POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 47 — ZESZYT 2

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 2001

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

Dyżury Redakcji: środy 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

WYDAWNICTWO NAUKOWE PWN SA
Warszawa, ul. Miodowa 10

Ark. wyd. 11,00 Ark. druk. 8,75	Podpisano do druku w lipcu 2001 r.
Papier offset. kl. III 80 g. B-1	Druk ukończono w lipcu 2001 r.

SKŁAD I ŁAMANIE: PHOTOTEXT, Warszawa, ul. Chmielna 120
DRUK I OPRAWA: Zakład Poligraficzny Oja Druk, ul. Sporna 2F, Warszawa

Szanowni Autorzy

„Kwartalnik Elektroniki i Telekomunikacji" — Electronics and Telecommunications Quartery jest kontynuatorem tradycji powstałego 47 lat temu kwartalnika p.t. „Rozprawy Elektrotechniczne".

Kwartalnik jest czasopismem Komitetu Elektroniki i Telekomunikacji Polskiej Akademii Nauk. Wydawany jest przez Wydawnictwo Naukowe PWN SA w Warszawie. 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, optoelektroniki, 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 specjalitów, co zapewnia że publikacje te są uznawane jako autorski dorobek naukowy. Opublikowanie w kwartalniku wyników 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 dotyczące 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

Fo
algorith
analysi
efficien
those
present

Keywo

For eve
· > such tha
the field. T
GF(p^m). Ev
possible rep
where p is
respectively
The mu
element g s
The elemen

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2001, 47, z. 2, ss. 145–160

Finite Fields and Fire Codes

TOMASZ OSTROWSKI

Technical University of Szczecin
Institute of Electronics, Communications and Computer Science
ul. 26 Kwietnia 10, PL 71-126 Szczecin, Poland
e-mail: ostrowsk@arcadia.tuniv.szczecin.pl

Otrzymano 2000.11.14
Autoryzowano 2001.03.28

Following the brief introduction into the algebra of the finite fields we develop the algorithm that supports the search for the efficient two-dimensional Fire cyclic codes. The analysis of the codes with generating polynomial of degree less than 12 demonstrated that the efficient codes are sparsely distributed in the set of all possible 2-D Fire codes. The examples of those codes are submitted. The case study of the parallel encoder and decoder design is presented.

Keywords: Finite field, two-dimensional Fire code, number theory.

1. INTRODUCTION TO FINITE FIELDS AND FIRE CODES

1.1. FINITE FIELDS [1, 2, 5, 6, 7]

For every prime number p and positive integer m there is a finite field $F = < F, +, \cdot >$ such that $\# F = p^m$, where $\# F$ denotes the cardinality of the set F, i.e. the order of the field. The field F is called Galois field with the characteristic p and is denoted by $GF(p^m)$. Every finite field is isomorphic with some Galois field. For instance, one of the possible representations for the field GF(p) is the field $Z_p = < Z_p = \{0,1,...,p-1\}, +, \cdot >$, where p is a prime, while + and $\cdot$ denote the addition and multiplication modulo p, respectively.

The multiplicative group $G = < G, \cdot >$ is said to be cyclic group if it contains an element g such that each element $\omega \in G$ can be represented as g^i where i is an integer. The element g is then called the generator of the group G.

If the field $F = < F, +, \cdot >$ is Galois field then the multiplicative group $G = < F - \{0\}, \cdot >$ is cyclic. The generator of the group G is called the primitive element of the Galois field F. The order of a non-zero element $\omega \in F$ is the least positive integer e such that $\omega^e = 1$.

If F is a field, then a polynomial in the indeterminate x over the field F is an expression of the form $f(x) = a_n x^n + \cdots + a_1 x + a_0$, where each coefficient $a_i \in F$ and $n \geq 0$. The polynomial is said to be monic if its leading coefficient is equal to 1.

The polynomial ring $F[x] = < F[x], +, \cdot >$ is the ring formed by the set of all polynomials in the indeterminate x having coefficients from F. The two operations are the standard polynomial addition and multiplication, with coefficient arithmetic performed in the field F. The polynomial is said to be irreducible if it cannot be written as the product of two polynomials, each of positive degree. If $g(x), h(x) \in F[x]$, and $h(x) \neq 0$, then ordinary polynomial long division of g(x) by h(x) yields unique polynomials q(x) and $r(x) \in F[x]$ such that $g(x) = q(x)h(x) + r(x)$, where deg $r(x) <$ deg $h(x)$. The polynomial q(x) is called the quotient, while r(x) is called the remainder. The remainder is referred to as g(x) mod h(x). The polynomials g(x) and h(x) are said to be congruent modulo polynomial f(x) if the division g(x) – h(x) by f(x) leaves the remainder 0. This is denoted by $g(x) \equiv h(x) \bmod f(x)$. The equivalence class of a polynomial g(x) is the set of all polynomials in $F[x]$ congruent to g(x) modulo f(x). The equivalence class is denoted by [r(x)]/f(x) where r(x) is the remainder of the division of g(x) by f(x). In the following $F[x]/f(x)$ denotes the set of all equivalence classes of the polynomials in $F[x]$ with the addition and multiplication performed modulo f(x). The equivalence classes are represented by the polynomials of degree less than deg f(x). If f(x) is irreducible then $F[x]/f(x)$ is a field.

A commonly used representation for the elements of the finite field $GF(p^m)$, where p is a prime, is a polynomial basis representation. It is henceforth assumed that $m \geq 2$, because if m = 1 then $GF(p^m)$ is just Z_p and the field arithmetic is performed modulo p. If $f(x) \in Z_p[x]$ is an irreducible polynomial of degree m, then $Z_p[x]/f(x)$ is a finite field isomorphic with $GF(p^m)$. Thus the elements of $GF(p^m)$ can be represented by the polynomials in $Z_p[x]$ of degree less than m. Furthermore, f(x) has a root $\alpha = [x]/f(x) \in GF(p^m)$ and $B = (\alpha^{m-1}, \ldots, \alpha^2, \alpha^1, 1)$ is a polynomial basis of $GF(p^m)$ over Z_p. The irreducible polynomial f(x) is called a primitive polynomial if the equivalence class [x]/f(x) is a generator of the multiplicative group of all non-zero elements in $GF(p^m) = Z_p[x]/f(x)$. Equivalently, the irreducible polynomial f(x) is said to be a primitive polynomial if its root α in $GF(p^m)$ is the primitive element of $GF(p^m)$.

