wiadomości do wysłania,
- należy także założyć, że wiadomości są poszeregowane od najdłuższej do najkrótszej, bowiem w takim przypadku występuje najczęściej problem niewykorzystania przez węzeł przyznanego mu czasu.

W węźle 1 czas trwania dodatkowej wiadomości m_0 (obrazującej zajęcie magistrali przez węzły 2, 3 i dane dodatkowe protokołu) będzie wynosić $cm_{10} = 1.119 [ms]$. Dla rozważanego modelu wiadomości wartość NWW wynosi 20. Przebiegi czasowe dla węzła 1, dla pierwszych 3 cykli obiegu znacznika, przedstawia rys. 6. Na rysunku tym podano także wartości zmiennej $NrPrzedz[i]$ dla poszczególnych grup wiadomości. Grupa, dla której zmienna $NrPrzedz[i]$ przyjmuje najmniejszą wartość, otrzymuje najwyższy priorytet. Jeżeli grupy uzyskają ten sam priorytet dynamiczny, to wybierana jest grupa z mniejszą wartością indeksu.

Z rys. 6 wynika, że dla wiadomości pm_{41} nie zostaną dotrzymane ograniczenia czasowe (w cyklu 4 zostaną wysłane znowu wiadomości pm_{11} i pm_{21}). Dla przydzielonych czasów $h_{j dop}$ nie jest możliwe wysłanie wiadomości pm_{41} (bez przekraczania czasów



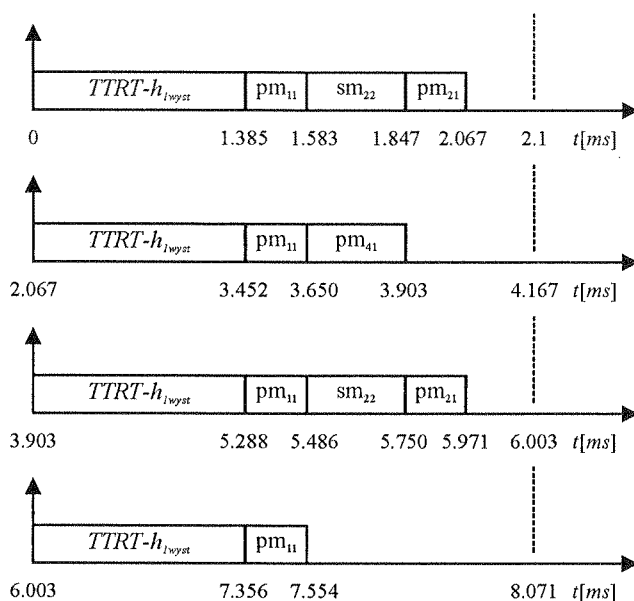
Rys. 6. Przebiegi czasowe w węźle 1

Fig. 6. Time behaviour for a node 1

$h_{j dop}$). Przeprowadzając podobną analizę dostajemy, że w węźle 2 ograniczenia nie zostaną dotrzymane dla wiadomości pm_{21} , a w węźle 3 dla wiadomości pm_{21} i sm_{22} . Wynika stąd wniosek, że system nie będzie spełniał ograniczeń RT. Dla przykładu przyjmijmy nowe wartości okresów występowania wiadomości. Niech nowe okresy wynoszą: $TP = \{2.1, 4.2, 8.4, 2.1, 4.2, 4.2\}$ [ms]. Postępując analogicznie jak poprzednio, otrzymamy 4 grupy wiadomości ograniczeniu bazowym $TTRT = 2.1$ [ms]. Czasy h_{jwyst} nie ulegną zmianie. Dokonajmy sprawdzenia warunku (7) dla nowego okresu $TTRT$.

$$\begin{aligned} \sum_{j=1}^3 h_{jwyst} &= 0.715 + 0.627 + 0.473 \\ &= 1.815 \text{ [ms]} < TTRT - w_T \\ &= 2.1 - 0.198 = 1.902 \text{ [ms]} \end{aligned}$$

Ponieważ powyższy warunek jest spełniony, to w systemie zostaną dotrzymane ograniczenia czasowe dla wszystkich wiadomości. Na rysunkach 7, 8 i 9 przedstawiono przebiegi czasowe dla poszczególnych węzłów dla najgorszego przypadku (dla kilku pierwszych cykli obiegu znacznika).



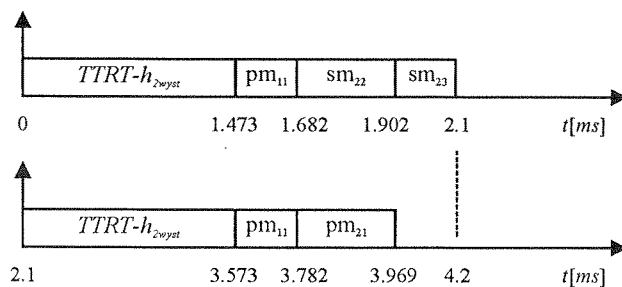
Rys. 7. Przebiegi czasowe w węźle 1

Fig. 7. Time behaviour for a node 1

Analiza powyższych przebiegów czasowych potwierdza, że w systemie spełnione są ograniczenia RT dla wszystkich wiadomości.

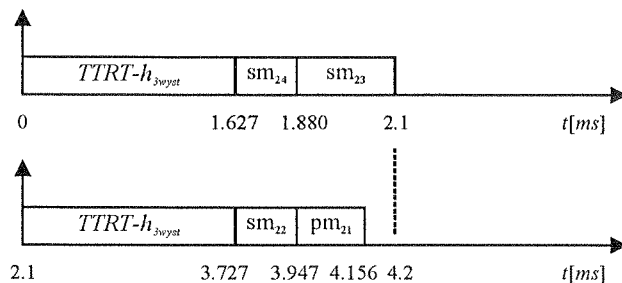
7.
O
DPGS
czasow
ru (6)
szczeg
grupy.
wartos
dotrzy
soki),
priory
wiadom
wolny
Opisan
przydz

nia nie
i sm_{22} .
zykładu
okresy
jak po-
1 [ms].
nowego



Rys. 8. Przebiegi czasowe w węźle 2

Fig. 8. Time behaviour for a node 2



Rys. 9. Przebiegi czasowe w węźle 3

Fig. 9. Time behaviour for a node 3

7. PROPOZYCJA METODY Z MIESZANYM SPOSOBEM PRZYDZIAŁU PRIORYTETU DLA GRUP WIADOMOŚCI

Opracowana metoda z mieszanym przydziałem priorytetu oparta została o metodę *DPGS*, gdyż dokonywany jest podział na grupy wiadomości o tym samym ograniczeniu czasowym, a priorytet dynamiczny grupy wiadomości GR_i wyznaczany jest ze wzoru (6). Różnica polega na wprowadzeniu dodatkowego, statycznego priorytetu dla poszczególnych grup wiadomości. Priorytet ten nazwany zostanie *priorytetem krytycznym grupy*. Priorytet ten jest ważniejszy od priorytetu przyporządkowanego na podstawie wartości zmiennej $NrPrzedz[i]$. Grupy wiadomości, dla których zawsze muszą być dotrzymane ograniczenia czasowe, otrzymują *priorytet krytyczny grupy PG_high* (wysoki), a pozostałe grupy *priorytet krytyczny grupy PG_low* (niski). Istota przydzielenia priorytetu krytycznego grupy polega na tym, że w pierwszej kolejności wysyłane są wiadomości z grup z przydzielonym priorytetem *PG_high*, a dopiero potem, w miarę wolnych zasobów czasowych, wiadomości z grup z przydzielonym priorytetem *PG_low*. Opisany powyżej sposób przydziału priorytetów należy do grupy metod z mieszanym przydziałem priorytetów i nazwany zostanie *Mixed Priority Group Scheduling (MPGS)*.

elnione

Przykład 4

Celem przykładu jest pokazanie różnicy pomiędzy szeregowaniem wiadomości przy użyciu metod *DGPS* i *MPGS*.

Założono, że w systemie po transformacji ograniczeń czasowych w węźle i systemu rozproszonego utworzone zostały 4 grupy wiadomości: GR_1 , GR_2 , GR_3 , i GR_4 , którym odpowiadają ograniczenia czasosowe następujących wartościach: $dm_1^* = TTRT$, $dm_2^* = 2 * TTRT$, $dm_3^* = 3 * TTRT$, $dm_4^* = 4 * TTRT$. Grupom GR_1 , GR_2 i GR_3 przyporządkowano priorytet krytyczny *PG.high*, a grupie GR_4 priorytet *PG.Low*. Tabele 5 i 6 przedstawiają sposób przydziału priorytetów odpowiednio dla metod *DGPS* i *MPGS*. Dla kolejnych cykli obiegu znacznika w tabelach przedstawiono wartość zmiennej $NrPrzedz[i]$ dla każdej grupy oraz uszeregowanie grup wg malejącego priorytetu.

Tabela 5

Przydział priorytetów dla grup wiadomości dla metody DGPS

Priorities assigning for message groups — DGPS method

Nazwa grupy	1 obieg znacznika		2 obieg znacznika		3 obieg znacznika		4 obieg znacznika	
	$NrPrzedz[i]$	Uszeregowanie	$NrPrzedz[i]$	Uszeregowanie	$NrPrzedz[i]$	Uszeregowanie	$NrPrzedz[i]$	Uszeregowanie
GR_1	1	GR_1	1	GR_1	1	GR_1	1	GR_1
GR_2	2	GR_2	1	GR_2	2	GR_3	1	GR_2
GR_3	3	GR_3	2	GR_3	1	GR_2	2	GR_4
GR_4	4	GR_4	3	GR_4	2	GR_4	1	GR_3

Tabela 6

Przydział priorytetów dla grup wiadomości dla metody MPGS

Priorities assigning for message groups — MPGS method

Nazwa grupy	Priorytet krytyczny	1 obieg znacznika		2 obieg znacznika		3 obieg znacznika		4 obieg znacznika	
		$NrPrzedz[i]$	Uszeregowanie	$NrPrzedz[i]$	Uszeregowanie	$NrPrzedz[i]$	Uszeregowanie	$NrPrzedz[i]$	Uszeregowanie
GR_1	<i>PG.high</i>	1	GR_1	1	GR_1	1	GR_1	1	GR_1
GR_2	<i>PG.high</i>	2	GR_2	1	GR_2	2	GR_3	1	GR_2
GR_3	<i>PG.high</i>	3	GR_3	2	GR_3	1	GR_2	2	GR_3
GR_4	<i>PG.low</i>	4	GR_4	3	GR_4	2	GR_4	1	GR_4

Jak wynika z tabel 5 i 6 dla 4 obiegu znacznika występuje różnica pomiędzy szeregowaniem wiadomości w oparciu o metodę *DGPS* i *MPGS*. W przypadku szeregowania metodą *MPGS* wiadomości z grupy GR_3 wysyłane są przed wiadomościami z grupy GR_4 pomimo iż na podstawie wartości zmiennej $NrPrzedz[i]$ grupa GR_4 uzyskuje wyższy priorytet. Jednakże, grupa GR_3 ma przydzielony priorytet krytyczny grupy *PG.high* i dlatego wiadomości z tej grupy wysyłane są przed wiadomościami z grupy GR_4 . Jeżeli w systemie zostaną dotrzymane ograniczenia dla wszystkich wiadomości, to różnica ta nie ma znaczenia. Natomiast w przypadku wystąpienia przeciążeń w sys-

temie w pierwszej kolejności dotrzymane zostaną ograniczenia dla wiadomości z grup z przydzielonym priorytetem krytycznym *PG_high*.

Dla metody *MPGS* można dodatkowo przewidzieć przydział *priorytetu użytkownika*. Priorytet użytkownika przydzielany byłby do poszczególnych wiadomości, nie grup wiadomości. Odgrywałby on rolę wówczas, gdy w wybranej grupie występowałyby wiadomości tej samej długości. W pierwsze kolejności wysyłane byłyby wiadomości z wyższym *priorytetem użytkownika*.

7.1. ANALIZA CZASOWA SYSTEMU ROZPROSZONEGO DLA METODY *MPGS*

W dokonywanych analizach zostało założone, że w węzłach systemu rozproszonego stosowany jest algorytm szeregowania wiadomości oparty metodą *MPGS*. Magistrala miejscowa może w warstwie *MAC* stosować metodę: odpytań, przekazywanie znacznika, *TDMA*. Proponowana metoda *MPGS* składa się z dwóch części sprawdzania dotrzymania ograniczeń *RT*. Metoda *MPGS_I* musi być wykonana zawsze, ponieważ wartości wyznaczone przy jej użyciu są wykorzystywane w metodzie *MPGS_II*. Poza tym, metoda *MPGS_I* jest metodą mniej czasochłonną. Sprawdzanie dotrzymania ograniczeń czasowych w systemie przy użyciu metody *MPGS_II* jest wykonywane wtedy, kiedy metoda *MPGS_I* nie da pożądanego efektów.

Metoda *MPGS_I* (podstawowa) obejmuje etapy opisane poniżej:

- **Etap I. Przydzielenie priorytetu krytycznego grupy**
Wszystkim grupom wiadomości w systemie zostaje przydzielony priorytet krytyczny grupy *PG_high*.
- **Etap II. Wyznaczenie czasu potrzebnego do nadawania wiadomości**
Czas przesłania wiadomości wyznaczany jest ze wzoru (4).
- **Etap III. Transformacja ograniczeń czasowych wiadomości**
Transformacja i podział na grupy wykonywane są jak opisano w rozdz. 5.1.
- **Etap IV. Przydział czasu h_{jwyst} do wyłącznej dyspozycji przez węzeł**
Wartość czasu h_{jwyst} , w którym węzeł j wyłącznie dysponuje magistralą, wyznaczana jest ze wzoru (9).
- **Etap V. Sprawdzenie warunku czasu rzeczywistego przydziału czasów h_{jwyst} dla całej sieci**
Jeżeli spełniony jest warunek (7), to w systemie zostaną dotrzymane ograniczenia czasowe dla wszystkich wiadomości.
- **Etap VI. Modyfikacja przydziału priorytetów**
W celu wykorzystania zalet metody należy odpowiednio zmodyfikować priorytet krytyczny *PG_high* przydzielony wszystkim grupom wiadomości. Priorytet *PG_high* powinien pozostać tylko przy tych grupach wiadomości, w których są wiadomości, dla których dotrzymanie ograniczeń czasowych w czasie przeciążeń systemu jest najistotniejsze. Pozostałym grupom wiadomości należy przydzielić priorytet *PG_low*. W czasie normalnej pracy systemu ograniczenia czasowe będą dotrzymywane dla wszystkich wiadomości (bo spełniony jest warunek 7), a podczas

wystąpienia chwilowych przeciążeń w pierwszej kolejności będą dotrzymywane ograniczenia dla wiadomości z zestawu krytycznego, czyli dla grup z przydzielonym priorytetem krytycznym PG_high . Należy także pamiętać o ewentualnym nadaniu wiadomościom priorytetów użytkownika.

Jeżeli warunek (7) nie zostanie dotrzymany, to należy dokonać dalszych sprawdzeń w oparciu o metodę $MPGS_II$.

Metoda $MPGS_II$ Każdemu węzłowi należy przydzielić maksymalny czas, w którym może on wyłącznie dysponować magistralą. Czas ten, dla węzła j oznaczony jest przez $h_{j dop}$ i obliczany jest na podstawie wzoru (10). System będzie spełniał wymagania RT, jeżeli każdy węzeł będzie je spełniał. Dlatego analiza systemu zostanie przeprowadzona dla każdego węzła osobno. Aby sprawdzić, czy obliczona wartość $h_{j dop}$ zapewnia spełnienie ograniczeń czasowych dla wszystkich wiadomości i dla każdego dowolnego układu wiadomości w węzle, należy dokonać sprawdzenia dla najgorszego przypadku:

1. Należy założyć, że węzeł utracił prawo do nadawania, a jednocześnie zostały mu przekazane wszystkie wiadomości do wysłania.
2. Należy dokonać sprawdzenia, czy dla takiego układu wiadomości, dotrzymane zostaną ograniczenia czasowe w czasie $t_{dyn, wysł}$ wyznaczanym ze wzoru (12). Sprawdzenie to odbywa się w następujący sposób:
 - Wszystkie grupy otrzymują priorytet krytyczny PG_high ,
 - dla każdej grupy wiadomości wyznaczamy jej priorytet dynamiczny zgodnie z zależnością (6),
 - jeżeli grupy uzyskają ten sam priorytet dynamiczny, to wybierana jest grupa z mniejszą wartością indeksu,
 - w obrębie wybranej grupy następuje wysyłanie wiadomości,
 - jeżeli wiadomościom przydzielone zostały priorytety użytkownika, wysyłana jest wiadomość z wyższym priorytetem użytkownika,
 - jeżeli brak jest gotowych wiadomości do wysłania z grup z wyższym prioryte-tem dynamicznym, wysyłane są wiadomości z grup z przydzielonym prioryte-tem niższym,
 - po każdorazowym wysłaniu wiadomości program szeregujący sprawdza, czy są gotowe wiadomości do wysłania poczynawszy od grupy z najwyższym priorytetem (wiadomości mają różne okresy występowania),
 - zakładając, że węzeł dysponuje czasem $h_{j dop}$, należy sprawdzić, czy wysyłanie wiadomości dotrzymają swoich ograniczeń czasowych.

Sprawdzenie dotrzymywania wg powyższego sposobu przeprowadzane jest dla najgor-szego przypadku. Jeżeli ograniczenia zostaną dotrzymane to zostaną one dotrzymane dla każdego dowolnego układu wystąpienia wiadomości w węzle. W celu wykorzy-stania zalet metody należy odpowiednio zmodyfikować priorytet krytyczny PG_high przydzielony wszystkim grupom wiadomości. Priorytet PG_high powinien pozostać tylko przy tych grupach wiadomości, dla których dotrzymanie ograniczeń czasowych w czasie przeciążeń systemu jest najistotniejsze. Pozostałym grupom wiadomości na-leży przydzielić priorytet PG_low .

V
GRM
stawi
• w
• lic
• lic
• dl
• m
Na os
λ. Par

gdzie:

W
wiado
miar
magis
czynn
liczby
można
współ
czane
szybk
wyzna
• wi
• za
Zmian
danych
wszyst
wiado

8. ZASADY I KRYTERIA PRZEPROWADZONYCH BADAŃ EKSPERYMENTALNYCH

W celu zaprezentowania zalet i wad metody DPGS porównano ją z metodami GRMS [2] i EDF [1]. Przeprowadzone badania bazują na modelu wiadomości przedstawionym w rozdz. 5.3, przy następujących, dodatkowych założeniach:

- wszystkie wiadomości są wiadomościami okresowymi,
- liczba węzłów w systemie $2 \dots 10$,
- liczba wiadomości w węźle $1 \dots 10$,
- długość pola danych wiadomości $1 \dots 32$ bajtów,
- minimalna wartość, o jaką zmieniany był okres występowania wiadomości to 1 [ms].

Na osi Y charakterystyk występuje parametr zwany *maksymalnym obciążeniem systemu*

λ . Parametr λ obliczany jest ze wzoru:

$$\lambda = \sum_{i=1}^n \sum_{j=1}^w \frac{(bp_{ij} * 8) * t_bit_500}{tp_{ij}} \quad (13)$$

gdzie: n — liczba węzłów w systemie
 w — liczba wiadomości w węźle
 bp_{ij} — długość pola danych wiadomości [B]
 tp_{ij} — okres występowania wiadomości [ms]
 t_bit_500 — czas przesłania 1 bitu dla szybkości transmisji 500 Kb/s.

Współczynnik λ obliczany jest na podstawie liczby bajtów pola 'dane' przesyłanych wiadomości, a nie na podstawie liczby bajtów całej wiadomości (czyli bez danych nadmiarowych i narzutów protokołu). Dane nadmiarowe i narzut protokołu poszczególnych magistral są oczywiście uwzględniane przy analizie systemu, ale zdefiniowanie współczynnika λ jw. umożliwia lepsze porównanie właściwości magistral (dla takiej samej liczby bajtów całej wiadomości, magistralą z mniejszą liczbą bajtów nadmiarowych można przesłać więcej bajtów danych w takim samym okresie czasu, a tym samym współczynnik λ będzie większy). Obciążenie systemu dla wszystkich magistral wyznaczone jest dla szybkości przesyłu danych 500 Kb/s, ponieważ jest to wspólna wartość szybkości dla badanych magistral. Przedstawione w niniejszej pracy charakterystyki wyznaczane były przy następujących dodatkowych założeniach:

- wiadomości w poszczególnych węzłach mają te same parametry czasowe,
- zadana jest liczba węzłów i liczba wiadomości w węźle.

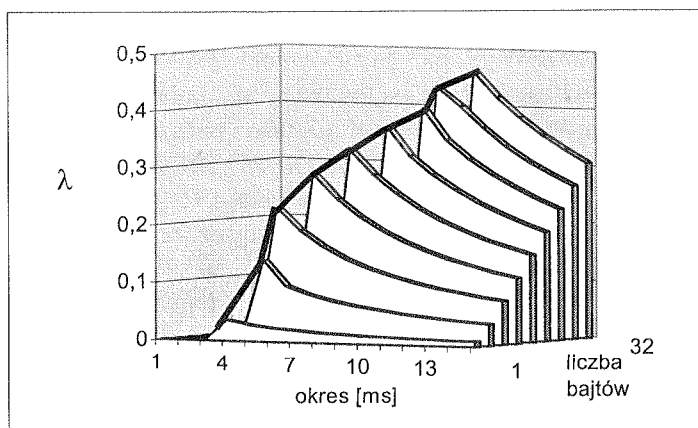
Zmianie ulegał okres występowania wiadomości i liczba przesyłanych bajtów pola danych wiadomości. Okresy występowania wiadomości w węźle były takie same dla wszystkich wiadomości lub tworzyły ciąg arytmetyczny. Jeżeli okresy występowania wiadomości w węźle były takie same, to wzór (13) przyjmie postać:

$$\lambda = n * w * \frac{(bp_r * 8) * t_bit_500}{tp_c} \quad (14)$$

Dla zadanej długości pola danych bp_r (jednakowej dla wszystkich wiadomości w systemie), poszukiwana była minimalna wartość okresu występowania tp_c (jednakowego dla wszystkich wiadomości), dla którego w systemie spełnione były jeszcze ograniczenia RT. Pierwszy punkt charakterystyki był wyznaczany dla sytuacji, kiedy w systemie spełnione są ograniczenia czasowe dla przesyłanych wiadomości zawierających 1 bajt danych. Ostatni punkt wykresu wyznaczany był dla takich parametrów systemu, kiedy spełnione są ograniczenia czasowe dla wiadomości zawierających 32 bajty danych. Jeżeli okresy występowania wiadomości w węźle tworzyły ciąg arytmetyczny, to λ wyznaczano ze wzoru:

$$\lambda = n * \sum_{j=1}^w \frac{(bp_r * 8) * t_{bit_500}}{j * tp_j} \quad (15)$$

Dla zadanej długości pola danych bp_r (jednakowej dla wszystkich wiadomości w systemie), poszukiwane były minimalne wartości okresów występowania tp_j tworzące ciąg arytmetyczny (jednakowy w każdym węźle), dla których w systemie spełnione były ograniczenia RT. Dla tego przypadku na osi X (odciętych) występują wartości odpowiadające najmniejszej wartości z ciągu. Przykładowo, jeżeli okresy występowania wiadomości tworzą ciąg arytmetyczny, a pierwsza wartość ciągu to 4 [ms], to wartość parametru λ zaznaczona jest na wykresie dla okresu równego 4 [ms]). Na rys. 10 przedstawiono charakterystykę trójwymiarową, która przedstawia sposób powstawania



Rys. 10. Sposób tworzenia charakterystyk typu pierwszego

Fig. 10. A method of creating charts of type one

opisywanych charakterystyk [10]. Poszczególne krzywe przedstawiają zmianę obciążenia systemu (oś Z) od zmian okresu występowania wiadomości (oś X) przy zachowaniu stałej liczby bajtów pola danych wiadomości (oś Y) - oczywiście rzecz dotyczy sytuacji, kiedy w systemie dotrzymane są ograniczenia RT. Aby nie zaciemniać rysunku przedstawione krzywe dotyczą przesłania tylko 1, 4, 8, 12, 16, 20, 24, 28 i 32 bajtów.

Ponieważ
temu dla
połączeni
liczbie p
malne ob

9. AN

Wyn
zależą o
modelu,
ległości
metody
grupie p
padku za
odbywa s
ści. Nato
każdoraz
mamy do
(różnica
od okresu
metody C
nego ind
metody o
dla nastę
• możn
• wiad
Na rys. 1
S, PROF
• różni
• różni

Uzys
wyniki u
metoda E
tej samej
kowuje w
Pods
• jeżeli
zultat
• dla ka
od GB

systeme-
go dla
czenia
stanie
1 bajt
kiedy
nych.
to λ

Ponieważ charakterystyka typu pierwszego przedstawia przebieg zmian obciążenia systemu dla minimalnych okresów występowania wiadomości, to powstaje ona w wyniku połączenia maksimów każdej z poszczególnych krzywych z rys. 10 (przy określonej liczbie przesyłanych bajtów dla minimalnego okresu występowania dostajemy maksymalne obciążenie systemu).

9. ANALIZA WŁAŚCIWOŚCI MAGISTRAL MIEJSCOWYCH DLA METODY DPGS

(15) Wyniki uzyskane dla metody DPGS w stosunku do metod GRMS i EDF silnie zależą od parametrów czasowych modelu strumienia wiadomości. Dla parametrów modelu, dla którego okresy występowania wiadomości tworzą ciąg arytmetyczny odległości t_{min} (t_{min} , $2 * t_{min}$, $3 * t_{min}$, ...) uzyskane wyniki są identyczne jak dla metody EDF. Wynika to z faktu, iż każda wiadomość tworzy osobną grupę, a każdej grupie przyporządkowany jest osobny priorytet. Należy zaznaczyć, że w takim przypadku zastosowanie metody DPGS jest lepsze, ponieważ przydział priorytetów do grup odbywa się zawsze tylko raz po otrzymaniu przez węzeł prawa do wysyłania wiadomości. Natomiast dla metody EDF przydział priorytetów do wiadomości odbywa się po każdorazowym wysłaniu wiadomości lub aktywacji nowej wiadomości. Z kolei, jeżeli mamy do czynienia z przypadkiem, że wszystkie wiadomości w węźle tworzą 1 grupę (różnica pomiędzy minimalnym okresem występowania, a maksymalnym jest mniejsza od okresu minimalnego) to dla metody DPGS uzyskujemy takie same rezultaty jak dla metody GRMS, bowiem wiadomości z grupy wysyłane są w kolejności od minimalnego indeksu (najmniejszego okresu) — patrz rozdz. 5.2. Aby sprawdzić zachowanie metody dla innych modeli niż omówione powyżej (skrajne) przeprowadzono badania dla następujących parametrów modelu:

- można utworzyć więcej niż jedną grupę,
- wiadomości w grupie mają różne okresy występowania.

Na rys. 11 do 16 przedstawiono dla wybranych magistral (PROFIBUS FMS, InterBus-S, PROFIBUS DP [5]) uzyskane wyniki badań dla dwóch przypadków:

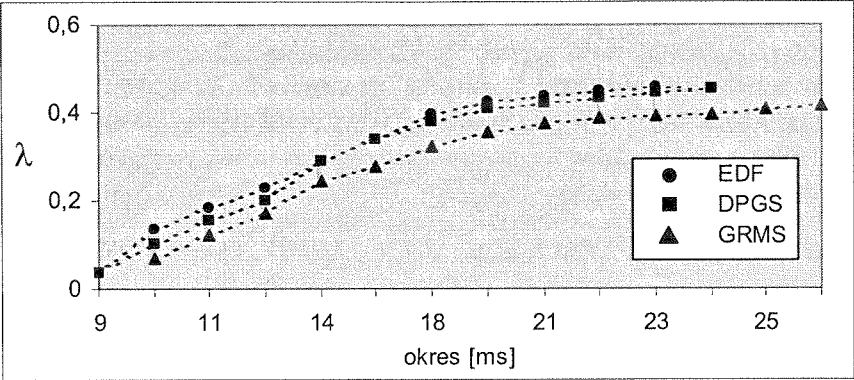
- różnica pomiędzy okresami w grupie wynosi 10% okresu minimalnego w grupie,
- różnica pomiędzy okresami w grupie wynosi 60% okresu minimalnego w grupie,

Uzyskane wyniki pokazują, że dla różnicy 10% pomiędzy okresami w grupie wyniki uzyskane dla EDF i DPGS są prawie identyczne. Natomiast dla różnicy 60% metoda EDF wyraźnie osiąga lepsze wyniki (dla metody DPGS wiadomości w obrębie tej samej grupy mają ten sam priorytet dynamiczny, natomiast metoda EDF przyporządkowuje wiadomościom różne wartości priorytetu, co pozwala uzyskać lepsze wyniki).

Podsumowując wyniki badań metody DPGS można stwierdzić:

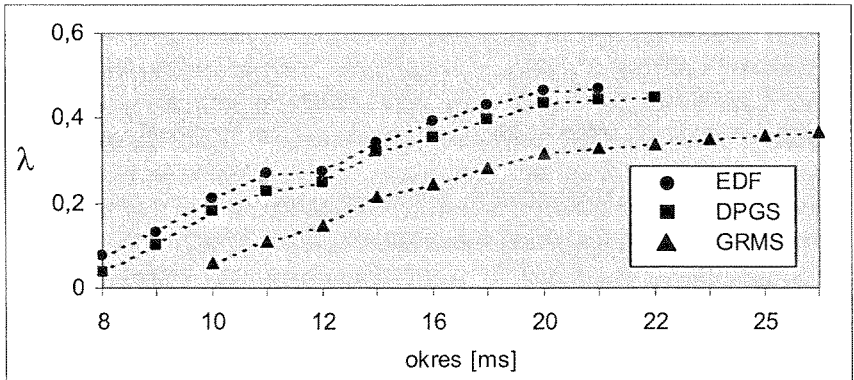
- jeżeli wiadomości tworzą jedną grupę, to metoda DPGS uzyskuje takie same rezultaty jak GRMS,
- dla każdego innego przypadku wyniki uzyskiwane przez metodę DPGS są lepsze od GRMS i mogą być zbliżone lub takie same jak dla metody EDF,

ociąże-
powaniu
y sytu-
wsunku
bajtów.



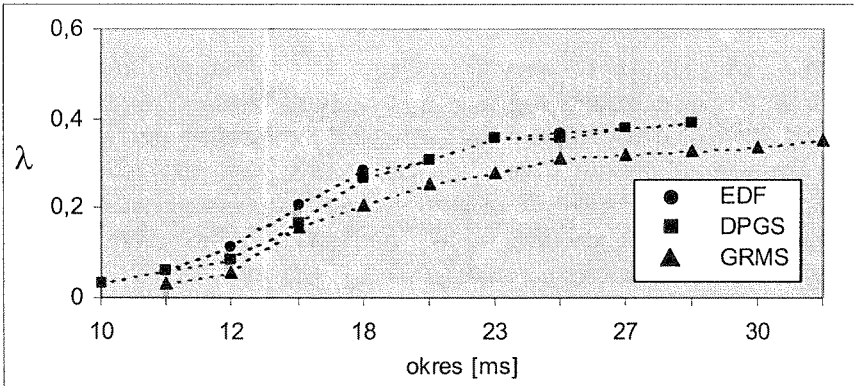
Rys. 11. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle (okresy 10%)

Fig. 11. PROFIBUS: 6 nodes, 6 messages in each node (arrival period 10%)



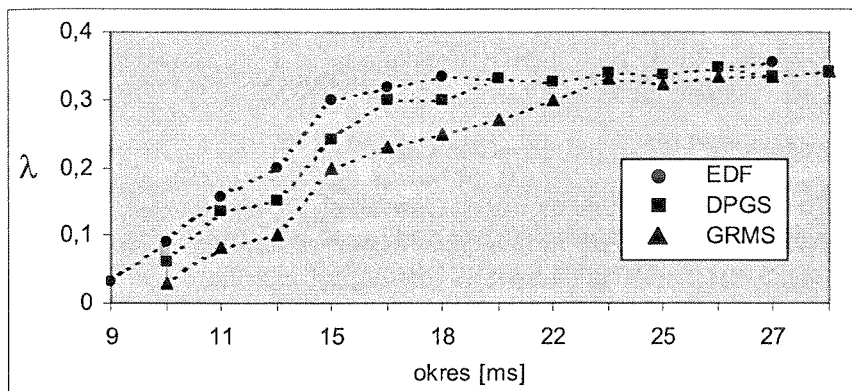
Rys. 12. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle (okresy 60%)

Fig. 12. PROFIBUS: 6 nodes, 6 messages in each node (arrival period 60%)



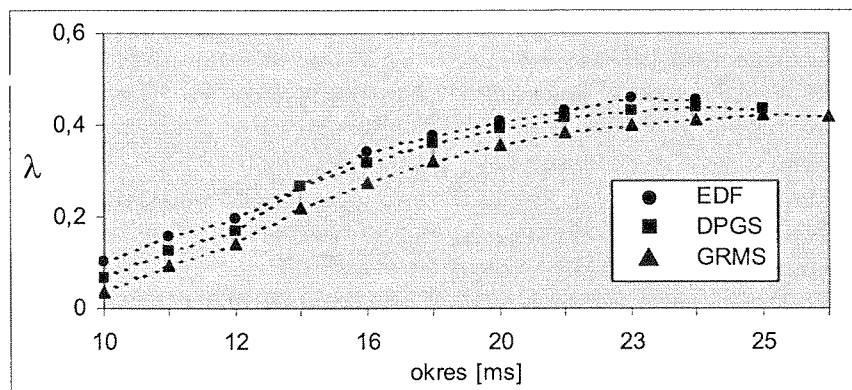
Rys. 13. InterBus-S: 6 węzłów, 6 wiadomości w węźle (okresy 10%)

Fig. 13. InterBus-S: 6 nodes, 6 messages in each node (arrival period 10%)



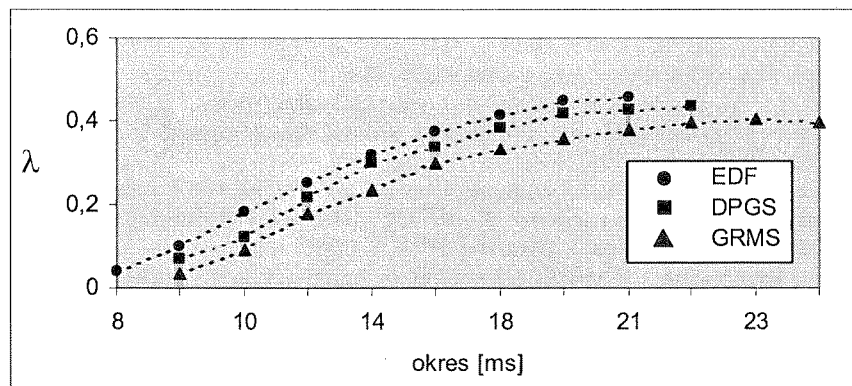
Rys. 14. InterBus-S: 6 węzłów, 6 wiadomości w węźle (okresy 60%)

Fig. 14. InterBus-S: 6 nodes, 6 messages in each node (arrival period 60%)



Rys. 15. PROFIBUS DP: 6 węzłów, 6 wiadomości w węźle (okresy 10%), akwizycja

Fig. 15. PROFIBUS DP: 6 nodes, 6 messages in each node (arrival period 10%), aquisition



Rys. 16. PROFIBUS DP: 6 węzłów, 6 wiadomości w węźle (okresy 60%), akwizycja

Fig. 16. PROFIBUS DP: 6 nodes, 6 messages in each node (arrival period 60%), aquisition

- ponieważ narzut obliczeniowy metody DPGS jest dużo mniejszy niż metody EDF, to w przypadku, kiedy obie metody uzyskują takie same wartości współczynnika obciążenia systemu, metoda DPGS jest metodą lepszą.

10. WNIOSKI

Opracowane metody szeregowania wiadomości uwzględniają specyfikę systemów rozproszonych z magistralą miejscową poprzez dostosowanie ograniczeń czasowych wysyłanych wiadomości do właściwości czasowych systemu rozproszonego. Należy podkreślić, że dużą zaletą opracowanych metod DPGS i MPGS jest mały narzut czasowy na pracę programu szeregującego. Przeprowadzone badania wykazały, że uzyskane rezultaty dla nowych metod, w odniesieniu do najpopularniejszych metod GRMS i EDF, zależą od parametrów przyjętego modelu wiadomości. Należy podkreślić, że uzyskane rezultaty dla opracowanych metod nie są nigdy gorsze niż wyniki uzyskane dla metody GRMS. Natomiast w sytuacji, gdy rezultaty te są zbliżone do wyników uzyskanych dla metody EDF, to uwzględniając mniejszy narzut obliczeniowy metod DPGS i MPGS, należy uznać je za metody bardziej wydajne [11].

11. BIBLIOGRAFIA

1. K. Jeffay, D. Stanat, C. Martel: *On Non Preemptive Scheduling of Periodic and Sporadic Tasks*. Proc. of IEEE Real Time Systems Symposium, 1991
2. L. Liu, J. Layland: *Scheduling algorithms for multiprogramming in a hard real - time environment*. J. Assoc. Comput. Mach. Vol. 20, No. 1, January 1973, pp. 46 - 61
3. P. Pleinevaux, J.D. Decotignie: *Time Critical Communication Networks: Field Buses*. IEEE Network Vol. 2, No. 3, May 1988, pp. 55-63
4. L. Sha, R. Rajkumar, S. Sathaye: *Generalised Rate-Monotonic Scheduling Theory: A Framework for Developing Real Time Systems*. Proc. of the IEEE. Vol. 82, No. 1, Jan. 1994, pp. 68-82
5. J. Werewka: *Systemy rozproszone sterowania i akwizycji danych. Sterowniki programowalne i magistrale miejscowe*. Kraków, Wydawnictwa CCATIE, 1998
6. F. Vasques, G. Juanolet: *Fieldbus MAC Mechanism for Hard Real Time Data Communication Support*. CNRS LAAS, LAAS Report 93396, 1993
7. S. Żaba: *Message scheduling in distributed real-time system based on fieldbus*. 5th International Symposium on Methods and Models in Automation and Robotics, Międzyzdroje, 25-29 August 1998, pp. 565-570
8. S. Żaba, J. Werewka: *Szeregowanie wiadomości w rozproszonych systemach czasu rzeczywistego opartycho magistrale miejscowe*. Kwartalnik Elektroniki i Telekomunikacji, Wydawnictwo Naukowe PWN, Tom 45, Zeszyt 1, Warszawa 1999, ss. 25-50
9. S. Żaba: *Szeregowanie wiadomości z dynamicznym przydziałem priorytetu w rozproszonych systemach sterujących*. VI Konferencja Systemy Czasu Rzeczywistego, 27-30 wrzesień, Zakopane 1999
10. S. Żaba: *Badania eksperymentalne własności czasowych wybranych magistral miejscowych*. W opracowaniu pod red. T. Szmuc i J. Werewka: *Analiza i projektowanie systemów komputerowych czasu rzeczywistego o różnym stopniu rozproszenia*. Polskie Towarzystwo Informatyczne, oddz. małopolski Kraków 2001, ss. 353-390
11. S. Żaba: *Analiza czasowa rozproszonych systemów sterowania bazujących na magistralach miejscowych*. Rozprawa doktorska, Kraków, Wydział EAIiE AGH, 2001

S. ŻABA

DYNAMIC MESSAGE GROUPS SCHEDULING IN DISTRIBUTED REAL TIME SYSTEMS

Summary

In this paper new methods of messages scheduling are presented. The first method, Dynamic Priority Group Scheduling method is based on assigning dynamic priorities not to individual message, but to group of messages. The second method, Mixed Priority Group Scheduling method enable to assign to group of messages static priorities as well. A procedure of checking RT-constraint fulfilment of messages in distributed real time system based on fieldbus for DPGS and MPGS methods are presented. For three fieldbuses PROFIBUS FMS, CAN and InterBus tests were performed in order to compare properties of created methods to GRMS and EDF methods. New methods take into consideration specificity of distributed real time system based on fieldbus and adapt message deadlines to time properties of distributed system. It is very important that new methods have less calculating time than GRMS and EDF methods. Tests which were performed give results as follow:

- Performances of new methods depend on message model parameters, but obtained results have never been worst than results for GRMS method,
- In the case, when results are almost the same as the results for EDF method, and it takes into consideration the smallest calculating time of new methods, it can say that new methods are most effectively.

Keywords: distributed control system, real time system, fieldbus network, message scheduling, message groups, GRMS, EDF methods

New

tur
is c
is t
Im
ma
CM
The
tal
con

Key

Incr
rvices, a
imaging
improve
analysis
suited to
nonlinea
preserve

New VLSI CMOS analogue median filter for real-time video applications

JACEK KOWALSKI, TOMASZ KACPRZAK AND KRZYSZTOF ŚLOT

*Institute of Electronics, Technical University of Łódź
ul. Wólczyńska 223, 90-924 Łódź, Poland
e-mail: jackowal@p.lodz.pl,
kacp45@p.lodz.pl, kslot@p.lodz.pl*

*Otrzymano 2003.04.02
Autoryzowano 2003.04.25*

Noise corruption is a problem common to many imaging applications, degrading picture quality and reducing the effectiveness of its utilisation. A new CMOS VLSI chip, which is designed to realise real-time median filtering of video images, is presented. The concept is based on Cellular Neural Network paradigm, which enables parallel analogue operation. Images, which are subject to processing, are loaded and produced after filtering in a serial manner through video signal. The circuit has been manufactured using MIETEC 2.4 μm CMOS technology and allows for processing of images with 64 pixels horizontal resolution. The resolution can be increased by cascading of several integrated circuits. Experimental verification using computer generated grey-level images has proved the validity of the concept.

Keywords: cellular neural networks, nonlinear filtering, image processing, ASIC

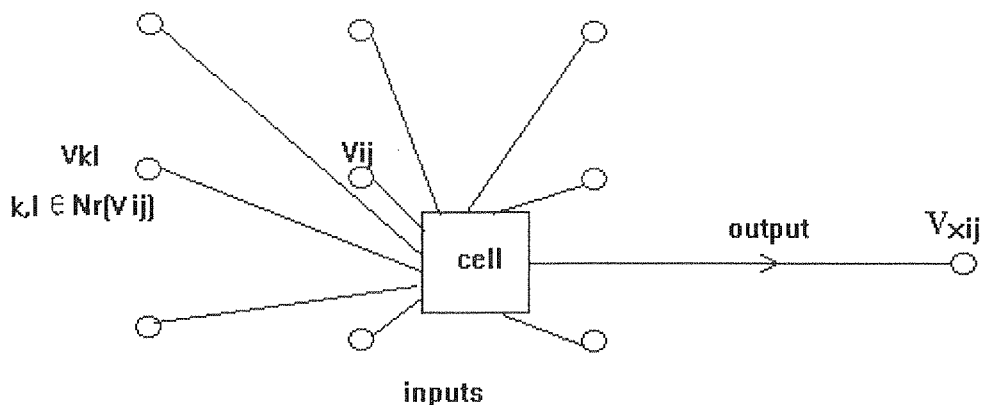
1. INTRODUCTION

Increasing demands for video applications in teleconferencing, multimedia services, autonomous navigation, object tracking, astronomical, technical and medical imaging make rigorous requirements for real-time image noise filtering algorithms to improve visual quality and performance of subsequent processing tasks, e.g. coding, analysis and interpretation. For these tasks, the median filter has proved to be very well suited to eliminate noises, especially impulsive ones. Median filters belong to a class of nonlinear filters and in comparison with linear ones show certain advantages, namely preserve edges without introducing new grey levels. While the theory of the median

filters is well developed [1–6], the digital or analogue hardware implementation especially for real-time processing of video images are still far from satisfactory solution. Digital approach (using for example ASICs, FPGA or μ P techniques) provides very accurate filtering but is time-consuming and requires large silicon area for VLSI implementation, which makes the real-time application — especially for video technology — difficult to adopt [7–11]. On the other hand, the analogue systems [12, 13, 18–22] demonstrate many advantages over digital systems in terms of power consumption and silicon area and even though this realisation is not so accurate as the digital one, it can be successfully accepted for many applications. Analogue circuits for rank order value detection have been proposed in [12, 18–20] but only motivated by application in video image processing. One can feel lack of integrating a whole system to perform median filtering on scanned images in real time. Taking into account these aspects a new analogue version of median filter based on the concept of nonlinear Cellular Neural Networks (CNN) [14, 15, 16] has been developed. Cellular Neural Networks are suitable for an analogue VLSI circuit implementation and they feature inherent parallelism in their operation, which results in very fast image processing. We show in this paper that proposed method is fast enough to filter out impulsive noise of video images in real time and is competitive with the digital methods utilised so far in terms of manufacturing cost. Therefore, the method presented here can find applications in telecommunication, information technology and multimedia services. In the next section, a fundamental theoretical analysis of CNN-based median filtering is summarised. Then, in Section 3 the median circuit realisation and respective CMOS VLSI layouts are described. Section 4 covers the computer simulation results, measurements of the filter dc and transient characteristics and experiments with real-time filtering of computer-generated grey-scale images of size 64×64 pixels. Finally, the results and conclusions of this study are presented in Section 5.