1.2. TWO-DIMENSIONAL FIRE CYCLIC CODE [3, 4]

Let m_1 and m_2 be the positive integers. Consider an irreducible polynomial p(x) of degree $m_1 m_2$ over GF(2) whose root α is of order e. Let e_1 and e_2 be the positive integers that fulfill the following conditions:

(i) $e_1 e_2 = e$,
(ii) m_1 is the least positive integer such that $2^{m_1} - 1$ is divisible by e_1,
(iii) e_1 and e_2 are relatively prime.

Define $\beta = \alpha^{e_2}$ and $\gamma = \alpha^{e_1}$. Thus the order of β equals e_1 and the order of γ equals e_2. Furthermore, β is the element of $GF(2^{m_1})$ because from condition (ii) the order of β divides $2^{m_1} - 1$. Let c_1 and c_2 be such positive integers that their product $c_1 c_2$ is not divisible by e. Define $n_1 = LCM(c_1, e_1)$ and $n_2 = LCM(c_2, e_2)$, where LCM denotes least common multiple.

The 2-D Fire cyclic code C_F with the row dimension n_1 and the column dimension n_2 is defined as the set of all bivariate polynomials V(x, y) over GF(2) that fulfill the following conditions:

(i) $V(x, y) \equiv 0 \bmod (x^{c_1} + 1, y^{c_2} + 1)$, i.e. $V(x, y) = a(x, y)(x^{c_1} + 1) + b(x, y)(y^{c_2} + 1)$, where a(x, y) and b(x, y) are the polynomials over GF(2),
(ii) $\deg_x V(x, y) < n_1$, $\deg_y V(x, y) < n_2$,
(iii) $V(\beta, \gamma) = 0$.

The elements of the binary codeword matrix are represented by the binary coefficients $v_{i,j}$ in terms $x^i y^j$ of the polynomial V(x, y). The code C_F is cyclic, i.e. the cyclic permutations of the rows or columns of the codeword matrix results in the valid codeword. In the following if not otherwise mentioned the arithmetic is performed modulo 2.

The check tensor of the code C_F is defined as $n_1 \times n_2$ matrix $H' = [h'_{i,j}]$, where each matrix element $h'_{i,j}$ is the column vector over GF(2) of length $c_1 c_2 + m_1 m_2$. The vectors $h'_{i,j}$ are determined as follows. Let $\Phi(k, l)$ be a bijection of the set $\{(k, l) | 0 \le k < c_1, 0 \le l < c_2\}$ onto the set $\{i | 0 \le i < c_1 c_2\}$ and $\mathbf{u}_i$ denotes the unit vector of length $c_1 c_2$ whose i-th coordinate is equal to 1. Then the check tensor $\mathbf{H}' = [h'_{i,j}]$ is computed in the following manner

$$h'_{i,j} = [\mathbf{u}_{\Phi(i \bmod c_1, j \bmod c_2)} \beta^i \gamma^j]^T \tag{1}$$

where the element $\beta^i \gamma^j$ that belongs to $GF(2^{m_1 m_2})$ is represented by the binary vector over GF(2) of length $m_1 m_2$ and T denotes the transpose. The number of the binary coordinates of $h'_{i,j}$ equals $c_1 c_2 + m_1 m_2$ thus the number of the check bits for the C_F is also $c_1 c_2 + m_1 m_2$. The polynomial V(x, y) is a valid C_F codeword if and only if the following equality holds

$$\sum_{\substack{0 \le i < n_1 \\ 0 \le j < n_2}} v_{i,j} h_{i,j} = \mathbf{0} \tag{2}$$

where $\mathbf{0}$ denotes a zero column vector.

The error correcting capability of the code C_F is as follows. If ν_1 and ν_2 are the positive integers such that $\nu_1 \nu_2 \le m_1 m_2$, and the set $S(\nu_1, \nu_2) = \{\beta^i \gamma^j | 0 \le i < \nu_1, 0 \le j < \nu_2\}$ that consists of $\nu_1 \nu_2$ elements of the field $GF(2^{m_1 m_2})$ is linearly independent over GF(2)

then C_F can correct every burst errors of dimension $b_1 \times b_2$ or less including the „end around" bursts where b_1 and b_2 are the positive integers such that:

(i) $b_1 \leq \min[(c_1 + 1)/2, \nu_1]$,

(ii) $b_2 \leq \min[(c_2 + 1)/2, \nu_2]$.

The integers ν_1 and ν_2 are not determined a priori, however it is always possible to choose as ν_1 and ν_2 the integers m_1 and m_2, respectively, because the set $S(m_1, m_2)$ is linearly independent. Alternatively, one can choose as ν_1 and ν_2 the integers m_2' and m_1', respectively, where m_1' is the least positive integer such that $2^{m_1'} - 1$ is divisible by e_2 and $m_2' = m_1 m_2 / m_1'$. If the irreducible polynomial p(x) is a primitive polynomial then one of the above cases is trivial, i.e. m_2 or m_2' is equal to 1.

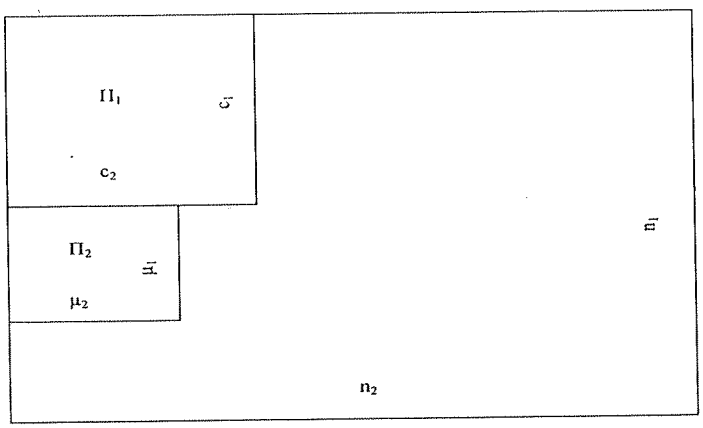


Fig. 1a. The set Π of the check positions

Fig. 1b. The set Π′ of the check positions

For the purpose of the encoder and decoder design one needs to know the positions of the $c_1 c_2 + m_1 m_2$ check bits of the code C_F. The set Π can be the set of check bit

positions if and only if the set $\{h'_{i,j} | (i,j) \in \Pi\}$ is linearly independent over GF(2). It is convenient to use as the positions of $c_1c_2 + m_1m_2$ check bits the set Π or the set Π' that are defined as follows:

(i) $\Pi = \Pi_1 + \Pi_2$,

(ii) $\Pi' = \Pi_1 + \Pi'_2$,

where $\Pi_1 = \{(i, j) | 0 \le i < c_1, \quad 0 \le j < c_2\}$, $\Pi_2 = \{(i,j) | c_1 \le i < c_1 + \mu_1, \quad 0 \le j < \mu_2\}$, $\Pi'_2 = (i,j) | 0 \le i < \mu_1, c_2 \le j < c_2 + \mu_2\}$. The μ_1 and μ_2 are the positive integers that fulfill the condition $\mu_1\mu_2 = m_1m_2$. The sets Π and Π' are depicted in Figs. 1a and 1b, respectively. There are following sufficient conditions for the Π and Π' to be the sets of the check bit positions.

Assume that c_1 is not divisible by e_1. The set Π is then the set of the check bit positions in the following cases:

(i) $\mu_1 \le c_1$, $\mu_2 \le c_2$, and the set $S(\mu_1, \mu_2)$ is linearly independent over GF(2),

(ii) $\mu_1 = m_1$, $\mu_2 = m_2$, and $m_2 \le c_2$,

(iii) $\mu_1 = m'_2$, $\mu_2 = m'_1$, and $m'_1 \le c_2$.

Assume that c_2 is not divisible by e_2. The set Π' is then the set of the check bit positions in the following cases:

(i) $\mu_1 \le c_1$, $\mu_2 \le c_2$, and the set $S(\mu_1, \mu_2)$ is linearly independent over GF(2),

(ii) $\mu_1 = m_1$, $\mu_2 = m_2$, and $m_1 \le c_1$,

(iii) $\mu_1 = m'_2$, $\mu = m'_2$, and $m'_2 \le c_1$.

The set $\{h'_{i,j} | (i,j) \in \Pi \text{ or } \Pi'\}$ can be transformed onto the set of $c_1c_2 + m_1m_2$ unit vectors. This form of the check tensor denoted as $\mathbf{H} = [h_{i,j}]$ is called the canonical form of the check tensor. In what follows we consider only the set Π_2 because in the case of the set Π'_2 one needs only to change the notation. To determine the check tensor in the canonical form it is necessary to define the bijection $\Psi(k,l)$ of the set $\Pi_2 = \{(k,l) | c_1 \le k < c_1 + \mu_1, \ 0 \le l < \mu_2\}$ onto the set $\{i | 0 \le i < m_1m_2$ and the elements $\xi_{i,j} = \beta^i\gamma^j + \beta^{i \bmod c_1}\gamma^{i \bmod c_2}$. Then for each vector $h_{i,j}$ one computes a binary vector $z_{i,j}$ of length m_1m_2 that represents $\xi_{i,j}$ as the linear combination over GF(2) of the linearly independent elements from the set $Z = \{\xi_{k,l} | (k,l) \in \Pi_2\}$. The bijection $\Psi(k,l)$ maps the matrix positions of the elements from Z onto the coordinates of vector $z_{i,j}$. It can be shown that the check tensor in the canonical form $\mathbf{H} = [h_{i,j}]$ can be written as follows

$$h_{i,j} = [t_{i,j} \ z_{i,j}]^T \tag{3}$$

where $t_{i,j} = \mathbf{u}_{\Phi(i \bmod c_1, j \bmod c_2)} + \mathbf{M} z_{i,j}$ and $\mathbf{M} = [m_{i,j}]$ denotes $c_1c_2 \times m_1m_2$ binary matrix whose $m_{i,j}$ matrix element is equal to 1 if there is a pair $(k,l) \in \Pi_2$ such that $i = \Phi(k \bmod c_1, l \bmod c_2)$ and $j = \Psi(k,l)$, otherwise the matrix element is equal to 0.

2. COMPUTATION OF THE CODE PARAMETERS

To find the efficient code parameters we analyzed numerically the possible codes C_F, assuming codeword matrix size which enables the construction of the squares

16×16, 32×32, 64×64, etc. that are typical for the computer science practice. Unfortunately, there is none $2^M\times2^N$ code C_F, where M and N are positive integers. However, the useful feature of the code C_F, is the possibility to assembly the large code matrix approximating desired C_F dimension from the several matrices of smaller size. To compute the C_F codes parameters we proposed the following algorithm.

Algorithm 1. C_F code analyzer.
For every ordered pair (n_1, n_2) of the positive integers such that $N_{min} < n_1 < N_{max}$, $N_{min} < n_2 < N_{max}$, **do**

1. Find the following all ordered pairs of positive integers
   (i) (e_1, c_1) such that $n_1 = LCM(e_1, c_1)$
   (ii) (e_2, c_2) such that $n_2 = LCM(e_2, c_2)$
2. Find all ordered four-tuples (e_1, e_2, c_1, c_2)
3. For each four-tuple (e_1, e_2, c_1, c_2) do
   3.1 Check if the conjunction of the following conditions is true
      (i) e_1 and e_2, are relatively prime
      (ii) product e_1e_2 is an odd integer
      (iii) product c_1c_2 is not divisible by e_1e_2
      (iv) $c_1 \neq n_1$ or $c_2 \neq n_2$
      If the conjunction is true accept the four-tuple and go to step 3.2, otherwise reject it and every other four-tuple that can be rejected on this stage and continue in step 3.1 with the next four-tuple
   3.2 Search for the least positive integer $m < M_{max}$ such that $2^m - 1$ is divisible by e_1e_2. If such integer exists go to step 3.3, otherwise reject all four-tuples with the first element e_1 and the second element e_2 and continue in step 3.1 with the next four-tuple
   3.3 Search the tables of irreducible polynomials for an irreducible polynomial of degree m over GF(2) that has a root of order e_1e_2 in $GF(2^m)$. If such polynomial exists go to step 3.4, otherwise reject all four-tuples tuples with the first element e_1 and the second element e_2 and continue in step 3.1 with the next four-tuple
   3.4 Compute least positive integer m_1 such that $2^{m_1} - 1$ is divisible by e_1 and the least positive integer m_1' such that $2^{m_1'} - 1$ is divisible by e_2
   3.5 Compute the following parameters of the code C_F
      (i) $e = e_1e_2$
      (ii) $m_2 = m/m_1$
      (iii) $m_2' = m/m_1'$
      (iv) $\lambda = m_1m_2 + c_1c_2$ — number of the check bits
      (v) $\eta = \lambda/n_1n_2$ — code efficiency
      (vi) $\tau_1 = \text{entier}[(c_1 + 1)/2]$
      (vii) $\tau_2 = \text{entier}[(c_2 + 1)/2]$
      (viii) $b_1 \times b_2 = \min(\tau_1, m_1) \times \min(\tau_2, m_2)$ — correcting capability if m_1 and m_2 are chosen as ν_1 and ν_2, respectively