2. THEORETICAL BACKGROUND

The function performed by median filter of an image is to replace the original pixel by the median value chosen from the set of all pixels surrounding the central pixel within a given neighbourhood, called window. Typically, this window is a rectangle of size $M \times K$, where M and K are the numbers of pixels (including the central one) along the horizontal and vertical axis, respectively. The median filter operation on the central pixel based on the 3×3 window is presented schematically in Fig. 1. Median filter algorithm is organised into two parts: in the first part pixels are sorted in ascending order by value while during the second one, the central pixel value is replaced by the value, which is located at the middle position (median) of the ascending stairs. Due to sorting task, especially when the observation window is large, the algorithm is time consuming because it requires analogue-to-digital conversion, coding and storing the pixel value in the digital form, comparing pixel values of the window and their ordering in a ascending way. There are many methods suggested for acceleration of

Fig. 1. CNN median filter cell with its control signals in 3×3 windowRys. 1. Komórka SNK filtru medianowego w oknie filtracji 3×3

this procedure, however none of them is satisfactory, especially when processing time and silicon area are under considerations. To eliminate the ordering stage, one can take advantage of the fact that the median value divides the set into two subsets of equal number of pixels, of which the first subset consists of pixels having lower values than the median value while the other one having the higher values than the median value. Mathematically, this is expressed by the following nonlinear algebraic equation:

$$f(V_{xij}) = \sum_{k,l \in Nr(V_{ij})} \text{sign}(V_{kl} - V_{xij}) = 0 \quad (1)$$

where Nr denotes the neighbourhood of the processed pixel located in the position (i,j) , V_{kl} and V_{ij} represent values of pixels inside the window, and $M_{ij} = V_{xij}$ for $f(V_{xij}) = 0$ is the median value of these pixels. The sign function is here defined as follows:

$$\text{sign}(x) = \begin{cases} +1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} \quad (2)$$

As was suggested in [14], the equation (1) can be thought of as a steady state solution of a nonlinear RC network, Fig. 2, consisting of one node capacitance C and several branches, by which input voltages (representing pixels values) are connected to this node via nonlinear resistors with I-V characteristics of the form $I = G \text{sign}(V)$, where G denotes absolute value of the current. The state equation for such a network can be expressed by:

$$C \frac{dV_{xij}}{dt} = G \sum_{k,l \in Nr(V_{ij})} \text{sign}(V_{kl} - V_{xij}) \quad (3)$$

A graphic solution of the equation (1) for the set of nine input pixels $V_1 \dots V_9$ is shown in Fig. 3. The horizontal axis V_{xij} is intersected by stepwise function at the point which

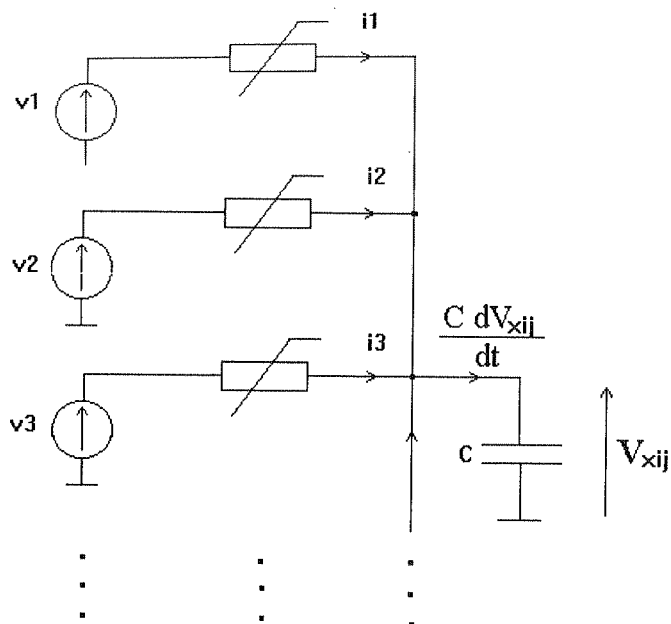


Fig. 2. Circuit representation of median image filter cell

Rys. 2. Reprezentacja obwodowa komórki filtru medianowego

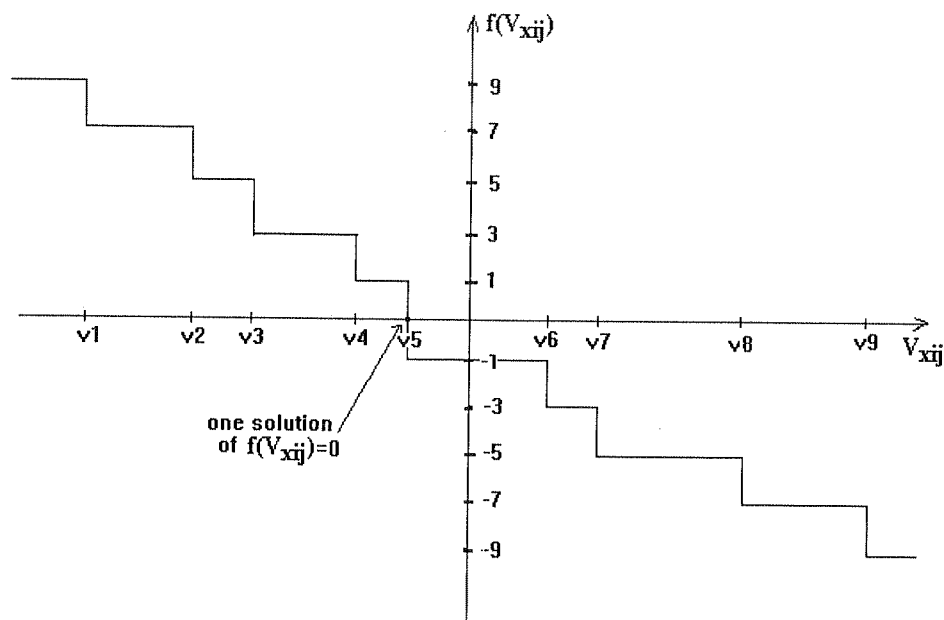


Fig. 3. Graphic solution of equation (1)

Rys. 3. Rozwiązanie graficzne równania (1)

is the
a dyna
the CI
the for
and V
therefo
curren
with a

transfor
equals

Bas
CMOS
in Fig.
circuit
line. To

is the median value of the set. The solution of (1) can be obtained as a steady state of a dynamic response of respective Cellular Neural Network (CNN) [14]. For this task the CNN template consists of the feedforward matrix with nonlinear coefficients of the form $G \text{sign}(V_n - V_{xij})$, where V_n is the input voltage of n-th pixel in the window, and V_{xij} is the state voltage of the central pixel. The CNN cell's state equation has therefore the same form as (3) while the CNN cell circuit corresponds to the Fig. 2. The current-voltage characteristics of a nonlinear resistor, which couples the central pixel with any of its neighbours is shown in Fig. 4. In an equilibrium state the equation (3)

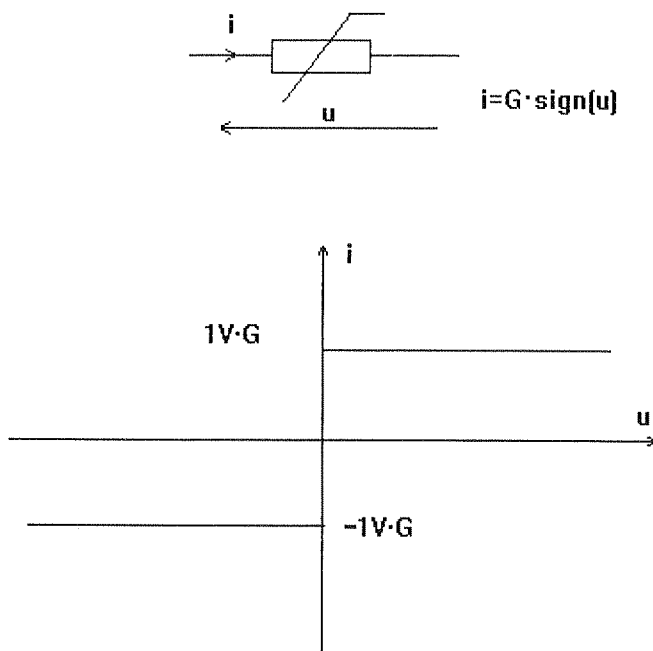


Fig. 4. Characteristic of nonlinear resistor of median filter

Rys. 4. Charakterystyka nieliniowego rezystora filtru medianowego

transforms into the equation (1) for which the value of the cell voltage $V_{xij}(t = \infty)$ equals to the median value M_{ij} of the set $\{V_{kl}\}$ in the filter window N_r .

3. MEDIAN FILTER HARDWARE IMPLEMENTATION

Based on the concept described above an integrated circuit implementation using CMOS 2.4 μm technology was developed. The overall system architecture is shown in Fig. 5, while the time-diagram of its operation is given in Fig. 6. The integrated circuit is designed for filtering of images with 64 pixels resolution in each horizontal line. To perform median filtering task of a given pixel within the window of size 3

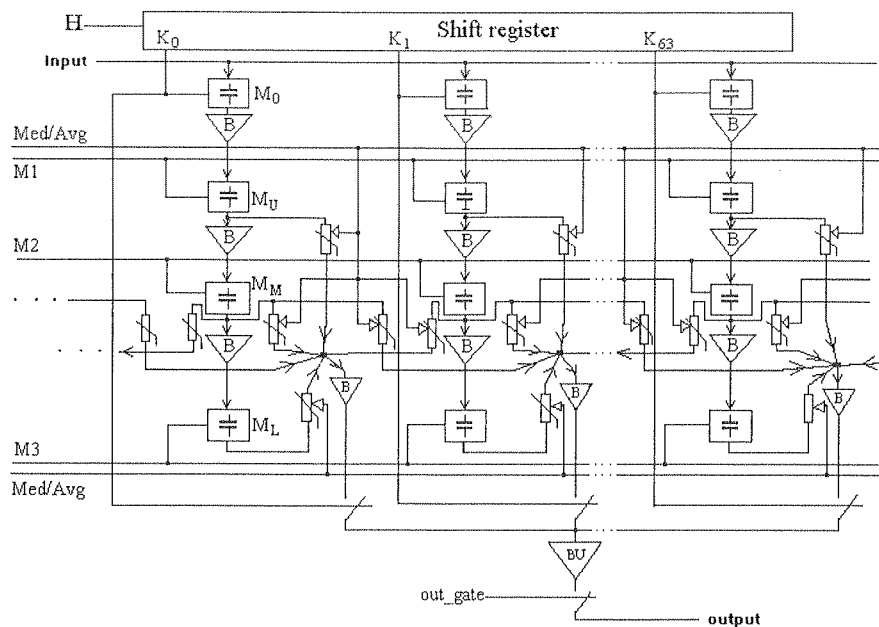


Fig. 5. System architecture of CNN-based median filter

Rys. 5. Architektura systemu filtru medianowego w koncepcji SNK

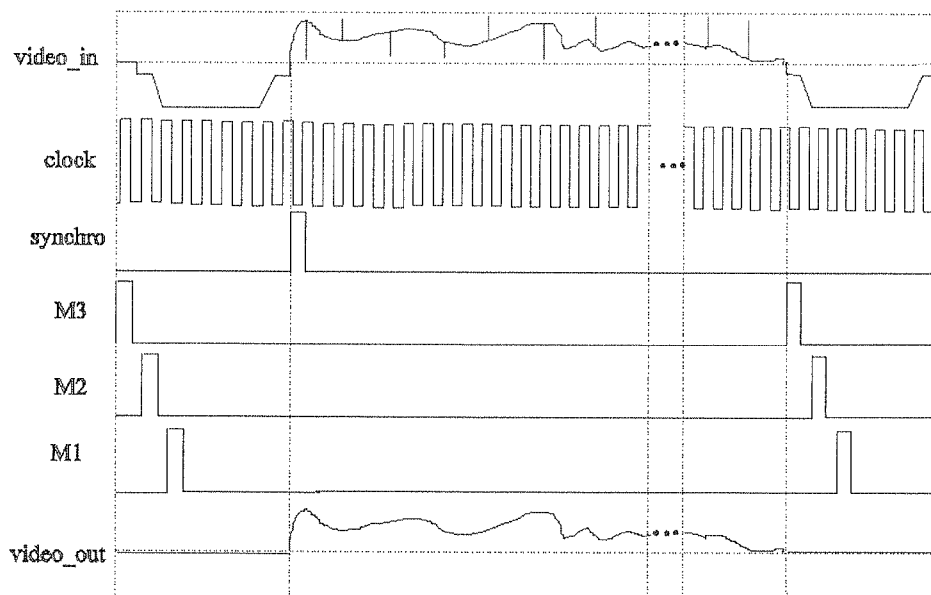


Fig. 6. Time-diagram of CNN-based median filter operation

Rys. 6. Przebiegi sterujące pracą filtru medianowego o architekturze SNK

by 3, the
the sy
by 64
additio
K₆₃ pl
line. In
middle
compo
pixel a
east. T
the syn
when t
M_U, M
signal l
video c
length
respect
are sam
time-sl
row do
discard
middle
first row
from th
line is p
transmi

A b
the cur
nent of
differen
compos
M5B. I
Current
capacito
output c
transisto
block. I
High im
polysilic
symmet
with po
feedback

by 3, it is required to have an access to pixels in three consecutive lines. Therefore the system is composed of an analogue memory array of pixels of structure 3 rows by 64 columns: M_U — upper line, M_M — middle line, and M_L — lower line. The additional line M_0 , which is controlled by the shift register signals $K_0, K_1, K_3, \dots, K_{63}$ plays the role of a temporary sample-and-hold memory for pixels of new input line. In this way three consecutive video lines are stored for real-time filtering of the middle one. The main element of the filter is a pixel processing section which is composed of five nonlinear resistors connected to controlling signals of the central pixel and four neighbour pixels with respect to the central one: north, south, west, and east. The filter operation is synchronised by an external signal Synchro, extracted from the synchronisation impulse of each video line. Let us start our observation at the time when the pixels of three consecutive lines are stored in a respective analogue memories M_U, M_M and M_L . Once the state transient voltage of a given cell settles down, its signal being the median pixel value is ready to be read-out from the circuit, giving the video output signal after filtering. This process is controlled by shift register of 64 bits length synchronised by the signal Synchro. During consecutive 64 cycles the pixels of respective columns are read-out and at the same time new pixels of input video signal are sampled and stored in the analogue memory M_0 . Next, during the synchronisation time-slot three consecutive short clock cycles M_3, M_2 , and M_1 initiate shifting one row down the respective voltages stored in analogue memories: the bottom row is discarded and replaced by the content of the middle row during the cycle M_3 , the middle row is replaced by the content of the upper one during the cycle M_2 , and the first row is replaced by a sampled voltages of a new line during the cycle M_1 . Starting from the next synchronisation pulse synchro, the median filtering of the next video line is performed in a real-time. In this way the total video frame is filtered during its transmission from the receiver input to the CRT display.

A basic block of the image median processing stage is a nonlinear resistor of which the current-to-voltage characteristic is of signum shape. This resistor realises component of sum in equations (1), (3) and it is designed using operational transconductance differential amplifier OTA coupled with current amplifier, Fig. 7. OTA amplifier is composed of the following transistors: $M1A, M1B, M2, M3, M4A, M4B, M5A$ and $M5B$. In Fig. 7 next to transistor's symbols dimensions W/L of transistors are shown. Current amplifier consists of $M6A, M6B, \dots, M15A, M15B$ transistors. Poly1–poly2 capacitor C_k is inserted to obtain an appropriate phase margin and stability. The first output of current amplifier is feeding a positive current feedback loop through a switch transistor MP . The second mirrored one is the output of this basic circuit building block. Layout of this circuit occupies $16128\mu\text{m}^2$ ($224\mu\text{m} \times 72\mu\text{m}$) of silicon area. High input impedance allows to connect pixel signals from analogue memory with polysilicon or MOS capacitor. In order to get a signum-like transfer characteristic with symmetrical current limits of both sides, we used current amplifier of gain equal to one with positive current feedback loop. When the transistor switch MP is off, the current feedback loop is disconnected and the current amplifier operates in a linear range with

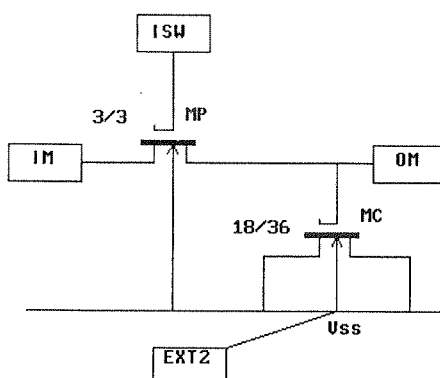


Fig. 8. CMOS realization of analogue memory cell

Rys. 8. Realizacja CMOS komórki pamięci analogowej

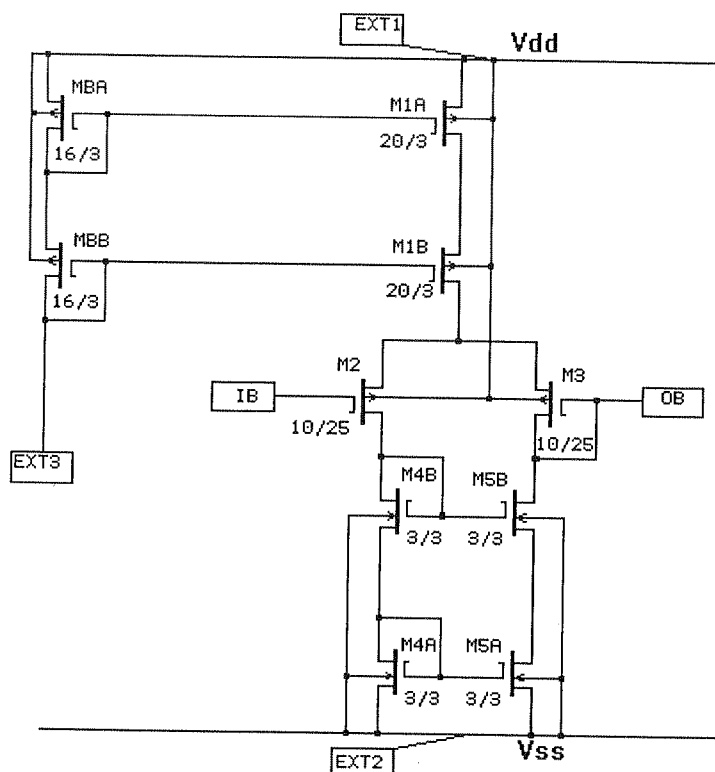


Fig. 9. CMOS realization of voltage buffer B

Rys. 9. Realizacja CMOS bufora napięciowego B

area. Output buffer BU has structure as shown in Fig. 10 and occupies $21792\mu\text{m}^2$ ($181.6\mu\text{m} \times 120\mu\text{m}$) silicon area.

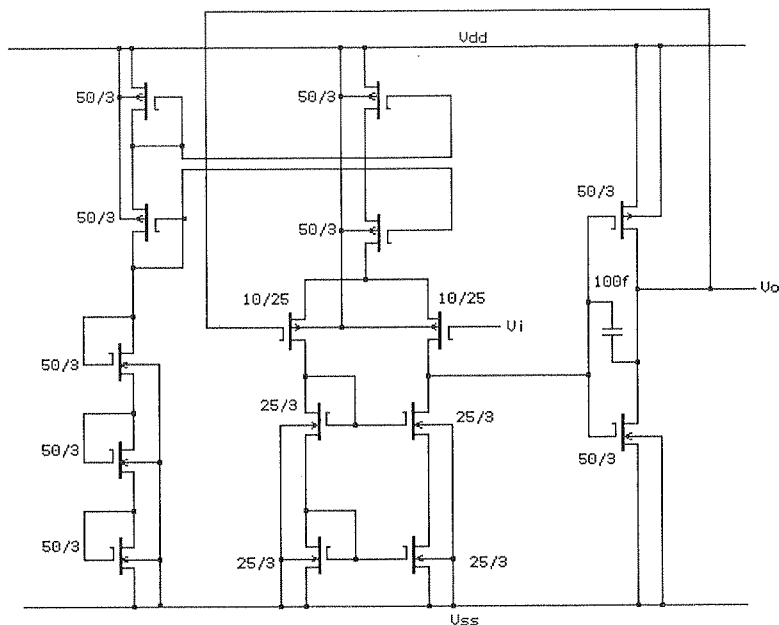


Fig. 10. CMOS realization of voltage buffer Bu

Rys. 10. Realizacja CMOS bufora napięciowego Bu

4. SIMULATION AND EXPERIMENTAL RESULTS

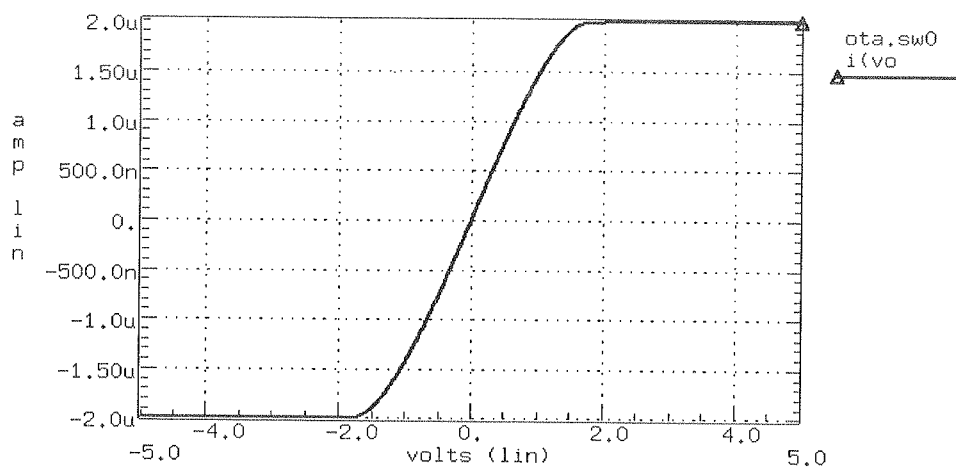
The CMOS nonlinear resistor of Fig. 7 has been simulated in HSPICE, using MIETEC $2.4\mu\text{m}$ technology MOS transistor model. The dc transfer characteristics simulation results of the filter in two modes, average mode and median mode are shown in Fig. 11.

The simulation result of transient response of median filter cell working in nonlinear mode of 9 pixels window is given in Fig. 12. The voltage pixel values $V_1, \dots, V_9$ (Fig. 12a) in the filter window are different. The cell transient response in Fig. 12b consists of very short phase of oscillatory form after which the cell output voltage is stabilised on the constant level corresponding to the median value of input voltages $V_1, \dots, V_9$, which is in this case equal to $V_5 = 0.1\text{ V}$.

The median filter chip has been manufactured in double poly, double metal, n-well CMOS MIETEC $2.4\mu\text{m}$ technology under EUROPRACTICE programme [17]. This chip, which occupies 24.8 mm^2 ($5.06\text{mm} \times 4.9\text{mm}$) silicon area and consists about 12000 MOS devices, encompasses 64 pixels median filter, input/output interfaces, and

2 μm^2

(a)



(b)

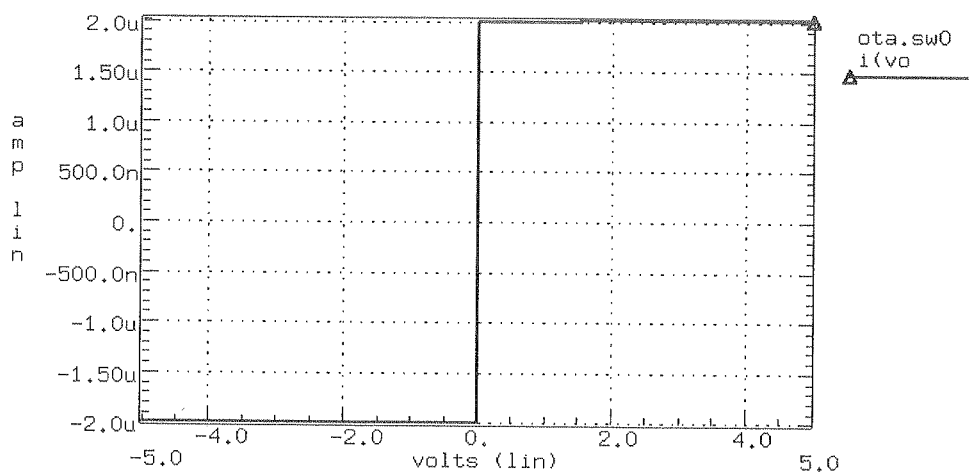


Fig. 11. DC $i(v)$ characteristic of CMOS resistor in linear mode (a), in nonlinear mode (b)

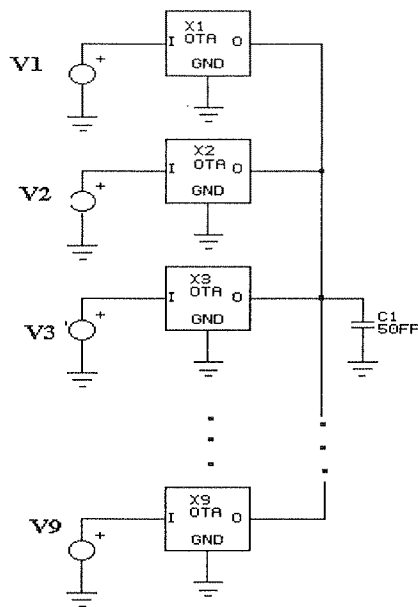
Rys. 11. Charakterystyka DC $i(v)$ rezystora CMOS pracującego w trybie liniowym (a), pracującego w trybie nieliniowym (b)

, using
eristics
ode are

a nonli-
... V_9
fig. 12b
ltage is
voltages

n-well
7]. This
s about
ces, and

(a)



(b)

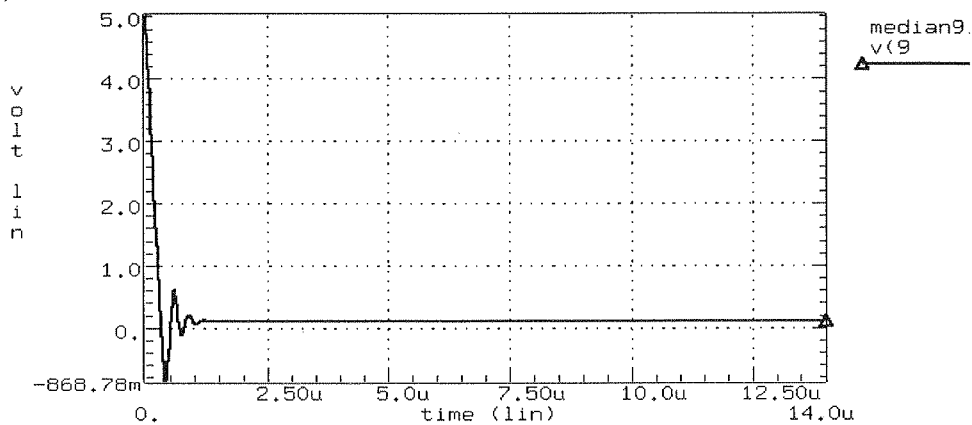


Fig. 12. CNN median filter cell of 9 pixels window (a), transient response in median filter cell (b)

Rys. 12. Komórka SNK filtru medianowego w oknie 9-cio pikselowym (a), stan przejściowy w komórce SNK filtru medianowego (b)

some inc
in Fig. 1

Layo
is shown
chip is e
The dc c
measured

Figur
Fig. 16b

is degrad
real slope
have som

Statio
(linear) fi

Here
one V_2 i
50 Hz. T
in Fig. 18

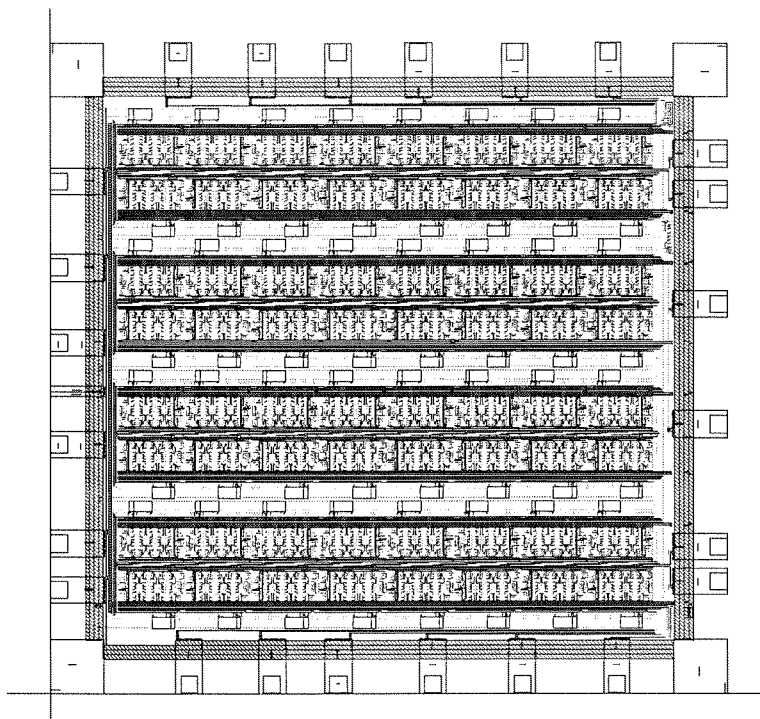


Fig. 13. Layout of the median filter chip

Rys. 13. Topografia układu scalonego filtru medianowego

some individual nonlinear coupling circuits. Layout of the median filter chip is shown in Fig. 13.

Layout of the main element of filter for individual pixel without shift register cell is shown in Fig. 14 and occupies $184550 \mu\text{m}^2$ ($518.5 \mu\text{m} \times 356 \mu\text{m}$) silicon area. The chip is encapsulated in DIL 24 package and its power consumption is about 100 mW. The dc characteristics of the nonlinear resistor in two modes, linear and nonlinear were measured using the set-up shown in the Fig. 15.

Figure 16a shows the measured characteristic of the average filter, while the Fig. 16b — the median one. The slope of the nonlinear characteristic shown in Fig. 16b is degraded by loading the filter output by the measuring resistor 51k, therefore the real slope is much greater as expected. Although curves are not ideally symmetric and have some minor offsets, they match the assumed requirements and constraints.

Static filtering characteristics of three pixels cell of median (nonlinear) and average (linear) filter have been measured using the set-up shown in Fig. 17.

Here, two pixels have a constant values $V_1 = -0.6 \text{ V}$, $V_3 = 0.6 \text{ V}$, while the third one V_2 is excited by sine waveform of 1.5V amplitude, zero bias and frequency of 50 Hz. The oscilloscope outputs for the average filter and the median one are shown in Fig. 18.

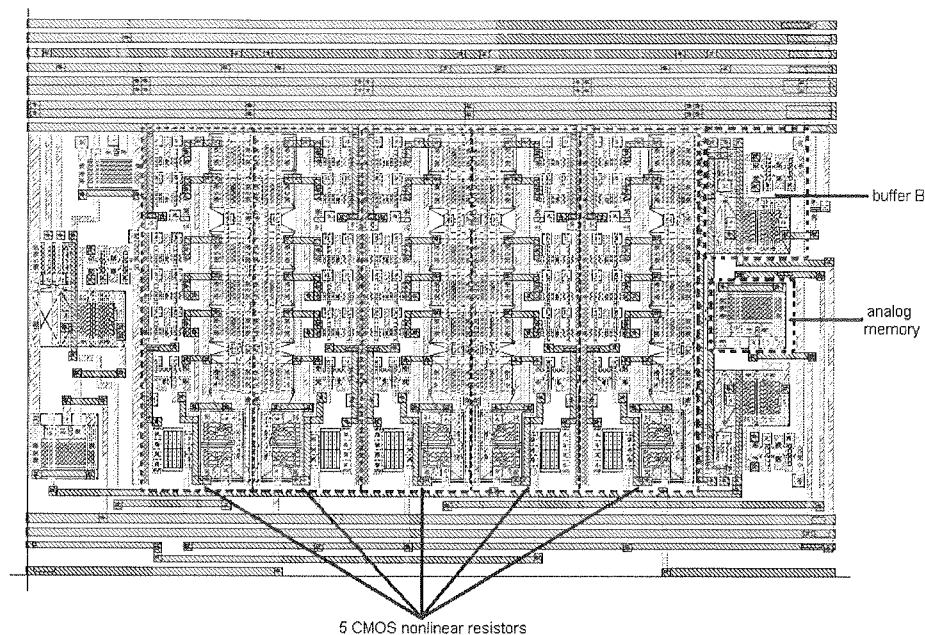


Fig. 14. Layout of the single CNN median filter cell

Rys. 14. Topografia pojedynczej komórki SNK filtru medianowego

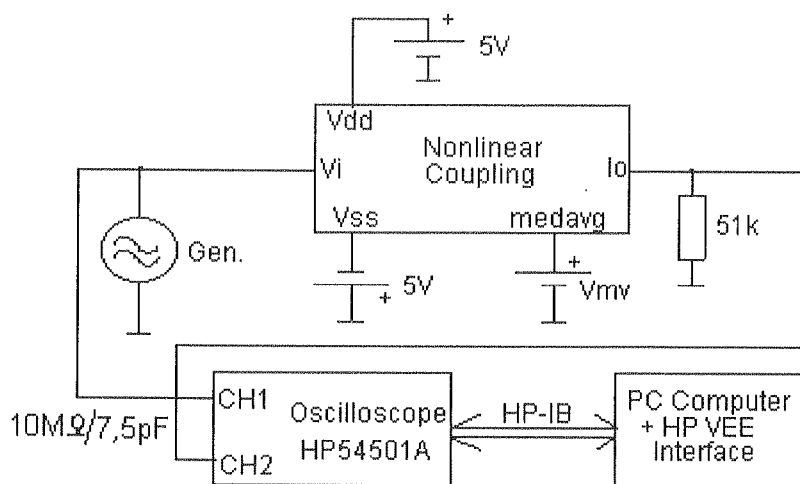


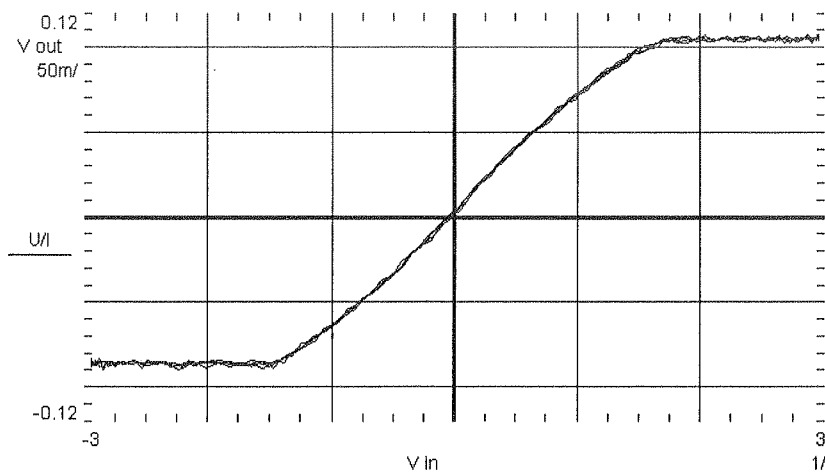
Fig. 15. Measurement set-up of the nonlinear resistor

Rys. 15. Układ pomiarowy nieliniowego rezystora CMOS

Fig.

Rys. 16

(a)



(b)

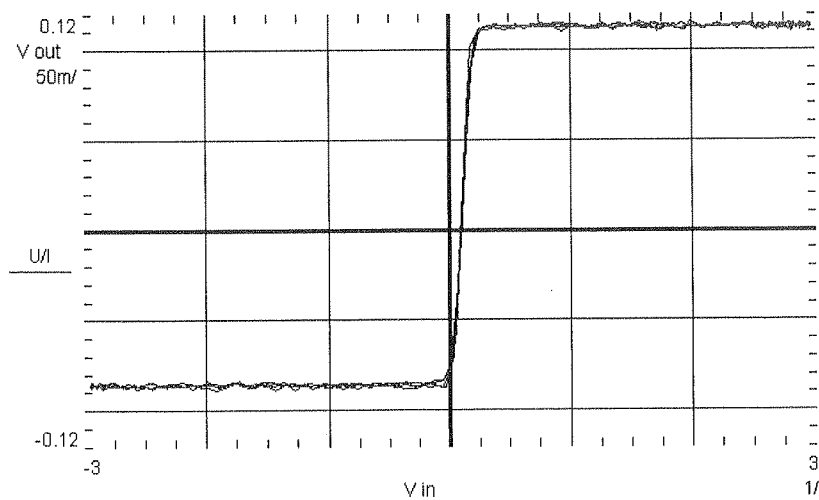


Fig. 16. Measured characteristics of nonlinear resistor in the linear mode (a), nonlinear mode (b)

Rys. 16. Zmierzone charakterystyki nieliniowego rezystora CMOS pracującego w trybie liniowym (a), pracującego w trybie nieliniowym (b)

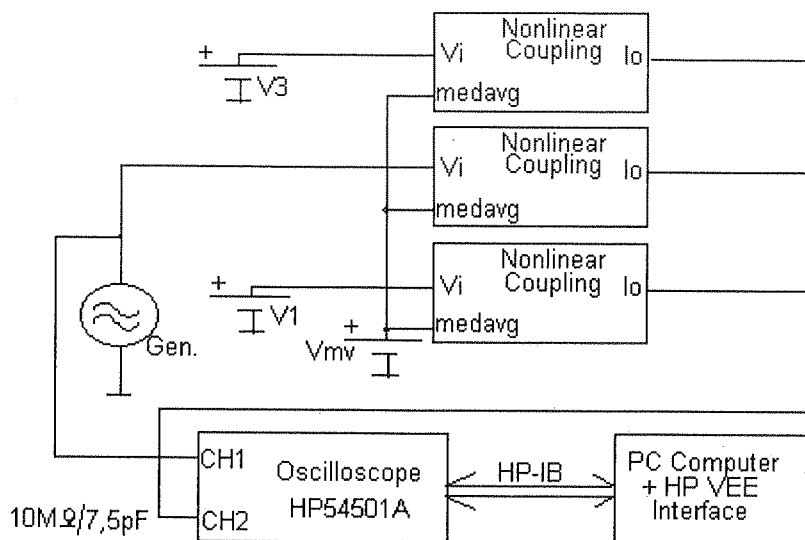


Fig. 17. Three pixels cell of median filter measurement set-up

Rys. 17. Układ pomiarowy 3-pikselowej komórki SNK filtru medianowego

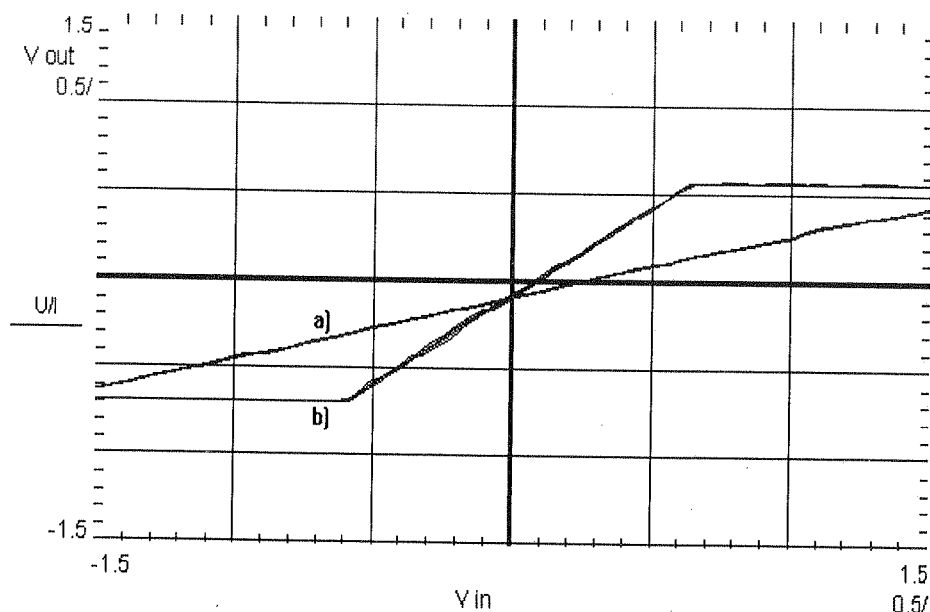


Fig. 18. Static filtering characteristics measured for linear mode a) and nonlinear mode b)

Rys. 18. Statyczna charakterystyka filtracji zmierzona dla trybu liniowego a) i trybu nieliniowego b)

It is evident that in the case of median filtering the output voltage is stabilised on the levels about -0.6 V and $+0.6\text{ V}$, which are the median values of the set of three pixels. Settling time of three pixels cell of median/average filter has not been measured because of too large oscilloscope capacitive load (7.5 pF), which is not realistic case inside IC.

The real-time filtering performance of the chip was tested using computer-based set-up shown in Fig. 19.

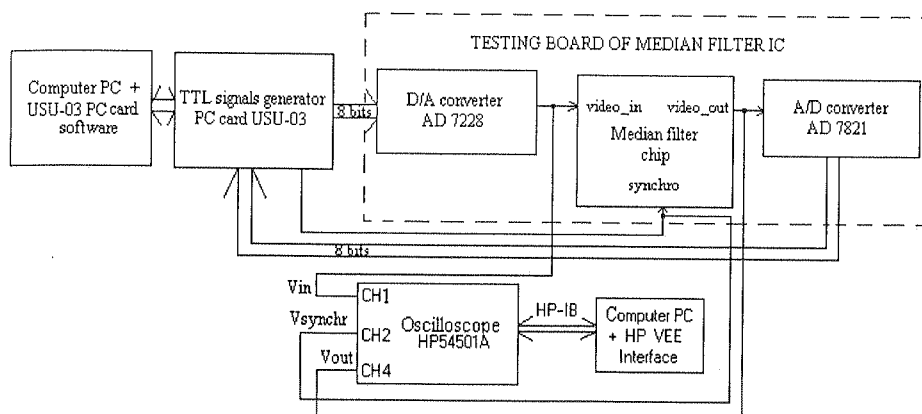


Fig. 19. The real-time filtering computer-based set-up of the median chip

Rys. 19. Schemat blokowy platformy testowej układu scalonego filtru medianowego do przetwarzania obrazów zadawanych z komputera

The upper trace of each screen represents two consecutive artificial video lines, the mid trace is a synchronisation signal while the lower trace represents the video output line after filtering. The first video line (the "odd" one) contains one object with four pulse artefacts while the second line (the "even" one) contains no object. It is evident from the Figures 19 and 21 that the median filter removes all pulse noise artefacts while the average filter not. This effect is even much more visible on the computer generated binary and grey-scale images of size 64×64 pixels, corrupted by pulse noise, shown in Fig. 22.

These images were serially inputted to the filtering system at a speed of $30\text{ }\mu\text{s}$ per line. The elemental median time response does not exceed $1\text{ }\mu\text{s}$ (Fig.12), and since the 64 medians are processed simultaneously, this reflects maximum processing speed of 64 megamedians per second. a basic median filter $2.4\text{ }\mu\text{m}$ chip parameters are presented in Table 1. In Table 2 a comparison of main characteristic of the proposed analogue chip and digital chips [9, 10, 11] is shown.

The set-up consists of programmable generator of video-like signals of binary and grey-scale images, D/A and A/D converters, and oscilloscope. The oscilloscope screen for two filtering modes, median and average (linear) are shown in Figures 20 and 21.