(ix) $b'_1 \times b'_2 = \min(\tau_1, m'_2) \times \min(\tau_1, m'_1)$ — correcting capability if m'_2 and m'_1 are chosen as ν_1 and ν_2, respectively

We analyzed approximately 3,000 cases of the 2-D Fire cyclic codes. It occurred that only small amount of them could be useful for the practical purposes. Some codes examples are collected in Table 1. For the reason of computational efficiency we limited our analysis to the codes C_F determined by the generating polynomial p(x) of degree m less than 12.

Fire code pareameters Table 1

n_1	n_2	e_1	e_2	e	c_1	c_2	m	m_1	m_2	m'_1	m'_2	λ	η	τ_1	τ_2	$b_1 \times b_2$	$b'_1 \times b'_2$
15	12	5	3	15	3	4	4	4	1	2	2	16	0.90	2	2	2×1	2×2
33	31	11	31	341	3	31	10	10	1	5	2	103	0.10	2	16	2×1	2×5
33	62	11	31	341	33	2	10	10	1	5	2	76	0.04	17	1	10×1	2×1
33	63	11	3	33	3	21	10	10	1	2	5	73	0.04	2	11	2×1	2×2
33	65	3	5	15	11	13	4	2	2	4	1	147	0.07	6	7	2×2	1×4
34	35	17	5	85	2	7	8	8	1	4	2	22	0.02	1	4	1×1	1×4

3. ENCODER AND DECODER DESIGN

We follow the parallel approach to linear error correcting codes design [7] that is modified here for the two-dimensional case. Reader interested in serial design can consult [3].

3.1. ENCODER DESIGN

Given the code C_F and the corresponding generating polynomial p(x) the encoder design can be performed as follows.

Algorithm 2. Encoder design.

1. Compute check tensor $\mathbf{H} = [h_{i,j}]$ in the canonical form according to the formula (3)
2. Define a bijection Θ of the set $\Omega = \{(i,j) | 0 \le i < n_1,\ 0 \le j < n_2\}$ onto the set $\{k | 0 \le k < n_1 n_2\}$ such that for each $(i,j) \in \Pi$, $0 \le \Theta(i,j) < c_1 c_2 + m_1 m_2$, and if $(i,j) \in \Pi_1$ then $\Theta(i,j) = \Phi(i,j)$, and if $(i,j) \in \Pi_2$ or Π'_2 then $\Theta(i,j) = \Psi(i,j) + c_1 c_2$. Then map the check tensor $\mathbf{H} = [h_{i,j}]$ onto $(c_1 c_2 + m_1 m_2) \times n_1 n_2$ matrix $\mathbf{H}^{\Theta} = [h^{\Theta}{}_{i,j}]$ such that (i,j)-th column vector of the check tensor $\mathbf{H}$ becomes the $\Theta(i,j)$-th column of the matrix $\mathbf{H}^{\Theta}$, i.e.

$$\mathbf{H}^{\Theta} = \begin{vmatrix} 1\ 0\ 0\ \ldots\ 0 & h^{\Theta}_{0,\ c_1c_2+m_1m_2} & \ldots & h^{\Theta}_{0,\ n_1n_2-1} \\ 0\ 1\ 0\ \ldots\ 0 & h^{\Theta}_{1,\ c_1c_2+m_1m_2} & \ldots & h^{\Theta}_{0,\ n_1n_2-1} \\ 0\ 0\ 1 \quad . & . & & . \\ . \quad . \quad . & . & & . \\ . \quad . \quad . & . & & . \\ . & . & & . \\ 0\ 0\ .\ .\ 0\ 0 & h^{\Theta}_{c_1c_2+m_1m_2-1,\,c_1c_2+m_1m_2} & & h^{\Theta}_{c_1c_2+m_1m_2-1,\,n_1n_2-1} \end{vmatrix} \tag{4}$$

3. Compute from the matrix $\mathbf{H}^{\Theta} = [h_{i,j}]$ the $[n_1 n_2 - (c_1 c_2 + m_1 m_2)] \times n_1 n_2$ generator matrix $\mathbf{G}^{\Theta} = [g^{\Theta}_{i,j}]$ of the orthogonal vectors as follows

$$\mathbf{G}^{\Theta} = \begin{vmatrix} h^{\Theta}_{0,\ c_1c_2+m_1m_2} & h^{\Theta}_{1,\ c_1c_2+m_1m_2} & \ldots & h^{\Theta}_{c_1c_2+m_1m_2-1,\,c_1c_2+m_1m_2} & 1\ 0\ \ldots\ 0 \\ h^{\Theta}_{0,\ c_1c_2+m_1m_2+1} & h^{\Theta}_{1,\ c_1c_2+m_1m_2+1} & \ldots & h^{\Theta}_{c_1c_2+m_1m_2-1,\,c_1c_2+m_1m_2+1} & 0\ 1\ \ldots\ 0 \\ . \quad . & . & & . & . \ .\ . \\ . \quad . & . & & . & . \ .\ \ . \\ . \quad . & . & & . & . \ .\ \ . \\ h^{\Theta}_{0,\ n_1n_2-1} & h^{\Theta}_{1,\ n_1n_2-1} & \ldots & h^{\Theta}_{c_1c_2+m_1m_2-1,\,n_1n_2-1} & 0\ 0\ \ldots\ 1 \end{vmatrix} \tag{5}$$