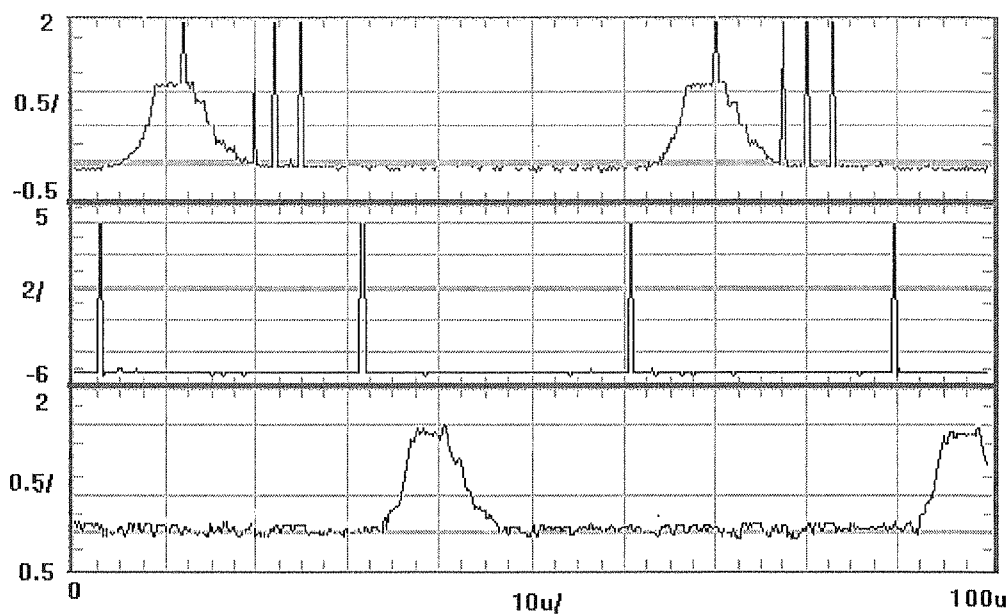


Fig. 20. The oscilloscope screen for median (nonlinear) filtering mode

Rys. 20. Rezultat obserwacji oscyloskopowej filtracji medianowej (nieliniowej)

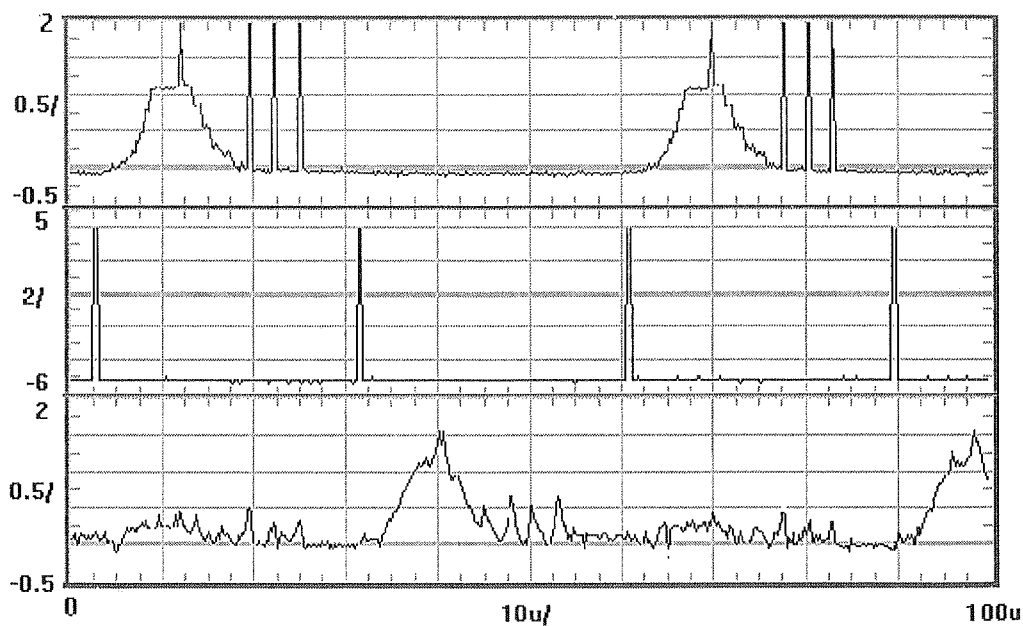


Fig. 21. The oscilloscope screen for average (linear) filtering mode

Rys. 21. Rezultat obserwacji oscyloskopowej filtracji uśredniającej (liniowej)

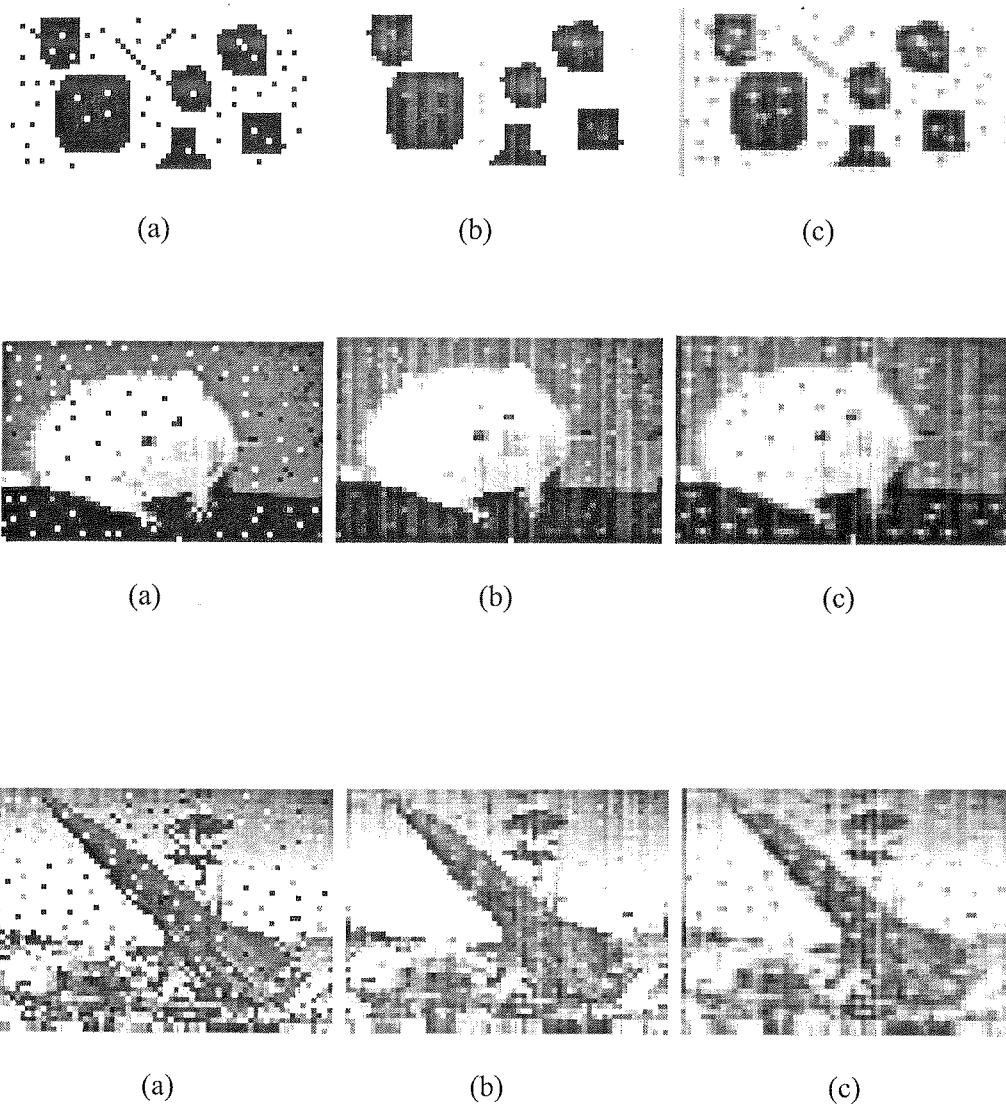


Fig. 22. Images processing results: input images (a), median filtering results (b), average filtering results (c)

Rys. 22. Wyniki przetwarzania obrazów: wejściowe obrazy (a), rezultaty filtracji medianowej (b), rezultaty filtracji uśredniającej (c)

Tabela 1

Basic median filter CMOS 2.4 μm chip parameters
Podstawowe parametry układu scalonego filtru medianowego
zaprojektowanego w technologii CMOS 2.4 μm

Parameter	Results in CMOS 2.4 μm MIETEC technology
Supply voltage	±5 V
Power consumption	100 mW
Silicon area	24.8 mm ²
Maximum speed of processing	64 megamedians per second

Tabela 2

Comparison of proposed analogue chip and digital chips (ref. [9, 10, 11]) parameters
Porównanie parametrów układów scalonych analogowego zaproponowanego w artykule i cyfrowych
przedstawionych w pracach [9, 10, 11]

Chip	Proposed in this paper	Ref [9]	Ref [10]	Ref [11]
Process	2.4 μm MIETEC	3 μm	2 μm	0.8 μm
Design technique	Full-custom	Full-custom	Full-custom	Semi-custom
Silicon area	24.8 mm ²	50 mm ²	36 mm ²	19.4 mm ²
Number of transistors	12 000	22 000	30 000	26 400
Transistors per mm ²	480	450	830	1360
Maximum speed of processing	64 megamedians per second	50 megamedians per second	72 megamedians per second	80 megamedians per second
Number of pins	24	40	—	68
Power dissipation	100 mW	800 mW	—	700 mW

Tabela 1

5. CONCLUSION

We have presented new CMOS VLSI circuit technique for real-time median filtering of video images corrupted by pulse noise due to imperfection in the channel or receiver. The main idea of the processing is based on Cellular Neural Network paradigm and enables to perform filtering in two modes, average and median. The validity of the approach has been demonstrated by experimental verification of noise removing from computer generated images performed by the chip manufactured in CMOS 2.4 (μm n-well technology). The processing speed of the proposed approach is similar to the other solutions known in the literature and the power consumption and silicon area occupation makes the approach competitive with the chips proposed so far. This enables to apply the proposed method in an advanced real-time video image filtering in different wireless communication systems including radio and TV broadcasting and internet.

Tabela 2

6. REFERENCES

1. G.R. Arce, N.C. Gallagher: *State Description for the Root-Signal Set of Median Filters*. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-30, No. 6, 1982, pp. 894-902.
2. G.R. Arce, R.L. Stevenson: *On the Synthesis of Median Filter Systems*. IEEE Transactions on Circuits and Systems, vol. CAS-34, No. 4, 1987, pp. 420-429.
3. J. Astola, P. Heinonen, Y. Neuvo: *Linear Median Hybrid Filters*. IEEE Transactions on Circuits and Systems, vol. 36, No. 11, 1989, pp. 1430-1437.
4. I. Pitas, A.N. Venetsanopoulos: *Nonlinear Digital Filters*. Boston, Kluwer Academic Publishers, 1990.
5. J. Astola, Y. Neuvo: *Matched Median Filtering*. IEEE Transactions on Communications, vol. 40, No. 4, 1992, pp. 722-729.
6. L.Y. Ruikang Yang, M. Gabbouj, Y. Neuvo: *Weighted Median Filters: a Tutorial*. IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing, vol. 43, no. 3, 1996, pp. 157-192.
7. N. Demassieux, F. Jutand, M. Dana: *A Low Cost Custom IC Real Time Image Median Filtering*. IEEE 1986 Custom Integrated Circuits Conference, Proc., pp. 166-169.
8. O. Vainio, P. Heinonen, Y. Neuvo: *Integrated FIR Median Hybrid Filter*. IEEE 1986 Custom Integrated Circuits Conference, Proc., pp. 170-172.
9. M. Karaman, L. Onural, A. Atalar: *Design and Implementation of a General-Purpose Median Filter Unit in CMOS VLSI*. IEEE Journal of solid-state Circuits, vol. 25, 1990, pp. 505-513.
10. R. Roncella, R. Saletti, P. Terreni: *70-MHz $2\mu\text{m}$ CMOS Bit-Level Systolic Array Median Filter*. IEEE Journal of Solid-State Circuits, vol. 28, 1993, pp. 530-535.
11. C.T. Chen, L.G. Chen, J.H. Hsiao: *VLSI Implementation of a Selective Median Filter*. IEEE Transactions on Consumer Electronics, vol. 42, No. 1, 1996, pp. 33-42.
12. P.H. Dietz, L.R. Carley: *Simple Networks for Pixel Plane Median Filtering*. IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing, vol. 40, 1993, pp. 799-801.
13. A. Diaz-Sanchez, J. Ramirez-Angulo, A. Lopez, E. Sanchez-Sinencio: *A Parallel Analog Median Filter*. IEEE International Conference on Electronics, Circuits and Systems, Vol. 1, Lisboa, Portugal, 7-10 September, 1998, pp. 381- 384.

14. B. E. Shi: *Order Statistic Filtering with Cellular Neural Networks*. CNNA-94 Third IEEE International Workshop on Cellular Neural Networks and their Applications, Rome, Italy, December 18-21, 1994, pp. 441-443.
15. J. Kowalski, T. Kacprzak, K. Ślot: *VLSI Implementation of Analog Image Median Filter with Average Filter Option Based on Cellular Neural Network Architecture*. XXI National Conference on Circuit Theory and Electronic Networks, Poznań-Kiekrz, Poland, October 22-24, 1998, Proc. vol. 2, pp. 643-648.
16. K. Ślot, J. Kowalski, A. Napieralski, T. Kacprzak: *Analogue median/average image filter based on cellular neural network paradigm*. Electronics Letters, 16th September 1999, vol. 35, No. 19, pp. 1619-1620.
17. C. Das: *MIETEC 2 μ m CMOS MPC*. MIE/F/06, IMEC/EUROCHIP, March 1993.
18. I. E. Opris, G. T. Kovacs: *A High-Speed Median Circuit*. IEEE Journal of Solid-State Circuits, 1997, vol. 32, No. 6, pp. 905-908.
19. G. Fikos, S. Vlassis, S. Siskos: *High-speed, accurate analogue CMOS rank filter*. Electronics Letters, 30th March 2000, vol. 36, No. 7, pp. 593-594.
20. A. Paasio, K. Halonen: *An analogue circuit for weighted rank order filtering*. European Conference on Circuit Theory and Design, August 28-31, 2001, Espoo, Finland, pp. I-125 - I-128.
21. J. Kowalski, T. Kacprzak, K. Ślot: *Analogue VLSI Chip of Image Median/Average Filter Based on Cellular Neural Network Paradigm*. Proceedings of the International Conference on Signals and Electronic Systems -ICSSES'2001, Łódź, Poland, September 18 - 21, 2001, pp. 457 - 462.
22. J. Kowalski, T. Kacprzak, K. Ślot, P. Dębiec: *Functional Tests Results of Analogue VLSI Chip of Image Median/Average Filter Based on Cellular Neural Network Paradigm*. Proceedings of the International Conference on Signals and Electronic Systems -ICSSES'2001, Łódź, Poland, September 18 - 21, 2001, pp. 221 - 226.

J. KOWALSKI, T. KACPRZAK, K. ŚLOT

NOWY UKŁAD SCALONY CMOS VLSI ANALOGOWEGO FILTRU MEDIANOWEGO DO APLIKACJI PRZETWARZANIA OBRAZU W CZASIE RZECZYWISTYM

Streszczenie

W artykule przedstawiono układ scalony CMOS VLSI analogowego filtru medianowego o architekturze sieci neuronowej komórkowej (SNK) do zadań przetwarzania obrazu w czasie rzeczywistym. Układ ten został zrealizowany w technologii CMOS MIETEC 2.4 μ m i jest zdolny przetwarzać obraz z rozdzielczością 64 pikseli na linię obrazu. Aby zwielokrotnić rozdzielczość przetwarzanego obrazu można wykorzystać jednocześnie kilka układów scalonych, łącząc je w wykorzystaniem specjalnych wprowadzeń. Dodatkową funkcją układu scalonego jest możliwość pracy jako filtru uśredniającego obrazu. W pracy przedstawiono podstawy teoretyczne filtrów medianowych o architekturze SNK, opis implementacji CMOS wszystkich bloków funkcjonalnych układu scalonego filtru medianowego oraz jego architekturę. Zamieszczono wyniki symulacji i pomiarów podstawowych bloków CMOS filtru. Posługując się specjalnie wykonaną do tego celu kartą do komputera PC wykonano badania funkcjonalne filtru medianowego dla przykładowych obrazów zakłóconych szumem impulsowym typu "sól i pieprz". Przedstawiono oscylogramy filtrowanych sygnałów oraz obrazy będące wynikiem pracy układu scalonego filtru medianowego. Zamieszczono tabele, które przedstawiają podstawowe parametry układu oraz porównanie do istniejących już implementacji cyfrowych filtrów medianowych.

Słowa kluczowe: sieci neuronowe komórkowe, filtry nieliniowe, przetwarzanie obrazów, ASIC

Szeregowanie niezależnych, wywłaszczalnych i periodycznych zadań wieloprocessorowych dla trzech procesorów

MIROSLAW GAJER

*Katedra Automatyki AGH
Al. Mickiewicza 30, 30-059 Kraków
email: mgajer@ia.agh.edu.pl*

*Otrzymano 2003.05.21
Autoryzowano 2003.09.18*

W artykule poruszono tematykę szeregowania zbioru niezależnych, wywłaszczalnych i periodycznych zadań wieloprocessorowych przeznaczonych do realizacji w systemie wieloprocessorowym zbudowanym z trzech procesorów. Rozważane zadania niezależne, wywłaszczalne i periodyczne stanowią najbardziej popularną klasę zadań występujących w komputerowych systemach czasu rzeczywistego. Do szeregowania zbioru tego typu zadań wykorzystywany jest najczęściej algorytm RMS (ang. *Rate Monotonic Scheduling*). Równie często stosowaną techniką w połączeniu z algorytmem RMS jest binaryzacja okresów zadań, mająca na celu skrócenie długości horyzontu czasowego systemu. Dotychczas algorytm RMS stosowany był wyłącznie do szeregowania zadań jednoprocessorowych. Natomiast autorowi artykułu nie jest wiadomo o próbach zastosowania algorytmu RMS do szeregowania zadań wieloprocessorowych. Stąd artykuł niniejszy stanowi propozycję takiej modyfikacji algorytmu RMS, która pozwoliłaby na jego zastosowanie również do szeregowania zadań wieloprocessorowych. Zamieszczone w artykule rozważania zostały zilustrowane na przykładzie szeregowania zbioru niezależnych, wywłaszczalnych i periodycznych zadań wieloprocessorowych przeznaczonych dla trzech procesorów. Rozważono zarówno przypadek procesorów dedykowanych, jak i arbitralnych.

Słowa kluczowe: szeregowanie zadań, zadania wieloprocessorowe, systemy czasu rzeczywistego, systemy wieloprocessorowe.

1. WPROWADZENIE

Systemy czasu rzeczywistego (ang. *real-time systems*) stanowią obecnie już dobrze zdefiniowaną i wyszczególnioną klasę systemów komputerowych [1], [2], [3], [4]. Istnieją trzy podstawowe czynniki odróżniające systemy czasu rzeczywistego od

systemów komputerowych ogólnego przeznaczenia, do których zaliczyć można czas obliczeń, bezpieczeństwo pracy systemu oraz bardzo silne powiązanie systemu komputerowego ze sterowanym przez niego obiektem [5]. W systemie czasu rzeczywistego nie wystarczy, że uzyskane wyniki obliczeń są poprawne pod względem ich analizy logicznej, ponieważ wyniki te muszą być jeszcze dostarczone na czas. W systemie o ostrych ograniczeniach czasowych (ang. *hard real-time*) wyniki poprawne lecz spóźnione są całkowicie bezużyteczne. Ponadto w przypadku sterowania złożonym obiektem technicznym takim, jak reaktor jądrowy, system elektroenergetyczny, samolot, rakieta kosmiczna, naruszenie ograniczeń czasowych realizowanych w systemie zadań może doprowadzić do katastrofy, pociągającej za sobą wielkie straty ekonomiczne i utratę życia ludzkiego. Z tego powodu systemy czasu rzeczywistego muszą być bardzo starannie projektowane, a przeprowadzona na etapie projektowania analiza formalna ma zagwarantować ich deterministyczne zachowanie we wszelkich możliwych do przewidzenia sytuacjach. Takie postawienie problemu wymusiło intensywny rozwój metod szeregowania zadań realizowanych w systemach czasu rzeczywistego.

Celem szeregowania zadań jest określenie w jakim momencie czasu, które z zadań może być realizowane, tak aby dla wszystkich zadań zostały dochowane ich ograniczenia czasowe (ang. *deadlines*) [6]. Najbardziej popularnym typem zadań występujących w systemach czasu rzeczywistego są zadania niezależne, wywłaszczalne i okresowe. Zadania takie są aktywowane w regularnie następujących po sobie momentach, oddległych od siebie o przedział czasu zwany okresem zadania. Kolejność realizacji poszczególnych zadań jest dowolna — stąd zadania takie nazywa się niezależnymi. Ponadto realizacja każdego z zadań może zostać przerwana w dowolnym momencie i później zostać wznowiona — stąd zadania takie nazywa się wywłaszczalnymi.

Najczęściej każde z zadań realizuje wybrany cel pracy komputerowego systemu sterowania w czasie rzeczywistym. To znaczy, na początku zadanie odczytuje wartości pochodzące z odpowiednich czujników, które umieszczone są na obiekcie sterowania, następnie wykonuje obliczenia związane z algorytmem sterowania, po czym wysyła sygnały sterujące do odpowiednich elementów wykonawczych, które mogą w sposób bezpośredni wpływać na zachowanie obiektu. Wymienione czynności powtarzane są cyklicznie — stąd nazwa zadania okresowe.

Najbardziej popularnym algorytmem służącym do szeregowania zbioru niezależnych, wywłaszczalnych i periodycznych zadań jest algorytm RMS (ang. *Rate Monotonic Scheduling*) [7].

2. PODSTAWOWE WŁAŚCIWOŚCI ALGORYTMU RMS

Działanie algorytmu RMS opiera się na kilku prostych regułach [8]. Po pierwsze, zadaniom przypisane zostają priorytety. Wysokości priorytetów są odwrotnie proporcjonalne do wartości okresów zadań. To znaczy, im zadanie posiada krótszy okres, czyli im aktywowane jest częściej, tym ma wyższy priorytet. Po drugie, w danym momencie wykonywane jest to zadanie, które znajduje się w stanie gotowości (ang.

ready)
najwyż
szym
tego z
jedyne
nie bę
Ni
nych. M
jego o
pracy
szerego
z zadan
W
zadań
go na
większ
koniec

Na
gowałn
następu

DL
0,69. Z
leżnych
pewnoś
szerego
Jab
aby zb
Nie jes
przesąc
może b
Ab
ale nie
zakońc
aktyw
z zadan

ready) oraz spośród wszystkich zadań będących w stanie gotowości posiada aktualnie najwyższy priorytet. Po trzecie, wejście w stan gotowości zadania o priorytecie wyższym niż zadanie aktualnie wykonywane powoduje natychmiastowe wywłaszczenie tego zadania. Wreszcie po czwarte, wywłaszczone zadanie może zostać wznowione jedynie w przypadku, gdy w danym momencie żadne z zadań o priorytecie wyższym nie będzie znajdowało się w stanie gotowości.

Niech będzie zatem dany zbiór N niezależnych, wywłaszczalnych zadań periodycznych. Niech dla każdego z zadań z_i znany będzie jego czas realizacji C_i oraz wartość jego okresu T_i . Należy zauważyć, że iloraz $\frac{C_i}{T_i}$ stanowi miarę wykorzystania czasu pracy procesora przez zadanie z_i . Powstaje pytanie, kiedy taki zbiór zadań będzie szeregowalny, to znaczy, kiedy ograniczenia czasowe będą dochowane dla każdego z zadań?

Warunek konieczny na szeregowalność zbioru N niezależnych, wywłaszczalnych zadań periodycznych jest bezpośrednią konsekwencją oczywistego faktu, polegającego na tym, że rozważany zbiór zadań nie może wykorzystywać procesora w stopniu większym niż 100%. Otrzymuje się zatem następującą nierówność stanowiącą warunek konieczny na szeregowalność zbioru zadań:

$$\sum_{i=1}^N \frac{C_i}{T_i} \leq 1 \quad (1)$$

Natomiast w roku 1973 Liu i Layland podali warunek wystarczający na szeregowalność zbioru N niezależnych, wywłaszczalnych zadań periodycznych, który ma następującą postać [8]:

$$\sum_{i=1}^N \frac{C_i}{T_i} \leq N \left(2^{\frac{1}{N}} - 1 \right) \quad (2)$$

Dla dużych wartości N prawa strona nierówności (2) zmierza do wartości $\ln 2 \approx 0,69$. Zatem jeżeli łączne wykorzystanie czasu pracy procesora przez zbiór N niezależnych, wywłaszczalnych zadań periodycznych nie przekracza 69%, wówczas istnieje pewność, że w przypadku zastosowania algorytmu RMS rozważany zbiór zadań będzie szeregowalny.

Jak już wspomniano nierówność (2) podaje jedynie warunek wystarczający na to, aby zbiór N niezależnych, wywłaszczalnych zadań periodycznych był szeregowalny. Nie jest to natomiast warunek konieczny i jego niespełnienie jeszcze o niczym nie przesądza. Zbiór zadań, który wykorzystuje procesor w stopniu większym niż 69% może być nadal szeregowalny.

Aby się przekonać, czy zbiór zadań, dla którego spełniony jest warunek konieczny, ale nie jest spełniony warunek wystarczający jest szeregowalny, należy policzyć czasy zakończenia realizacji poszczególnych zadań w sytuacji, gdy wszystkie zadania zostaną aktywowane w tym samym momencie. Jeżeli uda się pokazać, że realizacja każdego z zadań zostanie zakończona przed upływem jego ograniczenia czasowego, wówczas

na mocy odpowiedniego twierdzenia [7] zbiór rozważanych zadań będzie zawsze szeregowalny, dla dowolnych czasów aktywacji poszczególnych zadań. Czas zakończenia realizacji zadania z_i obliczany jest za pomocą procedury iteracyjnej. Zakłada się, że równocześnie z zadaniem z_i zostaje aktywowanych $M - 1$ zadań o wyższych priorytetach, niż priorytet zadania z_i , a zatem zadania te zostaną wykonane najpierw. Pierwsze przybliżenie czasu zakończenia realizacji zadania z_i otrzymuje się ze wzoru:

$$t_0 = \sum_{i=1}^M C_i \quad (3)$$

Wzór (3) wyznacza sumę czasów realizacji poszczególnych zadań. Kolejne przybliżenia czasu realizacji zadania z_i otrzymuje się ze wzoru opisującego procedurę iteracyjną:

$$t_{k+1} = \sum_{i=1}^M C_i \left\lceil \frac{t_k}{T_i} \right\rceil \quad (4)$$

Występujące we wzorze (4) oznaczenie $\lceil x \rceil$ stanowi najmniejszą liczbę całkowitą nie mniejszą niż x .

Warunkiem kończącym realizację procedury iteracyjnej jest spełnienie równości

$$t_{k+1} = t_k \quad (5)$$

W przypadku spełnienia równości (5), jako czas zakończenia realizacji zadania z_i przyjmuje się czas t_k . Jeżeli czas t_k jest nie większy niż wartość okresu T_i zadania z_i , wówczas oznacza to, że zadanie z_i zawsze dochowa swoje ograniczenie czasowe. Rozważaną procedurę należy powtórzyć dla każdego z zadań i jeśli uda się pokazać, że żadne z zadań nie narusza swojego ograniczenia czasowego, wówczas świadczy to o tym, że rozważany zbiór N niezależnych, wywłaszczalnych zadań periodycznych jest szeregowalny w przypadku zastosowania algorytmu RMS.

W przypadku implementacji algorytmu RMS często stosowaną techniką jest binaryzacja okresów zadań [9]. Dzięki binaryzacji okresów w systemie występują wyłącznie zadania, których okresy stanowią wielokrotność, będącą całkowitą potęgą liczby 2, pewnego okresu podstawowego r . Dzięki binaryzacji wydatnemu skróceniu ulega tzw. horyzont czasowy zbioru szeregowanych zadań. Horyzont czasowy jest równy najmniejszej wspólnej wielokrotności okresów szeregowanych zadań i po binaryzacji staje się po prostu równy wielkości najdłuższego z okresów [10].

W przypadku dokonywania binaryzacji okresów zadań o okresie bazowym r , wartości okresów poszczególnych zadań ulegają transformacji zgodnie ze wzorem:

$$T'_i = r \cdot 2^{\lceil \log_2(\frac{T_i}{r}) \rceil}, \quad \text{dla } i = 1, 2, \dots, N \quad (6)$$

Występujące we wzorze (6) oznaczenie $\lceil x \rceil$ stanowi największą liczbę całkowitą nie większą niż x .

W w
szeniu. C
rzystania
tak zdar
okresów

Poni
zadań me
binaryza
stania pr
która zap
binaryza
Dla
okresów

istnieje d
dla binar

Wyst
niemniejs

Jako
ra zapew
szeregow

W kl
szeregow
nych wył
jest znany
cesorowy

Zadan
wymaga j
konieczny
wcześniej
wanych, lu
dla proces

Rosną
nich latach
zadań wie
cje zawart
zostało po

wsze sze-
kończenia
da się, że
prioryte-
Pierwsze

W wyniku przeprowadzenia binaryzacji okresy niektórych z zadań ulegają zmniejszeniu. Oznacza to, że zadania te będą aktywowane częściej, a zatem stopień wykorzystania przez nie procesora wzrośnie. W pewnym szczególnym przypadku może się tak zdarzyć, że zbiór zadań który był szeregowalny, po przeprowadzeniu binaryzacji okresów zadań szeregowalność utracił.

(3)

jne przy-
procedurę

Ponieważ okres bazowy r może zostać wybrany dowolnie, binaryzacji okresów zadań można dokonać na nieskończenie wiele sposobów. Ponieważ każda z możliwych binaryzacji okresów zadań powoduje w różnym stopniu zwiększenie stopnia wykorzystania procesora przez zbiór szeregowanych zadań, należy wybrać taką binaryzację, która zapewni możliwie najmniejszy wzrost stopnia wykorzystania procesora. Taką binaryzację optymalną poszukuje się pośród tzw. binaryzacji istotnych [10].

Dla zbioru N niezależnych, wywłaszczalnych zadań periodycznych o wartościach okresów spełniających relację

(4)

$$T_1 < T_2 < T_3 < \dots < T_{N-1} < T_N \quad (7)$$

całkowitą

istnieje dokładnie N binaryzacji istotnych. Wartości okresów bazowych $r_1, r_2, \dots, r_N$ dla binaryzacji istotnych obliczane są ze wzoru [11]:

ówności

(5)

$$r_i = \frac{T_i}{2^{\lceil \log_2 \left(\frac{T_i}{r_1} \right) \rceil}}, \quad \text{dla } i = 1, 2, \dots, N \quad (8)$$

zadania z_i z_i zadania

czasowe.

pokażać,

wiadczy to

znych jest

Występujące we wzorze (8) oznaczenie $[x]$ stanowi najmniejszą liczbę całkowitą nie mniejszą niż x .

Jako binaryzacja optymalna wybierana jest spośród binaryzacji istotnych ta, która zapewnia najmniejszy wzrost stopnia wykorzystania czasu pracy procesora przez szeregowany zbiór zadań.

W klasycznej teorii szeregowania zadań algorytm RMS wykorzystywany jest do szeregowania zbioru niezależnych, wywłaszczalnych zadań okresowych przeznaczonych wyłącznie dla jednego procesora. Natomiast autorowi niniejszego artykułu nie jest znany przypadek zastosowania algorytmu RMS do szeregowania zadań wieloprocessorowych.

Zadaniem wieloprocessorowym nazywane jest takie zadanie, które do swej realizacji wymaga jednoczesnej dostępności co najmniej dwóch procesorów. Zbiór procesorów koniecznych do realizacji pewnego zadania może zostać zdefiniowany odpowiednio wcześniej — wówczas mówi się o zadaniach przeznaczonych dla procesorów dedykowanych, lub też może być dowolny — wówczas mówi się o zadaniach przeznaczonych dla procesorów arbitralnych [12].

Rosnąca nieustannie popularność rozwiązań wieloprocessorowych wymusiła w ostatnich latach intensywny rozwój prac teoretycznych ukierunkowanych na szeregowanie zadań wieloprocessorowych, czego wyrazem są między innymi następujące publikacje zawarte w specjalnym wydaniu miesięcznika *Parallel Computing*, które w całości zostało poświęcone tej tematyce [13]–[18].

(6)

całkowitą

W artykule niniejszym autor przedstawił opracowaną przez siebie oryginalną metodę polegającą na wykorzystaniu algorytmu RMS do szeregowania zbioru niezależnych, wywłaszczalnych i okresowych zadań wieloprocessorowych przeznaczonych dla trzech procesorów, przy czym uwzględniony został zarówno przypadek procesorów arbitralnych, jak i dedykowanych.

3. SZEREGOWANIE ZADAŃ DLA TRZECH PROCESORÓW Z WYKORZYSTANIEM ALGORYTMU RMS

Niech dany będzie zatem zbiór trzech procesorów $P = \{P_1, P_2, P_3\}$ oraz zbiór zadań Z przewidzianych do realizacji na dedykowanych procesorach. Zbiór zadań może zostać podzielony na rozłączne podzbiory zadań trójprocesorowych, dwuprocessorowych oraz jednoprocessorowych $Z = \{Z^{123}, Z^{12}, Z^{13}, Z^{23}, Z^1, Z^2, Z^3\}$. Górny indeks zawiera numery dedykowanych procesorów, które są wymagane do realizacji danego zadania. Na przykład symbol Z^{23} oznacza, że do realizacji rozważanego zbioru zadań wymagana jest dostępność procesorów P_2 i P_3 . Zbiór zadań Z^3 składa się z pewnej liczby zadań tego samego typu, tzn. takich, do realizacji których wymagany jest identyczny podzbiór procesorów. Na przykład zbiór zadań Z^3 może składać się z K zadań $Z^{23} = \{z_1^{23}, z_2^{23}, \dots, z_K^{23}\}$, przy czym indeks dolny wskazuje kolejny numer poszczególnego zadania.

Zagadnienia szeregowania zbioru niezależnych, wywłaszczalnych i okresowych zadań wieloprocessorowych przeznaczonych dla trzech dedykowanych procesorów zostaną zilustrowane na następującym przykładzie.

Niech zbiór zadań trójprocesorowych składa się z czterech zadań $Z^{123} = \{z_1^{123}, z_2^{123}, z_3^{123}, z_4^{123}\}$, których charakterystyki (wartości czasów wykonania oraz okresów zostały zamieszczone w tab. 1.

Tabela 1

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1, P_2 i P_3

The basic characteristics of three-processor tasks performed on the processors P_1, P_2 and P_3

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^{123}	123	3	0,024
z_2^{123}	245	4	0,016
z_3^{123}	437	5	0,011
z_4^{123}	472	9	0,019

Z kolei niech zbiór zadań dwuprocessorowych przeznaczonych dla procesorów P_1 i P_2 składa się z pięciu zadań $Z^{12} = \{z_1^{12}, z_2^{12}, z_3^{12}, z_4^{12}, z_5^{12}\}$, których charakterystyki zostały zamieszczone w tab. 2.

Tabela 2

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1 i P_2

The basic characteristics of three-processor tasks performed on the processors P_1 and P_2

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_3^{12}	268	7	0,026
z_4^{12}	229	3	0,013
z_5^{12}	441	2	0,005

Niech zbiór zadań dwuprocessorowych przeznaczonych dla procesorów P_1 i P_3 składa się z sześciu zadań $Z^{13} = \{z_1^{13}, z_2^{13}, z_3^{13}, z_4^{13}, z_5^{13}, z_6^{13}\}$, których charakterystyki zostały zamieszczone w tab. 3.

Tabela 3

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1 i P_3

The basic characteristics of three-processor tasks performed on the processors P_1 and P_3

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_4^{13}	267	5	0,018
z_5^{13}	489	12	0,025
z_6^{13}	501	24	0,048

Ponadto niech zbiór zadań dwuprocessorowych przeznaczonych dla procesorów P_2 i P_3 składa się z sześciu zadań $Z^{23} = \{z_1^{23}, z_2^{23}, z_3^{23}, z_4^{23}, z_5^{23}, z_6^{23}\}$, których charakterystyki zostały zamieszczone w tab. 4.

Tabela 4

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_2 i P_3

The basic characteristics of three-processor tasks performed on the processors P_2 and P_3

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_4^{23}	296	15	0,051
z_5^{23}	476	5	0,011
z_6^{23}	523	7	0,013

Z kolei niech zbiór zadań jednoprocessorowych przeznaczonych dla procesora P_1 składa się z siedmiu zadań $Z^1 = \{z_1^1, z_2^1, z_3^1, z_4^1, z_5^1, z_6^1, z_7^1\}$, których charakterystyki zostały zamieszczone w tab. 5.

Tabela 5

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1

The basic characteristics of three-processor tasks performed on the processors P_1

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_4^1	250	11	0,044
z_5^1	264	12	0,045
z_6^1	547	4	0,007
z_7^1	556	6	0,011

Podobnie niech zbiór zadań jednoprocessorowych przeznaczonych dla procesora P_2 składa się z ośmiu zadań $Z^2 = \{z_1^2, z_2^2, z_3^2, z_4^2, z_5^2, z_6^2, z_7^2, z_8^2\}$, których charakterystyki zostały zamieszczone w tab. 6.

Tabela 6

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_2

The basic characteristics of three-processor tasks performed on the processors P_2

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_4^2	325	2	0,006
z_5^2	456	13	0,029
z_6^2	478	8	0,017
z_7^2	486	2	0,004
z_8^2	497	3	0,006

Analogicznie niech zbiór zadań jednoprocessorowych przeznaczonych dla procesora P_3 składa się z ośmiu zadań $Z^3 = \{z_1^3, z_2^3, z_3^3, z_4^3, z_5^3, z_6^3, z_7^3, z_8^3\}$, których charakterystyki zostały zamieszczone w tab. 7.

Łączna wartość współczynnika wykorzystania czasu pracy systemu wieloprocessorowego η przez wszystkie zadania podlegające procesowi szeregowania liczona jest jako suma stopni wykorzystania procesorów przez poszczególne zadania C_i/T_i , z tym tylko zastrzeżeniem, że w przypadku zadań dwuprocessorowych wyliczona dla nich wartość współczynnika C_i/T_i mnożona jest przez dwa. Analogicznie w przypadku zadań trójprocesorowych wartość współczynnika C_i/T_i mnożona jest przez trzy.

W rozważanym przypadku współczynnik wykorzystania czasu pracy systemu wieloprocessorowego wynosi $\eta = 1,793$. Należy w tym miejscu zauważyć, że maksymalna wartość współczynnika może wynosić dla systemu trójprocesorowego 3. Jest to warunek konieczny na to, aby zbiór wieloprocessorowych zadań okresowych był szeregowalny. W przypadku rozważanego zbioru zadań (tab. 1–7) warunek ten jest spełniony.

Tabela 5

Tabela 7

Podstawowe charakterystyki zadań tróiprocessorowych przeznaczonych dla procesorów P_3

The basic characteristics of three-processor tasks performed on the processors P_3

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_4^3	219	8	0,037
z_5^3	311	7	0,023
z_6^3	313	11	0,035
z_7^3	357	6	0,017
z_8^3	589	3	0,005

procesora
erystyki

Tabela 6

2

W kroku kolejnym dla zbioru wartości okresów zadań występujących w systemie wyliczane są wartości okresów bazowych, zgodnie ze wzorem (8). Zbiór okresów zadań po jego posortowaniu od wartości najmniejszej do największej ma postać: {120, 123, 125, 133, 137, 139, 142, 145, 151, 156, 162, 187, 190, 219, 222, 229, 245, 250, 264, 267, 268, 289, 296, 300, 311, 313, 325, 357, 437, 441, 456, 472, 476, 478, 486, 489, 497, 501, 523, 547, 556, 589}. Z kolei wyliczony dla rozważanego zbioru okresów zadań zbiór okresów bazowych wygląda następująco: {120; 61,5; 62,5; 66,5; 68,5; 69,5; 71; 72,5; 75,5; 78; 81; 93,5; 95; 109,5; 111; 114,5; 61,25; 62,5; 66; 66,75; 67; 72,25; 74; 75; 77,75; 78,25; 81,25; 89,25; 109,25; 110,25; 114; 118; 119; 60,75; 61,125; 62,125; 62,625; 65,375; 68,375; 69,5; 73,625}.

W tab. 8 zamieszczono wyniki binaryzacji zbioru zadań dla każdego z okresów bazowych. Drukiem pogrubionym zaznaczono binaryzację, która powoduje najmniejsze zwiększenie współczynnika wykorzystania systemu wieloprocessorowego η .

Podstawowe charakterystyki zadań po przeprowadzeniu binaryzacji nr 6 zostały zamieszczone w tab. 9–15.

procesora
erystyki

oproce-
ona jest
, z tym
ich war-
u zadań

mu wie-
symalna
t to wa-
szerego-
ełniony.

W kroku kolejnym wszystkie zadanie o identycznym okresie łączone są w jedno tzw. superzadanie. Na przykład następujące zadania o wartości okresu 69,5 są łączone w jedno superzadanie: z_1^{123} , z_2^{12} , z_1^{13} , z_1^2 , z_1^1 , z_1^{23} . Sposób połączenia wymienionych zadań w superzadanie został zilustrowany na rys. 1.

Podobnie należy postąpić dla każdego z okresów zadań, tzn. dla okresu 139, 278 oraz 556. W ten sposób otrzymuje się cztery superzadania, których charakterystyki zostały zamieszczone w tab. 16.

Ponieważ sumaryczny stopień wykorzystania jednostki obliczeniowej (złożonej z trzech procesorów) wynosi 1,2. Rozważane superzadania nie są szeregowalne za pomocą algorytmu RMS, ponieważ nie jest spełniony warunek konieczny na szeregowanie zbioru tych zadań (1).

Zupełnie inaczej wygląda sytuacja w przypadku, gdy szeregowane zadania nie będą przypisane na sztywno do dedykowanych procesorów. Taki przypadek szeregowania nazywany jest szeregowaniem z procesorami arbitralnymi. Charakterystyki superzadań dla przypadku arbitralnych procesorów zostały zamieszczone w tab. 17.

Tabela 8

Zestawienie wartości współczynników wykorzystania systemu wieloprocesorowego η uzyskanych dla binaryzacji o różnych okresach bazowych r

The set of multiprocessor system usage coefficients η obtained for binarizations of different base periods r

Numer binaryzacji	Wartość okresu bazowego r	Wartość współczynnika wykorzystania systemu wieloprocesorowego η
1	120	2,517
2	61,5	2,696
3	62,5	2,339
4	66,5	2,824
5	68,5	2,432
6	69,5	2,252
7	71	2,329
8	72,5	2,718
9	75,5	2,595
10	78	2,285
11	81	2,751
12	93,5	2,252
13	95	2,329
14	109,5	2,252
15	111	2,329
16	114,5	2,824
17	61,25	2,824
18	62,5	2,339
19	66	2,915
20	66,75	2,758
21	67	2,696
22	72,25	2,513
23	74	2,915
24	75	2,696
25	77,75	2,281
26	78,25	2,274
27	81,25	2,945
28	89,25	2,945
29	109,25	2,294
30	110,25	2,274
31	114	2,915
32	118	2,285
33	119	2,329
34	60,75	2,976
35	61,125	2,893
36	62,125	2,482
37	62,625	2,294
38	65,375	2,936
39	68,375	2,457
40	69,5	2,252
41	73,625	2,938

Tabela 8

Tabela 9

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1 , P_2 i P_3 ,
po binaryzacji ich okresów

The basic characteristics of three-processor tasks performed on the processors P_1 , P_2 and P_3 ,
after binarisation of their periods

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^{123}	69,5	3	0,043
z_2^{123}	139	4	0,029
z_3^{123}	278	5	0,018
z_4^{123}	278	9	0,032

Tabela 10

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1 i P_2 ,
po binaryzacji ich okresów

The basic characteristics of three-processor tasks performed on the processors P_1 and P_2 ,
after binarisation of their periods

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^{12}	139	5	0,036
z_2^{12}	69,5	2	0,029
z_3^{12}	139	7	0,050
z_4^{12}	139	3	0,022
z_5^{12}	278	2	0,007

Tabela 11

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1 i P_3 ,
po binaryzacji ich okresów

The basic characteristics of three-processor tasks performed on the processors P_1 and P_3 ,
after binarisation of their periods

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^{13}	69,5	8	0,115
z_2^{13}	139	6	0,043
z_3^{13}	139	4	0,029
z_4^{13}	139	5	0,036
z_5^{13}	278	12	0,043
z_6^{13}	278	24	0,086

Tabela 12

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_2 i P_3 ,
po binaryzacji ich okresów

The basic characteristics of three-processor tasks performed on the processors P_2 and P_3 ,
after binarisation of their periods

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^{23}	69,5	5	0,072
z_2^{23}	139	6	0,043
z_3^{23}	278	12	0,043
z_4^{23}	278	15	0,054
z_5^{23}	278	5	0,018
z_6^{23}	278	7	0,025

Tabela 13

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_1 ,
po binaryzacji ich okresów

The basic characteristics of three-processor tasks performed on the processors P_1 ,
after binarisation of their periods

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^1	69,5	6	0,086
z_2^1	139	8	0,057
z_3^1	139	7	0,050
z_4^1	139	11	0,079
z_5^1	139	12	0,086
z_6^1	278	4	0,014
z_7^1	556	6	0,011

W przypadku szeregowania zdań dla arbitralnych procesorów sumaryczny stopień wykorzystania jednostki obliczeniowej (złożonej z trzech procesorów) wynosi 0,88. Z kolei prawa strona nierówności (2) dla $N = 4$ jest równa 0,76. Zatem jest to sytuacja, w której jest spełniony warunek konieczny, ale nie jest spełniony warunek wystarczający. Aby uzyskać pewność, że rozważany zbiór superzadań jest szeregowalny algorytmem RMS, należy policzyć czas zakończenia każdego z superzadań, zgodnie z procedurą iteracyjną (4), a następnie porównać uzyskane wartości z ograniczeniami czasowymi tych zadań.

Cz
jest obl

t_2

Tabela 12

 P_3 ,

3,

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_2 ,
po binaryzacji ich okresów

The basic characteristics of three-processor tasks performed on the processors P_2 ,
after binarisation of their periods

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^2	69,5	3	0,043
z_2^2	139	5	0,036
z_3^2	278	9	0,032
z_4^2	278	2	0,007
z_5^2	278	13	0,047
z_6^2	278	8	0,029
z_7^2	278	2	0,007
z_8^2	278	3	0,011

Tabela 13

1,

Podstawowe charakterystyki zadań trójprocesorowych przeznaczonych dla procesorów P_3 ,
po binaryzacji ich okresów

The basic characteristics of three-processor tasks performed on the processors P_3 ,
after binarisation of their periods

zadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1^3	139	2	0,014
z_2^3	139	5	0,036
z_3^3	139	4	0,029
z_4^3	139	8	0,058
z_5^3	278	7	0,025
z_6^3	278	11	0,040
z_7^3	278	6	0,022
z_8^3	556	3	0,005

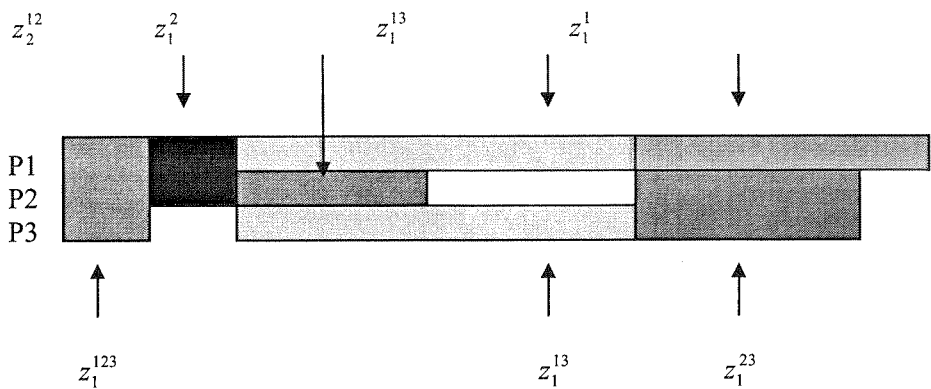
Tabela 14

Czas zakończenia superzadania o najniższym priorytecie, czyli superzadania Z_4
jest obliczany w sposób następujący:

$$t_0 = 18 + 48,7 + 72,3 + 6 = 145$$

$$t_1 = 18 \cdot \left\lceil \frac{145}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{145}{139} \right\rceil + 72,3 \cdot \left\lceil \frac{145}{278} \right\rceil + 6 \cdot \left\lceil \frac{145}{556} \right\rceil = 229,7$$

$$t_2 = 18 \cdot \left\lceil \frac{229,7}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{229,7}{139} \right\rceil + 72,3 \cdot \left\lceil \frac{229,7}{278} \right\rceil + 6 \cdot \left\lceil \frac{229,7}{556} \right\rceil = 247,7$$



Rys. 1. Superzadanie utworzone z połączenia sześciu różnych zadań o identycznym okresie równym 69,5.

Fig. 1. The supertask that was made by concatenation of six different tasks of the same period equal 69.5.

Tabela 16

Charakterystyki superzadań, które powstały po połączeniu zadań o takich samych wartościach okresów

The characteristics of the supertasks that were formed by joining together all the tasks with the same period

superzadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1	69,5	19	0,27
z_2	139	72	0,51
z_3	278	113	0,41
z_4	556	6	0,01

Tabela 17

Charakterystyki superzadań dla przypadku arbitralnych procesorów

The characteristics of the supertasks obtained for the case of arbitrary processors

superzadanie	okres zadania T_i [ms]	czas wykonania zadania C_i [ms]	współczynnik wykorzystania procesorów C_i/T_i
z_1	69,5	18	0,26
z_2	139	48,7	0,35
z_3	278	72,3	0,26
z_4	556	6	0,01

$$t_3 = 18 \cdot \left\lceil \frac{247,7}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{247,7}{139} \right\rceil + 72,3 \cdot \left\lceil \frac{247,7}{278} \right\rceil + 6 \cdot \left\lceil \frac{247,7}{556} \right\rceil = 247,7$$

Czas zakończenia realizacji superzadania Z_4 wynosi 247,7 co stanowi mniej niż okres tego zadania 556, czyli zadanie to dochowa zawsze swoje ograniczenie czasowe.

Z kolei czas zakończenia superzadania Z_3 jest wyznaczany ze wzorów:

$$t_0 = 18 + 48,7 + 72,3 + 6 = 145$$

$$t_1 = 18 \cdot \left\lceil \frac{139}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{139}{139} \right\rceil + 72,3 \cdot \left\lceil \frac{139}{278} \right\rceil = 157$$

$$t_2 = 18 \cdot \left\lceil \frac{157}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{157}{139} \right\rceil + 72,3 \cdot \left\lceil \frac{157}{278} \right\rceil = 223,7$$

$$t_3 = 18 \cdot \left\lceil \frac{223,7}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{223,7}{139} \right\rceil + 72,3 \cdot \left\lceil \frac{223,7}{278} \right\rceil = 241,7$$

$$t_4 = 18 \cdot \left\lceil \frac{241,7}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{241,7}{139} \right\rceil + 72,3 \cdot \left\lceil \frac{241,7}{278} \right\rceil = 241,7$$

Zatem czas zakończenia superzadania Z_3 wynosi 241,7 co stanowi mniej niż okres tego superzadania równy 278, czyli zadanie to jest szeregowalne.

W sposób analogiczny wyznaczony zostaje czas zakończenia superzadania Z_3 :

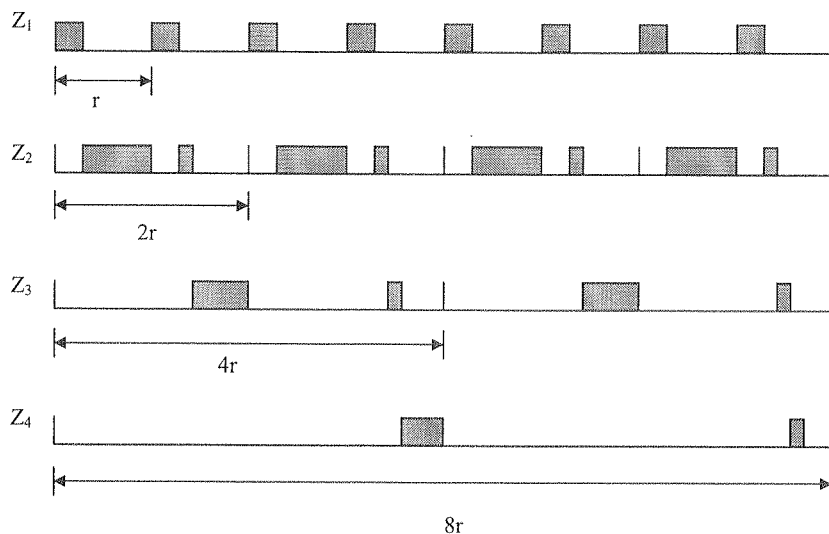
$$t = 18 + 48,7 = 66,7$$

$$t = 18 \cdot \left\lceil \frac{66,7}{69,5} \right\rceil + 48,7 \cdot \left\lceil \frac{66,7}{139} \right\rceil = 66,7$$

Super zadanie Z_3 zostanie zakończone po upływie 66,7 jednostek czasu, co stanowi mniej niż wartość okresu tego super zadania równa 139, czyli warunek szeregowalności jest w tym przypadku również spełniony.

Super zadanie Z_1 posiada najwyższy priorytet, czyli zostanie zawsze wykonane w pierwszej kolejności i zakończy się po upływie 18 jednostek czasu, co stanowi mniej niż jego okres równy 69,5. Wynika stąd, że cały zbiór superzadań jest szeregowalny.

Na rys. 2 przedstawiono w sposób poglądowy sposób szeregowania czterech zadań o okresach binarnych r , $2r$, $4r$ i $8r$.



Rys. 2. Ilustracja sposobu szeregowania czterech zadań o okresach binarnych.

Fig. 2. The illustration of the way in which four tasks with binary periods are scheduled.

4. ZAKOŃCZENIE

W artykule zamieszczono propozycję rozszerzenia obszaru stosowalności popularnego algorytmu RMS, służącego do szeregowania niezależnych, wywłaszczalnych i periodycznych zadań jednoprocessorowych na przypadek zbioru zadań wieloprocessorowych.

Zaproponowane podejście wykorzystuje binaryzację okresów zadań celem utworzenia tzw. superzadań będących połączeniem różnych zadań o takim samym okresie bazowym. Następnie otrzymane superzadania traktowane są jak zwykle zadania jednoprocessorowe, z tym, że wykorzystywana do ich realizacji jednostka obliczeniowa musi być klastrem złożonym z odpowiedniej liczby procesorów. Jednakże z punktu widzenia algorytmu RMS wewnętrzna organizacja jednostki obliczeniowej jest nieistotna i w związku z tym algorytm RMS może zostać użyty do szeregowania superzadań, w taki sam sposób jak gdyby były one zwykłymi zadaniami jednoprocessorowymi.

Rozważane podejście może zostać wykorzystane zarówno do szeregowania zadań wieloprocessorowych przeznaczonych dla procesorów dedykowanych, jak i arbitralnych. Należy zwrócić uwagę, że w przypadku dużej liczby zadań wieloprocessorowych M o takiej samej wartości okresu ich połączenie w superzadanie, w taki sposób, aby czas realizacji superzadania miał możliwie najmniejszą wartość, jest trudnym problemem optymalizacyjnym, ponieważ może istnieć nawet $M!$ sposobów uszeregowania rozważanych zadań wieloprocessorowych. Z tego powodu dalsze prace autora będą skupiały się na rozwiązaniu tego problemu, a do znalezienia planu szeregowania zadań wieloprocessorowych o najkrótszym czasie realizacji, w przypadku dużej liczby takich zadań,

1. T. Kral
2. T. Info
3. T. kow
4. T. Wyc
5. K. C and
6. K. port
7. Z a l part
8. S. 1997
9. C z kowe i pro
10. J. M Czas
11. W. the a pp. 5
12. O. mati
13. J. V mun
14. R. F com
15. A. Paral
16. K. for s pp. 4
17. C. sche
18. J. F Cono

zostaną zastosowane algorytmy genetyczne. Wybór ten podyktowany jest faktem, iż algorytmy genetyczne stanowią coraz częściej wykorzystywane uniwersalne i niezwykle wszechstronne narzędzie optymalizacyjne [19].

5. BIBLIOGRAFIA

1. T. Szmuc, G. Motet: *Specyfikacja i projektowanie oprogramowania czasu rzeczywistego*. Krakowskie Centrum Informatyki Stosowanej, Kraków, 1998.
2. T. Szmuc, G. Motet: *Programowanie systemów czasu rzeczywistego*. Krakowskie Centrum Informatyki Stosowanej, Kraków, 2000.
3. T. Szmuc: *Zaawansowane metody tworzenia oprogramowania systemów czasu rzeczywistego*. Krakowskie Centrum Informatyki Stosowanej, Kraków, 1998.
4. T. Szmuc: *Modele i metody inżynierii oprogramowania systemów czasu rzeczywistego*. Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków, 2001.
5. K.G. Shin, P. Ramanathan: *Real-time computing: A new discipline of computer science and engineering*. Proceedings of the IEEE, vol. 82, no. 1, January 1994, pp. 6-24.
6. K. Ramamritham, J.A. Stankovic: *Scheduling algorithms and operating systems support for real-time systems*. Proceedings of the IEEE, vol. 82, no. 1, January 1994, pp. 55-67.
7. Zalewski: *What every engineer needs to know about rate-monotonic scheduling: A tutorial*. Department of Computer Science, The University of Texas of the Permian Basin, Odessa, TX, 1995, pp. 321-335.
8. S. Tanenbaum: *Rozproszone systemy operacyjne*. Wydawnictwo Naukowe PWN, Warszawa, 1997.
9. Czajka, J. Nawrocki: *Szeregowanie zadań o okresach binarnych w systemach silnie uwarunkowanych czasowo*. I Krajowa Konferencja: Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim, Kraków, 1997, pp. 669-676.
10. J. Nawrocki, A. Czajka: *Binaryzacja okresów zadań cyklicznych*. VII Konferencja Systemy Czasu Rzeczywistego, Kraków, 2000, ss. 41-51.
11. W. Hsueh, K.J. Lin: *Scheduling real-time systems with end-to-end timing constraints using the distributed pinwheel model*. IEEE Transactions on Computers, vol. 50, no. 1, January, 2001, pp. 51-67.
12. O. Kwon, K.Y. Chwa: *Approximation algorithms for general parallel task scheduling*. Information Processing Letters 81 (2002), pp. 143-150.
13. J. Verriet: *Scheduling interval-ordered task with non-uniform deadlines subject to non-zero communications delays*. Parallel Computing 25, 1999, pp. 3-21.
14. R.H. Möhring, M.W. Schäffter: *Scheduling series-parallel orders subject to 0/1-communication delays*. Parallel Computing 25, 1999, pp. 23-40.
15. A. Munier: *Approximation algorithms for scheduling trees with general communication delays*. Parallel Computing 25, 1999, pp. 41-48.
16. K. Amoura, E. Bampis, Y. Manoussakis, Z. Tuza: *A comparison of heuristics for scheduling multiprocessor tasks on three dedicated processors*. Parallel Computing 25, 1999, pp. 49-61.
17. C. Boeres, V.E.F. Rebello: *A versatile cost modeling approach for multicomputer task scheduling*. Parallel Computing 25, 1999, pp. 63-86.
18. J. Błażewicz, M. Drozdowski, M. Markiewicz: *Divisible task scheduling — Concept and verification*. Parallel Computing 25, 1999, pp. 87-98.

19. Z. Kowalczyk: *Ewolucyjna optymalizacja wielokryterialna w automatyce i diagnostyce*. AUTOMATION 2003, Konferencja Naukowo-Techniczna Automatyzacja — Nowości i Perspektywy, Warszawa, 2–4 kwietnia 2003, pp. 23–38.

M. GAJER

THE SCHEDULING OF INDEPENDENT, PRE-EMPTIVE, AND PERIODIC MULTIPROCESSOR TASKS FOR THREE PROCESSORS

Summary

At the present moment real-time systems constitute a well-defined and separate class of computer systems. There exist three main factors that make the real-time systems to be totally different from the so-called general-purpose computer systems. First of all in the case of real-time systems the time of computations is the most precious resource, which must be used with a lot of care and attention. Even if the results of computations are correct but they are delivered too late, they are of no value in real-time system with hard time constraints. Second, any time constraint violation can make the controlled system unstable and thus can cause a catastrophe leading to great economic losses. And last but not least, real-time systems are in most cases integrated very closely with the object being controlled.

All these factors cause the necessity of formal analysis of real-time systems before they are put into operations. It must be proved that real-time systems will behave in a deterministic way under any possible conditions and in all the possible scenarios of its work. This necessity of formal analysis of real-time systems has caused the development of task scheduling theory and mathematical methods that allow prove in a formal way that time constraint will be met for each task under any circumstances.

The topic of the paper is about scheduling of the set of independent, pre-emptive, and periodic multiprocessor tasks that are to be performed in a multiprocessor system. Such independent, pre-emptive, and periodic tasks constitute the most popular class of the tasks that are to be met in real-time computer systems. In most cases the Rate Monotonic Scheduling (RMS) algorithm is used in order to schedule such set of tasks.

The rules of RMS algorithm are as follows:

- Each task is assigned a priority
- The shorter is the period of a task the higher priority is assigned to that task
- In every moment only the task with the highest priority from all the tasks that are ready for execution is performed
- The task can be pre-empted by any other task with a higher priority always when such a task enters into a ready state
- The pre-empted task can be resumed only when there is no other task with a higher priority that is ready for the execution

Very often a binarisation of the periods of tasks is used together with the RMS algorithm. The purpose of this is to make shorter the time horizon of the scheduled set of tasks. The time horizon is the least common multitude of the periods of all tasks. The binarisation of the periods of tasks transforms the periods in such a way that they are the multitudes of some base period. Only the multitudes that are the powers of two are allowable. According with this author knowledge, the RMS algorithm has been applied only for tasks that are executed on a single processor up till now. This author has found nothing about the trials of implementing the RMS algorithm also for scheduling multiprocessor tasks.

ce. AU-
ektywy,

The paper is a proposition of new approach in which the RMS algorithm is modified in such a way that it can also be applied for the purpose of scheduling a set of multiprocessor independent, pre-emptive, and periodic tasks. The proposition of using the RMS algorithm for multiprocessor task scheduling is the following. At the beginning the binarisation of the periods of tasks is performed. Then the tasks are concatenated into so called supertasks in such a way so that the total execution time of the supertasks was as short as possible. In the next step the supertasks are treated as normal uniprocessor tasks, so the RMS algorithm can be directly applied.

In the paper both the case of task for dedicated processors and for arbitrary processors is considered. These are illustrated by the scheduling of an example multiprocessor task set.

Keywords: Task Scheduling, Multiprocessor Tasks, Real-Time Systems, Multiprocessor Systems

computer
nt from
he time
ttention.
value in
he con-
And last
ng con-

are put
nder any
alysis of
methods
circum-

periodic
emptive,
computer
ule such

execution

k enters

y that is

nm. The
orizon is
ts trans-
ultitudes
gorithm
thor has
rocessor

nej w
w tra
szar
łączn
właś
mięć
wpły
leps
post
zosta

Słow

Prop
gęstą i w
stał się s
oraz tech
kalne roz
te zostały
propagac
Tak v
lotorowy

Zakłócenia bierne modelowane rozkładem Suzuki

ANDRZEJ JAKUBIAK, ARKADIUSZ CHRUSTOWSKI

*Wydział Elektroniki i Technik Informacyjnych Politechniki Warszawskiej
ul. Nowowiejska 15/19, 00-665 Warszawa
e-mail: a.jakubiak@rekt.pw.edu.pl*

Otrzymano 2003.03.11

Autoryzowano 2003.12.30

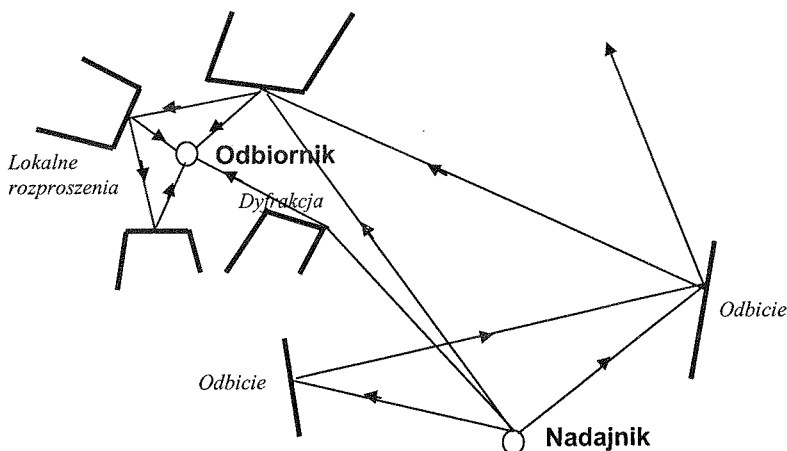
Zakłócenia bierne, powstające w wyniku przypadkowych odbić fali elektromagnetycznej w przestrzeni pomiędzy nadajnikiem a odbiornikiem, stanowią poważne utrudnienie w transmisji sygnałów drogą radiową. Zjawisko to występuje szczególnie intensywnie w obszarze zabudowy miejskiej, powodując istotne zakłócenia w systemach telefonii komórkowej, łączności radiowej służb specjalnych czy też w systemach radiolokacyjnych. Znalezienie właściwego modelu probabilistycznego (ze względu na losowy charakter zjawiska) pozwala między innymi na efektywne projektowanie układów eliminujących bądź minimalizujących wpływ tego typu zakłóceń. Opisany w artykule tak zwany model Suzuki jest jednym z najlepszych modeli, potwierdzonych empirycznie, tyle tylko, że ze względu na skomplikowaną postać analityczną trudny do aplikacji. Zaproponowany sposób symulacji komputerowej został przetestowany i pozwala na efektywne generowanie próbek zakłóceń Suzuki.

Słowa kluczowe: zakłócenia bierne, modelowanie sygnałów, generacja ciągów pseudolosowych

1. WPROWADZENIE

Propagacja fali elektromagnetycznej w obszarze miejskim, charakteryzującym się gęstą i wysoką zabudową, stanowi od dziesiątek lat przedmiot badań i analiz. Problem stał się szczególnie aktualny w momencie burzliwego rozwoju telefonii komórkowej oraz technik precyzyjnego namiaru radiolokacyjnego. Odbicia, załamania czy też lokalne rozproszenia powodują zakłócenia w odbiorze sygnału użytecznego. Zjawiska te zostały schematycznie zilustrowane na rys. 1, gdzie przedstawiono wielotorowość propagacji fali elektromagnetycznej między budynkami.

Tak więc kanał propagacji fali elektromagnetycznej w obszarze miejskim jest wielotorowy i zazwyczaj modelowany jako filtr liniowy [1]. Przyjmując, że $Re\{s(t)e^{j\omega_0 t}\}$



Rys. 1. Model propagacji wielotorowej w obszarze miejskim

Fig. 1. Multipath urban propagation model

jest sygnałem nadawanym oraz $Re\{y(t)e^{j\omega_0 t}\}$ jest sygnałem odbieranym, gdzie $s(t)$ i $y(t)$ reprezentują sygnał dolnopasmowy a ω_0 jest pulsacją nośną, można na podstawie rys. 1 zapisać:

$$y(t) = \sum_{k=0}^N x_k s(t - t_k) e^{j\theta_k} = X s(t) e^{j\theta}. \quad (1)$$

Właściwości propagacyjne ośrodka są charakteryzowane poprzez wielkości losowe x_k , t_k i θ_k . Zmienne x_k , t_k , θ_k reprezentują odpowiednio wartość amplitudy losowej, opóźnienie oraz fazę sygnału w pojedynczym torze. Założenia o właściwościach statystycznych tych zmiennych, a także o liczbie torów propagacji oraz lokalnych rozprośzeniach, determinują model probabilistyczny wypadkowej amplitudy X oraz fazy θ , będących multiplikatywną postacią zakłóceń biernych. w przypadku aktywnego systemu radiolokacyjnego zależność (1) modeluje odbierane zakłócenie bierne, będące wektorową sumą sygnałów odbitych od przypadkowych obiektów (budynków). Jeżeli chodzi o fazę θ , to powszechnie przyjmuje się założenie o jej rozkładzie równomiernym w przedziale $(0, 2\pi)$. Rozkład prawdopodobieństwa zmiennej X , reprezentującej moc zakłóceń, dąży do rozkładu Rayleigha przy dużej liczbie N — mówi się wówczas o tak zwanych zakłóceniach gaussowskich. Dane pomiarowe wskazują, że warunek „dużego N ”, konieczny do zastosowania centralnego twierdzenia granicznego, w większości przypadków nie jest spełniony — zakłócenia noszą wówczas nazwę niegaussowskich. Pośród wielu modeli probabilistycznych zakłóceń niegaussowskich szczególną rolę odgrywają rozkłady mieszane, w których zakłada się, że zmienna losowa X zależy od pewnego czynnika, będącego losowym parametrem funkcji gęstości prawdopodobieństwa tej zmiennej. Czynnikiem ten modeluje dodatkowe zjawiska losowe, towarzyszące propagacji wielotorowej, jak na przykład lokalne rozprośzenia. Oznaczając losowy parametr jako

σ można wyznaczyć funkcję gęstości prawdopodobieństwa zmiennej X albo z rozkładu łącznego, albo z rozkładu warunkowego, korzystając ze znanych zależności:

$$f(x) = \int_0^{+\infty} f(x, \sigma) d\sigma = \int_0^{+\infty} f(x|\sigma) f(\sigma) d\sigma \quad (2)$$

Przykładem takiego złożonego modelu zakłóceń biernych jest rozkład typu K, w którym zmienna losowa X ma warunkowy rozkład Rayleigha z losowym parametrem o rozkładzie gamma [2].

2. MODEL SUZUKI

$s(t)$ i $y(t)$
stawie rys.

(1)

losowe x_k ,
y losowej,
iach staty-
rozproszeń,
będących
emu radio-
wektorową
odzi o fazę
przedziale
óceń, dąży
nych zakłó-
konieczny
adków nie
śród wielu
ywają roz-
d pewnego
ieństwa tej
propagacji
ametr jako

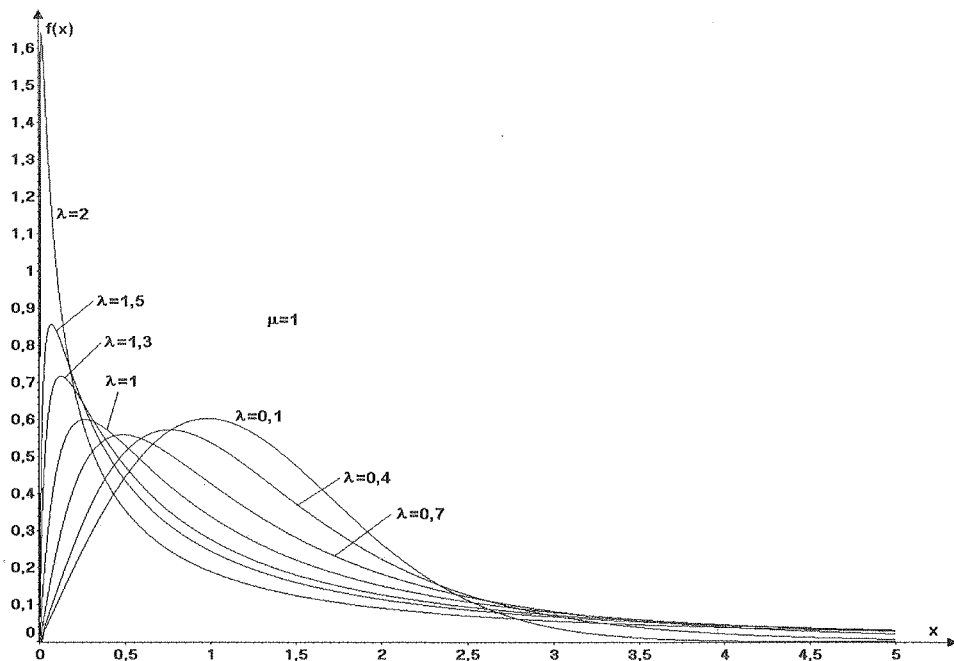
Jednym z modeli używanych do analizy amplitud zakłóceń biernych powstających podczas propagacji fal radiowych w zabudowie wielkomiejskiej jest rozkład zaproponowany przez Hirofumi Suzuki w [1]. Łączy on cechy dwóch innych rozkładów: log-normalnego i Rayleigha. Pierwszy z nich sprawdza się w sytuacji, gdy budynki są rozmieszczone niezbyt gęsto lub tworzą odseparowane grupy. Drugi dobrze modeluje amplitudy zakłóceń w obszarach bardzo gęsto zabudowanych, np. centrach miast, gdzie budynki są rozdzielone niewielkimi przestrzeniami, a fale radiowe ulegają wielokrotnym odbiciom. Rozkład Suzuki łączy opisane cechy, pozwalając zawrzeć w jednym modelu statystykę amplitud zakłóceń powstających w dowolnym typie zabudowy miejskiej, co ułatwia ich analizę w systemach telefonii komórkowej i radiolokacyjnych. Podobnie jak inne rozkłady używane do modelowania zakłóceń biernych, model Suzuki jest dwuparametrowy. Gęstość prawdopodobieństwa $f(x)$ i dystrybuenta $F(x)$ rozkładu Suzuki mają postać:

$$f(x) = \int_0^{\infty} \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma\lambda} \exp\left(-\frac{\log^2(\sigma/\mu)}{2\lambda^2}\right) d\sigma \quad (3)$$

$$F(x) = \int_0^x f(x) dx = \int_0^{\infty} \left(1 - \exp\left(-\frac{x^2}{2\sigma^2}\right)\right) \frac{1}{\sqrt{2\pi}\sigma\lambda} \exp\left(-\frac{\log^2(\sigma/\mu)}{2\lambda^2}\right) d\sigma \quad (4)$$

gdzie μ oraz λ są parametrami.

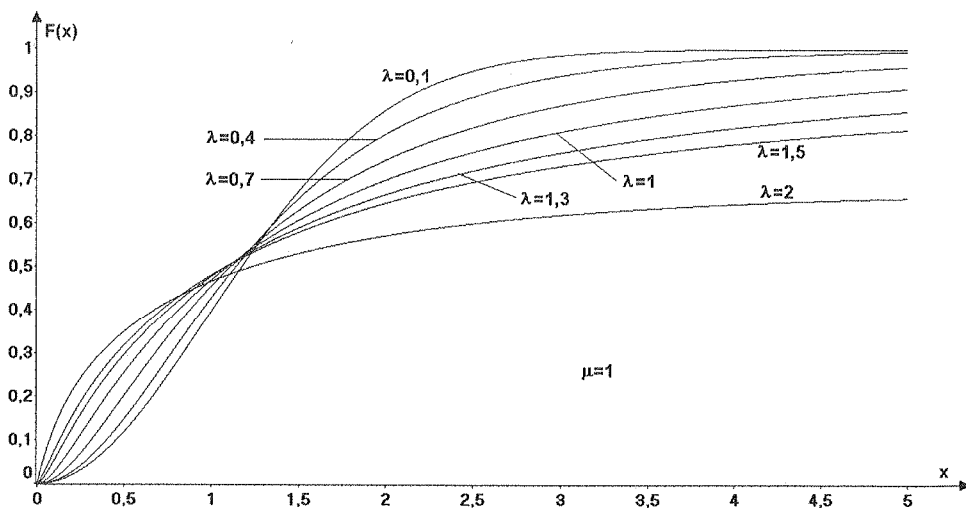
W zależności (3) rozkład warunkowy Rayleigh'a $\left(\frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)\right)$ modeluje rozpraszanie w grupie budynków, natomiast rozkład log-normalny $\left(\frac{1}{\sqrt{2\pi}\sigma\lambda} \exp\left(-\frac{\log^2(\sigma/\mu)}{2\lambda^2}\right)\right)$ charakteryzuje propagację ze źródła do grupy budynków,



Rys. 2. Funkcja gęstości $f(x)$ rozkładu Suzuki dla różnych wartości parametru λ przy ustalonym $\mu = 1$

Fig. 2. Suzuki probability density function $f(x)$ for different parameters λ and $\mu = 1$

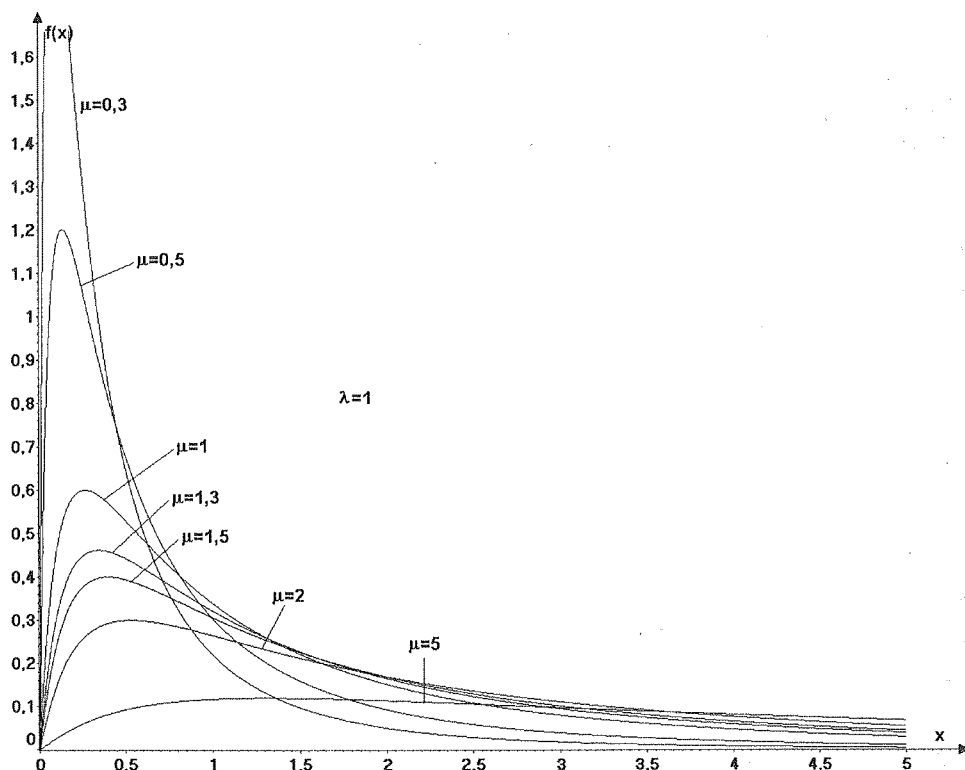
Rys. 4.



Rys. 3. Dystrybuanta $F(x)$ rozkładu Suzuki dla różnych wartości parametru λ przy ustalonym $\mu = 1$

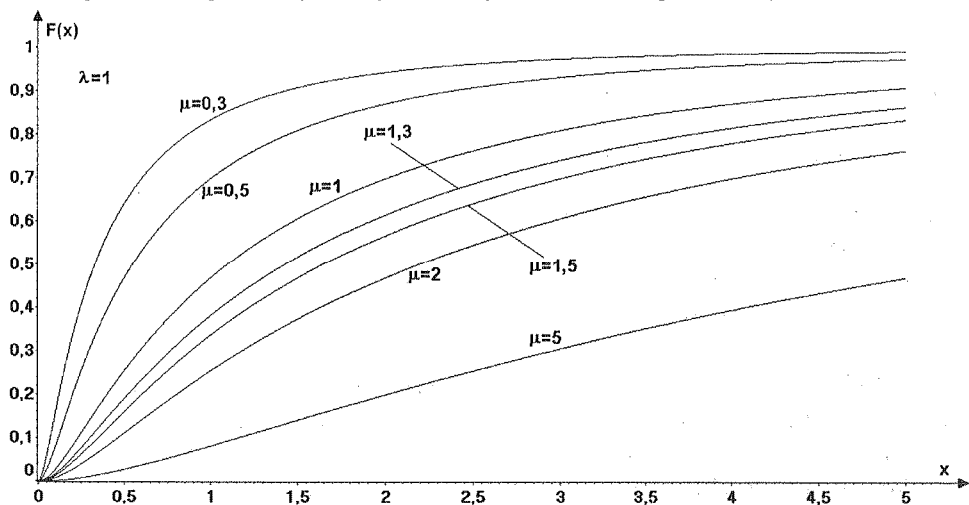
Fig. 3. Suzuki cumulative distribution function $F(x)$ for different parameters λ and $\mu = 1$

Rys.



Rys. 4. Funkcja gęstości $f(x)$ rozkładu Suzuki dla różnych wartości parametru μ przy ustalonym $\lambda = 1$

Fig. 4. Suzuki probability density function $f(x)$ for different parameters μ and $\lambda = 1$

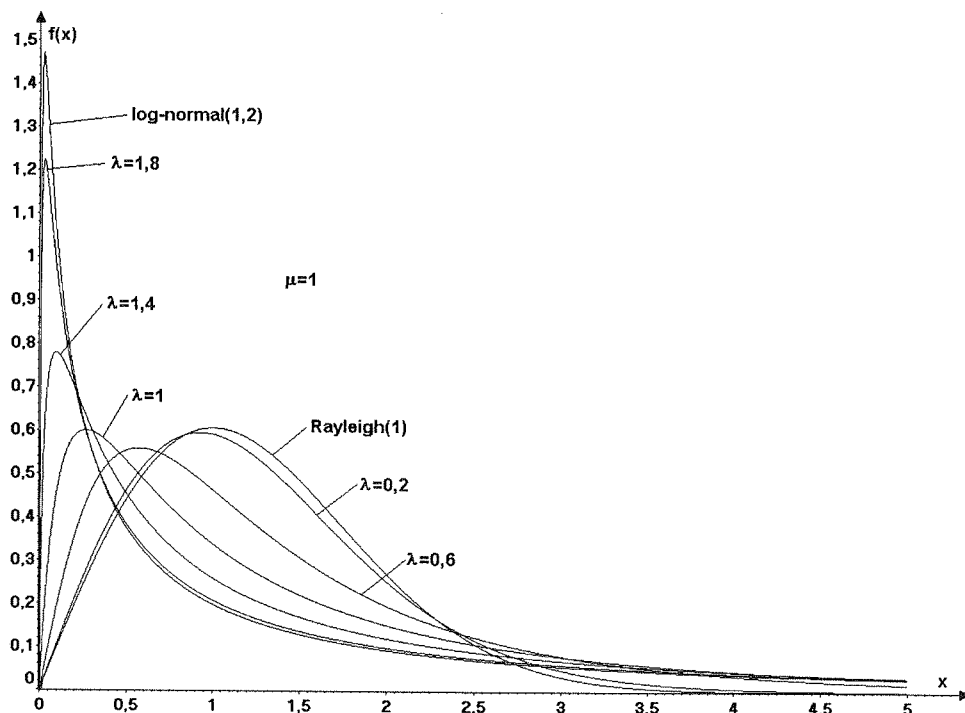


Rys. 5. Dystrybuanta $F(x)$ rozkładu Suzuki dla różnych wartości parametru μ przy ustalonym $\lambda = 1$

Fig. 5. Suzuki cumulative distribution function $F(x)$ for different parameters μ and $\lambda = 1$

dla wszystkich możliwych sposobów tej propagacji. Na rysunkach 2–5 przedstawiono funkcje gęstości i dystrybuanty rozkładu Suzuki w zależności od parametrów.

Cechą decydującą o dużej przydatności rozkładu Suzuki w modelowaniu zakłóceń biernych jest zbieżność do rozkładu log-normalnego lub Rayleigha dla pewnych wartości parametru kształtu λ . Przedstawiono to na rys. 6. Przy ustalonym $\mu = 1$



Rys. 6. Zbieżność funkcja gęstości $f(x)$ rozkładu Suzuki do rozkładu Rayleigha i log-normalnego ($\mu = 1$)

Fig. 6. Convergence of Suzuki pdf to Rayleigh and log-normal pdf ($\mu = 1$)

dla parametru $\lambda \rightarrow 0$ rozkład Suzuki dąży do rozkładu Rayleigh'a z parametrem 1, natomiast dla $\lambda \rightarrow \infty$ dąży do rozkładu log-normalnego z parametrami 1 i 2.

Zarówno funkcja gęstości, jak i dystrybuanta rozkładu Suzuki są dane w skomplikowanej postaci analitycznej, dlatego sporządzając wykresy zamieszczone na rysunkach 2 — 6 do obliczeń całek wykorzystano metody numeryczne [6]. Przy wyznaczaniu całki metodą numeryczną należy określić zakres argumentu x , w którym liczone będzie przybliżenie oraz podział tego zakresu na takie podprzedziały, które pozwoliłyby możliwie dokładnie policzyć wartość całki. Funkcje podcałkowe mają postać podobną do funkcji gęstości rozkładu Suzuki — są to krzywe z jednym maksimum, posiadające tylko dodatnie wartości, równe 0 dla argumentu x mniejszego od 0 i dążące do zera dla $x \rightarrow \infty$. Przeprowadzone badania metodą symulacji Monte Carlo wykazały, iż najlepszym zakresem jest przedział $(0, a)$, gdzie a to iloczyn wartości argumentu, dla

TOM 50

jakich
(co okWyzna
argume

gdzie:

Suma w
wiście
przedzi
tości p
— duża
względ
podczas
go przy
przedzi

Mo
ny w p
ciągów
tycznie
losowa
na

ma rozk
Suzuki j
1. Wyt
2. Wyt
3. Obli
Algorytm
oraz por

jakich dystrybuanta części log-normalnej przyjmuje wartość 0,95 oraz argumentu $x * 10$ (co określono eksperymentalnie badając część Rayleigh'a funkcji podcałkowej):

$$a = F_{\log-normal}^{-1}(0,95) * x * 10. \quad (5)$$

Wyznaczony w ten sposób zakres jest dzielony na n rozłożonych równomiernie na osi argumentu przedziałów, a w każdym przedziale wyznaczana jest następująca wartość:

$$b = \frac{f(x_{\min}) + f(x_{\max})}{2} * (x_{\max} - x_{\min}) \quad (6)$$

gdzie:

$f(\cdot)$ — funkcja podcałkowa

$x_{\min}$ — minimalna wartość argumentu x dla i -tego przedziału,

$x_{\max}$ — maksymalna wartość argumentu x dla i -tego przedziału,

Suma wartości b dla wszystkich przedziałów daje przybliżenie obliczanej całki. Oczywiście masa funkcji podcałkowej nie jest rozłożona równomiernie w tak określonych przedziałach, co powoduje, iż dla ustalonej ilości podprzedziałów oraz pewnych wartości parametrów (gdy w rozkładzie Suzuki zaczyna dominować część log-normalna — duża wartość parametru λ), obliczone przybliżenie staje się dość niedokładne. Ze względu na fakt, iż wartości dystrybuanty teoretycznej rozkładu Suzuki są używane podczas obliczania testu χ^2 (a więc wiarygodność testu zależy od dokładności tego przybliżenia), w czasie wykonywania wszystkich symulacji używano dużej liczby przedziałów ($n = 100\,000$), zapewniającej wystarczającą dokładność.

3. SYMULACJA KOMPUTEROWA

Model Suzuki, ze względu na skomplikowaną postać analityczną, jest również trudny w procesie modelowania. Zastosowanie najbardziej znanych technik generowania ciągów pseudolosowych, jak na przykład metoda odwracania dystrybuanty, jest praktycznie niemożliwe [4]. Jednak stosunkowo prosto można wykazać, że jeżeli zmienna losowa u ma rozkład Rayleigh, a zmienna losowa v ma rozkład log-normalny, to zmienna

$$x = uv \quad (7)$$

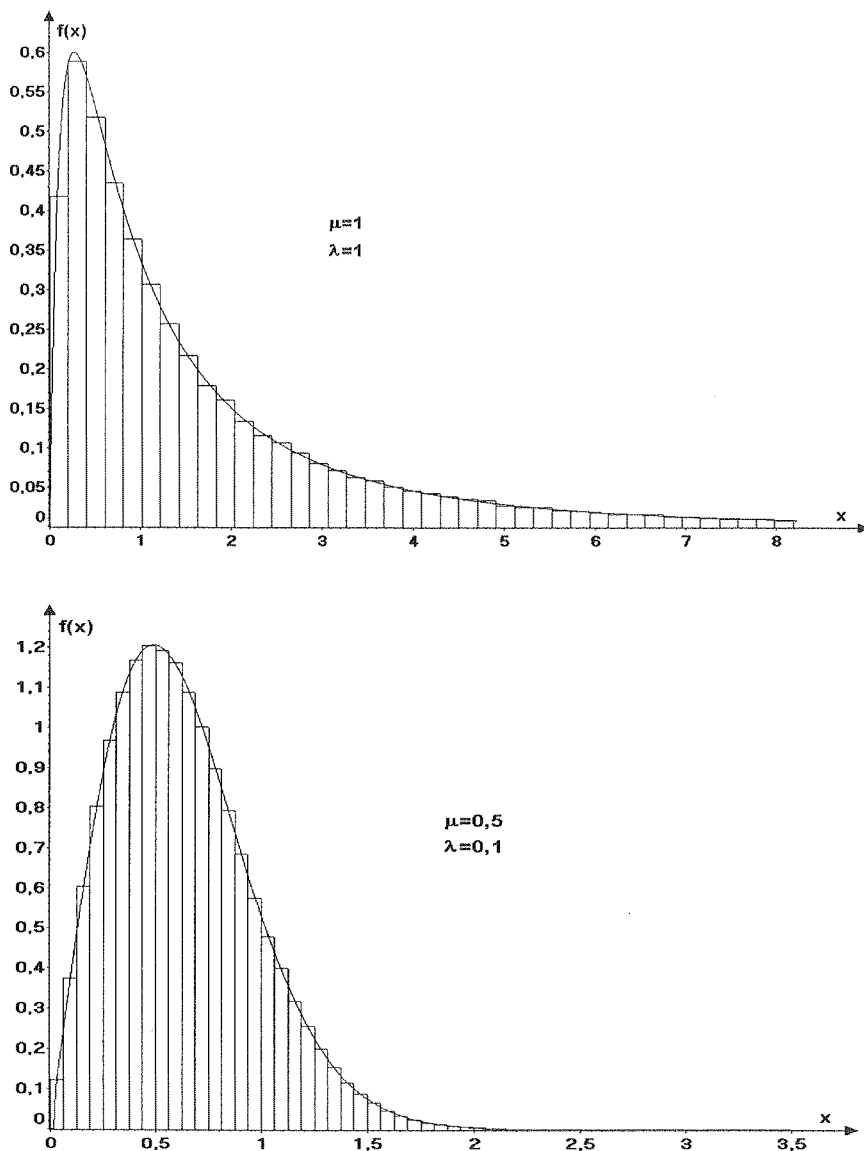
ma rozkład Suzuki [5]. Na tej podstawie algorytm generacji próbek zakłóceń o modelu Suzuki jest następujący:

1. Wygeneruj próbkę v o rozkładzie log-normalnym, o zadanych parametrach μ i λ .
2. Wygeneruj próbkę u o rozkładzie Rayleigha, o parametrze $\sigma = v$.
3. Oblicz $x = uv$.

Algorytm ten został potwierdzony doświadczalnie poprzez wygenerowanie ciągu próbek oraz porównaniu ich empirycznego rozkładu z rozkładem teoretycznym przy użyciu

testu χ^2 . Ciągi o liczności 100 000 próbek generowano, wykorzystując metody opisane w [3], dla różnych zestawów parametrów μ i λ . Na poziomie ufności 0,95, przy 40 stopniach swobody nie było podstaw do odrzucenia hipotezy o postaci rozkładu prawdopodobieństwa.

Na rys. 7 porównano unormowane histogramy wygenerowanych próbek zakłóceń Suzuki z rozkładami teoretycznymi dla różnych wartości parametrów.



Rys. 7. Rozkład teoretyczny (krzywa ciągła) i unormowany histogram próbek zakłóceń Suzuki

Fig. 7. Theoretical Suzuki pdf (solid line) and normalised histogram of generated samples

ly opi-
5, przy
zkładu
złóceń

4. PODSUMOWANIE

Wielotorowa propagacja fali elektromagnetycznej w obszarze zabudowy wielkomiejskiej powoduje powstawanie zakłóceń, szczególnie uciążliwych w systemach radiokomunikacyjnych i radiolokacyjnych. Proste modele probabilistyczne odzwierciedlają właściwości tych zakłóceń w pewnych ograniczonych przypadkach. Modele złożone mają znacznie szerszy zakres zastosowań, są jednak skomplikowane i trudne do symulacji. Zaproponowany przez H. Suzuki model należy do tej kategorii i mimo uwikłanej postaci całkowitej można, stosując metody numeryczne, generować ciągi pseudolosowe, o zadanych parametrach. Przedstawiony w niniejszym artykule algorytm został wszechstronnie przetestowany i ze względu na stosunkowo prostą postać jest użytecznym narzędziem, możliwym do wykorzystania w projektowaniu systemów odpornych na zakłócenia bierne.

5. BIBLIOGRAFIA

1. H. Suzuki: *A statistical model for urban radio propagation*. IEEE Transactions On Communications, vol. COM-25, No 7, July 1977, pp. 673-680.
2. A. Abdi, M. Kaveh: *K distribution: an appropriate substitute for Rayleigh-lognormal distribution in fading-shadowing wireless channels*. Electronics Letters, Vol. 34, No 9, 30 April 1998, pp. 851-852.
3. A. Jakubiak: *Metody i algorytmy symulacji radiolokacyjnych zakłóceń biernych*. Kwartalnik Elektroniki i Telekomunikacji, t. 46, z. 2, 2000, s. 237-253.
4. R. Wieczorkowski, R. Zieliński: *Komputerowe generatory liczb losowych*. Warszawa, WNT, 1997.
5. A. Papoulis: *Prawdopodobieństwo, zmienne losowe i procesy stochastyczne*. Warszawa, WNT, 1972.
6. S. Rosłonec: *Wybrane metody numeryczne z przykładami zastosowań w zadaniach inżynierskich*. Warszawa, Oficyna Wydawnicza Politechniki Warszawskiej, 2002.

A. JAKUBIAK, A. CHRUSTOWSKI

URBAN CLUTTER MODELLED BY SUZUKI DISTRIBUTION

S u m m a r y

Signal modelling plays an important role in analytic design studies and computer simulation of many systems, like mobile communications, radar etc. The constructions of suitable signal models are commonly concentrated on finding a mathematical description of experimental data and on generation of artificial signal draws with assumed properties. The urban radio channel is essentially a multipath channel, due to reflections, refractions and scattering by buildings, commonly modeled as a linear filter.

A mixture of Raleigh and lognormal distribution, called Suzuki distribution, appropriately describes local and global spatial variations of a signal envelope and/or short-term and long-term temporal fluctuations of a signal envelope in fading-shadowing wireless channel. Suzuki probability density function $f(x)$

zuki
es

and cumulative distribution function $F(x)$ for different parameters are shown as well as the convergence of Suzuki pdf to Rayleigh and log-normal pdf.

The Suzuki model of a signal envelope in fading-shadowing wireless channel has a complicated integral forms of probability density function and cumulative distribution. From this reason it is very difficult to generate pseudorandom Suzuki draws by using the most widely disseminated inverse distribution method. However it is not difficult to prove that if: R is a Rayleigh distributed random variable, V is lognormal distributed random variable, R and V are statistical independent, than $S = RV$ is Suzuki distributed. Thus an algorithm of generation is simple:

1. Generate a lognormal distributed sample v with assumed parameters μ and λ .
2. Generate a Rayleigh distributed sample r with parameter $\sigma = v$.
3. Multiply r and v .

This algorithm was successfully verified by Chi-square test for different parameters values.

The proposed algorithm of Suzuki random samples generation is effective and permits design studies by using the computers simulation methods in order to avoid costly hardware tests of wireless communication systems.

Keywords: urban clutter, signal modelling, computer simulation

N
legając
dobrze
system
strale
przypa
tych m
go [10
najgor
ograni

Badania eksperymentalne magistral miejscowych dla wybranych metod szeregowania wiadomości

ŚLAWOMIR ŻABA

*Instytut Elektrotechniki Teoretycznej i Automatyki, Politechnika Krakowska
ul. Warszawska 24, 31-155 Kraków
e-mail: szaba@usk.pk.edu.pl*

Otrzymano 2003.02.19

Autoryzowano 2003.06.18

W artykule przedstawiono właściwości i parametry czasowe trzech wybranych magistral miejscowych: PROFIBUS FMS, CAN i InterBus-S. Prezentowane wyniki uzyskano na podstawie badań, polegających na sprawdzeniu dotrzymania ograniczeń czasowych dla wiadomości w rozproszonym systemie czasu rzeczywistego z magistralą miejscową przy założeniu, że wiadomości szeregowane są przy użyciu metod szeregowania: FIFO, GRMS, EDF. Badania przeprowadzone były dla najgorszego przypadku, tzn. zostało założone, że w chwili $t = 0$ zostały aktywowane wszystkie zadania w systemie. Uwzględniono także czasy blokowania wiadomości. Wyniki badań zaprezentowano w postaci graficznej i tabelarycznej.