4. Compute the outputs of encoder as the Boolean functions of its inputs as follows. Let $Q = [Q_0, Q_1, \ldots, Q_{(n_1n_2-1)-(m_1m_2+c_1c_2)}]$ be the vector of the encoder inputs. Then the elements of the C_F code matrix $\mathbf{V} = [v_{i,j}]$, i.e. the outputs of the encoder are determined as

$$v_{\Theta^{-1}(k)} = R^{\Theta}{}_k \tag{6}$$

where $R^{\Theta}{}_k$ is the value of the k-th coordinate of the vector R^{Θ} of length $n_1 n_2$, and $R^{\Theta} = Q \cdot \mathbf{G}^{\Theta}$

3.2. DECODER DESIGN

Given the matrix $\mathbf{H}^{\Theta}$ of the code C_F the decoder design can be performed as follows.

Algorithm 3. Decoder design.

1. Let $R^{\Theta} = [R^{\Theta}_0, R^{\Theta}_1, \ldots, R^{\Theta}_{n_1 n_2 - 1}]$ be the vector of the decoder inputs, where $R^{\Theta}_k = v_{\Theta^{-1}(k)}$, and $v_{\Theta^{-1}(k)}$ is an element of the C_F code matrix $\mathbf{V} = [v_{i,j}]$. Compute syndrome vector $s = [s_0, s_1, \ldots, s_{c_1 c_2 + m_1 m_2 - 1}]$ as the Boolean functions of the decoder inputs according to the formula

$$s = R^{\Theta} \cdot \mathbf{H}^{\Theta T} \tag{7}$$

where $\mathbf{H}^{\Theta T}$ denotes the transpose of the matrix $\mathbf{H}^{\Theta}$

2. Given the switching logic for the computation of the syndrome vector s design logic to determine the elements of the error vector $e^{\Theta} = [e^{\Theta}_0, e^{\Theta}_1, \ldots, e^{\Theta}_{n_1 n_2 - 1}]$ from s. This is performed by computing the values of s for each allowable in given code C_F value of e^{Θ}, followed by the design of the Boolean functions to compute e^{Θ} from s. The design of the Boolean function is required only for the coordinates e^{Θ}_i such that $c_1 c_2 + m_1 m_2 \le i < n_1 n_2$
3. Recover the original message vector $Q = [Q_0, Q_1, \ldots, Q_{(n_1 n_2 - 1) - (c_1 c_2 + m_1 m_2)}]$ as

$$Q_i = R^{\Theta}_{i + c_1 c_2 + m_1 m_2} + e^{\Theta}_{i + c_1 c_2 + m_1 m_2} \tag{8}$$

3.3. THE DESIGN EXAMPLE

Consider 2-D Fire cyclic code defined by the following parameters: $n_1 = 3$, $n_2 = 5$, $e_1 = 3$, $e_2 = 5$, $e = 15$, $c_1 = 3$, $c_2 = 1$, $m = 4$, $m_1 = 2$, $m_2 = 2$, $\lambda = 7$, $\tau_1 = 2$, $\tau_2 = 1$, and the irreducible polynomial $p(x) = x^4 + x + 1$ over GF(2). The polynomial p(x) is the primitive polynomial in $GF(2^4)$, thus it has a root α of order $2^4 - 1 = 15$. The root α is the primitive element of $GF(2^4)$ and the elements of $GF(2^4)$ can be represented as the linear combination of the elements of the polynomial basis $B = (\alpha^3, \alpha^2, \alpha^1, 1)$ over GF(2), e.g. $\alpha^{12} = \alpha^3 + \alpha^2 + \alpha + 1$. The structure of the field $GF(2^4)$ is given in Table 2 [5].

Define $\beta = \alpha^5$ and $\gamma = \alpha^3$. If m_1 and m_2 are chosen as v_1 and v_2, respectively, then the correcting capability of the code is $b_1 \times b_2 = 2 \times 1$ or less. The set of the check positions is the set $\Pi' = \Pi_1 + \Pi'_2$, where $\Pi'_1 = \{(0,0), (1,0), (2,0)\}$ and $\Pi'_2 = \{(0,1), (0,2), (1,1), (1,2)\}$. The check tensor $\mathbf{H} = [h_{i,j}]$ in the canonical form is computed as follows. Define the bijections Φ and Ψ as: $\Phi(0,0) = 0$, $\Phi(1,0) = 1$, $\Phi(2,0) = 2$; $\Psi(0,1) = 0$, $\Psi(0,2) = 1$, $\Psi(1,1) = 2$, $\Psi(1,2) = 3$. Find the $c_1 c_2 \times m_1 m_2$ matrix M that in our case has dimension 3×4, i.e.

$$\mathbf{M} = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

Compute the elements of the set $Z = \{\xi_{k,l} | (k,l) \in \Pi'_2\}$. The computations are performed in the field $GF(2^4)$ using the polynomial basis representation given in Table 2.

Table 2

The elements of GF(2^4)

i	$\alpha^i \bmod x^4+x+1$	vector notation
0	1	(0001)
1	α	(0010)
2	α^2	(0100)
3	α^3	(1000)
4	$\alpha+1$	(0011)
5	$\alpha^2+\alpha$	(0110)
6	$\alpha^3+\alpha^2$	(1100)
7	$\alpha^3+\alpha+1$	(1011)
8	α^2+1	(0101)
9	$\alpha^3+\alpha$	(1010)
10	$\alpha^2+\alpha+1$	(0111)
11	$\alpha^3+\alpha^2+\alpha$	(1110)
12	$\alpha^3+\alpha^2+\alpha+1$	(1111)
13	$\alpha^3+\alpha^2+1$	(1101)
14	α^3+1	(1001)

$$\xi_{0,1} = \beta^0\gamma^1 + \beta^{0 \bmod 3}\gamma^{1 \bmod 1} = \alpha^3+1$$
$$\xi_{0,2} = \beta^0\gamma^2 + \beta^{0 \bmod 3}\gamma^{2 \bmod 1} = \alpha^3+\alpha^2+1$$
$$\xi_{1,1} = \beta^1\gamma^1 + \beta^{1 \bmod 3}\gamma^{1 \bmod 1} = \alpha+1$$
$$\xi_{1,2} = \beta^1\gamma^2 + \beta^{1 \bmod 3}\gamma^{2 \bmod 1} = \alpha^3$$