Słowa kluczowe: rozproszony system komputerowy, system czasu rzeczywistego, magistrala miejscowa, szeregowanie wiadomości, metoda FIFO, GRMS, EDF

1. WSTĘP

Nowoczesne koncepcje sterowania bazują na decentralizowanej automatyzacji polegającej na rozproszeniu sprzętu i decentralizacji inteligencji. Rozproszenie sprzętu jest dobrze sprawdzone w systemach automatyzacji i polega na zastosowaniu odpowiednich systemów transmisji danych bazujących na magistrali miejscowej (*fieldbus*) [5]. Magistrale miejscowe tym różnią się od zwykłych sieci komputerowych, że w większości przypadków gwarantują deterministyczny czas odpowiedzi [2]. Jednakże różnorodność tych magistral sprawia, że nie jest możliwe opracowanie wspólnego cyklu projektowego [10]. Poprawne zaprojektowanie systemu RT (*real time*) dla różnych magistral dla najgorszego przypadku jest trudne i czasochłonne. Dodatkowo, w systemach o ostrych ograniczeniach czasowych, niezbędne staje się stosowanie metod szeregowania wiado-

regowania wiadomości [7,8,9]. Dlatego też pomocnym narzędziem są charakterystyki czasowe magistral miejscowych dla różnych metod szeregowania wiadomości w funkcji zmian wielu parametrów systemu. W artykule przedstawiono wyniki badań dla trzech magistral miejscowych (PROFIBUS FMS, CAN, InterBus-S [6]) obejmujących trzy etapy:

- Analiza czasowa magistral miejscowych, dla wiadomości szeregowanych kolejkowo (metoda FIFO [1]).
- Analiza czasowa magistral miejscowych, dla wiadomości szeregowanych przy użyciu metody ze statycznym przydziałem priorytetu (metoda GRMS [4]),
- Analiza czasowa magistral miejscowych, dla wiadomości szeregowanych przy użyciu metody z dynamicznym przydziałem priorytetu (metoda EDF [3]).

Wyniki badań uzyskane zostały na bazie autorskich algorytmów zaimplementowanych na komputerze. Algorytmy te, uwzględniające specyfikę działania poszczególnych magistral miejscowych, sprawdzały dotrzymanie ograniczeń czasowych dla wszystkich wiadomości w systemie. Badania przeprowadzone były dla najgorszego przypadku. Uwzględniono także czasy blokowania wiadomości. Prezentowany artykuł nie zawiera opisu zasad działania i funkcjonowania magistral miejscowych oraz opisu metod szeregowania zadań. Wiadomości te można znaleźć w literaturze podanej w artykule.

2. ZASADY I KRYTERIA PRZEPROWADZONYCH BADAŃ

Przeprowadzone badania bazują na następującym modelu systemu:

- wszystkie wiadomości są wiadomościami okresowymi,
- liczba węzłów w systemie 2 ... 10,
- liczba wiadomości w węźle 1 ... 10,
- długość pola danych wiadomości 1 ... 32 bajtów,
- minimalna wartość, o jaką zmieniany był okres występowania wiadomości to 1 [ms].

Na osi Y wszystkich charakterystyk występuje parametr zwany *maksymalnym obciążeniem systemu* λ . Parametr λ obliczany jest ze wzoru:

$$\lambda = \sum_{i=1}^n \sum_{j=1}^w \frac{(bp_{ij} * 8) * t_{bit_500}}{tp_{ij}} \quad (1)$$

gdzie:

- n — liczba węzłów w systemie
- w — liczba wiadomości w węźle
- bp_{ij} — długość pola danych wiadomości [B]
- tp_{ij} — okres występowania wiadomości [ms]
- t_{bit_500} — czas przesłania 1 bitu dla szybkości transmisji 500 Kb/s.

Współczynnik λ obliczany jest na podstawie liczby bajtów pola 'dane' przesyłanych wiadomości, a nie na podstawie liczby bajtów całej wiadomości (czyli bez danych nadmiarowych i narzutów protokołu). Dane nadmiarowe i narzut protokołu poszczególnych

magistral miejscowych dla ws...
czynności...
liczby...
można...
współ...
dla ws...
Dział...
przy s...
magist...
Kb/s...
system...
szybko...
wych...
własno...

□
wystę...
założe...
• wi...
• zac...
Zmian...
danych...
wszyst...
wiadom...

DL...
w syst...
koweg...
ogran...
w syste...
cych 1...
temu, l...
danych...
działu...
to λ w...

DL...
w syste...
ce ciąg...

ystyki
unkcji
trzech
h trzy
jkowo
y uży-
y uży-

ranych
h ma-
stkich
padku.
zawie-
metod
ykule.

magistral są oczywiście uwzględniane przy analizie systemu, ale zdefiniowanie współczynnika λ jw. umożliwia lepsze porównanie właściwości magistral (dla takiej samej liczby bajtów całej wiadomości, magistralą z mniejszą liczbą bajtów nadmiarowych można przesłać więcej bajtów danych w takim samym okresie czasu, a tym samym współczynnik λ będzie większy). Wyjaśnienia wymaga, dlaczego obciążenie systemu dla wszystkich magistral wyznaczane jest dla szybkości przesyłu danych 500 Kb/s. Działanie takie ma celu umożliwienie porównania właściwości różnych protokołów przy stałej szybkości przesyłu danych, a przy szybkości 500 Kb/s pracują wszystkie magistrale. Praca magistrali przy większych szybkościach (PROFIBUS FMS - 1500 Kb/s, CAN - 1000 Kb/s) spowoduje, że osiągane wartości współczynnika obciążenia systemu danymi będą odpowiednio mniejsze lub większe obrazując tym samym wpływ szybkości przesyłu danych na osiągi systemu. Ponieważ przebadanie magistral miejscowych dla wszystkich wartości parametrów jest nadmierne, w kolejnych podrozdziałach własności magistral przedstawione zostaną przy użyciu trzech typów charakterystyk.

□ *Charakterystyka typu pierwszego obrazuje zmiany parametru λ od zmian okresu występowania wiadomości.* Ten typ charakterystyki wyznaczany był przy następujących założeniach:

- wiadomości w poszczególnych węzłach mają te same parametry czasowe,
- zadana jest liczba węzłów i liczba wiadomości w węźle.

Zmianie ulegał okres występowania wiadomości i liczba przesyłanych bajtów pola danych wiadomości. Okresy występowania wiadomości w węźle były takie same dla wszystkich wiadomości lub tworzyły ciąg arytmetyczny. Jeżeli okresy występowania wiadomości w węźle były takie same, to wzór (1) przyjmie postać:

$$\lambda = n * w * \frac{(bp_r * 8) * t_bit_500}{tp_c} \quad (2)$$

[ms].
obciąż-

Dla zadanej długości pola danych bp_r (jednakowej dla wszystkich wiadomości w systemie), poszukiwana była minimalna wartość okresu występowania tp_c (jednakowego dla wszystkich wiadomości), dla którego w systemie spełnione były jeszcze ograniczenia RT. Pierwszy punkt charakterystyki był wyznaczany dla sytuacji, kiedy w systemie spełnione są ograniczenia czasowe dla przesyłanych wiadomości zawierających 1 bajt danych. Ostatni punkt wykresu wyznaczany był dla takich parametrów systemu, kiedy spełnione są ograniczenia czasowe dla wiadomości zawierających 32 bajty danych (szczegóły dotyczące powstawania charakterystyki przedstawiono na końcu rozdziału). Jeżeli okresy występowania wiadomości w węźle tworzyły ciąg arytmetyczny, to λ wyznaczano ze wzoru:

$$\lambda = n * \sum_{j=1}^w \frac{(bp_r * 8) * t_bit_500}{j * tp_j} \quad (3)$$

Dla zadanej długości pola danych bp_r (jednakowej dla wszystkich wiadomości w systemie), poszukiwane były minimalne wartości okresów występowania tp_j tworzące ciąg arytmetyczny (jednakowy w każdym węźle), dla których w systemie spełnione

ranych
n nad-
5lnych

były ograniczenia RT. Dla tego przypadku na osi X (odciętych) występują wartości odpowiadające najmniejszej wartości z ciągu. Przykładowo, jeżeli okresy występowania wiadomości tworzą ciąg arytmetyczny, a pierwsza wartość ciągu to 4 [ms], to wartość parametru λ zaznaczona jest na wykresie dla okresu równego 4 [ms]).

□ *Charakterystyka typu drugiego obrazuje zmiany parametru λ od zmian liczby wiadomości w węźle.* Ten typ charakterystyki wyznaczany był przy następujących założeniach:

- wiadomości w poszczególnych węzłach mają te same parametry czasowe,
- zadana jest stała liczba węzłów i jednakowa długość pola danych wiadomości.

Wartość współczynnika λ wyznaczana była ze wzoru:

$$\lambda = n * w * \frac{(bp_r * 8) * t_{bit_500}}{tp_c} \quad (4)$$

gdzie:

$$n \in (1 \dots 10)$$

$$w \quad \quad \quad \text{— } const$$

$$bp_r \quad \quad \quad \text{— } const$$

Dla zadanej liczby wiadomości w węźle w (z przedziału od 1 do 10), przy określonej liczbie węzłów n i długości pola danych bp_r , poszukiwana była minimalna wartość okresu występowania wiadomości tp_c (jednakowego dla wszystkich wiadomości), dla której spełnione były ograniczenia RT.

□ *Charakterystyka typu trzeciego obrazuje zmiany parametru λ od zmian liczby węzłów w systemie.* Ten typ charakterystyki wyznaczany był przy następujących założeniach:

- wiadomości w poszczególnych węzłach mają te same parametry czasowe,
- zadana jest stała liczba wiadomości w węźle i jednakowa długość pola danych wiadomości.

Wartość współczynnika λ wyznaczana była ze wzoru:

$$\lambda = n * w * \frac{(bp_r * 8) * t_{bit_500}}{tp_c} \quad (5)$$

gdzie:

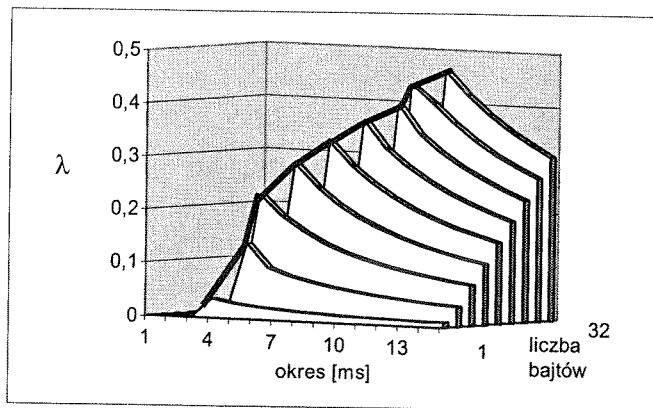
$$w \in (2 \dots 10)$$

$$n \quad \quad \quad \text{— } const$$

$$bp_r \quad \quad \quad \text{— } const$$

Dla zadanej liczby węzłów n (z przedziału od 2 do 10), przy określonej liczbie wiadomości w węźle w i długości pola danych bp_r , poszukiwana była minimalna wartość okresu występowania wiadomości tp_c (jednakową dla wszystkich wiadomości), dla której spełnione były ograniczenia RT.

Na rys. 1 przedstawiono charakterystykę trójwymiarową, która przedstawia sposób powstawania charakterystyk typu pierwszego. Poszczególne krzywe przedstawiają zmianę obciążenia systemu (oś Z) od zmian okresu występowania wiadomości (oś X) przy zachowaniu stałej liczby bajtów pola danych wiadomości (oś Y) - oczywiście rzecz dotyczy sytuacji, kiedy w systemie dotrzymane są ograniczenia RT. Aby nie zaciemniać rysunku przedstawione krzywe dotyczą przesłania tylko 1, 4, 8, 12, 16, 20, 24, 28 i 32 bajtów. Ponieważ charakterystyka typu pierwszego przedstawia przebieg zmian obciążenia systemu dla minimalnych okresów występowania wiadomości, to powstaje ona w wyniku połączenia maksimów każdej z poszczególnych krzywych z rys. 1 (przy określonej liczbie przesyłanych bajtów dla minimalnego okresu występowania dostajemy maksymalne obciążenie systemu).



Rys. 1. Sposób tworzenia charakterystyk typu pierwszego

Fig. 1. A method of creating charts of type one

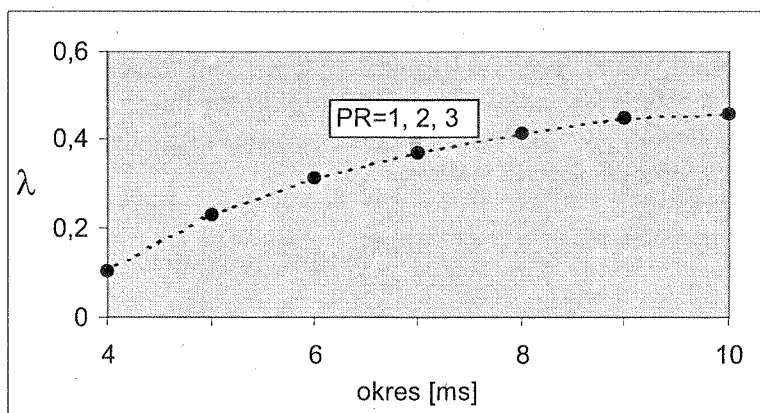
3. ANALIZA WŁAŚCIWOŚCI MAGISTRAL MIEJSCOWYCH DLA METODY FIFO

W rozdziale przedstawione zostaną właściwości trzech wybranych magistral miejscowych przy założeniu, że w węzłach wiadomości szeregowane są przy użyciu metody FIFO. Sprawdzenie spełnienia warunków RT dla metody FIFO przeprowadzono dla najgorszego przypadku, tzn. w systemie aktywowane zostały wszystkie wiadomości, a wiadomość z najkrótszym okresem występowania (najkrótszym ograniczeniem) znalazła się na końcu kolejki wiadomości. W przyjętym modelu założono, że wszystkie wiadomości mają ten sam okres występowania (wyjątek stanowią modele przyjęte dla PROFIBUS FMS).

Magistrala PROFIBUS FMS

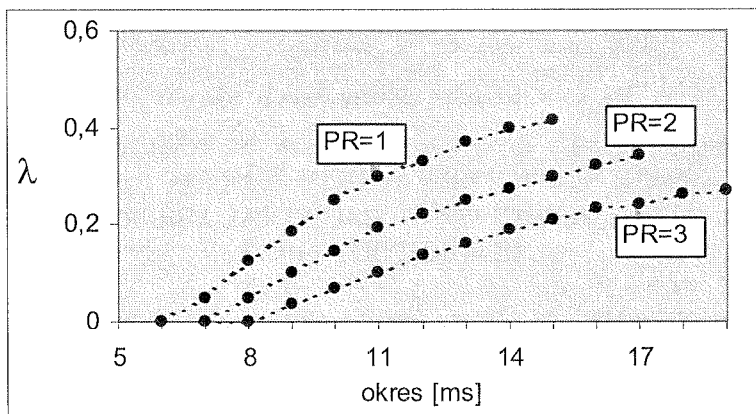
Ponieważ dla magistrali PROFIBUS FMS istnieje możliwość przyporządkowania priorytetów *wysoki* lub *niski* do poszczególnych wiadomości, więc badania przepro-

wadzono dla kilku wariantów przyporządkowania priorytetów. Na rys. 2, 3 i 4 przedstawiono wykresy pierwszego typu. Poszczególnym krzywym przyporządkowany jest parametr PR , który oznacza liczbę wiadomości (w każdym węźle z osobna) z przyporządkowanym priorytetem *wysoki* (np. $PR = 1$ oznacza, że jedna wiadomość w każdym węźle ma przyporządkowany priorytet *wysoki*). Jeżeli wiadomości w węźle miały róż-



Rys. 2. PROFIBUS FMS: 3 węzły, 3 wiadomości w węźle

Fig. 2. PROFIBUS FMS: 3 nodes, 3 messages in each node



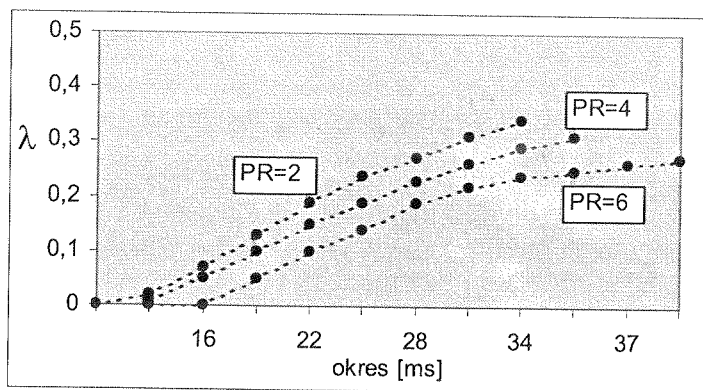
Rys. 3. PROFIBUS FMS: 3 węzły, 3 wiadomości w węźle (ciąg arytmetyczny)

Fig. 3. PROFIBUS FMS: 3 nodes, 3 messages in each node (arithmetical process)

ne okresy występowania, to przydzielanie priorytetu *wysoki* odbywało się począwszy od wiadomości z najkrótszym okresem występowania. Jeżeli wszystkie wiadomości w węźle mają ten sam okres występowania, to dla dowolnej wartości parametru PR osiągnięto te same wartości (rys. 2). Zróznicowanie charakterystyk uzyskujemy, gdy

wiadomoś
dla przyp
metryczny
gdzie t_m
tego rodz
ny"). Naj
otrzyman
przypadku
czasowym
soki), nato
wiadomoś
z najkróts

Magis
wych — l
użytkowni
odznaczaj
stu współ
wiadomoś
magania E
być równy
użytkowni
tości dany
Tym samy
danych wz
magistral z
dowe char
praktyczni

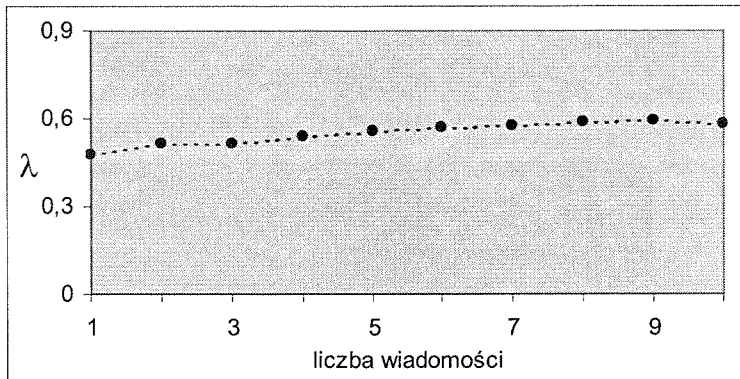


Rys. 4. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle (ciąg arytmetyczny)

Fig. 4. PROFIBUS FMS: 6 nodes, 6 messages in each node (arithmetical process)

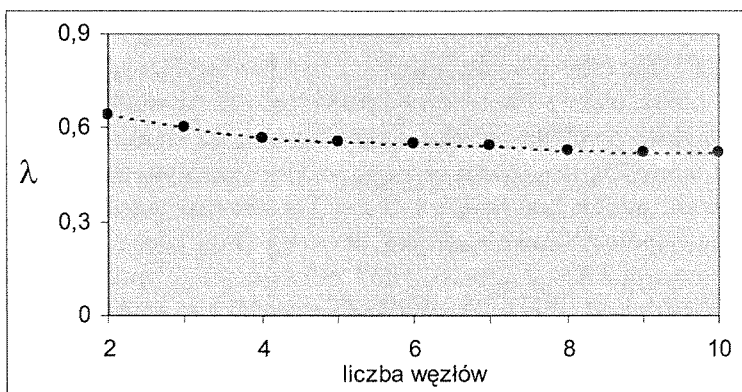
wiadomości mają różne okresy występowania. Charakterystyki z rys. 3 i 4 wyznaczono dla przypadku, kiedy okresy występowania wiadomości w węźle tworzą ciąg arytmetyczny, a odległość pomiędzy wartościami poszczególnych okresów wynosi t_{min} , gdzie t_{min} jest wartością minimalną w ciągu, czyli pierwszym elementem ciągu (dla tego rodzaju badań, podpis pod wykresem zawiera sformułowanie „ciąg arytmetyczny”). Największe współczynniki obciążenia, dla których spełnione są ograniczenia RT, otrzymano dla krzywej $PR = 1$ (rys. 3) i $PR = 2$ (rys. 4). Wynika to z faktu, że dla przypadku $PR = 1$ (rys. 3) i $PR = 2$ (rys. 4) wiadomości z najkrótszym ograniczeniem czasowym są wysyłane jako pierwsze (dzięki przyporządkowanemu priorytetowi wysoki), natomiast dla przypadku $PR = 3$ (rys. 3) i $PR = 6$ (rys. 4), tzn. kiedy wszystkie wiadomości mają przyporządkowany ten sam priorytet, przyjmuje się, że wiadomości z najkrótszym okresem występowania wysyłane są jako ostatnie (najgorszy przypadek).

Magistrala PROFIBUS FMS jest magistralą ze stałym narzutem danych nadmiarowych — liczba tych danych w ramce wiadomości nie zależy od liczby bajtów danych użytkownika w przesyłanej wiadomości. Dlatego też wykresy w początkowej fazie odznaczają się szybkim narastaniem, przechodząc następnie w łagodniejszą fazę wzrostu współczynnika obciążenia systemu danymi w funkcji zmian okresu występowania wiadomości. Wynika to z następujących faktów: założmy, że aby system spełnił wymagania RT dla długości pola danych 1 bajt, okres występowania wiadomości musi być równy minimum 2 [ms]. W tym czasie przesyłane są dane nadmiarowe i dane użytkownika. Tak więc, dla okresu występowania wiadomości 3 [ms], przy stałej wartości danych nadmiarowych, system spełni wymagania RT dla 5 bajtów pola danych. Tym samym przy wzroście okresów występowania z 2 do 3 [ms] liczba przesłanych danych wzrosła z 1 do 5 bajtów (każdej wiadomości). Sytuacja taka jest typowa dla magistral ze stałą wartością danych nadmiarowych. Na rys. 5 i 6 przedstawiono przykładowe charakterystyki drugiego i trzeciego typu. Dla tego typu charakterystyk system praktycznie utrzymuje stałą wartość współczynnika obciążenia systemu, dla którego



Rys. 5. PROFIBUS FMS: 5 węzłów, długość pola danych 32[B]

Fig. 5. PROFIBUS FMS: 5 nodes, data length 32[B]



Rys. 6. PROFIBUS FMS: 5 wiadomości w węźle, długość pola danych 32[B]

Fig. 6. PROFIBUS FMS: 5 messages, data length 32[B]

spełnione są ograniczenia RT. Wartość współczynnika zależy od liczby bajtów przesyłanych w wiadomościach. Przykładowo, dla tego typu charakterystyk, przy przesyłaniu 1 bajtu danych, wartość współczynnika obciążenia systemu wynosiła 0.04. Należy także zauważyć, że przy stałej liczbie węzłów i wzroście liczby wiadomości, charakterystyka posiada tendencję do lekkiego wzrostu współczynnika obciążenia systemu (rys. 5). Natomiast dla stałej liczby wiadomości w węźle i wzroście liczby węzłów, charakterystyka ma tendencję do lekkiego spadku (rys. 6). Wynika to z faktu, że przy zwiększaniu liczby węzłów, rosną straty na przekazywanie znacznika (czas $t_w T$). Ponieważ sprawdzanie spełnienia ograniczeń RT dla metody FIFO odbywa się dla jednego cyklu obiegu znacznika (wiadomości mają ten sam okres występowania równy czasowi obiegu znacznika), to omawiana właściwość będzie dokładniej zbadana dla innych metod (GRMS, EDF) i dla innych parametrów modelu strumienia wiadomości, kiedy to wystąpi więcej niż jeden obieg znacznika (straty na przekazywanie znacznika

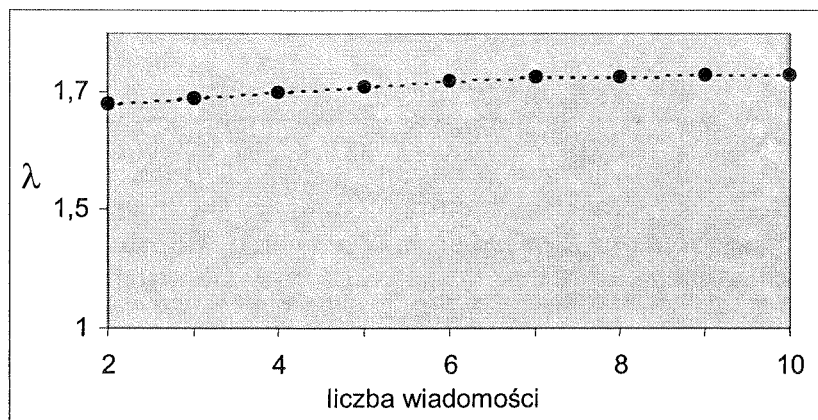
będą bar
gistrali
przesyła
współcz

R

Ma
Ma
trybach
proces
wynos
wzrost
odwrotn
a najdl
2 bajto
całej wi
bajtowe
związ
czas pr
bajtowe
różnej
magistr

- uzy
FIB
odb
wią
każ

będą bardziej zauważalne). Na rys. 7 przedstawiono przykładową charakterystykę magistrali PROFIBUS FMS dla szybkości przesyłu danych 1500 Kb/s. Wzrost szybkości przesyłanych danych z 500 Kb/s do 1500 Kb/s powoduje trzykrotny wzrost wartości współczynnika obciążenia systemu.



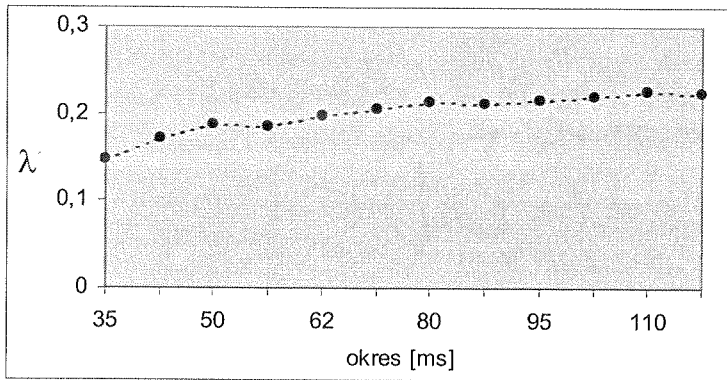
Rys. 7. PROFIBUS FMS: 5 węzłów, długość pola danych 32[B], szybkość 1500 [Kb/s]

Fig. 7. PROFIBUS FMS: 5 nodes, data length 32 [B], throughput 1500 [Kb/s]

Magistrala InterBus-S

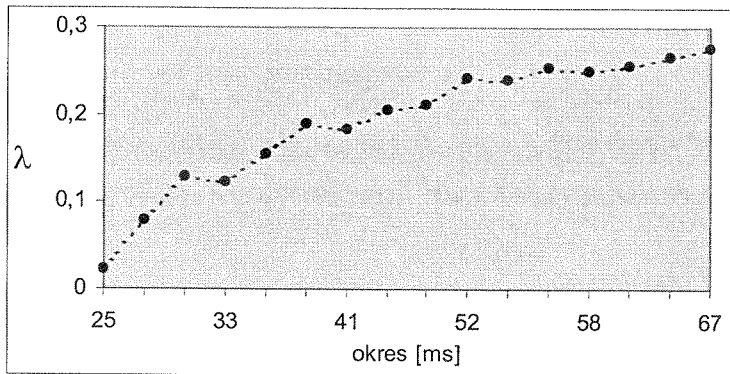
Magistrala InterBus-S jest magistralą pierścieniową i może pracować w dwóch trybach: tylko kanał procesowy (każdy węzeł może przesłać 2 bajty danych) lub kanał procesowy i kanał PCP (długość szczeliny kanału PCP musi być z góry ustalona i może wynosić 2, 4 lub 8 bajtów). Czas jednego cyklu przesyłu danych wzrasta wraz ze wzrostem długości szczeliny. Jednakże całkowity czas przesyłu danych zmienia się odwrotnie w stosunku do zmian długości szczeliny PCP: jest najkrótszy dla 8 bajtów, a najdłuższy dla 2 bajtów (rys. 5 do 10). Wynika to z faktu, iż pomimo dla szczeliny 2 bajtowej czas jednego cyklu przesyłu danych jest najkrótszy, to w celu przesłania całej wiadomości musi być wykonane więcej cykli przesyłu danych niż dla szczeliny 8 bajtowej. Każdy cykl przesłania danych wiąże się z przesyłaniem danych dodatkowych związanych z działaniem magistrali InterBus-S (narzut protokołu). W konsekwencji czas przesłania danych dla szczelin 2 i 4 bajtowych jest dłuższy niż dla szczeliny 8 bajtowej. Dla magistrali InterBus-S wyznaczono charakterystyki pierwszego typu dla różnej długości szczelin (rys. 8, 9, 10). W stosunku do charakterystyk uzyskanych dla magistrali PROFIBUS FMS występuje kilka istotnych różnic:

- uzyskane wartości współczynnika obciążenia systemu są mniejsze niż dla PROFIBUS FMS - wynika to głównie z faktu, że dla InterBus-S przesyłanie danych odbywa się w kilku lub kilkunastu cyklach, a każdy kolejny cykl przesyłu danych wiąże się z przesyłaniem kolejnych danych nadmiarowych (5 bitów sterujących po każdym przesłanym bajcie, 6 bajtów dodatkowych w każdym cyklu),



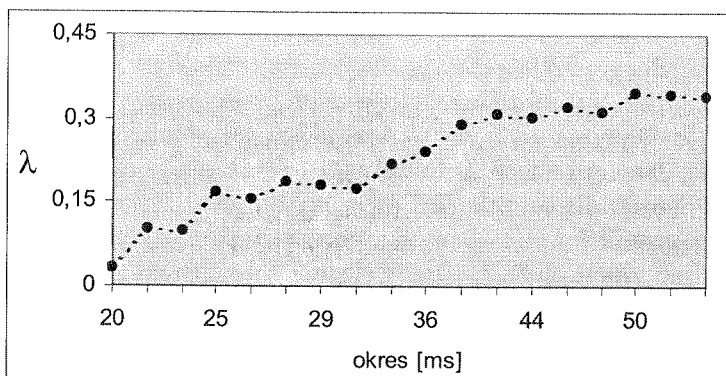
Rys. 8. InterBus-S: 6 węzłów, 6 wiadomości w węzle, szczelina 2[B]

Fig. 8. InterBus: 6 nodes, 6 messages in each node, slot 2[B]



Rys. 9. InterBus-S: 6 węzłów, 6 wiadomości w węzle, szczelina 4[B]

Fig. 9. InterBus: 6 nodes, 6 messages in each node, slot 4[B]



Rys. 10. InterBus-S: 6 węzłów, 6 wiadomości w węzle, szczelina 8[B]

Fig. 10. InterBus: 6 nodes, 6 messages in each node, slot 8[B]

• ks
sie
cia
sta
ko
og
a
ob
w
je
je

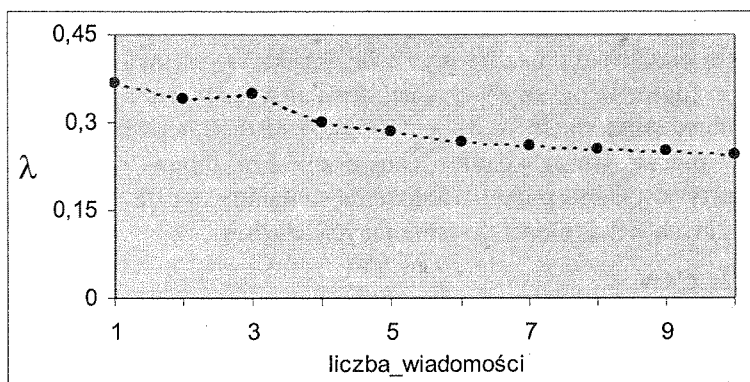
Na ry
widoc
wystę
dla st
do wz

obser
nie ob
skład
Każd
cyklu
[ms]
dla sz
każde
dwie

• d
• d

- kształt charakterystyk jest nieco inny — dla PROFIBUS FMS krzywe odznaczały się stałym wzrostem, a dla InterBus-S występują spadki wartości współczynnika obciążenia lub utrzymywanie się go na tym samym poziomie. Wynika to z faktu, iż dla stałej szczeliny, przesłanie kolejnego, jednego bajtu danych może wiązać się z wykonaniem nowego cyklu obiegu danych. Tym samym mogą zostać przekroczone ograniczenia czasowe. Jeżeli więc okres występowania danych został zwiększony, a przesyłana jest taka sama ilość danych jak poprzednio, to wartość współczynnika obciążenia maleje. Ten efekt jest najbardziej widoczny dla szczeliny 8 bajtowej, bowiem czas przesłania danych w jednym cyklu jest najdłuższy, a najmniej widoczny jest on dla szczeliny 2 bajtowej, ponieważ czas przesłania danych w jednym cyklu jest najkrótszy.

Na rys. 11 i 12 przedstawiono charakterystyki drugiego i trzeciego typu. Wyraźnie widoczna jest właściwość, że przy stałej liczbie węzłów i wzroście liczby wiadomości, występuje tendencja do spadku współczynnika obciążenia systemu (rys. 11). Natomiast dla stałej liczby wiadomości w węźle i wzroście liczby węzłów, występuje tendencja do wzrostu wartości tegoż współczynnika (rys. 12). Jest to więc sytuacja odwrotna do

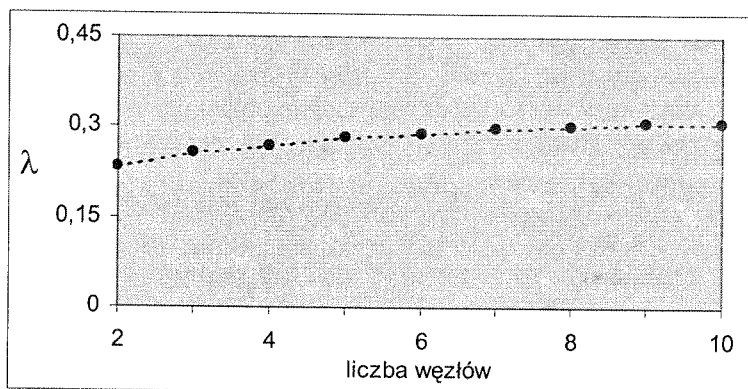


Rys. 11. InterBus-S: 5 węzłów, długość pola danych 32[B], szczelina 4[B]

Fig. 11. InterBus: 5 nodes, data length 32[B], slot 4[B]

obserwowanej dla magistrali PROFIBUS FMS. Przedstawiona powyżej sytuacja zostanie objaśniona na przykładzie. Załóżmy, że dany jest system z magistralą pierścieniową składający się z dwóch węzłów, w każdym węźle są dwie wiadomości do wysłania. Każdy węzeł ma przydzieloną szczelinę 0,4 ms, a czas narzutu protokołu w jednym cyklu wynosi 0,2 ms. Tym samym jeden cykl obiegu danych wynosi 1 ms ($2 \cdot 0,4$ [ms] + 0,2 [ms] = 1 [ms]). Załóżmy, że przesłanie 1 wiadomości (z każdego węzła) dla szczeliny 0,4 ms wymaga dwóch cykli. Tym samym przesłanie 2 wiadomości (z każdego węzła) to 4 cykle czyli 4 [ms]. Załóżmy, że do systemu zostaną dodane kolejne dwie wiadomości. Możliwe są dwa przypadki:

- do istniejących węzłów dołożone zostanie po jednej wiadomości,
- dołożony zostanie kolejny węzeł z dwoma nowymi wiadomościami.



Rys. 12. InterBus-S: 5 wiadomości w węźle, długość pola danych 32[B], szczelina 4[B]

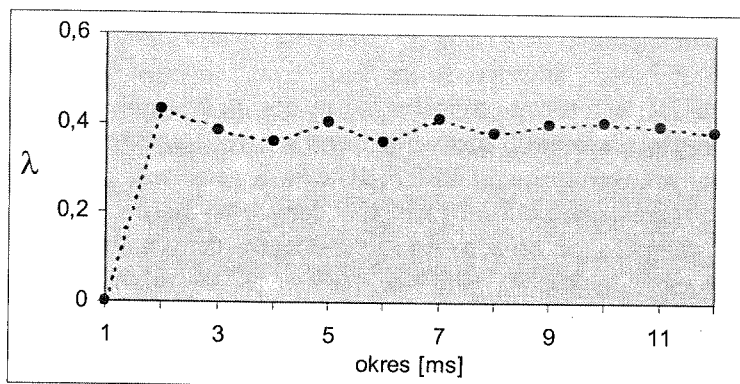
Fig. 12. InterBus: 5 messages in each node, data length 32[B], slot 4[B]

W pierwszym przypadku, ze względu na fakt, że przesłanie 1 wiadomości (z każdego węzła) zajmuje dwa cykle, to przesłanie trzech wiadomości wyniesie 6 cykli czyli 6 ms. W drugim przypadku dołożenie kolejnego węzła spowoduje, że zmieni się czas pojedynczego cyklu przesłania danych i wyniesie 1.4 ms ($3 * 0.4[\text{ms}] + 0.2[\text{ms}] = 1.4[\text{ms}]$). Ponieważ w każdym węźle są dwie wiadomości, to przesłanie ich zajmuje 4 cykle (przesłanie jednej wiadomości wymaga 2 cykli) czyli 5.6 [ms]. Czas ten jest więc krótszy niż w pierwszym przypadku. Oznacza to, że z punktu widzenia zagadnień RT lepsze efekty uzyskano poprzez dołożenie kolejnego węzła do systemu niż przez dołożenie kolejnych wiadomości do istniejących węzłów.

Magistrala CAN

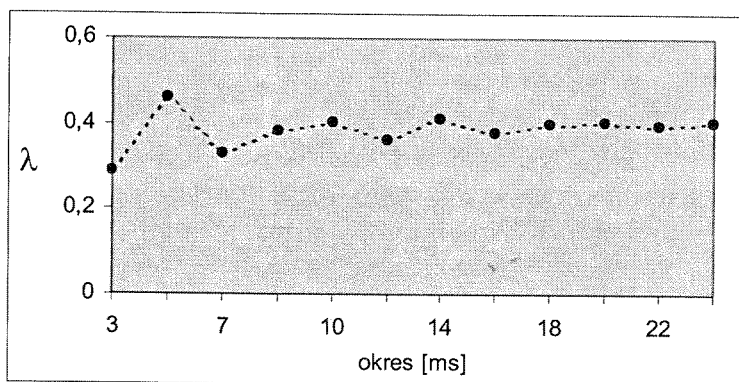
Magistrala CAN umożliwia przydzielenie każdej wiadomości priorytetu (bity 7 do 10 numeru ID). Tym samym przeprowadzone badania de facto nie dotyczą metody FIFO, ale metody ze statycznym przydziałem priorytetu. Jednak wyniki badań prezentowane są w bieżącym rozdziale, ponieważ badania przeprowadzono dla tych samych modeli, co dla innych magistral w przypadku metody FIFO. Postępowanie takie ma na celu możliwość porównania osiągnięć magistrali CAN w stosunku do innych badanych magistral dla tych samych parametrów modelu. Badania przeprowadzono przy założeniu, że przyporządkowane priorytety są odwrotnie proporcjonalne do długości ograniczenia czasowego (największy priorytet dla wiadomości z najkrótszym ograniczeniem czasowym). Ponieważ dla przyjętego modelu ograniczenie czasowe równe jest okresowi występowania, wiadomości szeregowane są według metody GRMS.

Na rys. 13, 14, 15 przedstawiono charakterystyki pierwszego typu. Są one inne niż charakterystyki uzyskane dla PROFIBUS FMS i InterBus-S. CAN jest magistralą z proporcjonalnym narzutem danych nadmiarowych (do liczby przesyłanych bajtów danych). W jednej ramce może być przesłane do 8 bajtów danych. Jeżeli dane te są dłuższe od 8 bajtów, to dzielone są one na 7 bajtowe bloki danych. Każdy przesyłany 7-mio bajtowy blok danych (plus 1 bajt dodatkowy z informacją o podziale) przesyłany



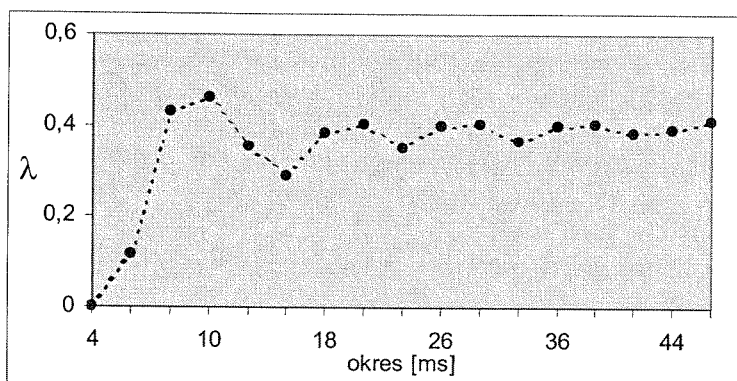
Rys. 13. CAN: 3 węzły, 3 wiadomości w węźle

Fig. 13. CAN: 3 nodes, 3 messages in each node



Rys. 14. CAN: 3 węzły, 6 wiadomości w węźle

Fig. 14. CAN: 3 nodes, 6 messages in each node

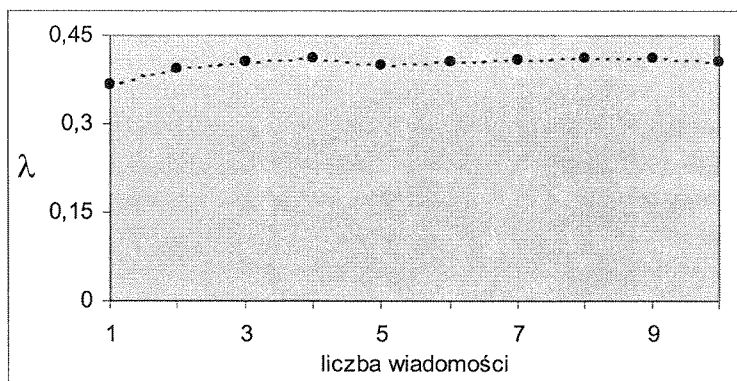


Rys. 15. CAN: 6 węzłów, 6 wiadomości w węźle

Fig. 15. CAN: 6 nodes, 6 messages in each node

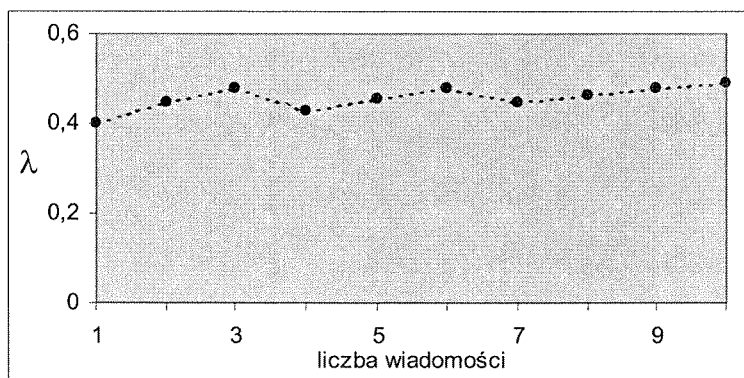
jest w formie osobnej ramki, czyli z osobną porcją danych nadmiarowych. Tym samym rośnie liczba przesyłanych danych nadmiarowych wraz ze wzrostem liczby bajtów przesyłanych danych. Dlatego też maksymalna wartość współczynnika obciążenia systemu osiągana jest przy przesyłaniu 8 bajtów danych (wartość ta jest większa niż osiągana dla tej liczby danych wartość przez PROFIBUS FMS i InterBus-S). Następnie charakterystyka opada (przesłanie od 8 do 15 bajtów danych wymaga dwóch ramek), rośnie, a następnie znowu spada (przesłanie od 16 do 21 bajtów danych wymaga trzech ramek), itd. Przesłanie każdej dodatkowej ramki to przesłanie kolejnych danych nadmiarowych, czyli efektywność systemu spada.

Na rys. 16, 17 przedstawiono charakterystyki drugiego typu. Dla magistrali CAN, ze względu na brak narzutu protokołu (czas t_{-w} $T = 0$) charakterystyki drugiego i trzeciego typu są identyczne, dlatego też te ostatnie nie są prezentowane. Przedstawiono natomiast charakterystyki drugiego typu dla dwóch wartości długości pola danych: 8



Rys. 16. CAN: 5 węzłów, długość pola danych 32[B]

Fig. 16. CAN: 5 nodes, data length 32[B]



Rys. 17. CAN: 5 węzłów, długość pola danych 8[B]

Fig. 17. CAN: 5 nodes, data length 8[B]

i 32 bajtów. Wzrost liczby danych nadmiarowych wraz ze wzrostem liczby bajtów przesyłanych danych. Dlatego też maksymalna wartość współczynnika obciążenia systemu osiągana jest przy przesyłaniu 8 bajtów danych (wartość ta jest większa niż osiągana dla tej liczby danych wartość przez PROFIBUS FMS i InterBus-S). Następnie charakterystyka opada (przesłanie od 8 do 15 bajtów danych wymaga dwóch ramek), rośnie, a następnie znowu spada (przesłanie od 16 do 21 bajtów danych wymaga trzech ramek), itd. Przesłanie każdej dodatkowej ramki to przesłanie kolejnych danych nadmiarowych, czyli efektywność systemu spada.

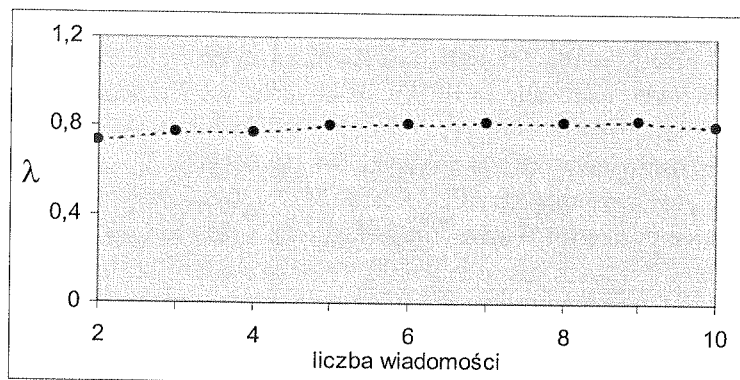
Przeanalizowano jak działają tak zwane GRMS n... nych pri... kowano... w poszczególnych minimalnych... wynosi 2... to kolejn... 8 ms... I... wyznacz... charakter...

Magistrala Jeżeli... złoży j... magistral...

...samym
...ów prze-
...systemu
...osiągana
...charak-
..., rośnie,
...ramek),
...rowych,

...li CAN,
...o i trze-
...stawiono
...anych: 8

i 32 bajty. Wyraźnie widać, że dla 8 bajtów system osiąga większe wartości współczynnika obciążenia niż dla 32 bajtów. Na rys. 18 przedstawiono przykładową charakterystykę dla szybkości przesyłu danych 1000 Kb/s. Wzrost szybkości przesyłanych danych z 500 Kb/s do 1000 Kb/s powoduje dwukrotny wzrost wartości współczynnika obciążenia systemu.



Rys. 18. CAN: 5 węzłów, długość pola danych 32[B], szybkość 1000 [Kb/s]

Fig. 18. CAN: 5 nodes, data length 32[B], throughput 1000 [Kb/s]

4. ANALIZA WŁAŚCIWOŚCI MAGISTRAL DLA METODY GRMS

Przeprowadzone analizy dla metody GRMS dla identycznych parametrów modelu jak dla metody FIFO (wszystkie wiadomości mają ten sam okres występowania) dały takie same wyniki jak dla metody FIFO. Oznacza to, że stosowanie metody GRMS nie ma sensu, jeżeli wiadomościom nie można przyporządkować zróżnicowanych priorytetów na podstawie różnych okresów występowania. Dlatego też zmodyfikowano parametry modelu strumienia wiadomości: okresy występowania wiadomości w poszczególnych węzłach tworzą ciąg arytmetyczny, o odległości równej okresowi minimalnemu (pierwszemu z elementów ciągu). Przykładowo, jeżeli okres minimalny wynosi 2 ms (ta wartość jestznaczona na osi odciętych wykresów pierwszego typu), to kolejne okresy (w zależności od liczny wiadomości w węźle) wynoszą 4 ms, 6 ms, 8 ms... Dla wykresów drugiego typu poszczególne punkty na charakterystyce były wyznaczane też dla powyższego modelu. W celu porównania metody GRMS i FIFO charakterystyki dla powyższego modelu wyznaczono także dla metody FIFO.

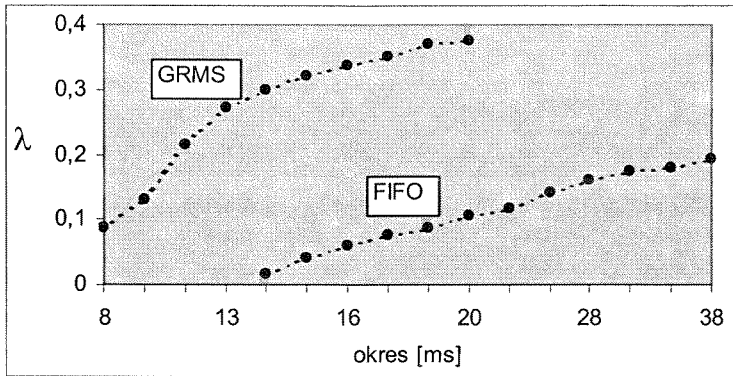
Magistrala PROFIBUS FMS

Jeżeli w czasie wykonywania transakcji zostanie przekroczony przydzielony węzłowi j czas h_j (h_j — maksymalny czas, w którym węzeł j wyłącznie dysponuje magistralą), to pomimo tego węzeł kontynuuje wykonywanie transakcji aż do jej peł-

nego zakończenia. Ponieważ jest to istotna różnica w stosunku do innych magistral, przeprowadzono analizy dla dwóch wariantów działania magistrali PROFIBUS FMS:

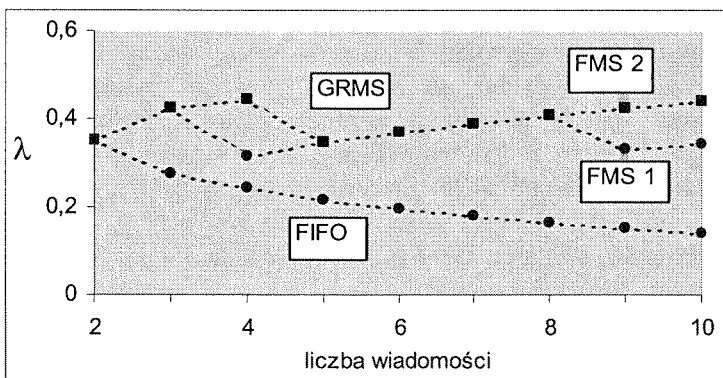
- wariant 1, węzeł j może przekraczać przydzielony czas h_j ,
- wariant 2, węzeł j nie może przekraczać przydzielonego czasu h_j (w praktycznym rozwiązaniu zadanie to przypadałoby na warstwę aplikacji, która w przypadku, gdyby wysłanie wiadomości miało spowodować przekroczenie zadanego czasu h_j nie wysyłałyby danych do bufora sieciowego).

Na rys. 19 przedstawiono charakterystyki pierwszego typu. Wyraźnie widać przewagę metody GRMS nad metodą FIFO. W obydwu wariantach działania magistrali (możliwość przekraczania i brak przekraczania czasu h_j) uzyskano te same wyniki. Na rys. 20 przedstawiono charakterystyki drugiego typu. W tym przypadku także widać wyraźną przewagę metody GRMS nad FIFO. Dodatkowo zwraca uwagę fakt, że występują różnice pomiędzy wariantem 1 (FMS 1), a wariantem 2 (FMS 2) działania



Rys. 19. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle

Fig. 19. PROFIBUS FMS: 6 nodes, 6 messages in each node



Rys. 20. PROFIBUS FMS: 5 węzłów, długość pola danych 32[B]

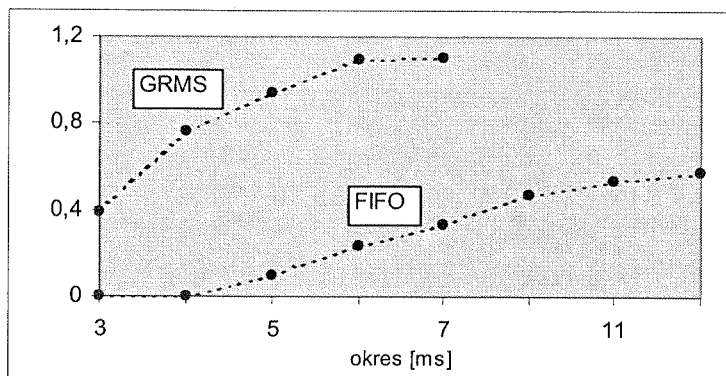
Fig. 20. PROFIBUS FMS: 5 nodes, data length 32 [B]

magistral,
FMS:

ycznym
ypadku,
zasu h_j

ć prze-
magistrali
wyniki.
u także
fakt, że
ziałania

magistrali. Lepsze wyniki uzyskano w przypadku braku przekraczania przydzielonego czasu h_i . Wynika to z faktu, że jeżeli węzeł i może przekroczyć przydzielony mu czas h_i to wysłane mogą zostać te wiadomości, które mogłyby zostać wysłane później, a i tak dotrzymałyby ograniczeń czasowych. Natomiast inny węzeł ma mniej czasu do swojej dyspozycji i może dojść do przekroczenia ograniczeń czasowych.



Rys. 21. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle, szybkość 1500 [Kb/s]

Fig. 21. PROFIBUS FMS: 6 nodes, 6 messages in each node, throughput 1500 [Kb/s]

Na rys. 21 przedstawiono przykładową charakterystykę magistrali PROFIBUS FMS dla szybkości przesyłu danych 1500 Kb/s. Wzrost szybkości przesyłanych danych z 500 Kb/s do 1500 Kb/s powoduje trzykrotny wzrost wartości współczynnika obciążenia systemu danymi przy zachowaniu tych samych własności charakterystyk. Przeprowadzono także analizy w celu potwierdzenia wyników, jakie otrzymano dla metody FIFO dotyczące wpływu liczby węzłów na osiągi systemu. Badania przeprowadzono w następujący sposób: wyznaczano wartość współczynnika obciążenia systemu dla przypadku, gdy system składał się z n węzłów, a w każdym węźle było m wiadomości ($m = 2 * n$). Następnie wyznaczano wartość tegoż współczynnika, gdy system składał się z m węzłów, a każdym węźle było n wiadomości ($n = m/2$). Parametry czasowe wiadomości w obu przypadkach były identyczne. Uzyskane wyniki przedstawia tabela 1. Wartość współczynnika obciążenia systemu jest większa dla przypadku mniejszej liczby węzłów w systemie (dla tej samej liczby wiadomości w systemie).

Tabela 1

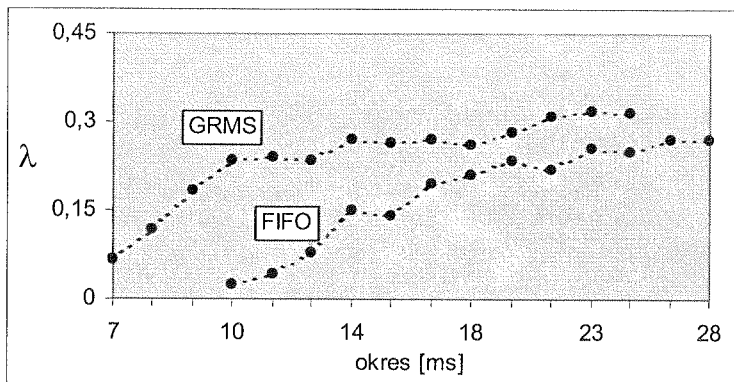
Właściwości systemu z magistralą PROFIBUS FMS

Properties of system based on PROFIBUS FMS

Liczba węzłów	Liczba wiadomości w węźle	Obciążenie systemu	Liczba węzłów	Liczba wiadomości w węźle	Obciążenie systemu
2	4	0.439	4	2	0.341
3	6	0.433	6	3	0.402
4	8	0.449	8	4	0.328
5	10	0.433	10	5	0.354

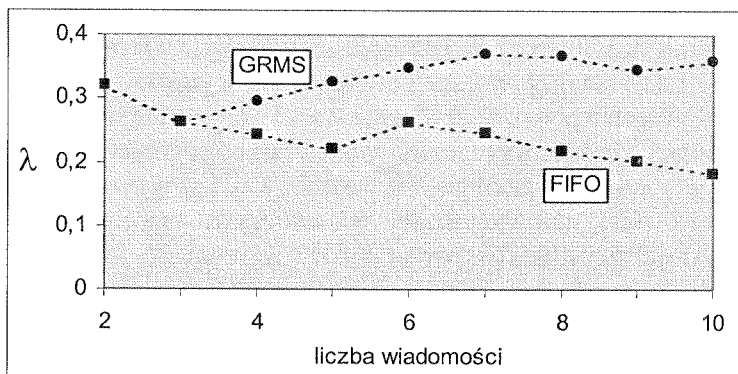
Magistrala InterBus-S

Na rys. 22, 23 przedstawiono charakterystyki uzyskane dla magistrali InterBus-S. Podobnie jak dla PROFIBUS FMS, dla metody GRMS uzyskano lepsze wyniki niż dla metody FIFO. Przeprowadzono także badania mające na celu potwierdzenia wyników, jakie otrzymano dla metody FIFO dotyczące wpływu liczby węzłów na osiągi systemu. Badania przeprowadzono w analogiczny sposób jak badania dla magistrali PROFIBUS FMS. Wyniki badań przedstawia tabela 2. Uzyskane wyniki nie potwierdzają wniosków



Rys. 22. InterBus-S: 6 węzłów, 6 wiadomości w węźle, szczelina 8[B]

Fig. 22. InterBus: 6 nodes, 6 messages in each node, slot 8[B]



Rys. 23. InterBus-S: 5 węzłów, długość pola danych 32[B], szczelina 8[B]

Fig. 23. InterBus: 5 nodes, data length 32[B], slot 8[B]

uzyskanych dla metody FIFO, gdyż zwiększanie liczby węzłów kosztem zwiększenia liczby wiadomości w węźle nie zawsze przynosi lepsze efekty. Przeprowadzono następujące, dodatkowe badania, których celem było stwierdzenie, czy dodawać nowe węzły, czy też dodawać wiadomości do istniejących węzłów w przypadku zwiększenia liczby wiadomości w systemie. Wybrano trzy przykłady:

Liczba węzłów
2
3
4
5

I. Para...
ści, okre...
nika ob...
Problem...
węzła d...
kolejny...
Wynik:...
obciążen...
0.296. I...
liczby w...
szano li...
węzle, z...
liczbę w...
II. Para...
wiadom...
współcz...
Problem...
dwie wi...
węzłowa...
Wynik:...
obciążen...
0.305.
III. Para...
domośc...
czynnik...
Problem...
wiadom...
wiadom...
Wynik:...
nika ob...
wartość

Tabela 2

Właściwości systemu z magistralą InterBus-S

Properties of system based on InterBus-S

Liczba węzłów	Liczba wiadomości w węźle	Obciążenie systemu	Liczba węzłów	Liczba wiadomości w węźle	Obciążenie systemu
2	4	0.279	4	2	0.307
3	6	0.375	6	3	0.268
4	8	0.371	8	4	0.316
5	10	0.377	10	5	0.344

I. Parametry systemu: system składa się z 4 węzłów, w każdym węźle jest 4 wiadomości, okresy występowania wiadomości tworzą ciąg arytmetyczny. Wartość współczynnika obciążenia wynosi 0.284.

Problem: system rozbudowano o 4 wiadomości. Powstaje pytanie, czy do każdego węzła dołożyć po jednej wiadomości (4 węzły/5 wiadomości w węźle), czy dołożyć kolejny węzeł (5 węzłów/4 wiadomości w węźle).

Wynik: dla przypadku 4 węzły/5 wiadomości w węźle uzyskano wartość współczynnika obciążenia 0.267, a dla przypadku 5 węzłów/4 wiadomości w węźle uzyskano wartość 0.296. Lepsze efekty uzyskano więc zwiększając liczbę węzłów kosztem zwiększenia liczby wiadomości dla istniejącej liczby węzłów. Dla podobnych sytuacji, gdy zwiększano liczbę węzłów o 1 kosztem zwiększenia liczby wiadomości o 1 w każdym węźle, zawsze uzyskiwano lepsze wyniki. Natomiast sytuacja ulegała zmianie, gdy liczbę węzłów zwiększano o więcej niż 1 węzeł.

II. Parametry systemu: system składa się z 4 węzłów, w każdym węźle są 4 typy wiadomości, okresy występowania wiadomości tworzą ciąg arytmetyczny. Wartość współczynnika obciążenia wynosi 0.284.

Problem: system rozbudowano o 8 wiadomości. Czy do każdego węzła dołożyć po dwie wiadomości (4 węzły/6 wiadomości w węźle), czy dołożyć 2 kolejne węzły (6 węzłów/4 wiadomości w węźle).

Wynik: dla przypadku 4 węzły/6 wiadomości w węźle uzyskano wartość współczynnika obciążenia 0.320, a dla przypadku 6 węzłów/4 wiadomości w węźle uzyskano wartość 0.305.

III. Parametry systemu: system składa się z 4 węzłów, w każdym węźle jest 4 wiadomości, okresy występowania wiadomości tworzą ciąg arytmetyczny. Wartość współczynnika obciążenia wynosi 0.284.

Problem: system rozbudowano o 16 wiadomości. Czy do każdego węzła dołożyć po 4 wiadomości (4 węzły/8 wiadomości w węźle), czy dołożyć 4 kolejne węzły (8 węzłów/4 wiadomości w węźle).

Wynik: dla przypadku 4 węzły/8 wiadomości w węźle uzyskano wartość współczynnika obciążenia 0.371, a dla przypadku 8 węzłów/4 wiadomości w węźle uzyskano wartość 0.316.

Lepsze efekty uzyskano więc zwiększając liczbę wiadomości dla istniejącej liczby węzłów kosztem zwiększenia liczby węzłów.

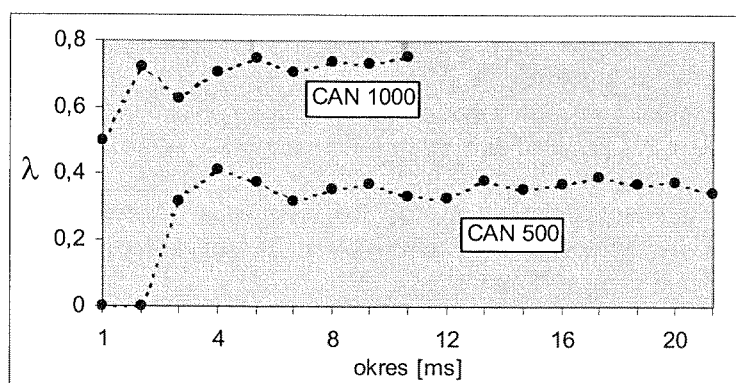
Na podstawie uzyskanych wyników badań proponowany jest więc następujący sposób projektowania systemu z magistralą InterBus-S:

- wyznaczamy minimalny okres występowania wiadomości w systemie t_{min} ,
- wyznaczamy czas obiegu danych T_{cycle} najbardziej zbliżony do t_{min} , ale nie większy niż t_{min} ,
- zakładając, że węzły mają przydzieloną szczelinę 8 bajtów, obliczamy maksymalną liczbę węzłów dla obliczonego czasu T_{cycle} ,
- jeżeli obliczona liczba węzłów jest zbyt mała do praktycznego zrealizowania systemu, obliczamy liczbę węzłów dla szczeliny 4 lub 2 bajty,
- na ile to możliwe, należy równomiernie obciążyć każdy węzeł wiadomościami. Wynika to z faktu, że przydzielona wartość szczeliny jest stała nawet jeżeli węzeł w danym momencie nie wysyła danych. Problem ten nie jest istotny tak bardzo dla innych magistral, np. Dla PROFIBUS FMS, bowiem jeżeli węzeł nie ma wiadomości do wysłania, to oddaje prawo do wysłania wiadomości następnemu węzłowi,
- sprawdzane jest dotrzymanie ograniczeń RT w systemie. Jeżeli ograniczenia nie są dotrzymane, należy zmniejszyć liczbę węzłów, a wiadomości ze zlikwidowanego węzła przenieść do pozostałych. Następnie należy znowu sprawdzić dotrzymanie ograniczeń RT.

Postępowanie takie należy prowadzić do momentu spełnienia warunków RT lub osiągnięcia granicy, poniżej której nie można zmniejszyć liczby węzłów. Jeżeli ograniczenia nie będą spełnione, należy wybrać inną magistralę (nie są tutaj rozważane inne możliwości jak np. grupowanie wiadomości).

Magistrala CAN

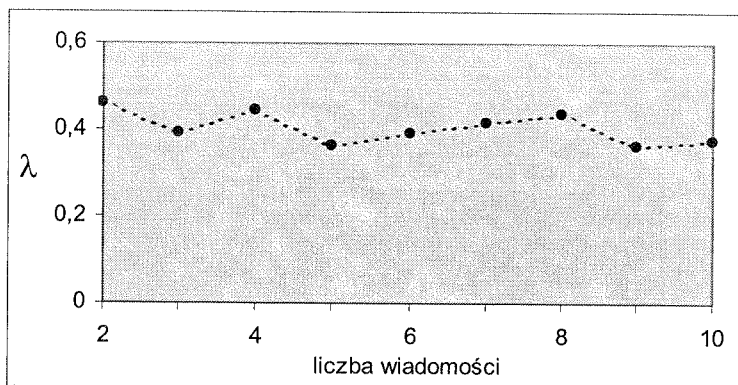
Charakterystyki uzyskane dla CAN przedstawiono na rys. 24, 25, 26. Dla zastosowanych modeli (okresy występowania wiadomości tworzą ciąg arytmetyczny) uzyskano



Rys. 24. CAN: 6 węzłów, 6 wiadomości w węźle, szybkość 500 [Kb/s] i 1000 [Kb/s]

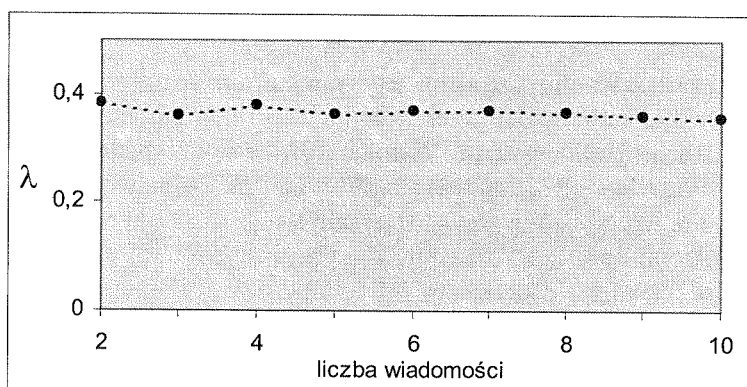
Fig. 24. CAN: 6 nodes, 6 messages in each node, throughput 500 [Kb/s] and 1000 [Kb/s]

zbliżony przebieg charakterystyk jak na rys. 13 do 18. Największą wartość współczynnika obciążenia system uzyskuje przy przesyłaniu 8 bajtów danych. Wzrost szybkości przesyłanych danych z 500 Kb/s do 1000 Kb/s powoduje dwukrotny wzrost wartości współczynnika obciążenia systemu przy zachowaniu tych samych własności charakterystyk (rys. 24).



Rys. 25. CAN: 5 węzłów, długość pola danych 8[B]

Fig. 25. CAN: 5 nodes, data length 8[B]



Rys. 26. CAN: 5 węzłów, długość pola danych 32[B]

Fig. 26. CAN: 5 nodes, data length 32[B]

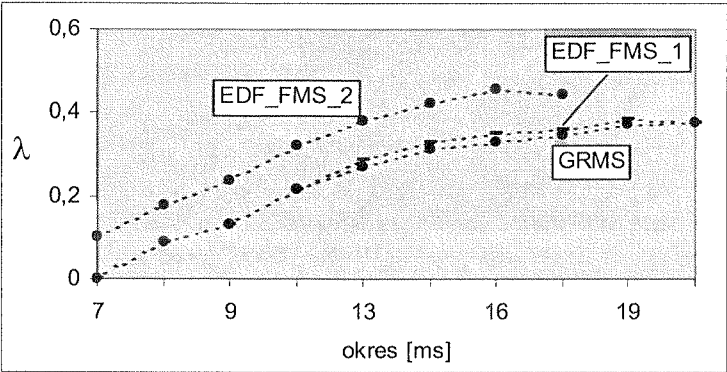
5. ANALIZA MAGISTRAL DLA METODY EDF

Przeprowadzone analizy dla metody EDF dla parametrów modelu identycznych jak dla metody FIFO (tzn. wszystkie wiadomości mają ten sam okres występowania) dały takie same wyniki jak metoda FIFO (podobna sytuacja wystąpiła dla metody GRMS). Dlatego też wykresy te nie zostały pokazane. Przedstawiono natomiast charakterystyki

uzyskane dla modeli stosowanych dla metody GRMS. Dla celów porównawczych, na każdym rysunku zamieszczono charakterystyki uzyskane dla metody GRMS i EDF.

Magistrala PROFIBUS FMS

Na rys. 27 przedstawiono charakterystyki uzyskane dla modelu stosowanego dla metody GRMS — okresy występowania tworzą ciąg arytmetyczny o odległości równej t_{min} (ciąg arytmetyczny 1). Dla wariantu 1 (EDF_FMS_1) działania magistrali

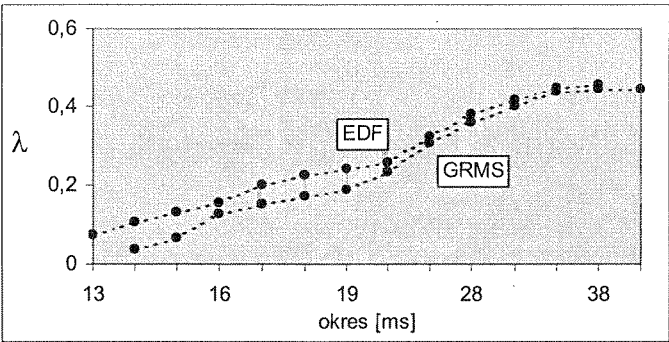


Rys. 27. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle (ciąg arytmetyczny 1)

Fig. 27. PROFIBUS FMS: 6 nodes, 6 messages in each node (arithmetical process 1)

(z możliwością przekraczania czasu h_j) uzyskano gorsze rezultaty niż dla wariantu 2 (EDF_FMS_2) (bez możliwości przekraczania czasu h_j). Dla metody GRMS wyniki w obu przypadkach były identyczne. Wynika stąd, że w przypadku metod dynamicznych szeregowania wiadomości system jest dużo bardziej wrażliwy na wybór wariantu działania sieci niż w przypadku metod statycznych.

Przeprowadzono dodatkowe badania, dla innych parametrów modelu. Na rys. 28 zamieszczono przykładowe charakterystyki, kiedy to dla wariantu 1 i 2 uzyskano te sa-



Rys. 28. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle (ciąg arytmetyczny 2)

Fig. 28. PROFIBUS FMS: 6 nodes, 6 messages in each node (arithmetical process 2)

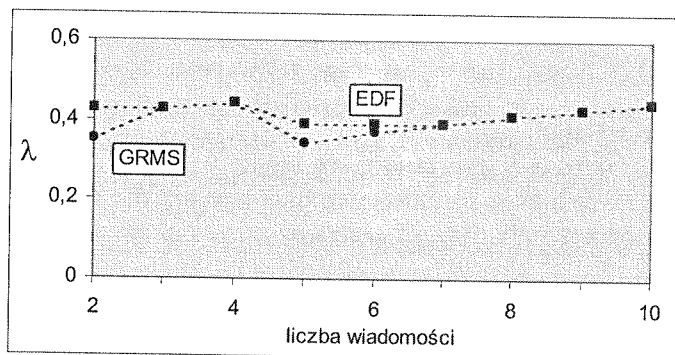
ych, na
EDF.

ego dla
ci rów-
magistrali

me wyniki. W zastosowanym modelu, okresy występowania tworzą ciąg arytmetyczny o stałej odległości 1 [ms] (ciąg arytmetyczny 2).

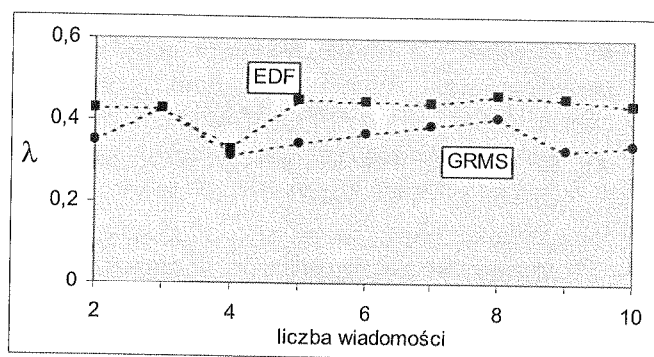
Z przeprowadzonych badań płyną następujące wnioski:

- magistrala PROFIBUS FMS przy zastosowaniu metod dynamicznych, jest wrażliwa na zastosowany wariant działania magistrali,
- dla wariantu 2 (bez możliwości przekraczania czasu h_j) dla metody EDF uzyskuje się zawsze lepsze wyniki niż dla GRMS (rys. 27, 28, 30),
- dla wariantu 1 (z możliwością przekraczania czasu h_j) wyniki uzyskane dla EDF są zbliżone do GRMS (rys. 27, 29).



Rys. 29. PROFIBUS FMS (możliwe przekraczanie czasów h_j): 5 węzłów, długość pola danych 32[B]

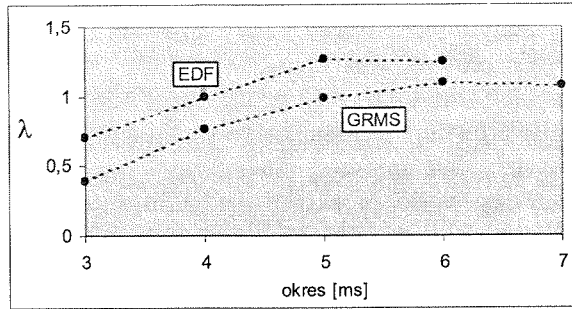
Fig. 29. PROFIBUS FMS (possibility of exceeding h_j): 5 nodes, data length 32[B]



Rys. 30. PROFIBUS FMS (bez przekraczania czasów h_j): 5 węzłów, długość pola danych 32[B]

Fig. 30. PROFIBUS FMS (without exceeding h_j): 5 nodes, data length 32[B]

Na rys. 31 przedstawiono przykładową charakterystykę magistrali PROFIBUS FMS dla szybkości przesyłu danych 1500 Kb/s. Wzrost szybkości przesyłanych danych powoduje wzrost wartości współczynnika obciążenia systemu danymi przy zachowaniu tych samych własności charakterystyk.

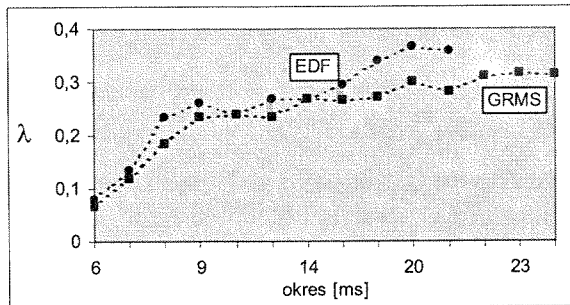


Rys. 31. PROFIBUS FMS: 6 węzłów, 6 wiadomości w węźle, szybkość 1500 [Kb/s]

Fig. 31. PROFIBUS FMS: 6 nodes, 6 messages in each node, throughput 1500 [Kb/s]

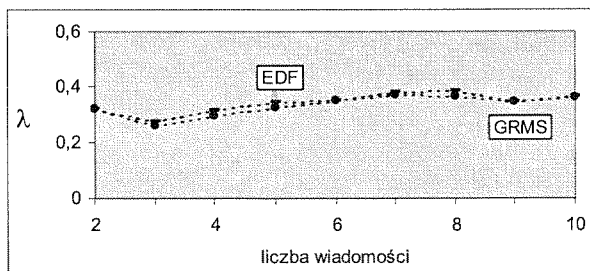
Magistrala InterBus-S

Dla magistrali InterBus-S uzyskano także lepsze wyniki dla metody EDF niż dla GRMS (rys. 32, 33). Jednakże różnice te są mniejsze niż dla magistrali PROFIBUS FMS. Powodem tego jest stała długość szczeliny przydzielona do każdego węzła.



Rys. 32. InterBus-S: 6 węzłów, 6 wiadomości w węźle

Fig. 32. InterBus-S: 6 nodes, 6 messages in each node



Rys. 33. InterBus-S: 5 węzłów, długość pola danych 32[B]

Fig. 33. InterBus-S: 5 nodes, data length 32[B]

Magistrala CAN

Zastosowanie metod z dynamicznym przydziałem priorytetu dla magistrali CAN, powoduje, że w różnych węzłach sieci wiadomości mogą otrzymać ten sam priorytet (ten sam numer ID). Spowoduje to z kolei wystąpienie błędu przy próbie wysłania dwóch wiadomości o tym samym numerze ID. W związku z powyższym stosowanie metod dynamicznych dla magistrali CAN jest niemożliwe.

6. WNIOSKI

Projektowanie rozproszonych systemów czasu rzeczywistego o ostrych ograniczeniach czasowych opartych o magistrale miejscowe jest trudne i czasochłonne. Pomocnym narzędziem są wyniki badań przedstawiające wypowiedzi magistral miejscowych w funkcji zmian różnych parametrów systemu i dla różnych metod szeregowania wiadomości. W szczególności, z uzyskanych wyników badań dla metod szeregowania można wyciągnąć następujące wnioski:

- metoda FIFO jest metodą bardzo prostą i z małymi narzutami, ale charakteryzuje się najmniejszą wartością współczynnika obciążenia systemu, dla którego w systemie dotrzymane zostają ograniczenia RT w stosunku do innych metod.
- dla metod dynamicznych uzyskano lepsze wyniki niż dla metod statycznych. Należy jednak zauważyć, że w przeprowadzonych badaniach nie uwzględniano narzutów obliczeniowych poszczególnych metod. Ponieważ narzuty te są większe dla metod dynamicznych, należy sądzić, że różnice dla metod dynamicznych i statycznych byłyby mniejsze po uwzględnieniu tychże narzutów.

Wartość współczynnika λ dla różnych magistral zależy w dużej mierze od liczby bajtów danych przesyłanych w jednej ramce wiadomości. Przeprowadzone dodatkowe badania pokazały, że przy przesyłaniu 1 lub 2 bajtów danych najlepsze efekty uzyskujemy w przypadku magistrali InterBus-S (przesyłanie kanałem procesowym), dla przesyłu danych od 3 do 8 bajtów najlepiej wykorzystać magistralę CAN, z kolei PROFIBUS FMS uzyskuje najlepsze efekty dla przesyłu danych powyżej 8 bajtów [11].

Praca została wykonana w ramach grantu KBN Nr 8T11A00718: *Analiza czasowa rozproszonych systemów sterowania bazujących na magistralach miejscowych.*

7. BIBLIOGRAFIA

1. K. Chen, L. Decreasefond: *Just How Bad is the FIFO Discipline for Handling Randomly Arriving Time-Critical Messages*. Proc. of 1995 IEEE International Workshop on Factory Communication Systems, pp. 183-190
2. A. Drwal, J. Werewka, S. Żaba: *Protokoły dostępu i charakterystyki czasowe magistral miejscowych*. Pomiary Automatyka Robotyka, wrzesień 1997 Vol. 7, 1997, pp. 4-11
3. K. Jeffay, D. Stanat, C. Martel: *On Non Preemptive Scheduling of Periodic and Sporadic Tasks*. Proc. of IEEE Real Time Systems Symposium, 1991

4. L. Liu, J. Layland: *Scheduling algorithms for multiprograming in a hard real - time environment*. J. Assoc. Comput. Mach. Vol. 20, No. 1, 1973, pp. 46-61
5. P. Pleinevaux, J.D. Decotignie: *Time Critical Communication Networks: Field Buses*. IEEE Network Vol. 2, No. 3, May 1988, pp. 55-63
6. J. Werewka: *Systemy rozproszone sterowania i akwizycji danych. Sterowniki programowalne i magistrale miejscowe*. Kraków, Wydawnictwa CCATIE, 1998
7. S. Żaba: *Message scheduling in distributed real-time system based on fieldbus*. 5th International Symposium on Methods and Models in Automation and Robotics, Międzyzdroje, 25-29 August 1998, pp. 565-570
8. S. Żaba, J. Werewka: *Szeregowanie wiadomości w rozproszonych systemach czasu rzeczywistego opartych o magistrale miejscowe*. Kwartalnik Elektroniki i Telekomunikacji, T. 45, Z. 1, 1999, ss. 25-50
9. S. Żaba: *Szeregowanie wiadomości z dynamicznym przydziałem priorytetu w rozproszonych systemach sterujących*. VI Konferencja Systemy Czasu Rzeczywistego, 27-30 wrzesień, Zakopane 1999
10. S. Żaba: *Analiza czasowa rozproszonych systemów sterowania bazujących na magistralach miejscowych*. W opracowaniu pod red. T. Szmuc i J. Werewka: *Analiza i projektowanie systemów komputerowych czasu rzeczywistego o różnym stopniu rozproszenia*. Polskie Towarzystwo Informatyczne, oddz. małopolski Kraków 2001
11. S. Żaba: *Analiza czasowa rozproszonych systemów sterowania bazujących na magistralach miejscowych*. Rozprawa doktorska, Kraków, Wydział EAIiE AGH, 2001

S. ŻABA

EXPERIMENTAL TESTING OF FIELDBUSES FOR SOME MESSAGE SCHEDULING METHODS

S u m m a r y

In this paper properties and time parameters of three fieldbuses PROFIBUS FMS, CAN, InterBus are presented as the result of research based on checking of RT - constraint fulfilment of messages in distributed real time system based on fieldbus. Research was done for FIFO method and real time methods scheduling GRMS and EDF. Tests were performed for the worst case - it was assumed that all messages in system were activated in instant zero. The message blocking is also considered. Results of testing were obtain by using computer algorithms created by author. On the basis of performing tests we can draw conclusions as follow:

- FIFO method is very simple and requires small calculating time, but its performance are the least from among priorities method,
- Dynamic methods have better performance than static priority for each fieldbus (except for CAN fieldbus, because there is impossible to apply dynamic method to this fieldbus),
- InterBus-S fieldbus has the best effects for transmitting 1 or 2 bytes data length, CAN fieldbus has the best effects for transmitting 3 to 8 bytes data length and PROFIBUS FMS fieldbus has the best effects for transmitting messages with data length above 8 bytes.

Keywords: distributed control system, real time system, fieldbus network, message scheduling, FIFO, GRMS, EDF methods

oraz
z na
się p
wać
do c
mien
Pozw
mec
zowa
kosz
na r
char

Analiza pełnofalowa anteny skanującej na podłożu ferroelektrycznym

EUGENIUSZ JASZCZYSZYN

*Instytut Radioelektroniki, Politechnika Warszawska, Nowowiejska 15/19, 00-665, Warszawa
e-mail: e.jaszczyszyn@ire.pw.edu.pl*

*Otrzymano 2003.04.17
Autoryzowano 2003.06.02*

W pracy przedstawiono koncepcję i opracowano nowy rodzaj skanującej anteny paskowej na podłożu ferroelektrycznym bez konieczności wykorzystania przesuwników fazy. Ponadto opracowano uogólnioną metodę analizy elektrodynamicznej rozpatrywanej struktury promieniującej wraz z metodą analizy numerycznej. Wyniki badań anteny skanującej wskazują na możliwość szerokiego jej zastosowania, a przedstawiona metoda analizy pełnofalowej do projektowania struktur promieniujących.

Słowa kluczowe: anteny paskowe, anteny skanujące, ferroelektryki

1. WPROWADZENIE

W związku z gwałtownym rozwojem systemów telekomunikacji bezprzewodowej oraz „wyrafinowanych” systemów radarowych, technika antenowa jest obecnie jedną z najważniejszych dziedzin elektroniki i telekomunikacji. W takich systemach pojawia się potrzeba stosowania anten inteligentnych. Anteny inteligentne mogą dostosowywać się do zewnętrznych warunków w przestrzeni elektromagnetycznej, na przykład, do dynamicznie powstających zakłóceń. Problem kształtowania charakterystyki promieniowania można rozwiązać poprzez zastosowanie fazowanych szyków antenowych. Pozwalają one sterować kierunkiem nadawania i odbioru sygnałów bez konieczności mechanicznej zmiany położenia systemu antenowego w przestrzeni. Możliwości fazowanych szyków antenowych są dobrze znane, a ich główną wadę stanowią znaczne koszty podzespołów (przesuwników fazy, tłumików, dzielników mocy, itd.). Ze względu na rozmiary, objętość, koszty systemu antenowego oraz możliwości łatwego sterowania charakterystykami anteny, konieczne jest opracowanie i badanie nowych, tanich i łatwo

sterowanych elementów promieniujących oraz szyków antenowych. Celem niniejszej pracy jest opracowanie nowego rodzaju skanującej anteny paskowej (odpowiedniej do zastosowań w systemach z przemiataaniem wiązki) bez konieczności stosowania przesuwników fazy, dzielników mocy itd. Przedstawiona koncepcja ma na celu znaczne obniżenie kosztów budowy anteny skanującej.

Pomimo ogromnych wysiłków, obniżenie kosztów fazowanych szyków antenowych odbywa się znacznie wolniej, niż można było przewidzieć. Jednak, przedstawione w literaturze światowej wyniki badań różnych rodzajów anten skanujących wskazują na możliwość otrzymania praktycznych rozwiązań i szerokiego ich zastosowania w nowoczesnych systemach radiokomunikacji oraz radiolokacji.

Wśród dużej ilości proponowanych tanich rozwiązań można wyróżnić cztery podstawowe sposoby skanowania (pomijając szyki fazowane): anteny z elektronicznym skanowaniem, ze sterowaniem optycznym, z wykorzystaniem materiałów ferromagnetycznych oraz z wykorzystaniem materiałów ferroelektrycznych. Są znane także anteny ze skanowaniem częstotliwościowym [1], jednak liczne zastosowania wymagają anten skanujących, działających na pewnej, ustalonej częstotliwości.

Najczęściej wykorzystywane są skanujące anteny z elektronicznym skanowaniem. W [2] autorzy zaproponowali wykorzystanie diod p-i-n do modulacji wymiarów falowodu z falą wypływającą. W taki sposób udało się otrzymać odchylenie wiązki o 10° . Podobne rozwiązanie było przedstawione w [3], gdzie elektronicznie zmieniano odległość między niejednorodnościami. Prostokątny falowód dielektryczny był obciążony okresowym układem pasków metalowych, które zawierały diody p-i-n. Diody te działały jako przełączniki. W taki sposób można zmieniać odległość między paskami. Przedstawiony model anteny mikrofalowej, pracujący na częstotliwości 3.5 GHz, pozwala na odchylenie wiązki o 35° . Jednak, antena nie była pozbawiona wad. Na przykład, poziom listków bocznych wynosił tylko -7 dB.

Zasada działania anten ze sterowaniem optycznym polega na wykorzystaniu otwartego falowodu albo z półprzewodnika, albo z dielektryka, zawierającego światłoczułe warstwy [4], [5]. Oświetlając te warstwy przez odpowiednią maskę zmienia się przewodność powierzchni falowodu. Podobnie do anteny z falą wypływającą, obciążonej paskami metalowymi, przedstawiona antena jest „obciążona” zmiennymi przewodnościami, których kształt odpowiada zastosowanej masce. Wadą takiego rozwiązania jest duże tłumienie fali propagującej wzdłuż struktury. W [6] autorzy przedstawili nową antenę, która charakteryzuje się lepszymi parametrami. Przedstawiona antena ma możliwość odchylenia wiązki w zakresie $\pm 15^\circ$. Oprócz tego antena może pracować na polaryzacji liniowej, jak i kołowej.

Kilka przykładów anten, wykorzystujących materiały ferromagnetyczne, przedstawiono w [7] i [8]. Antena przedstawiona w [7] jest liniową anteną 20-szczelinową działającą na częstotliwości 36.25 GHz. Pozwala ona na przemiataanie wiązki w zakresie od 29° do 43° . Większy sektor przemiataania (do 24°) może być uzyskany przez powiększenie powierzchni przekroju falowodu ferromagnetycznego. Uzyskanie takiego sektora pokrywa się za cenę zmniejszenia zysku i zwiększenia poziomu listków

bocznych. Podstawowym elementem anteny przedstawionej w [8] jest trzy warstwowy falowód: ferryt-dielektryk-ferryt. Na górnej warstwie ferrytowej znajduje się liniowy szczyk dipoli. W najprostszym przypadku antena cechowała się sektorem $\pm 20^\circ$ oraz poziomem listków bocznych około -12 dB. Moc sterowania wynosiła około 1 W, a prąd rzędu ± 0.7 A.

Główna zaleta anten wykorzystujących materiały ferroelektryczne polega na możliwości zmiany przenikalności elektrycznej struktury anteny, poprzez zmianę zewnętrznego sterującego pola elektrycznego E_0 . Porównanie właściwości dwóch typów materiałów nieliniowych — ferroelektryków i ferromagnetyków, ze względu na ich wykorzystanie w procesie sterowania, pozwala wyciągnąć wnioski na korzyść ferroelektryków [9], [10]. W szczególności moc sterowania ferroelektryków jest o rząd wielkości (lub nawet dwa) niższa niż w przypadku ferromagnetyków. Anteny, wykorzystujące ferroelektryki nadają się do szybszego przestrajania, są mniejsze i mają prostszą strukturę oraz mogą pracować z wykorzystaniem większych mocy.

Podstawowym powodem, dla którego materiały ferroelektryczne nie były wykorzystywane w zakresie mikrofal, były duże straty oraz wymagane znaczne wartości pola polaryzującego.

W [11] przedstawiono dwie konfiguracje ferroelektrycznych szczyków antenowych. Dla jednej anteny zastosowano materiał ferroelektryczny, którego straty wynosiły około 2 dB/ λ (λ — długość fali) w paśmie X. Jeśli wymagana długość anteny wynosiłaby 50λ , wtedy straty uniemożliwiłyby stworzenie pożądanego rozkładu pola na aperturze. Z tego powodu skoncentrowano się na konfiguracji soczewkowej. Wykorzystując materiał ferroelektryczny, z zakresem zmiany przenikalności od 80 do 120 oraz z tangensem strat rzędu 0.01 , autorzy pokazali, że pożądana grubość soczewki dla uzyskania przesunięcia fazy o 360° jest około λ , co powoduje, że straty w soczewce nie przekraczają 2 dB. Jednak, autorom nie udało się zmniejszyć bardzo wysokiego napięcia sterowania (rzędu 13.5 kV).

Celem niniejszej pracy jest przedstawienie koncepcji oraz opracowanie nowego rodzaju skanującej anteny paskowej na podłożu ferroelektrycznym, bez konieczności wykorzystania przesuwników fazy. Ponadto opracowano uogólnioną metodę analizy elektrodynamicznej rozpatrywanej struktury promieniującej wraz z metodą analizy numerycznej.