Compute the vectors $h_{i,j}$. The vectors are computed only for the information positions because in the check positions the vectors $h_{i,j}$ are the unit vectors, e.g. for $h_{0,3}$

$$\xi_{0,3} = \beta^0\gamma^3 + \beta^{0 \bmod 3}\gamma^{3 \bmod 1} = \alpha^3+\alpha+1 = \xi_{1,1}+\xi_{1,2} \text{ thus}$$

$$z_{0,3} = \begin{vmatrix} 0 \\ 0 \\ 1 \\ 1 \end{vmatrix}$$

and $t_{0,3} = \mathbf{u}_{\Phi(0 \bmod 3, 3 \bmod 1)} + \mathbf{M} \cdot z_{0,3} = \mathbf{u}_0 + \mathbf{M} \cdot z_{0,3}$, i.e.

$$t_{0,3} = \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix} + \begin{vmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix} \cdot \begin{vmatrix} 0 \\ 0 \\ 1 \\ 1 \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix} + \begin{vmatrix} 0 \\ 0 \\ 0 \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix}$$

Finally,

$$h_{0,3} = \begin{vmatrix} t_{0,3} \\ z_{0,3} \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{vmatrix}$$

In the similar manner we obtain for, e.g. $h_{2,4}$

$\xi_{2,4} = \beta^2 \gamma^4 + \beta^{2 \bmod 3} \gamma^{4 \bmod 1} = \alpha^3 + \alpha^2 = \xi_{0,1} + \xi_{0,2} + \xi_{1,2}$ thus

$$z_{2,4} = \begin{vmatrix} 1 \\ 1 \\ 0 \\ 1 \end{vmatrix}$$

and $t_{2,4} = \tilde{\mathbf{u}}_{\Phi(2 \bmod 3, 4 \bmod 1)} + \mathbf{M} \cdot z_{2,4} = \mathbf{u}_2 + \mathbf{M} \cdot z_{2,4}$, i.e.

$$t_{2,4} = \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix} + \begin{vmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix} \cdot \begin{vmatrix} 1 \\ 1 \\ 0 \\ 1 \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix} + \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix} = \begin{vmatrix} 0 \\ 1 \\ 1 \end{vmatrix}$$

tion positions
$h_{0,3}$

Finally,

$$h_{2,4} = \begin{vmatrix} t_{2,4} \\ z_{2,4} \end{vmatrix} = \begin{vmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{vmatrix}$$

After computing all column vectors $h_{i,j}$ according to the procedure explained above define the bijection Θ fulfilling the conditions: $\Theta(i,j) = \Phi(i,j)$ if (i,j) belongs to Π_1, and $\Theta(i,j) = \Psi(i,j) + 3$ if (i,j) belongs to Π_2' as follows:

$\Theta(0,0) = 0$	$\Theta(0,1) = 3$	$\Theta(0,3) = 7$	$\Theta(2,1) = 11$
$\Theta(1,0) = 1$	$\Theta(0,2) = 4$	$\Theta(0,4) = 8$	$\Theta(2,2) = 12$
$\Theta(2,0) = 2$	$\Theta(1,1) = 5$	$\Theta(1,3) = 9$	$\Theta(2,3) = 13$
	$\Theta(1,2) = 6$	$\Theta(1,4) = 10$	$\Theta(2,4) = 14$

The image of the check tensor $\mathbf{H}$ under the bijection Θ is the 7×15 matrix $\mathbf{H}^{\Theta}$

$$\mathbf{H}^{\Theta} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \end{vmatrix}$$

Given the matrix $\mathbf{H}^{\Theta}$ compute the 8×15 generator matrix $\mathbf{G}^{\Theta}$

$$\mathbf{G}^{\Theta} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

In our example there are 8 information bits thus the vector of the input message is $Q = [Q_0, Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7]$. The Boolean functions for the output matrix $\mathbf{V} = [v_{i,j}]$ are obtained by multiplying the input vector Q by the matrix $\mathbf{G}^{\Theta}$ and then applying the inverse mapping Θ^{-1} to the vector $R^{\Theta} = Q \cdot \mathbf{G}^{\Theta}$. This gives for the check positions:

$$\begin{aligned} v_{0,0} &= Q_0 + Q_3 + Q_5 \\ v_{1,0} &= Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_7 \\ v_{2,0} &= Q_4 + Q_5 + Q_6 + Q_7 \\ v_{0,1} &= Q_2 + Q_3 + Q_4 + Q_6 + Q_7 \\ v_{0,2} &= Q_1 + Q_2 + Q_4 + Q_5 + Q_6 + Q_7 \\ v_{1,1} &= Q_0 + Q_1 + Q_2 + Q_3 + Q_4 \\ v_{1,2} &= Q_0 + Q_2 + Q_3 + Q_5 + Q_7 \end{aligned}$$

and for the information positions:

$$\begin{aligned} v_{0,3} &= Q_0 \\ v_{0,4} &= Q_1 \\ v_{1,3} &= Q_2 \\ v_{1,4} &= Q_3 \\ v_{2,1} &= Q_4 \\ v_{2,2} &= Q_5 \\ v_{2,3} &= Q_6 \\ v_{2,4} &= Q_7 \end{aligned}$$

The decoder design starts from the computation of the switching logic for the syndrome vector $s = [s_0, s_1, s_2, s_3, s_4, s_5, s_6]$. The transpose of the matrix $\mathbf{H}^{\Theta}$ is

$$\mathbf{H}^{\Theta T} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 \end{vmatrix}.$$

and the vector R^{Θ} of the decoder inputs is $R^{\Theta} = [R_0^{\Theta}, R_1^{\Theta}, \ldots, R_{14}^{\Theta}]$ where $R_k^{\Theta} = v_{\Theta}{}^{-1}{}_{(k)}$, e.g. $R_{10}^{\Theta} = v_{\Theta}{}^{-1}{}_{(10)} = v_{1,4}$. The Boolean functions for the computation of the syndrome vector are determined as the result of the multiplication $R^{\Theta} \cdot \mathbf{H}^{\Theta T}$, i.e.