Podstawowym materiałem ferroelektrycznym zastosowanym w przeprowadzonych badaniach były tworzywa oparte na modyfikowanym różnego rodzaju dodatkami tytanianem baru lub roztworze stałym tytanianu baru i strontu ($\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$) [12]. Tytanian barowy jako podstawowe ceramiczne tworzywo ferroelektryczne bardzo rzadko stosowane jest w postaci czystej. Wiele jego parametrów ulega korzystnym zmianom, jeśli wprowadza się do niego niewielkie domieszki różnego typu substancji (np. tlenki żelaza, niklu czy manganu). Domieszki te powinny tworzyć roztwór stały z tytanianem barowym lub też modyfikować granice międzyziarnowe w polikrystalicznym spieku ceramicznym. Czysty tytanian barowy charakteryzuje się temperaturą Curie w granicach 120 – 127°C . Poprzez zmianę składu roztworu stałego oraz wprowadzenie

niewielkich ilości domieszek możliwe jest jednak obniżenie tej temperatury nawet do -95°C . W przypadku zastosowania tego typu materiałów do budowy anten i innych przyrządów działających w zakresie mikrofalowym, byłoby to niezwykle korzystne z powodu tego, że zakres temperatury ich stosowania wynosi na ogół $-40 \div +60^{\circ}\text{C}$, czyli w zakresie ich właściwości paraelektrycznych. Do badań zastosowano warstwy kompozytowe otrzymane z proszku $\text{Ba}_{0,65}\text{Sr}_{0,35}\text{TiO}_3$ syntezowanego w reakcji w fazie stałej. W celu modyfikacji jego parametrów elektrycznych domieszkowano go tlenkami niklu, manganu i żelaza. Odpowiedni do zastosowania rozmiar ziarna uzyskiwano poprzez zmielenie spieczonego proszku. Kompozytowe warstwy ceramiczno-polimerowe projektowano w ten sposób, że objętość proszku ferroelektrycznego wynosiła od 75 do 97,5% obj., natomiast objętość polimeru od 2,5 do 25% obj. Grubość warstwy kompozytu zmieniano w zakresie od 100 do 500 μm .

2. ANALIZA PEŁNOFALOWA ANTENY FERROELEKTRYCZNEJ

2.1. MODEL PASKOWEJ ANTENY FERROELEKTRYCZNEJ

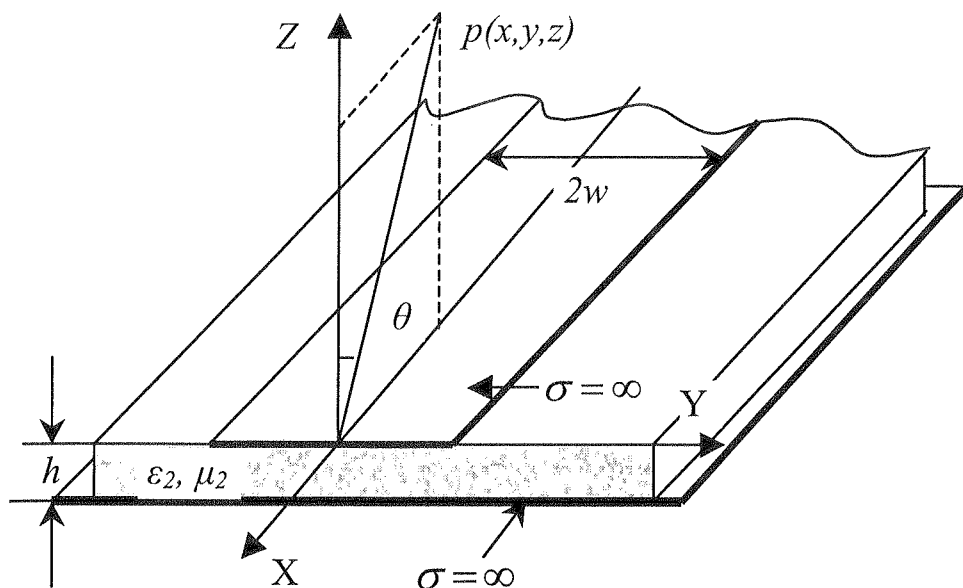
Najczęściej stosowanym promiennikiem paskowym jest rezonansowy element paskowy, ze wzbudzonym pierwszym wyższym rodzajem pola [13]. Podobny rozkład pola jest w linii paskowej, pracującej z wykorzystaniem pierwszego wyższego rodzaju pola. Na krawędziach linii paskowej komponenty pola elektrycznego, równoległe do podłoża, mają ten sam kierunek. W wyniku ich złożenia uzyskujemy pole promieniowania. W pierwszym przybliżeniu ten rodzaj może być traktowany jako rodzaj hybrydowy HE_{01} [14]. W porównaniu z rezonatorem rodzaj HE_{01} w linii istnieje w szerokim zakresie częstotliwości. Dla częstotliwości, która leży około częstotliwości krytycznej, co znaczy, że $\beta/k < 1$ (β — stała fazowa, $k = 2\pi/\lambda$ — liczba falowa), pole pierwszego wyższego rodzaju słabo skupia się w linii paskowej, powodując znaczne straty na promieniowanie.

Model paskowej anteny ferroelektrycznej opiera się na modelu niesymetrycznej linii paskowej, która może pracować na wyższych rodzajach fal [14], z tą różnicą, że podłoże jest wykonane z ferroelektryka. Idea sterowanej anteny ferroelektrycznej, w tym przypadku, polega na możliwości zmiany skutecznej przenikalności elektrycznej podłoża anteny paskowej, poprzez zmianę zewnętrznego sterującego pola elektrycznego E_0 . Zmiana E_0 wywołuje zmianę stałej fazowej β i związanego z nią kierunku promieniowania $\theta_m = \sin^{-1}(\beta)$.

Do analizy struktury pokazanej na Rys. 1 wykorzystano metodę równań całkowych oraz metodę momentów do ich rozwiązania.

W przypadku ogólnym jednorodne równania całkowe dla rozkładów prądu $J_x(x, y)$ i $J_y(x, y)$ na powierzchni linii paskowej mogą być przedstawione w postaci:

$$\iint [G_{xx}(x, y; x', y') \cdot J_x(x', y') + G_{xy}(x, y; x', y') \cdot J_y(x', y')] \cdot dx' dy' = 0, \quad (1)$$



Rys. 1. Przekrój struktury paskowej

Fig. 1. Cross section of open microstrip

$$\iint [G_{yx}(x, y; x', y') \cdot J_x(x', y') + G_{yy}(x, y; x', y') \cdot J_y(x', y')] \cdot dx' dy' = 0, \quad (2)$$

gdzie G_{xx} , $G_{xy} = G_{yx}$, G_{yy} — składowe funkcji Greena (dla przykładu funkcja G_{xy} może być traktowana jako składowa styczna pola elektrycznego w punkcie (x, y) , wywołana prądem $J_y(x, y)$ istniejącym w punkcie (x', y')).

Równania (1) i (2) mogą być rozwiązane dla zespolonej częstotliwości rezonansowej względem nieznanymi rozkładów prądu $J_x(x, y)$ i $J_y(x, y)$. W strukturze promieniującej częstotliwości rezonansowe stają się zespolone, gdzie część urojona określa straty na promieniowanie.

Poniżej jest przedstawiona ogólna metoda rozwiązywania układu równań całkowych przy zastosowaniu metody momentów.

Założmy, że wszystkie zmiany prądu (podłużnego jak i poprzecznego) wzdłuż osi x mogą być przedstawione w postaci $e^{-jk_x x}$, gdzie $k_x(\epsilon_2) = \beta(\epsilon_2) - j \cdot \alpha(\epsilon_2)$ — stała propagacji od której zależy kierunek promieniowania rozpatrywanej anteny.

Gęstości prądów na linii paskowej mogą być przedstawione w sposób przybliżony jako liniowa kombinacja liniowo niezależnych funkcji bazowych

(1)

$$J_x = \sum_{i=1}^n I_{xi} \psi_{xi} e^{-jk_x x}, \quad (3)$$

$$J_y = \sum_{i=1}^n I_{yi} \psi_{yi} e^{-jk_x x}, \quad (4)$$

gdzie I_{xi} i I_{yi} są nieznanymi (poszukiwanymi) stałymi zespolonymi, a ψ_{xi} i ψ_{yi} funkcjami bazowymi.

Składowe styczne natężenia pola elektrycznego można zapisać w postaci

$$\sum_{m=1}^M \left\{ I_{xm} \iint G_{xx}(x, y; x', y') \cdot \psi_{xi}(x', y') dx' dy' + \right. \\ \left. + I_{ym} \iint G_{xy}(x, y; x', y') \cdot \psi_{yi}(x', y') dx' dy' \right\} = E_{xi}, \quad (5)$$

$$\sum_{m=1}^M \left\{ I_{xm} \iint G_{yx}(x, y; x', y') \cdot \psi_{xi}(x', y') dx' dy' + \right. \\ \left. + I_{ym} \iint G_{yy}(x, y; x', y') \cdot \psi_{yi}(x', y') dx' dy' \right\} = E_{yi}. \quad (6)$$

Wymnażając prawą i lewą strony tych wyrażeń przez odpowiednie funkcje wagowe i całkując po powierzchni paska możemy przejść do układu równań liniowych. Część prawa będzie w tym przypadku reprezentować impedancje uogólnione.

Relacja pomiędzy składowymi stycznymi pola elektrycznego a prądami może być zapisana w postaci równania macierzowego

$$\begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{xx} & \mathbf{Z}_{xy} \\ \mathbf{Z}_{yx} & \mathbf{Z}_{yy} \end{bmatrix} \begin{bmatrix} \mathbf{I}_x \\ \mathbf{I}_y \end{bmatrix}. \quad (7)$$

Ponieważ składowe styczne pola elektrycznego na powierzchni przewodnika znikają, powyższe równanie macierzowe przyjmuje postać:

$$[\mathbf{Z}] [\mathbf{I}] = 0, \quad (8)$$

gdzie macierz $[\mathbf{Z}]$ jest macierzą rzędu $2m$, $[\mathbf{I}]$ — kolumna nieznanymi (poszukiwanymi) współczynników znanych rozkładów prądów.

Równanie charakterystyczne dla określenia stałej propagacji k_x uzyskuje się przez przyrównywanie wyznacznika macierzy $\mathbf{Z}$ do zera

$$\det [\mathbf{Z}(k_x)] = 0, \quad (9)$$

gdzie $[\mathbf{Z}]$ jest macierzą impedancji uogólnionych, której składniki można obliczyć wykorzystując wzór dla impedancji wzajemnej

$$Z_{mn} = -\frac{1}{I_m I_n^*} \int_x \int_y E_t \{I_m \psi_m\} \cdot I_n^* \psi_n^* \cdot dx dy, \quad (10)$$

gdzie $E_t \{I_m \psi_m\}$ jest składową styczną pola elektrycznego do powierzchni $x \times y$ wygenerowaną przez prąd $J_m = I_m \cdot \psi_m$

Elektrodynamiczna analiza ferroelektrycznej anteny paskowej na pierwszym etapie sprowadza się do obliczenia macierzy impedancji uogólnionych oraz rozwiązania równania charakterystycznego (9). Po określeniu stałej propagacji k_x można znaleźć wektor prądów $J_x(x, y)$ i $J_y(x, y)$ a następnie obliczyć pole promieniowania znalezionych rozkładów prądu.

Rozpatrzmy na początku ogólną metodę wyznaczania pola elektromagnetycznego wygenerowanego dowolnym rozkładem podłużnych i poprzecznych prądów.

2.2. FALE ELEKTRYCZNE I MAGNETYCZNE

W naszym przypadku korzystne wydaje się przedstawienie rozwiązania równań Maxwella w postaci superpozycji fal elektrycznych i magnetycznych [15]. Załóżmy, że kierunek propagacji fal elektrycznych i magnetycznych odpowiada kierunkowi osi z .

W ogólnym przypadku natężenia pola elektrycznego i pola magnetycznego są związane z elektrodynamicznymi potencjałami wektorowymi ($\mathbf{A}^e$ — dla prądów elektrycznych oraz $\mathbf{A}^m$ — dla prądów magnetycznych) układem równań:

$$\left. \begin{aligned} \mathbf{E} &= -j\omega\mu'_a \mathbf{A}^e + \frac{1}{j\omega\epsilon'_a} \text{grad} \cdot \text{div} \mathbf{A}^e - \text{rot} \mathbf{A}^m \\ \mathbf{H} &= -j\omega\epsilon'_a \mathbf{A}^m + \frac{1}{j\omega\mu'_a} \text{grad} \cdot \text{div} \mathbf{A}^m + \text{rot} \mathbf{A}^e \end{aligned} \right\}, \quad (11)$$

gdzie ϵ'_a — bezwzględna zespolona stała dielektryczna ośrodka, μ'_a — bezwzględna zespolona stała magnetyczna ośrodka.

Z kolei podłużna składowa natężenia pola elektrycznego E_z może być przedstawiona wzorem

$$E_z = \frac{k^2}{j\omega\epsilon'_a} A_z^e + \frac{1}{j\omega\epsilon'_a} \cdot \frac{\partial}{\partial z} \left[\frac{\partial A_x^e}{\partial x} + \frac{\partial A_y^e}{\partial y} + \frac{\partial A_z^e}{\partial z} \right] - \left[\frac{\partial A_y^m}{\partial x} - \frac{\partial A_x^m}{\partial y} \right]. \quad (12)$$

Składowe potencjału wektorowego wyznacza się dla określonego punktu obserwacji (x, y, z) , który może znajdować się wewnątrz objętości, gdzie znajdują się prądy pierwotne, jak i na zewnątrz tej objętości

$$\begin{aligned} \mathbf{A}(p) &= \int_V \mathbf{j}^z(q) \cdot G(p, q) dv; \\ A_i(x, y, z) &= \int_V j^z(x', y', z') \cdot G(x, y, z; x', y', z') dx' dy' dz', \end{aligned} \quad (13)$$

gdzie przez $p(x, y, z)$ jest oznaczony punkt obserwacji, a przez $q(x', y', z')$ — punkt źródła pola; $G(x, y, z; x', y', z')$ — funkcja Greena dla wolnej przestrzeni. Funkcja Greena

jest funkcją dwóch punktów — punktu źródła pola q oraz punktu obserwacji pola p . Jest ona symetryczna względem tych dwóch punktów. Funkcji Greena może być przedstawiona w różnych postaciach. We współrzędnych prostokątnych może być zapisana następująco:

$$\begin{aligned} G(x, y, z; x', y', z') &= \frac{1}{(2\pi)^2} \iiint_{-\infty}^{\infty} \frac{e^{-j\chi_1(x-x') - j\chi_2(y-y') - j\chi_3(z-z')}}{\chi_1^2 + \chi_2^2 + \chi_3^2 - k^2} d\chi_1 d\chi_2 d\chi_3 = \\ &= \frac{1}{8\pi^2} \iint_{-\infty}^{\infty} \frac{e^{-j\chi_1(x-x') - j\chi_2(y-y') \pm \gamma(z-z')}}{\gamma} d\chi_1 d\chi_2, \end{aligned} \quad (14)$$

gdzie $\gamma = \sqrt{\chi_1^2 + \chi_2^2 - k^2}$. Podwójny znak w wykładniku uwzględnia różną postać funkcji Greena dla obszarów przestrzeni: $(z - z') < 0$ (znak górny) oraz $(z - z') > 0$ (znak dolny).

Analiza wzoru (14) pokazuje, że dla ośrodka bezstratnego i dla przypadku, kiedy $\chi_1^2 + \chi_2^2 < k^2$ zaś γ jest wartością urojoną, otrzymujemy widmo fal bieżących, oddalających się od płaszczyzny $z = z'$ w obydwóch kierunkach osi z . W przypadku $\chi_1^2 + \chi_2^2 > k^2$ wartość γ jest wartością rzeczywistą i propagacji fali w kierunku osi z nie obserwuje się. Ma wówczas miejsce widmo drgań zanikających, czyli drgań, których amplituda zmniejsza się wykładniczo przy oddalaniu się w obie strony od płaszczyzny $z = z'$.

Podstawiając wyrażenie (14) do (13) potencjał wektorowy można zapisać w postaci:

$$\mathbf{A} = \iint_{-\infty}^{\infty} \frac{e^{-j\chi_1 x - j\chi_2 y \pm \gamma z}}{\gamma} d\chi_1 d\chi_2 \cdot \frac{1}{8\pi^2} \int_V \mathbf{j}^p e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dV, \quad (15)$$

gdzie $\mathbf{j}^p$ — rozkład gęstości pierwotnych prądów elektrycznych lub magnetycznych. Podstawiając wyrażenie (15) do wyrażenia (12) opisującego pole E_z otrzymujemy:

$$E_z = \iint_{-\infty}^{\infty} E_{zx} d\chi_1 d\chi_2 = \iint_{-\infty}^{\infty} \frac{F^e(\chi)}{\gamma} e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} d\chi_1 d\chi_2, \quad (16)$$

gdzie

$$\begin{aligned} F^e(\chi) &= \frac{1}{8\pi^2} \int_{V'} \left\{ \frac{\chi_1 + \chi_2^2}{j\omega \epsilon'_a} J_z^e \right\} - j\chi_1 \cdot \left\{ \frac{(\pm\gamma)}{j\omega \epsilon'_a} \cdot J_x^e - J_y^m \right\} - \\ &- j \cdot \chi_2 \left\{ \frac{(\pm\gamma)}{j\omega \epsilon'_a} \cdot J_y^e + J_x^m \right\} \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dV'. \end{aligned} \quad (17)$$

Analogicznie można przedstawić wyrażenie dla podłużnej składowej natężenia pola magnetycznego H_z :

$$H_z = \iint_{-\infty}^{\infty} H_{zx} d\chi_1 d\chi_2 = \iint_{-\infty}^{\infty} \frac{F^m(\chi)}{\gamma} e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} d\chi_1 d\chi_2, \quad (18)$$

gdzie

$$F^m(\chi) = \frac{1}{8\pi^2} \int_{V'} \left[\left\{ \frac{\chi_1^2 + \chi_2^2}{j\omega\mu'_a} J_z^m \right\} - j \cdot \chi_1 \left\{ \frac{\pm\gamma}{j\omega\mu'_a} J_x^m + J_y^e \right\} - \right. \\ \left. - j \cdot \chi_2 \left\{ \frac{\pm\gamma}{j\omega\mu'_a} J_y^m - J_x^e \right\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dv. \quad (19)$$

Składowe poprzeczne pola elektromagnetycznego E_x, E_y, H_x oraz H_y mogą być przedstawione (dla punktów przestrzeni w których nie ma prądów pierwotnych) przez składowe podłużne w postaci:

$$E_i = \iint_{-\infty}^{+\infty} E_{ix} d\chi_1 d\chi_2, \quad H_i = \iint_{-\infty}^{+\infty} H_{ix} d\chi_1 d\chi_2, \quad (i = x, y) \quad (20)$$

Podstawiając wyrażenia (16), (18) oraz (20) w jednorodne równania Maxwella otrzymamy:

$$\begin{aligned} \text{rot} \mathbf{H}_x &= j\omega\epsilon'_a \mathbf{E}_x; \\ \text{rot} \mathbf{E}_x &= -j\omega\mu'_a \mathbf{H}_x. \end{aligned} \quad (21)$$

Zapisując równania (21) we współrzędnych prostokątnych otrzymamy: dla $H_z = 0$

$$\begin{aligned} E_{xx}^e &= \frac{\pm\gamma}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial E_{zx}}{\partial x}; & H_{xx}^e &= \frac{j\omega\epsilon'_a}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial E_{zx}}{\partial y}; \\ E_{yx}^e &= \frac{\pm\gamma}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial E_{zx}}{\partial y}; & H_{yx}^e &= -\frac{j\omega\epsilon'_a}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial E_{zx}}{\partial x}. \end{aligned} \quad (22)$$

dla $E_z = 0$

$$\begin{aligned} E_{xx}^m &= -\frac{i\omega\mu_a}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial H_{zx}}{\partial y}; & H_{xx}^m &= \frac{\pm\gamma}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial H_{zx}}{\partial x}; \\ E_{yx}^m &= +\frac{i\omega\mu_a}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial H_{zx}}{\partial x}; & H_{yx}^m &= \frac{\pm\gamma}{\chi_1^2 + \chi_2^2} \cdot \frac{\partial H_{zx}}{\partial y}. \end{aligned} \quad (23)$$

Pełne pole elektromagnetyczne jest sumą wektorową (22) i (23).

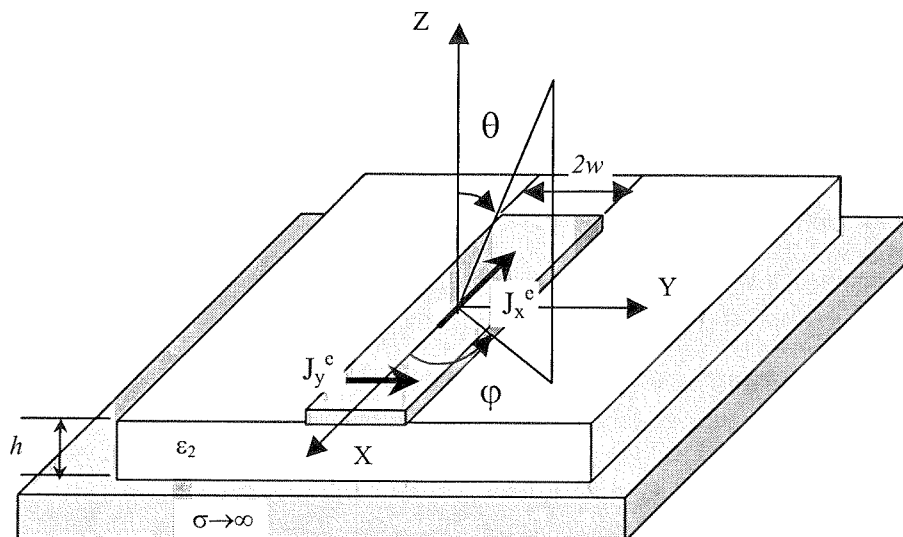
Zakładając znane rozkłady prądów pierwotnych możemy obliczyć przy pomocy wzorów (16) i (18) podłużne składowe natężenia pola elektrycznego oraz magnetycznego, a wtedy z kolei przy pomocy wzorów (22) i (23) składowe poprzeczne.

2.3. IMPEDANCJA UOGÓLNIIONA

Do określenia impedancji uogólnionej konieczne jest oszacowanie składowych stycznych na powierzchni anteny paskowej wygenerowanych przez założony rozkład prądu. W tym celu założymy, że części metalowe anteny są bezstratne a ich grubości są

równe zero. Przyjmijmy, że otoczeniem rozpatrywanej struktury jest próżnia, a podłoże anteny paskowej zostało wykonane z dielektryka niemagnetycznego ($\mu_{a1} = \mu_{a2} = \mu_0$). Zakładamy również, że elektryczne prądy pierwotne skupione są na powierzchni górnego przewodnika ($z \geq 0$).

Pole będziemy przedstawiać w postaci sumy fal elektrycznych i magnetycznych bieżących wzdłuż osi z . Założymy także, że są zadane podłużne J_x^e i poprzeczne J_y^e rozkłady prądu elektrycznego. Na Rys. 2 pokazana jest rozpatrywana antena.



Rys. 2. Rozpatrywana struktura i prądy pierwotne

Fig. 2. Basic configuration and the source currents

Na pole nad warstwą dielektryka składa się pole padające oraz pole odbite od powierzchni dielektryka. Pole padające można przedstawić w postaci wzorów (16) ÷ (23). Pole odbite przedstawia się analogicznie, z tą różnicą, że wprowadza się nie znany (na razie) współczynnik odbicia od warstwy dielektryka z ekranem. Przykładowo wzór (16) dla pola odbitego należy zapisać w postaci:

$$E_z^R = \iint_{-\infty}^{+\infty} E_{zx}^R d\chi_1 d\chi_2 = \iint_{-\infty}^{+\infty} \frac{F^e(\chi_1, \chi_2) \cdot R^{e/m}(\chi_1, \chi_2)}{\gamma} e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} d\chi_1 d\chi_2, \quad (24)$$

gdzie $R^{e/m}(\chi_1, \chi_2)$ — współczynnik odbicia dla fali elektrycznej/magnetycznej.

Na pole wewnątrz warstwy dielektryka składa się pole fali przechodzącej z próżni do dielektryka oraz pole fali odbitej od ekranu. Sumaryczne pole wewnątrz dielektryka ma spełniać warunki brzegowe na powierzchni ekranu $E_{t/z=-h} = 0$. W płaszczyźnie

dołożę
 $= \mu_0$).
 ni gór-

cznych
 zne J_y^e

$z = 0$ pole ma spełniać warunki ciągłości składowych stycznych **E** i **H**. Oznacza to, że powinniśmy przyrównać wartości E_t i H_t na zewnątrz warstwy dielektryka jak i wewnątrz tej warstwy przy $z = 0$. Stosując odwrotną transformatę Fouriera otrzymamy dwa równania dla nieznanymi gęstości widmowych współczynników odbicia i załamania.

Interesuje nas pole nad anteną (czyli pole promieniowania), dlatego przedstawiamy tylko współczynniki odbicia w wygodnej dla dalszego rozpatrywania postaci:

$$1 + R^e = \frac{2 \cdot \gamma_2}{\varepsilon_2 \cdot \gamma \cdot \text{cth}(h \cdot \gamma_2) + \gamma_2} \quad (25)$$

$$1 + R^m = \frac{2 \cdot \gamma}{\gamma_2 \cdot \text{cth}(h \cdot \gamma_2) + \gamma} \quad (26)$$

gdzie $\gamma_2 = \sqrt{k_x^2 + \chi_2^2 - k_0^2} \cdot \varepsilon_2$, ε_2 — względna przenikalność elektryczna warstwy dielektryka.

Korzystając ze wzorów (16)÷(23) możemy zapisać składowe styczne pola elektrycznego w postaci

$$\begin{aligned} E_{xx} &= E_{xx}^e + E_{xx}^m = \\ &= \frac{(\pm\gamma) \cdot (-j\chi_1)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^e(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^e] + \\ &+ \frac{(j\omega\mu) \cdot (-j\chi_2)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^m(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^m] = \\ &= \left\{ \frac{(\mp j\chi_1)}{\chi_1^2 + \chi_2^2} \cdot F^e(\chi) [1 + R^e] - \right. \\ &\quad \left. - \frac{(\omega\mu) \cdot (\chi_2)}{(\chi_1^2 + \chi_2^2) \cdot \gamma} \cdot F^m(\chi) [1 + R^m] \right\} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z}, \end{aligned} \quad (27)$$

ite od
 v (16)
 ię nie
 adowo

(24)

próżni
 elektry-
 zycznie

$$\begin{aligned} E_{yx} &= E_{yx}^e + E_{yx}^m = \\ &= \frac{(\pm\gamma) \cdot (-j\chi_2)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^e(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^e] + \\ &+ \frac{(j\omega\mu) \cdot (-j\chi_1)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^m(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^m] = \\ &= \left\{ \frac{(\mp j\chi_2)}{\chi_1^2 + \chi_2^2} \cdot F^e(\chi) [1 + R^e] + \right. \\ &\quad \left. + \frac{(\omega\mu) \cdot (\chi_1)}{(\chi_1^2 + \chi_2^2) \cdot \gamma} \cdot F^m(\chi) [1 + R^m] \right\} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z}, \end{aligned} \quad (28)$$

gdzie

$$F^e(\chi) = \frac{1}{8\pi^2} \iint_{x' y'} \left[\left\{ \frac{-j \cdot \chi_1 \cdot (\pm\gamma)}{j\omega\epsilon_a} \cdot J_x^e \right\} - \left\{ j \cdot \chi_2 \cdot \frac{(\pm\gamma)}{j\omega\epsilon_a} \cdot J_y^e \right\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy', \quad (29)$$

$$F^m(\chi) = \frac{1}{8\pi^2} \iint_{x' y'} [-j \cdot \chi_1 \cdot J_y^e + j \cdot \chi_2 \cdot J_x^e] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy'. \quad (30)$$

Poszukujemy rozkładu gęstości prądu na powierzchni apertury anteny w postaci: dla fal parzystych

$$J_x(x, y, z) = \frac{\delta(z - H)}{\sqrt{1 - (y/w)^2}} \sum_{m=1}^M I_{xm} (-1)^m T_{2m}(y/w) \cdot e^{-jk_x x}, \quad (31)$$

$$J_y(x, y, z) = \delta(z - H) \cdot \sum_{m=1}^M \frac{I_{ym}}{2m} \sin[2m \cdot \arcsin(y/w)] \cdot e^{-jk_x x}, \quad (32)$$

dla fal nieparzystych

$$J_x(x, y, z) = \frac{\delta(z - H)}{\sqrt{1 - (y/w)^2}} \sum_{m=1}^M I_{xm} (-1)^{m-1} T_{2m-1}(y/w) \cdot e^{-jk_x x}, \quad (33)$$

$$J_y(x, y, z) = \delta(z - H) \cdot \sum_{m=1}^M \frac{I_{ym}}{(2m-1)} \cos[(2m-1) \cdot \arcsin(y/w)] \cdot e^{-jk_x x}, \quad (34)$$

gdzie $\delta(z - H)$ — delta Diraca, $T_{2m/2m-1}$ — funkcja Chebyszewa, H — poziom płaszczyzny prądów.

Rozpatrzmy przypadek fal nieparzystych co odpowiada pierwszemu wyższemu rodzajowi pola w linii paskowej. Podstawiając zależności, określające rozkłady gęstości prądów, (33) oraz (34) w wyrażenia (29) i (30) otrzymamy

$$\begin{aligned} F_1^e(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[\left\{ \frac{-j \cdot \chi_1 \cdot (\pm\gamma)}{j\omega\epsilon_a} \cdot J_x^e \right\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy' dz' = \\ &= \frac{\mp j \cdot \chi_1 \cdot \gamma}{8\pi^2 \cdot j\omega\epsilon_a} eF\gamma H \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{xm} (-1)^{m-1} \int_{-w}^w e^{j\chi_2 y'} \frac{T_{2m-1}(y'/w)}{\sqrt{1 - (y'/w)^2}} dy' = \\ &= \frac{\mp j \cdot \delta(\chi_1 - k_x) \cdot \chi_1 \cdot \gamma \cdot w}{4 \cdot \omega\epsilon_a} eF\gamma H \sum_{m=1}^M I_{xm} \cdot J_{2m-1}(\chi_2 w), \end{aligned} \quad (35)$$

dy',

(29)

(30)

ostaci:

(31)

(32)

(33)

(34)

płasz-

szemu

ęstości

=

(35)

$$\begin{aligned}
 F_2^e(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[\left\{ \frac{-j \cdot \chi_2 \cdot (\pm \gamma)}{j\omega\epsilon_a} \cdot J_y^e \right\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy' dz' = \\
 &= \frac{\mp j \cdot \chi_2 \cdot \gamma}{8\pi^2 \cdot j\omega\epsilon_a} eF\gamma H \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{ym} \int_{-w}^w e^{j\chi_2 y'} \cdot \frac{\cos[(2m-1) \arcsin(y'/w)]}{(2m-1)} dy' = \\
 &= \frac{\mp \delta(\chi_1 - k_x) \cdot \chi_2 \cdot \gamma \cdot w}{8 \cdot \omega\epsilon_a \cdot (2m-1)} eF\gamma H \sum_{m=1}^M I_{ym} \cdot [J_{2m}(\chi_2 w) + J_{2m-2}(\chi_2 w)], \quad (36)
 \end{aligned}$$

gdzie **J** jest funkcja Bessela.

Wybór znaku uwzględnia różną postać funkcji Greena dla obszarów przestrzeni nad (znak dolny) i pod (znak górny) płaszczyzną z prądami pierwotnymi. Ponieważ rozpatrujemy pole nad anteną (pole promieniowania) funkcja F^e może być przedstawiona w postaci (dla $H=0$)

(32)

$$\begin{aligned}
 F^e(\chi) &= F_1^e + F_2^e = \frac{\delta(\chi_1 - k_x) \cdot \gamma \cdot w}{4 \cdot \omega\epsilon_a} \sum_{m=1}^M \left[j \cdot \chi_1 \cdot I_{xm} \cdot J_{2m-1}(\chi_2 w) + \right. \\
 &\quad \left. + \frac{\chi_2}{2 \cdot (2m-1)} I_{ym} \{ J_{2m}(\chi_2 w) + J_{2m-2}(\chi_2 w) \} \right]. \quad (37)
 \end{aligned}$$

(33)

W analogiczny sposób można przedstawić funkcję F^m .

(34)

$$\begin{aligned}
 F_1^m(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[\{-j \cdot \chi_1 \cdot J_y^e\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' + \gamma z'} dx' dy' dz' = \\
 &= \frac{-j \cdot \chi_1}{8\pi^2} \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{ym} \int_{-w}^w e^{j\chi_2 y'} \cdot \frac{\cos[(2m-1) \arcsin(y'/w)]}{(2m-1)} dy' = \\
 &= -j \frac{\delta(\chi_1 - k_x) \cdot \chi_1 \cdot w}{8} \sum_{m=1}^M \frac{I_{ym}}{(2m-1)} \cdot [J_{2m}(\chi_2 w) + J_{2m-2}(\chi_2 w)], \quad (38)
 \end{aligned}$$

=

(35)

$$\begin{aligned}
 F_2^m(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[\{j \cdot \chi_2 \cdot J_x^e\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' + \gamma z'} dx' dy' dz' = \\
 &= \frac{j \cdot \chi_2}{8\pi^2} \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{xm} (-1)^{m-1} \int_{-w}^w e^{j\chi_2 y'} \frac{T_{2m-1}(y'/w)}{\sqrt{1 - (y'/w)^2}} dy' = \\
 &= \frac{-\delta(\chi_1 - k_x) \cdot \chi_2 \cdot w}{4} \sum_{m=1}^M I_{xm} \cdot J_{2m-1}(\chi_2 w), \quad (39)
 \end{aligned}$$

$$F^m(\chi) = F_1^m + F_2^m = -\frac{\delta(\chi_1 - k_x) \cdot w}{4} \sum_{m=1}^M \left[\chi_2 \cdot I_{xm} \cdot \mathbf{J}_{2m-1}(\chi_2 w) + \right. \\ \left. + j \frac{\chi_1}{2 \cdot (2m-1)} I_{ym} \{ \mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w) \} \right]. \quad (40)$$

Podstawiając wzory (37) i (40) do wyrażeń (27) oraz (28) otrzymamy gęstości widmowe składowych stycznych pola elektrycznego w postaci

$$E_{xx\chi} = \frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{4 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{-\chi_1^2 \cdot \gamma}{k} \cdot [1 + R^e] + \right. \\ \left. + \frac{k \cdot \chi_2^2}{1} \cdot [1 + R^m] \right\} \cdot \sum_{m=1}^M I_{xm} \cdot \mathbf{J}_{2m-1}(\chi_2 w) \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}, \quad (41)$$

$$E_{xy\chi} = j \frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{8 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{\chi_1 \cdot \chi_2 \cdot \gamma^2}{k} \cdot [1 + R^e] + \right. \\ \left. + \frac{k \cdot \chi_1 \cdot \chi_2}{1} \cdot [1 + R^m] \right\} \cdot \sum_{m=1}^M I_{ym} \cdot \left[\frac{\mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w)}{(2m-1)} \right] \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}, \quad (42)$$

$$E_{yx\chi} = -\frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{4 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{\chi_1 \cdot \chi_2 \cdot \gamma^2}{k} \cdot [1 + R^e] + \frac{k \cdot \chi_1 \cdot \chi_2}{1} \cdot [1 + R^m] \right\} \cdot \\ \cdot \sum_{m=1}^M I_{xm} \cdot [\mathbf{J}_{2m-1}(\chi_2 w)] \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}, \quad (43)$$

$$E_{yy\chi} = j \frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{8 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{\chi_2^2 \cdot \gamma^2}{k} \cdot [1 + R^e] - \right. \\ \left. - \frac{k \cdot \chi_1^2}{1} \cdot [1 + R^m] \right\} \cdot \\ \cdot \sum_{m=1}^M I_{ym} \cdot \left[\frac{\mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w)}{(2m-1)} \right] \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}. \quad (44)$$

gdzie $Z_0 = 120\pi$.

$z = 0$ pole ma spełniać warunki ciągłości składowych stycznych $\mathbf{E}$ i $\mathbf{H}$. Oznacza to, że powinniśmy przyrównać wartości E_t i H_t na zewnątrz warstwy dielektryka jak i wewnątrz tej warstwy przy $z = 0$. Stosując odwrotną transformatę Fouriera otrzymamy dwa równania dla nieznanymi gęstości widmowych współczynników odbicia i załamania.

Interesuje nas pole nad anteną (czyli pole promieniowania), dlatego przedstawiamy tylko współczynniki odbicia w wygodnej dla dalszego rozpatrywania postaci:

$$1 + R^e = \frac{2 \cdot \gamma_2}{\varepsilon_2 \cdot \gamma \cdot \text{cth}(h \cdot \gamma_2) + \gamma_2} \quad (25)$$

$$1 + R^m = \frac{2 \cdot \gamma}{\gamma_2 \cdot \text{cth}(h \cdot \gamma_2) + \gamma} \quad (26)$$

gdzie $\gamma_2 = \sqrt{k_x^2 + \chi_2^2 - k_0^2 \cdot \varepsilon_2}$, ε_2 — względna przenikalność elektryczna warstwy dielektryka.

Korzystając ze wzorów (16)÷(23) możemy zapisać składowe styczne pola elektrycznego w postaci

$$\begin{aligned} E_{xx} &= E_{xx}^e + E_{xx}^m = \\ &= \frac{(\pm\gamma) \cdot (-j\chi_1)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^e(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^e] + \\ &+ \frac{(-j\omega\mu) \cdot (-j\chi_2)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^m(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^m] = \end{aligned} \quad (27)$$

$$\begin{aligned} &= \left\{ \frac{(\mp j\chi_1)}{\chi_1^2 + \chi_2^2} \cdot F^e(\chi) [1 + R^e] - \right. \\ &\left. - \frac{(\omega\mu) \cdot (\chi_2)}{(\chi_1^2 + \chi_2^2) \cdot \gamma} \cdot F^m(\chi) [1 + R^m] \right\} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z}, \end{aligned} \quad (43)$$

$$\begin{aligned} E_{yx} &= E_{yx}^e + E_{yx}^m = \\ &= \frac{(\pm\gamma) \cdot (-j\chi_2)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^e(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^e] + \\ &+ \frac{(j\omega\mu) \cdot (-j\chi_1)}{\chi_1^2 + \chi_2^2} \cdot \frac{F^m(\chi)}{\gamma} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z} [1 + R^m] = \\ &= \left\{ \frac{(\mp j\chi_2)}{\chi_1^2 + \chi_2^2} \cdot F^e(\chi) [1 + R^e] + \right. \\ &\left. + \frac{(\omega\mu) \cdot (\chi_1)}{(\chi_1^2 + \chi_2^2) \cdot \gamma} \cdot F^m(\chi) [1 + R^m] \right\} \cdot e^{-j\chi_1 x - j\chi_2 y \pm \gamma z}, \end{aligned} \quad (28)$$

gdzie

$$F^e(\chi) = \frac{1}{8\pi^2} \iint_{x' y'} \left[\left\{ \frac{-j \cdot \chi_1 \cdot (\pm\gamma)}{j\omega\epsilon_a} \cdot J_x^e \right\} - \left\{ j \cdot \chi_2 \cdot \frac{(\pm\gamma)}{j\omega\epsilon_a} \cdot J_y^e \right\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy', \quad (29)$$

$$F^m(\chi) = \frac{1}{8\pi^2} \iint_{x' y'} [-j \cdot \chi_1 \cdot J_y^e + j \cdot \chi_2 \cdot J_x^e] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy'. \quad (30)$$

Poszukujemy rozkładu gęstości prądu na powierzchni apertury anteny w postaci: dla fal parzystych

$$J_x(x, y, z) = \frac{\delta(z - H)}{\sqrt{1 - (y/w)^2}} \sum_{m=1}^M I_{xm} (-1)^m T_{2m}(y/w) \cdot e^{-jk_x x}, \quad (31)$$

$$J_y(x, y, z) = \delta(z - H) \cdot \sum_{m=1}^M \frac{I_{ym}}{2m} \sin[2m \cdot \arcsin(y/w)] \cdot e^{-jk_x x}; \quad (32)$$

dla fal nieparzystych

$$J_x(x, y, z) = \frac{\delta(z - H)}{\sqrt{1 - (y/w)^2}} \sum_{m=1}^M I_{xm} (-1)^{m-1} T_{2m-1}(y/w) \cdot e^{-jk_x x}, \quad (33)$$

$$J_y(x, y, z) = \delta(z - H) \cdot \sum_{m=1}^M \frac{I_{ym}}{(2m-1)} \cos[(2m-1) \cdot \arcsin(y/w)] \cdot e^{-jk_x x}, \quad (34)$$

gdzie $\delta(z - H)$ — delta Diraca, $T_{2m/2m-1}$ — funkcja Chebyszewa, H — poziom płaszczyzny prądów.

Rozpatrzmy przypadek fal nieparzystych co odpowiada pierwszemu wyższemu rodzajowi pola w linii paskowej. Podstawiając zależności, określające rozkłady gęstości prądów, (33) oraz (34) w wyrażenia (29) i (30) otrzymamy

$$\begin{aligned} F_1^e(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[\left\{ \frac{-j \cdot \chi_1 \cdot (\pm\gamma)}{j\omega\epsilon_a} \cdot J_x^e \right\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy' dz' = \\ &= \frac{\mp j \cdot \chi_1 \cdot \gamma}{8\pi^2 \cdot j\omega\epsilon_a} eF\gamma H \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{xm} (-1)^{m-1} \int_{-w}^w e^{j\chi_2 y'} \frac{T_{2m-1}(y'/w)}{\sqrt{1 - (y'/w)^2}} dy' = \\ &= \frac{\mp j \cdot \delta(\chi_1 - k_x) \cdot \chi_1 \cdot \gamma \cdot w}{4 \cdot \omega\epsilon_a} eF\gamma H \sum_{m=1}^M I_{xm} \cdot J_{2m-1}(\chi_2 w), \end{aligned} \quad (35)$$

dy',

(29)

(30)

ostaci:

(31)

(32)

(33)

(34)

płasz-

zemu

stości

=

(35)

$$\begin{aligned}
 F_2^e(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[\left\{ \frac{-j \cdot \chi_2 \cdot (\pm \gamma)}{j\omega\epsilon_a} \cdot J_y^e \right\} \right] \cdot e^{j\chi_1 x' + j\chi_2 y' \mp \gamma z'} dx' dy' dz' = \\
 &= \frac{\mp j \cdot \chi_2 \cdot \gamma}{8\pi^2 \cdot j\omega\epsilon_a} eF\gamma H \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{ym} \int_{-w}^w e^{j\chi_2 y'} \frac{\cos[(2m-1) \arcsin(y'/w)]}{(2m-1)} dy' = \\
 &= \frac{\mp \delta(\chi_1 - k_x) \cdot \chi_2 \cdot \gamma \cdot w}{8 \cdot \omega\epsilon_a \cdot (2m-1)} eF\gamma H \sum_{m=1}^M I_{ym} \cdot [\mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w)], \quad (36)
 \end{aligned}$$

gdzie $\mathbf{J}$ jest funkcja Bessela.

Wybór znaku uwzględnia różną postać funkcji Greena dla obszarów przestrzeni nad (znak dolny) i pod (znak górny) płaszczyzną z prądami pierwotnymi. Ponieważ rozpatrujemy pole nad anteną (pole promieniowania) funkcja F^e może być przedstawiona w postaci (dla $H=0$)

$$\begin{aligned}
 F^e(\chi) = F_1^e + F_2^e &= \frac{\delta(\chi_1 - k_x) \cdot \gamma \cdot w}{4 \cdot \omega\epsilon_a} \sum_{m=1}^M \left[j \cdot \chi_1 \cdot I_{xm} \cdot \mathbf{J}_{2m-1}(\chi_2 w) + \right. \\
 &\quad \left. + \frac{\chi_2}{2 \cdot (2m-1)} I_{ym} \{ \mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w) \} \right]. \quad (37)
 \end{aligned}$$

W analogiczny sposób można przedstawić funkcję F^m .

$$\begin{aligned}
 F_1^m(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[[-j \cdot \chi_1 \cdot J_y^e] \right] \cdot e^{j\chi_1 x' + j\chi_2 y' + \gamma z'} dx' dy' dz' = \\
 &= \frac{-j \cdot \chi_1}{8\pi^2} \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{ym} \int_{-w}^w e^{j\chi_2 y'} \frac{\cos[(2m-1) \arcsin(y'/w)]}{(2m-1)} dy' = \\
 &= -j \frac{\delta(\chi_1 - k_x) \cdot \chi_1 \cdot w}{8} \sum_{m=1}^M \frac{I_{ym}}{(2m-1)} \cdot [\mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w)], \quad (38)
 \end{aligned}$$

$$\begin{aligned}
 F_2^m(\chi) &= \frac{1}{8\pi^2} \iiint_{x' y' z'} \left[[j \cdot \chi_2 \cdot J_x^e] \right] \cdot e^{j\chi_1 x' + j\chi_2 y' + \gamma z'} dx' dy' dz' = \\
 &= \frac{j \cdot \chi_2}{8\pi^2} \int_{-\infty}^{\infty} e^{j(\chi_1 - k_x)x'} dx' \sum_{m=1}^M I_{xm} (-1)^{m-1} \int_{-w}^w e^{j\chi_2 y'} \frac{T_{2m-1}(y'/w)}{\sqrt{1 - (y'/w)^2}} dy' = \\
 &= \frac{-\delta(\chi_1 - k_x) \cdot \chi_2 \cdot w}{4} \sum_{m=1}^M I_{xm} \cdot \mathbf{J}_{2m-1}(\chi_2 w), \quad (39)
 \end{aligned}$$

$$F^m(\chi) = F_1^m + F_2^m = -\frac{\delta(\chi_1 - k_x) \cdot w}{4} \sum_{m=1}^M \left[\chi_2 \cdot I_{xm} \cdot \mathbf{J}_{2m-1}(\chi_2 w) + \right. \\ \left. + j \frac{\chi_1}{2 \cdot (2m-1)} I_{ym} \{ \mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w) \} \right]. \quad (40)$$

Podstawiając wzory (37) i (40) do wyrażeń (27) oraz (28) otrzymamy gęstości widmowe składowych stycznych pola elektrycznego w postaci

$$E_{xx\chi} = \frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{4 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{-\chi_1^2 \cdot \gamma}{k} \cdot [1 + R^e] + \right. \\ \left. + \frac{k \cdot \chi_2^2}{1} \cdot [1 + R^m] \right\} \cdot \sum_{m=1}^M I_{xm} \cdot \mathbf{J}_{2m-1}(\chi_2 w) \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}, \quad (41)$$

$$E_{xy\chi} = j \frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{8 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{\chi_1 \cdot \chi_2 \cdot \gamma^2}{k} \cdot [1 + R^e] + \right. \\ \left. + \frac{k \cdot \chi_1 \cdot \chi_2}{1} \cdot [1 + R^m] \right\} \cdot \sum_{m=1}^M I_{ym} \cdot \left[\frac{\mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w)}{(2m-1)} \right] \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}, \quad (42)$$

$$E_{yx\chi} = -\frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{4 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{\chi_1 \cdot \chi_2 \cdot \gamma^2}{k} \cdot [1 + R^e] + \frac{k \cdot \chi_1 \cdot \chi_2}{1} \cdot [1 + R^m] \right\} \cdot \\ \cdot \sum_{m=1}^M I_{xm} \cdot [\mathbf{J}_{2m-1}(\chi_2 w)] \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}, \quad (43)$$

$$E_{yy\chi} = j \frac{Z_0 \cdot \delta(k_x - \chi_1) \cdot w}{8 \cdot (\chi_1^2 + \chi_2^2) \cdot \gamma} \left\{ \frac{\chi_2^2 \cdot \gamma^2}{k} \cdot [1 + R^e] - \right. \\ \left. - \frac{k \cdot \chi_1^2}{1} \cdot [1 + R^m] \right\} \cdot \\ \cdot \sum_{m=1}^M I_{ym} \cdot \left[\frac{\mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w)}{(2m-1)} \right] \cdot e^{-j\chi_1 x - j\chi_2 y - \gamma z}. \quad (44)$$

gdzie $Z_0 = 120\pi$.