$$s_0 = R_0^{\Theta} + R_7^{\Theta} + R_{10}^{\Theta} + R_{12}^{\Theta}$$
$$s_1 = R_1^{\Theta} + R_8^{\Theta} + R_9^{\Theta} + R_{10}^{\Theta} + R_{11}^{\Theta} + R_{12}^{\Theta} + R_{14}^{\Theta}$$
$$s_2 = R_2^{\Theta} + R_{11}^{\Theta} + R_{12}^{\Theta} + R_{13}^{\Theta} + R_{14}^{\Theta}$$
$$s_3 = R_3^{\Theta} + R_9^{\Theta} + R_{10}^{\Theta} + R_{11}^{\Theta} + R_{12}^{\Theta} + R_{13}^{\Theta} + R_{14}^{\Theta}$$
$$s_4 = R_4^{\Theta} + R_8^{\Theta} + R_9^{\Theta} + R_{11}^{\Theta} + R_{12}^{\Theta} + R_{13}^{\Theta} + R_{14}^{\Theta}$$
$$s_5 = R_5^{\Theta} + R_7^{\Theta} + R_8^{\Theta} + R_9^{\Theta} + R_{10}^{\Theta} + R_{11}^{\Theta}$$
$$s_6 = R_6^{\Theta} + R_7^{\Theta} + R_9^{\Theta} + R_{10}^{\Theta} + R_{12}^{\Theta} + R_{14}^{\Theta}$$

Multiplication of each allowable values of the error vector $e^{\Theta} = [e_0^{\Theta}, e_1^{\Theta}, \ldots, e_{14}^{\Theta}]$ by the matrix $\mathbf{H}^{\Theta T}$ results in the correspondence between the values of the error vector e^{Θ} and the syndrome vector $s = e^{\Theta} \cdot \mathbf{H}^{\Theta T}$, depicted in Table 3. This enables the computation of Boolean functions that determine the error vector e^{Θ} from s. It is enough to compute the Boolean functions only for e_7^{Θ}, e_8^{Θ}, e_9^{Θ}, e_{10}^{Θ}, e_{11}^{Θ}, e_{12}^{Θ}, e_{13}^{Θ}, and e_{14}^{Θ}, e.g.

$$e_{12}^{\Theta} = s_0 s_1 s_2 \bar{s}_3 s_4 \bar{s}_5 s_6 + s_0 s_1 s_2 \bar{s}_3 s_4 \bar{s}_5 \bar{s}_6 + s_0 s_1 s_2 \bar{s}_3 s_4 s_5 s_6.$$

Table 3

The correspondence between error and syndrome vectors

e_0^Θ	e_1^Θ	e_2^Θ	e_3^Θ	e_4^Θ	e_5^Θ	e_6^Θ	e_7^Θ	e_8^Θ	e_9^Θ	e_{10}^Θ	e_{11}^Θ	e_{12}^Θ	e_{13}^Θ	e_{14}^Θ	s_0	s_1	s_2	s_3	s_4	s_5	s_6
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	1	1	1
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0	1	1
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	0	1
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0
0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0	1	1	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	1
0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	1	0	1
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	1	0
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	1	1
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	1	0
0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	1	1
0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	1	1

To recover the bit Q_i of the original message it is enough to add modulo 2 bits $R^\Theta_{i+c_1c_2+m_1m_2}$ and $e^\Theta_{i+c_1c_2+m_1m_2}$, e.g., $Q_5 = R^\Theta_{12} + e^\Theta_{12} = v_{2,2} + e^\Theta_{12}$.

4. CONCLUSIONS

We developed the algorithm to support the search for the efficient 2-D Fire cyclic codes. The code analysis was performed by the synthesis of the several thousand cases. The codes useful for the practical applications were collected in Table 1. These codes are characterized by the efficient parameters of the error correction, e.g. there was found 2-D Fire code of size 33×31 that is able to correct the bursts of dimensions $2 \times 5 = 10$ bits or less. The check bits in this case occupy 10 per cent of the codeword matrix. Other

codes in Table 1 are also characterized by the decent parameters. The parameters of the codeword matrices allow to assembly the codeword matrix from two or four matrices of smaller size. The example of the logic design for the small codeword matrix showed that for the practical cases computer aided approach is required.

ACKNOWLEDGMENT

This work was supported by the research grant – Komitet Badań Naukowych 108/E-364/BW/2000.

REFERENCES

1. A. Gill: *Applied algebra for the computer sciences*. Prentice Hall, 1976.
2. T. L. Booth: *Sequential machines and automata theory*. Wiley, 1967.
3. H. Imai: *Two-dimensional Fire Codes*. IEEE Trans. on Information Theory, vol. 19, no. 6, pp. 796-806, Nov., 1973.
4. H. Imai: *A theory of two-dimensional cyclic codes*. Information and Control, vol. 34, no. 1, pp. 1-21, 1977.
5. A. Menezes, P. C. van Oorschot, S. A. Vanstone: *Handbook of applied cryptography*. CRC Press, 1996.
6. M. Rosing: *Implementing elliptic curve cryptography*. Manning Publications, 1998.
7. S. A. Vanstone, P. C. Van Oorschot: *An Introduction to error correcting codes with applications*. Kluwer Academic Publishers, 1989.

T. OSTROWSKI

CIAŁA SKOŃCZONE I KODY FIRE'A

Streszczenie

W artykule przedstawiono podstawy teorii ciał skończonych, algorytm wspomagający poszukiwanie korzystnych dwuwymiarowych kodów cyklicznych Fire'a oraz szczegółową metodykę projektowania równoległych układów kodera i dekodera. Bazując na teorii dwuwymiarowych kodów Fire'a opracowano algorytm wspierający dobór kodów o wysokiej sprawności. Badania komputerowe dla kodów o wielomianach generujących do stopnia 11 włącznie pokazały, że tylko niewielka część wszystkich dwuwymiarowych kodów Fire'a posiada znaczenie praktyczne. W pracy podano znalezione numerycznie przykłady tych kodów. Charakteryzują się one dobrymi parametrami korekcji. Między innymi znaleziony został kod, którego wymiary macierzy słowa kodowego wynoszą 33×31, posiadający zdolność korekcji wszystkich błędów obszarowych o wymiarach $2 \times 5 = 10$ bitów lub mniejszych, przy czym pozycje kontrolne zajmują zaledwie 10 procent powierzchni macierzy słowa kodowego. Składanie macierzy słowa kodowego z dwóch lub czterech macierzy o mniejszych wymiarach, umożliwia dobranie układu kodującego efektywnego z technicznego i ekonomicznego punktu widzenia.