Przedstawione wyrażenia dla określenia gęstości widmowej składowych stycznych pola elektrycznego (41)÷(44) pozwalają wyznaczyć składowe styczne pola elektrycznego a następnie impedancję uogólnioną. Zgodnie z wzorem (10) otrzymamy:

(40)

stości

$$Z_{xx}^{mn} = -j \frac{Z_0 \cdot w^2 \pi^2}{2} \sum_{n=1}^N \int_{-\infty}^{\infty} \left\{ \frac{k \cdot \chi_2^2}{1} \cdot [1 + R^m] - \frac{k_x^2 \cdot \gamma^2}{k} \cdot [1 + R^e] \right\} \cdot \frac{1}{(k_x^2 + \chi_2^2) \cdot \gamma} \cdot$$

(41)

$$\cdot \mathbf{J}_{2m-1}(\chi_2 w) \cdot \mathbf{J}_{2n-1}(\chi_2 w) d\chi_2,$$

$$m = 1, \dots, M$$

(42)

$$Z_{xy}^{mn} = \frac{Z_0 \cdot w^2 \pi^2}{4} \sum_{n=1}^N \int_{-\infty}^{\infty} \left\{ \frac{k \cdot \chi_2 \cdot k_x}{1} \cdot [1 + R^m] + \frac{k_x \cdot \chi_2 \cdot \gamma^2}{k} \cdot [1 + R^e] \right\} \cdot \frac{1}{(k_x^2 + \chi_2^2) \cdot \gamma \cdot (2n-1)} \cdot$$

(43)

$$\cdot \mathbf{J}_{2m-1}(\chi_2 w) \cdot [\mathbf{J}_{2n}(\chi_2 w) + \mathbf{J}_{2n-2}(\chi_2 w)] d\chi_2,$$

$$m = 1, \dots, M$$

(44)

$$Z_{yy}^{mn} = j \frac{Z_0 \cdot w^2 \pi^2}{8} \sum_{n=1}^N \int_{-\infty}^{\infty} \left\{ \frac{\chi_2^2 \cdot \gamma^2}{k} \cdot [1 + R^e] - \frac{k \cdot k_x^2}{1} \cdot [1 + R^m] \right\} \cdot \frac{1}{(k_x^2 + \chi_2^2) \cdot \gamma \cdot (2n-1)^2} \cdot$$

$$\cdot [\mathbf{J}_{2n}(\chi_2 w) + \mathbf{J}_{2n-2}(\chi_2 w)] \cdot [\mathbf{J}_{2m}(\chi_2 w) + \mathbf{J}_{2m-2}(\chi_2 w)] d\chi_2,$$

$$m = 1, \dots, M$$

$$Z_{xy}^{mn} = -Z_{yx}^{mn}. \quad (48)$$

3. ANALIZA NUMERYCZNA

Zbadamy obecnie zachowanie się funkcji podcałkowej w wyrażeniach (45)÷(48). Badania występujących funkcji podcałkowych wykazują istnienie punktów osobliwych typu:

1. rozgałęzienia, dla $\gamma = \sqrt{k_x^2 + \chi_2^2 - k^2} = 0|_{\chi_2=\chi_0}$;
2. bieguna pozornego, dla $k_x^2 + \chi_2^2 = 0$ (jest to osobliwość usuwalna);
3. biegunów izolowanych, dla fal powierzchniowy typu E:

$$\varepsilon_2 \cdot \gamma \cdot \text{cth}(h \cdot \gamma_2) + \gamma_2 = 0|_{\chi_2=\chi_s^E}, \quad (49a)$$

oraz/lub dla fal powierzchniowych typu H:

$$\gamma_2 \cdot \text{cth}(h \cdot \gamma_2) + \gamma = 0|_{\chi_2=\chi_s^H}, \quad (49b)$$

gdzie

$$\gamma = \sqrt{k_x^2 + \chi_2^2 - k^2}; \gamma_2 = \sqrt{k_x^2 + \chi_2^2 - \varepsilon_2 \cdot k^2}.$$

Wyznaczenie miejsca występowania biegunów i gałęzi pierwiastków na płaszczyźnie zespolonej zmiennej χ_2 jest podstawą do numerycznego oszacowania składowych pola elektromagnetycznego oraz impedancji uogólnionej. Załóżmy w związku z tym, że dielektryk podłoża ε_2 jak i dielektryk na zewnątrz anteny ε_0 są stratne:

$$\varepsilon_0 = \varepsilon'_0 - j \cdot \varepsilon''_0, \varepsilon_2 = \varepsilon'_2 - j \cdot \varepsilon''_2, \quad (50)$$

gdzie $\varepsilon'_0 > 0$; $\varepsilon''_0 > 0$; $\varepsilon'_2 > 0$; $\varepsilon''_2 > 0$.

Założenie takie powoduje, że liczba falowa k , wartości własne fal powierzchniowych ξ_s oraz poszukiwana stała propagacji k_x będą zespolone:

$$k = k_r - jk_i; \xi_s = \xi_{rs} - j\xi_{is}; k_x = \beta - j\alpha. \quad (51)$$

Wartości własne fal powierzchniowych korespondują z biegunami funkcji podcałkowej oraz zależą od stałej propagacji k_x ($\xi_s^2 = k_x^2 + \chi_s^2$). Bieguna na płaszczyźnie zmiennej zespolonej występuje w punkcie, który można określić jako

$$\chi_s^2 = \xi_s^2 - k_x^2 = (\xi_{rs}^2 - \xi_{is}^2 - \beta^2 + \alpha^2) - 2j \cdot (\xi_{rs} \cdot \xi_{is} - \beta \cdot \alpha) \quad (52)$$

Założmy dla wygody, że rozpatrywana struktura może podtrzymywać tylko jedną falę powierzchniową najniższego rodzaju. Założenie to pozwala nam brać pod uwagę tylko jedną parę biegunów na płaszczyźnie zmiennej zespolonej, co nie zmniejsza ogólności analizy.

Dla punktu rozgałęzienia, który powoduje wieloznaczność funkcji, analizę przeprowadza się analogicznie. Wykorzystując zależność

$$\gamma^2 = k_x^2 + \chi_2^2 - k^2 = 0|_{\chi_2=\chi_0}, \quad (53)$$

oraz uwzględniając (51), wzór (53) można zapisać w postaci

$$\chi_0^2 = (k_r^2 - k_i^2 - \beta^2 + \alpha^2) - 2j \cdot (k_r \cdot k_i - \beta \cdot \alpha). \quad (54)$$

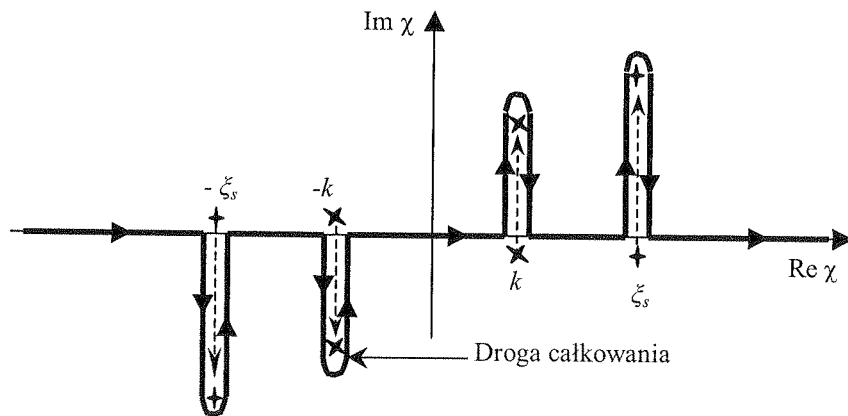
Wzory (52) i (54) mogą być wykorzystane do identyfikacji różnych rodzajów propagacji fali (rodzaju falowodowego, rodzaju fali powierzchniowej oraz rodzaju promieniującego) w rozpatrywanej strukturze. Identyfikacja ta bazuje na określeniu lokalizacji punktów bieguna i rozgałęzienia oraz odpowiedniej zmiany drogi całkowania na płaszczyźnie zmiennej zespolonej [16].

Rozpatrzmy rodzaj promieniujący. Rodzaj ten jest możliwy jeśli stała propagacji k_x spełnia nierówność

$$\operatorname{Re}\{k_x^2\} > \operatorname{Re}\{k^2\} \quad \text{lub} \quad k_r^2 - k_i^2 > \beta^2 - \alpha^2 \quad (55)$$

Założmy, że straty w dielektryku są małe, czyli, że $\beta\alpha > \xi_r\xi_i$ lub $\beta\alpha > k_r k_i$. W tym przypadku ze wzorów (52) i (54) widać, że przy zmianie stałej propagacji k_x punkt bieguna i rozgałęzienia migruje z kwadrantu II i IV przez kontur całkowania do kwadrantu III oraz I (Rys. 3).

Deformacja drogi całkowania z uwzględnieniem migracji punktów biegunów i rozgałęzienia jest konieczna. Zdeformowany kontur całkowania, pokazany na Rys. 3, nie jest jedynym możliwym. Zasady wyboru drogi całkowania są ogólnie znane z matematyki. Obliczenie całek może odbywać się w różny sposób. Jednym z podejść jest



Rys. 3. Droga całkowania z uwzględnieniem migracji punktów biegunów i rozgałęzienia

Fig. 3. Spectral integration contour and migration paths of pole and branch points

zastosowanie twierdzenia Cauchy'ego oraz wykorzystanie metody karbowania i lematu Jordana. Innym sposobem jest przeprowadzenie całkowania numerycznego wzdłuż pokazanego konturu całkowania. Residua funkcji w punkcie $\pm\xi_s$ odpowiadają fizycznemu wzbudzeniu się fal powierzchniowych w podłożu dielektrycznym. Całki wzdłuż części konturów, które omijają punkty rozgałęzienia odpowiadają fizycznemu wypromienianiu w przestrzeń, czyli są związane ze stratami na promieniowanie.

4. WYNIKI

Przedstawiona w niniejszym artykule metoda została w pierwszej kolejności przetestowana na znanych charakterystykach dyspersyjnych [16]. Wysoka zgodność wyników pokazuje użyteczność tej metodyki do rozwiązywania postawionego zagadnienia analizy antenowej.

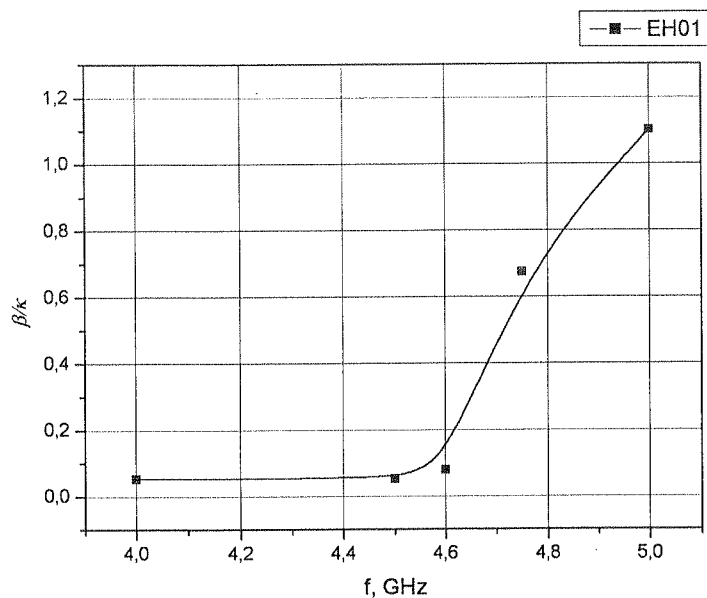
Na Rys. 4 przedstawiono zależności dyspersyjne (β — stała fazowa; α — współczynnik strat promieniowania) dla pierwszego wyższego rodzaju pola, tj. EH_{01} , w linii paskowej o następujących parametrach: $h = 0.2$ mm, $2w = 10$ mm, $\varepsilon/\varepsilon_0 = 10$. Podobne zależności przedstawiono na Rys. 5 dla trzeciego wyższego rodzaju pola, tj. EH_{03} . Rysunki te pokazują, że każdy rodzaj ma swoją własną częstotliwość krytyczną, czyli częstotliwość na której zanika zjawisko promieniowania fali. Dla przyjętych parametrów linii częstotliwość krytyczna pierwszego rodzaju wynosi około 5 GHz, zaś trzeciego rodzaju około 14 GHz. Dla częstotliwości znacznie niższych od częstotliwości krytycznej fala ulega silnemu rozproszeniu do otaczającej przestrzeni dzięki szybkiemu wzrostowi wartości współczynnika strat. Jednak w pobliżu częstotliwości krytycznej zachodzi bardzo korzystne dla nas zjawisko: szybka zmiana stałej fazowej przy względnie małej wartości współczynnika strat. Zjawisko to wykorzystuje się do skanowania częstotliwościowego [14].

W przypadku rozpatrywanej anteny użyteczna jest analiza zmian stałej propagacji w funkcji przenikalności dielektrycznej. Na Rys. 6 ($h = 0,2$ mm; $2w = 10$ mm) i Rys. 7 ($h = 0,15$ mm; $2w = 8$ mm) przedstawiono stałe fazowe oraz współczynniki strat w funkcji przenikalności dielektrycznej podłoża anteny. Widać, że w pobliżu częstotliwości krytycznej można uzyskać zakres zmian stałej fazowej od 0 do 1 (co powoduje zmianę kierunku kąta promieniowania od 0° do 90°) przy zmianie przenikalności dielektrycznej mniejszej od 20%. W tym samym czasie zmienia się współczynnik strat, co powoduje, że rozkład gęstości prądów na powierzchni apertury anteny będzie bardziej nierównomierny dla małych kątów promieniowania. W wyniku tego można oczekiwać niesymetrycznej charakterystyki kierunkowej, co w naszym przypadku nie jest istotne.

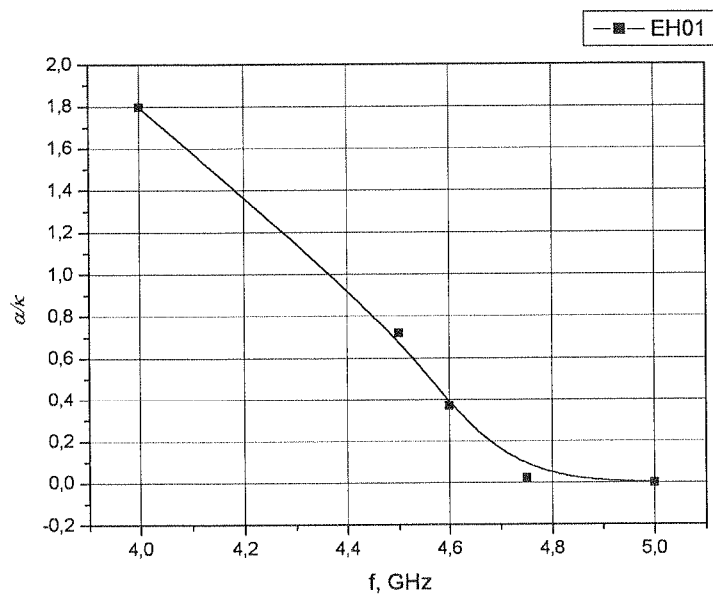
Na Rys. 8 pokazano wpływ szerokości linii paskowej na wartości stałej fazowej i współczynnika strat w funkcji przenikalności dielektrycznej dla różnych grubości podłoża ferroelektrycznego. Szerokość linii $2w = 10$ mm, częstotliwość $f = 6$ GHz.

Po określeniu stałej propagacji można oszacować wektor prądów $J_x(x, y)$ i $J_y(x, y)$ a następnie obliczyć pole promieniowania znalezionych rozkładów prądów.

a)



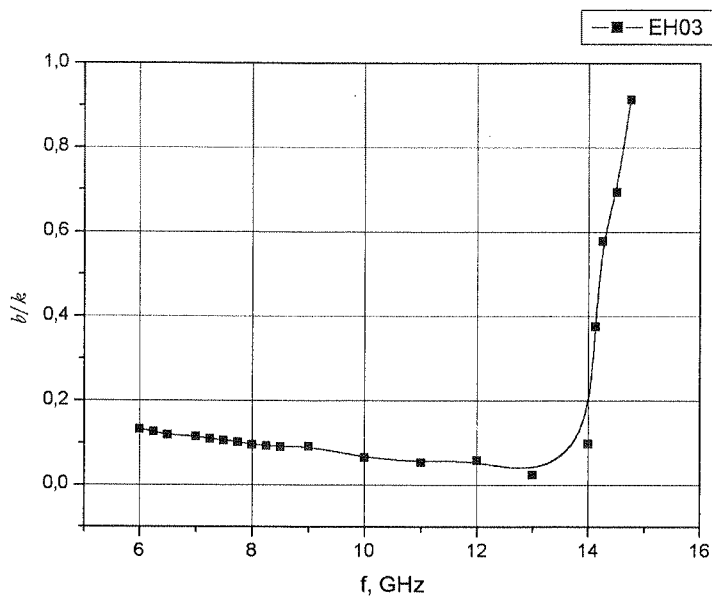
b)



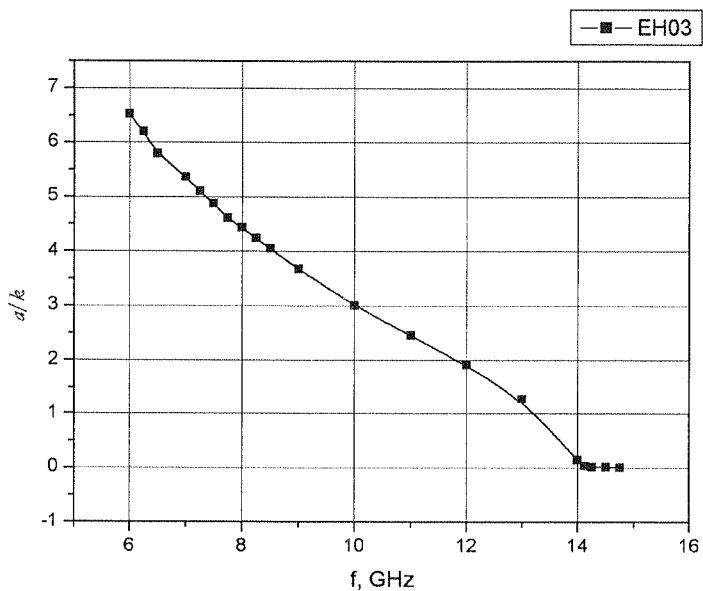
Rys. 4. Zależności dyspersyjne pierwszego wyższego rodzaju EH_{01} ($\epsilon/\epsilon_0 = 10$; $h = 0.2$ mm; $2w = 10$ mm)

Fig. 4. Dispersion curves for the first higher order modes EH_{01} ($\epsilon/\epsilon_0 = 10$; $h = 0.2$ mm; $2w = 10$ mm)

a)



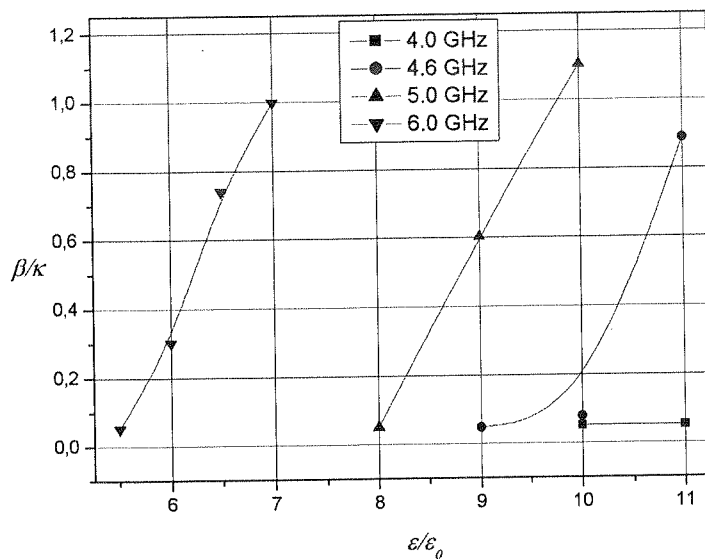
b)



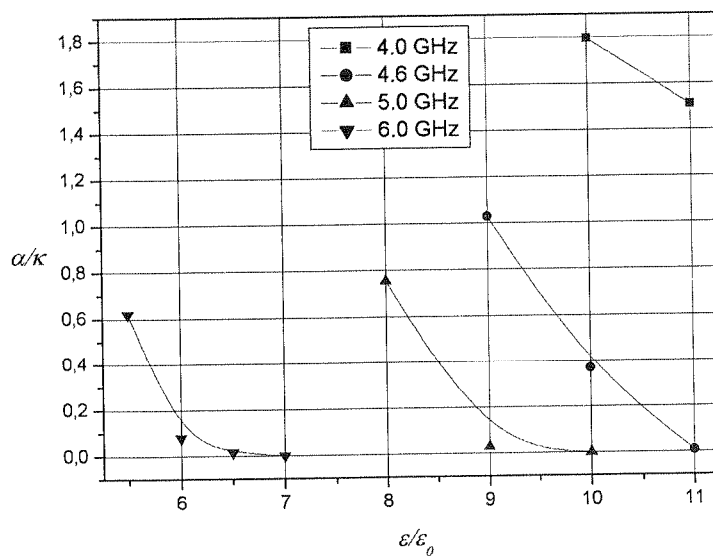
Rys. 5. Zależności dyspersyjne pierwszego wyższego rodzaju EH_{03} ($\epsilon/\epsilon_0 = 10$; $h = 0.2$ mm; $2w = 10$ mm)

Fig. 5. Dispersion curves for the first higher order modes EH_{03} ($\epsilon/\epsilon_0 = 10$; $h = 0.2$ mm; $2w = 10$ mm)

a)



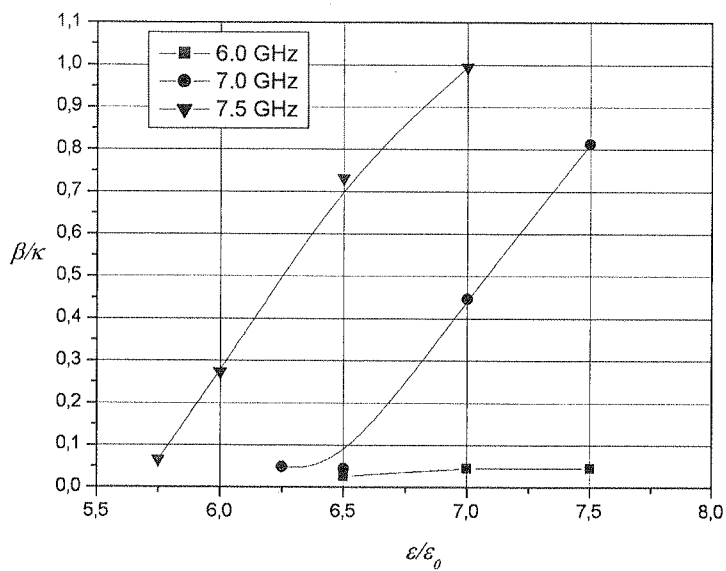
b)



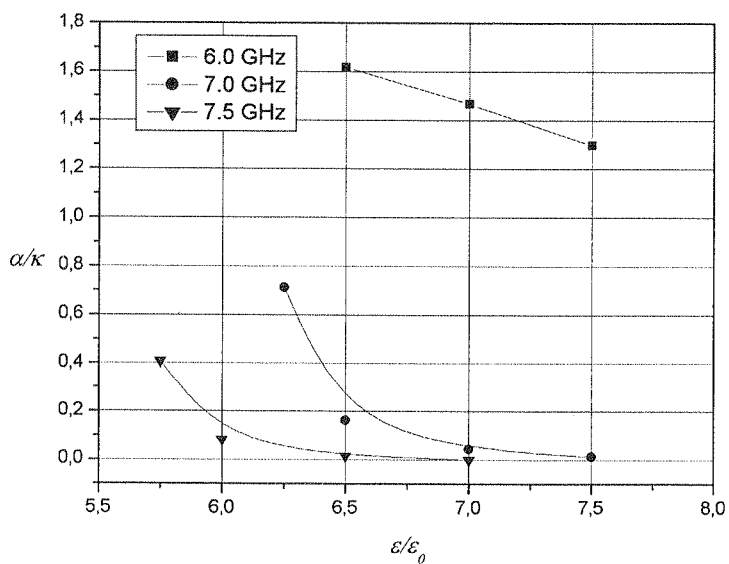
Rys. 6. Stała propagacji w funkcji przenikalności dielektrycznej dla różnych częstotliwości
($h = 0.2$ mm; $2w = 10$ mm)

Fig. 6. Complex axial propagation constant versus permittivity for different frequency
($h = 0.2$ mm; $2w = 10$ mm)

a)



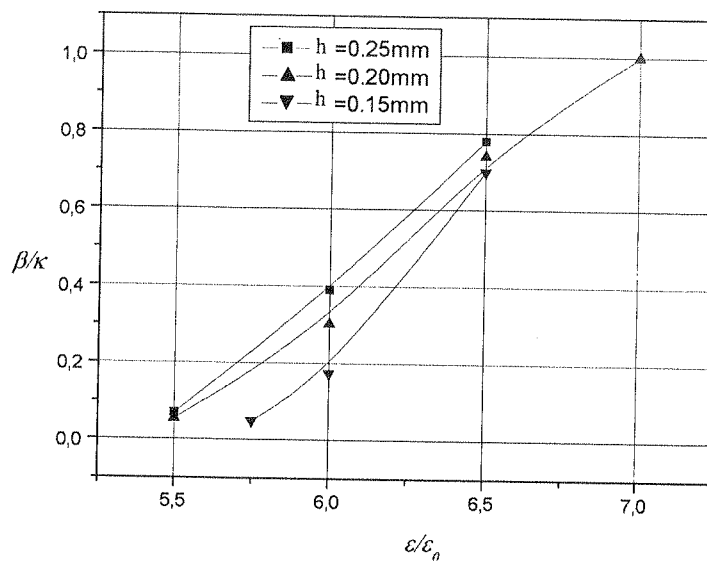
b)



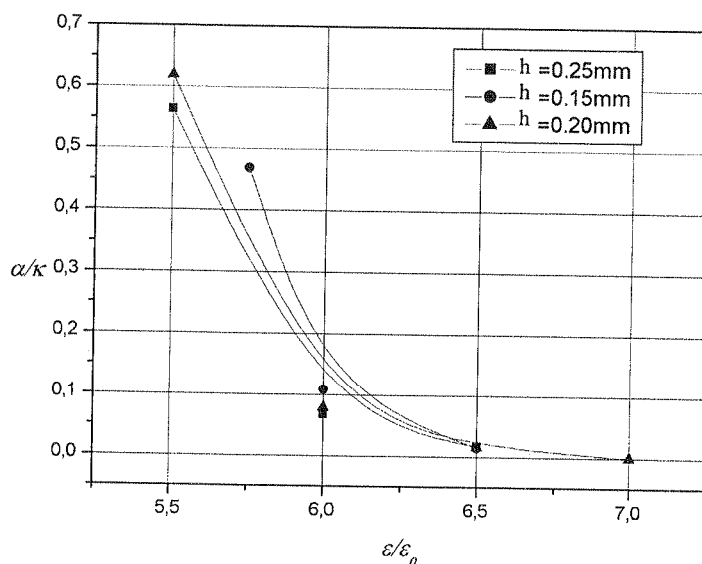
Rys. 7. Stała propagacji w funkcji przenikalności dielektrycznej dla różnych częstotliwości
($h = 0.15$ mm; $2w = 8$ mm)

Fig. 7. Complex axial propagation constant versus permittivity for different frequency
($h = 0.15$ mm; $2w = 8$ mm)

a)



b)



Rys. 8. Stała propagacji w funkcji przenikalności dielektrycznej dla różnych grubości podłoża ferroelektrycznego ($f = 6\text{ GHz}$; $2w = 8\text{ mm}$)

Fig. 8. Complex axial propagation constant versus permittivity for different thick of a ferroelectric substrate ($f = 6\text{ GHz}$; $2w = 8\text{ mm}$)

Na podstawie przedstawionej analizy struktury promieniującej została opracowana i zbadana skanująca antena ferroelektryczna. Przeprowadzono pomiary kierunku promieniowania w funkcji napięcia polaryzowania podłoża ferroelektrycznego. Zmiana tego napięcia wywołuje zmianę przenikalności dielektrycznej [12]. Wyniki pomiarów w pełni potwierdzają koncepcję skanującej anteny ferroelektrycznej. Aktualnie prowadzone są dalsze badania.

5. WNIOSKI

W artykule przedstawiono koncepcję skanującej anteny na podłożu ferroelektrycznym oraz metodę jej analizy pełnofalowej. Zaprezentowana metoda analizy pozwala wyznaczyć wszystkie charakterystyki anteny skanującej: stałe propagacji różnych rodzajów fal propagujących w linii paskowej, gęstości prądów na powierzchni linii, pole promieniowania, impedancję wejściową itd. Wyniki symulacji oraz badań eksperymentalnych w pełni potwierdzają słuszność wybranych rozwiązań i wskazują na możliwość szerokiego zastosowania metody analizy pełnofalowej przy projektowaniu struktur antenowych.

6. BIBLIOGRAFIA

1. R. Fralich, J. Litva: *Beam steerable active array antenna*. Electron. Lett., vol. 28, Jan. 1992, pp. 184-185.
2. R.E. Horn, H. Jacobs, E. Freibergs, K.L. Kohn: *Electronic modulated beam-steerable silicon waveguide array antenna*. IEEE Trans. Microwave Theory Tech., vol. 28, June 1980, pp. 647-655.
3. L. Huang, J.C. Chiao, M.P. De Lisio: *An electronically switchable leaky wave antenna*. IEEE Trans. On Antennas and Propag., vol. 48, No. 11, November 2000, pp. 1769-1772.
4. C.H. Lee, P.S. Mark, A.P. DeFonzo: *Optical control of millimeter-wave propagation in dielectric waveguides*. IEEE J. Quantum Electron., vol. QE-16, 1990, pp. 277-288.
5. A. Rosen, P.J. Stabile, P. Herczfeld, A. Daryoush, J.K. Butler: *Optically controlled IMPATT diodes and subsystems*. In Proc. 1989 SBMO Int. Microwave Symp., vol. II, 1989, pp. 589-594.
6. V.A. Manasson, L.S. Sadovnik, V.A. Yepishin, D. Marker: *An Optically Controlled MMW Beam-Steering Antenna Based on a Novel Architecture*. IEEE Trans. Microwave Theory Tech., vol. 45, August 1997, pp. 1497-1500.
7. R.A. Stern, J. Borowick: *A mm-wave Homogeneous Ferrite Phase Scan Antenna*. Microwave Journal, April 1987, pp. 101-108.
8. A.S. Cherepanov A.B. Gouskov, at. al: *Innovative integrated ferrite phased array technologies for EHF radar and communication applications*. IEEE International Symposium on Phased Array Systems and Technology. Oct. 15-18, 1996, Boston, US, pp. 74-77.
9. J. Modelski, E. Jaszczyzyn: *Sterowane anteny paskowe na podłożu z dielektrykiem nieliniowym*. Prace naukowe politechniki warszawskiej. Elektronika. Z.129. 2001. ss. 33-50.
10. E. Jaszczyzyn, J. Modelski: *Synteza paskowych szyków antenowych na podłożu ferroelektrycznym*. Kwartalnik elektroniki i telekomunikacji, 2002, 48, z. 1, ss. 67-80.

11. J.B.L. Rao, G.V. Trunk, D.P. Patel: *Two Low-Cost Phased Arrays*. IEEE AES Systems Magazine, June 1997, pp. 39-44.
12. E. Jaszczyżyn, M. Szafran, J. Modelski, E. Bobryk: *Badanie właściwości materiałów ferroelektrycznych w zakresie mikrofal*. Kwartalnik elektroniki i telekomunikacji. 2003. 49. z.2. ss. 245 -254.
13. D.M. Pożar, D.H. Schaubert: *Microstrip Antennas*. IEEE Press. New York. 1995.
14. W. Menzel: *A New Traveling-Wave Antenna in Microstrip*. Electronics and communication. Band 33, Heft 4, April 1979, pp. 137 - 140.
15. Г.Т. Марков, А.Ф. Чаплин: *Возбуждение электромагнитных волн*. Энергия, М.–Л., 1967, с. 376.
16. J.S. Bagby, C.H. Lee, D.P. Nyquist, Y. Yuan: *Identification of propagation regimes on integrated microstrip transmission lines*. IEEE Trans. on MTT, Vol. 41, November 1993.

Y. YASHCHYSHYN

FULL WAVE ANALYSIS OF SCAN ANTENNA ON FERROELECTRIC SUBSTRATE

Summary

Electronically scanned microwave antennas is accomplished by electronically altering the phase across the antenna radiating aperture (by means of discrete phase shifting elements or by changing frequency). This conventional inertialess scanning techniques in the microwave region are not cheap. A new scan antenna concept that realizable design using low-cost fabrication processes is presented.

This concept is based on the first higher order mode on a microstrip line, which substrate is made using ceramic-polymer composite with modified ferroelectric powder $\text{Ba}_{0.65}\text{Sr}_{0.35}\text{TiO}_3$ and appropriate polymer (grains of the powder have been sprayed into polymer with specific method). This ceramic-polymer composite was designed to change permittivity in response to an applied electric control field. The property of ferroelectric materials having a dielectric constant which can be modulated at high frequencies is very attractive.

The feasibility of dielectric scanning (changing the direction of the beam by varying the relative permittivity of the electrically controlled ceramic-polymer composite) has been investigated.

The radiation characteristics of such antenna have been accurate evaluated. Both longitudinal and transverse currents are treated. The method of solution used is an extension of the Fourier integral method which is used in conjunction with the method of moments (Galerkin solution).

The propagation constant k_x for higher order modes in open microstrip were calculated. In the present structure the complex propagation constant can be changed by changing the frequency or by changing the dielectric constant of the substrate (ϵ_2) as in the case of electrically controlled ferroelectric ceramic-polymer composite. In general, the direction of the radiated beam is given by $\theta_m = \sin^{-1}(\beta)$, where $k_x(\epsilon_2) = \beta(\epsilon_2) + j \cdot \alpha(\epsilon_2)$. For frequencies not far above the cut-off frequency, that means for $\beta/k_0 < 1$ (β is phase constant of the first higher order mode), the fields of the first higher order mode on the microstrip line are coupled to the line very loosely resulting in a considerable radiation.

Keywords: electronically scanned microwave antennas, ferroelectric materials, ceramic-polymer composite, full wave analysis, method of moments

Reda
szero
mem
się p
miki
Arty
lub
rozw
Elek
Obje

Wyr
Arty
w fc
w w
pow
Ukl
- T
- A
- M
- Z
- T

- U

Ob

Str
wz
w
-

Sp
Te
po
(n)
po
str

INFORMACJE DLA AUTORÓW

Redakcja przyjmuje do publikowania prace oryginalne, przeglądowe i monograficzne wchodzące w zakres szeroko pojętej elektroniki. Ponieważ KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI jest czasopismem Komitetu Elektroniki i Telekomunikacji Polskiej Akademii Nauk, w związku z tym na jego łamach znajdują się prace naukowe dotyczące podstaw teoretycznych i zastosowań z zakresu elektroniki, telekomunikacji, mikroelektroniki, optoelektroniki, radiotechniki i elektroniki medycznej.

Artykuły powinny charakteryzować oryginalne ujęcie zagadnienia, własna klasyfikacja, krytyczna ocena (teorii lub metod), omówienie aktualnego stanu, lub postępu danej gałęzi techniki oraz omówienie perspektyw rozwojowych. Artykuły publikowane w innych czasopismach nie mogą być kierowane do druku w Kwartalniku Elektroniki i Telekomunikacji w drugiej kolejności zgłoszenia.

Objętość artykułu nie powinna przekraczać 30 stron po około 1800 znaków na stronie, w tym rysunki i tabele.

Wymagania podstawowe.

Artykuły należy nadsyłać na wyraźnym, jednostronnym, czarno-białym wydruku komputerowym. Wydruk w formacie A4 powinien mieć znormalizowaną liczbę wierszy i znaków w wierszu (30 wierszy po 60 znaków w wierszu), w dwóch egzemplarzach, w języku polskim lub angielskim wybranym przez autora. Do wydruku powinna być dołączona dyskietka z elektronicznym tekstem artykułu. Preferowane edytory to WORD 6 lub 8. Układ artykułu (w wersji podstawowej) musi być następujący:

- Tytuł.
- Autor (imię i nazwisko autora/ów).
- Miejsce pracy (nazwa instytucji, miejscowość, adres. + ew. adres elektroniczny (e-mail)).
- Zwięzłe streszczenie powinno być w języku takim, w jakim jest pisany artykuł (wraz ze słowami kluczowymi).
- Tekst podstawowy powinien mieć następujący układ:

1. WPROWADZENIE

2. np. TEORIA

3. np. WYNIKI NUMERYCZNE

3.1.

3.2.

4.

5.

6. PODSUMOWANIE

7. ew. PODZIĘKOWANIA

8. BIBLIOGRAFIA

- Układ streszczenia w dodatkowej wersji językowej powinien być następujący:

AUTOR (inicjał imienia i nazwisko).

TYTUŁ (w języku angielskim – o ile artykuł pisany jest w języku polskim i na odrwót).

Obszerne do 3600 znaków streszczenia (wraz z słowami kluczowymi) w języku:

a. angielskim, gdy artykuł pisany jest w języku polskim.

b. polskim, gdy artykuł pisany jest w języku angielskim.

Streszczenie to powinno pozwolić czytelnikowi na uzyskanie istotnych informacji zawartych w pracy. Z tego względu w streszczeniu tym mogą być cytowane numery istotnych wzorów, rysunków i tabel zawartych w podstawowej wersji językowej.

- Wszystkie strony muszą mieć numerację ciągłą.

Sposób pisania tekstu.

Tekst powinien być pisany bez używania wyróżnień, a w szczególności nie dopuszcza się spacjiowania, podkreślania i pisania tekstu dużymi literami z wyjątkiem wyrazów, które umownie pisze się dużymi literami (np. FORTRAN). Proponowane wyróżnienia Autor może zaznaczyć w maszynopisie zwykłym ołówkiem za pomocą przyjętych znaków adjustacyjnych, np. podkreślenie linią przerywaną oznacza spacjiowanie (rozstrzelanie), podkreślenie linią ciągłą – pogrubienie, podkreślenie wężymkiem — kursywa.

Tekst powinien być napisany z podwójnym odstępem między wierszami, tytuły i podtytuły małymi literami. Marginesy z każdej strony powinny mieć około 35 mm. Wielkość czcionki wydruku powinna być zbliżona co najmniej co wielkości czcionki maszyny do pisanie (minimum 12 punktów). Przy podziale pracy na rozdziały i podrozdziały cyfrowe ich oznaczenia nie powinny być większe niż II stopnia (np. 4.1.1.).

Sposób pisania tabel.

Tabele powinny być pisane na oddzielnych stronach. Tytuły rubryk pionowych i poziomych powinny być napisane małymi literami z podwójnym odstępem między wierszami. Przypisy (notki) dotyczące tabel należy pisać bezpośrednio pod tabelami. Tabele należy numerować kolejno liczbami arabskimi, u góry każdej tabeli podać tytuł dwujęzyczny. W pierwszej kolejności w podstawowej wersji językowej, a później w dodatkowej wersji językowej. Tabele umieścić na końcu maszynopisu. Przyjmowane są tabele algorytmów i programy na wydrukach komputerowych. W tym przypadku zachowany jest ich oryginalny układ. Tabele powinny być cytowane w tekście.

Sposób pisania wzorów matematycznych.

Rozmieszczenie znaków, cyfr, liter i odstępów powinno być zbliżone do rozmieszczenia elementów druku. Wskaźniki i wykładniki potęg powinny być napisane wyraźnie i być prawidłowo obniżone lub podwyższone w stosunku do linii wiersza podstawowego. Znaki nad literami i cyframi, całkami i in. symbolami (strzałki, linie, kropki, daszki) powinny być umieszczone dokładnie nad tymi elementami, do których się odnoszą. Numery wzorów cyframi arabskimi powinny być kolejne i umieszczone w nawiasach okrągłych z prawej strony. Nazwy jednostek, symbole literowe i graficzne powinny być zgodne z wytycznymi IEE (International Electrical Commission) oraz ISO (International Organization of Standardization).

Powołania.

Powołania na publikacje powinny być umieszczone na ostatnich stronach tekstu pod tytułem „Bibliografia”, opatrzone numeracją kolejną bez nawiasów. Numeracja ta powinna być zgodna z odnośnikami w tekście artykułu. Przykłady opisu publikacji:

- periodycznej 1. F. Valdoni: A new milimetre wave satelite. E.T.T. 1990, vol. 2, no 5, pp. 141–148
- nieperiodyczne 2. K. Andersen: A resource allocation framework. XVI International Symposium, Stockholm (Sweden), may 1991, paper A 2,4
- książki 3. Y.P. Tvidis: Operation and modeling of the MOS transistors. New York, McGraw-Hill, 1987, p. 553

Materiały ilustracyjne.

Rysunki powinny być wykonane wyraźnie, na papierze gładkim lub milimetrowym w formacie nie mniejszym niż 9 × 12 cm. Mogą być także w postaci wydruku komputerowego (preferowany edytor Corel Draw). Fotografie lub diapozytywy przyjmowane są raczej czarno-białe w formacie nie przekraczającym 10 × 15 cm. Na marginesie każdego rysunku i na odwrocie fotografii powinno być napisane ołówkiem imię i nazwisko Autora oraz skrót tytułu artykułu, do którego są przeznaczone oraz numer rysunku. Spis podpisów pod rysunki i fotografie powinien być umieszczony na oddzielnej stronie. Podpisy pod rysunkami (fotografiami) powinny być dwujęzyczne: w pierwszej kolejności w podstawowej wersji językowej, a później w dodatkowej wersji językowej. Rysunki powinny być cytowane w tekście.

Uwagi końcowe.

Na odrębnej stronie powinny być podane następujące informacje:

- adres do korespondencji z kodem pocztowym (domowy lub do miejsca pracy),
- telefon domowy i/lub do miejsca pracy,
- adres e-mailowy (jeśli autor posiada).

Autorowi przysługuje bezpłatnie 20 odbitek artykułu. Dodatkowe egzemplarze odbitek, lub cały zeszyt Autor może zamówić u wydawcy na własny koszt.

literami.
żona co
ozdziały

nnny być
l należy
ej tabeli
atkowej
ramy na
nnny być

y druku.
yższone
ki, linie,
Numery
Nazwy
tronical

grafia",
rtykułu.

148
Stock-
w-Hill,

ym niż
afie lub
ginesie
z skrót
winien
yczne:
ysunki

Autor

Autora obowiązują korekta autorska, którą powinien wykonać w ciągu 3 dni od daty otrzymania tekstu z Redakcji oraz zwrócić osobiście lub listownie pod adres Redakcji. Korekta powinna być naniesiona na przekazanych Autorowi szpaltach na marginesach ew. na osobnym arkuszu w przypadku uzupełnień tekstu większych niż dwa wiersze. W przypadku nie zwrócenia korekty w terminie, korektę przeprowadza Redakcja Techniczna Wydawcy. Redakcja prosi Autorów o powiadomienie ją o zmianie miejsca pracy i adresu prywatnego.

INFORMATION FOR AUTHORS OF K.E.T.

The editorial staff will accept for publishing only original monographic and survey papers concerning widely understood electronics. Because of the fact that KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI is a journal of the Committee for Electronics and Telecommunications of Polish Academy of Science, it presents scientific works concerning theoretical bases and applications from the field of electronics, telecommunications, microelectronics, optoelectronics, radioelectronics and medical electronics.

Articles should be characterised by original depiction of a problem, its own classification, critical opinion (concerning theories or methods), discussion of an actual state or a progress of a given branch of a technique and discussion of development perspectives.

An article published in other magazines can not be submitted for publishing in K.E.T. The 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 staff as a one side, clear, black and white computer printout in two copies. The article should be prepared in English or Polish. 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 and vice versa)
 - Extensive summary, hawever not exceeching 3600 characters (along with keywords) in Polish, if artide was prepared in English and vice versa). 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 can not 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 consist 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 are attributed to. Formulas should be numbered with Arabic numbers placed 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 Electrotechnical Commission) and ISO (International Organisation of Standardisation).

References

References should be placed at the end of the main text with the subtitle „References“. References should be numbered (without brackets) adequately to references placed in the text. Examples of periodical [1], non-periodical [2] and book [3] references:

1. F. Valdoni: A new millimetre wave satellite. 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. Tividis: Operation and modeling of the MOS transistors. New York. McGraw-Hill. 1987. p. 553

Figures

Figures should be clearly drawn on plain or millimetre 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 white format not greater than 10×15 cm. On the margin of each drawing and on the back side of each photo author's name and abbreviation of the title of article should be placed. Figure's captions should be given in two languages (first in the language the article is written in and then in additional language). Figure's captions 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 entitled to free of charge 20 copies of article. Additional copies or the whole magazine can be ordered at publisher at the author's expense.

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

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