Słowa kluczowe: Ciało skończone, dwuwymiarowy kod Fire'a, teoria liczb.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2001, 47, z. 2, ss. 161–183

Adaptacja częstotliwości próbkowania jako metoda kompandowania w modulacjach różnicowych

RYSZARD GOLAŃSKI

Katedra Elektroniki, Akademia Górniczo-Hutnicza

Al. Mickiewicza 30, 30-059 Kraków
e-mail: golanski@uci.agh.edu.pl

Otrzymano 2000.12.12
Autoryzowano 2001.02.23

Dokonano charakterystyki dotychczasowych osiągnięć w dziedzinie badań modulacji różnicowych z nierównomiernym próbkowaniem. Na bazie reguły maksymalizacji SQNR Abate'a wyznaczono zależności opisujące poszczególne parametry wyjściowe modulatora NSDM w funkcji zmian częstotliwości próbkowania. Udowodniono, że dla modulacji NSDM istnieje krok kwantyzacji Δ, przy którym, zmieniając częstotliwość próbkowania, można uzyskać minimum szumów kwantyzacyjnych zarówno przy minimalnym jak i maksymalnym poziomie mocy sygnału wejściowego. W oparciu o funkcję Lambert W wyznaczono graniczne wartości częstotliwości próbkowania, które zapewniają wymaganą dynamikę przy minimalnych szumach kwantyzacji. Przedstawiono dwa sposoby wyznaczania kroku kwantyzacji dla granicznych poziomów sygnału wejściowego. Dokonano analizy wzajemnych relacji pomiędzy parametrami wewnętrznymi modulatora NSDM i przedstawiono je w postaci graficznej. Zależności te pozwalają na ustalenie logicznego sposobu postępowania podczas doboru parametrów modulatorów z adaptacją częstotliwości próbkowania. Zaprezentowano kilka przykładowych diagramów wskazujących na poprawną kolejność i sposób doboru parametrów modulatorów różnicowych ze zmienną szybkością próbkowania.

Słowa kluczowe: modulacje różnicowe, próbkowanie nierównomierne, adaptacja częstotliwości próbkowania, maksymalizacja stosunku S/N, funkcja Lambert W, dynamika modulatora, dobór parametrów modulatora Δ.

1. CHARAKTERYSTYKA DOTYCHCZASOWYCH OSIĄGNIĘĆ W BADANIACH MODULACJI RÓŻNICOWYCH Z NIERÓWNOMIERNYM PRÓBKOWANIEM

Pomysły zastosowania zmiennej częstotliwości próbkowania były prezentowane w zasadzie od momentu, kiedy podczas badań nad modulacjami różnicowymi wykazano analitycznie (5), iż stosowanie zmian nachylenia krzywej aproksymującej (predykcji,

estymacji) sygnał wejściowy zwiększa zakres dynamiki (zapewnia kompandację), w którym sygnał ten może być przetwarzany.

Autorzy pierwszych aplikacji zmiennej częstotliwości próbkowania pragną wykazać, że zastosowanie tej techniki poprawia właściwości szumowe lub redukuje szybkość transmisji natomiast stopień skomplikowania rozwiązania sprzętowego jest umiarkowanie mały. Spośród badaczy, którzy zajmowali się systemami pracującymi ze zmienną szybkością próbkowania (kodowania) ze względu na ich osiągnięcia należy wymienić kilku najważniejszych zespołów.

W 1974 r. **T.A. Hawkes, P.A. Simonpieri** w artykule *„Signal Coding Using Asynchronous Delta Modulation"* [1] uzasadniają stosowanie zmian częstotliwości próbkowania transmisją (przechowywaniem) mniejszej liczby próbek w okresach niewielkich zmian sygnału wejściowego a więc poprawą współczynnika kompresji bitowej. Omawiają założenia algorytmu działania prostego systemu ADM ze zmienną częstotliwością próbkowania i prezentują wyniki zastosowania tego systemu do przetwarzania obrazu. Algorytm używany do zmiany okresu próbkowania jest identyczny w części nadawczej i odbiorczej systemu i dlatego możliwa jest w części odbiorczej rekonstrukcja sygnału kodowanego bez konieczności przesyłania dodatkowej informacji o długości okresu próbkowania.

Poniżej przedstawiono zaproponowany przez Hawkes'a [1] algorytm zmian okresu próbkowania[1]:

Opisuje go następującymi zależnościami:

$$y(t_{i-1}) = \sum_{k=1}^{i-1} SGN_k \Delta + Y(t_0)$$

gdzie $SGN_i = \operatorname{sgn}[x(t_i) - y(t_{i-1})]$

oznaczenia:

x(t) — sygnał kodowany (wejściowy)

y(t) — sygnał aproksymujący (sygnał wyjściowy predyktora)

$Y(t_0)$ — wartość sygnału aproksymującego w chwili t_0

Δ — krok kwantyzacji

W systemie modulacji ze zmianami okresu próbkowania bieżący okres próbkowania jest funkcją poprzedniego okresu i poprzednich wartości SGN_i, co można zapisać:

$$t_{i+1} - t_i = (t_i - t_{i-1}) F(SGN_i, SGN_{i-1}, ...)$$

gdy pamięć układu modulacji ograniczy się do wartości bieżącej i poprzedniej, zależność powyższa przybiera postać:

$$t_{i+1} - t_i = (t_i - t_{i-1}) F(SGN_i, SGN_{i-1})$$

Zaproponowano koincydencyjną logikę adaptacji (funkcję F) wzorowaną na opracowaniach Jayanta:

$$\left.\begin{aligned} F(+1, +1) = F(-1, -1) = A < 1 \\ F(+1, -1) = F(-1, +1) = B > 1 \end{aligned}\right\}$$

[1] Na identycznej zasadzie dokonywano adaptacji kroku kwantyzacji

Z praktycznych względów zarówno minimalny jak i maksymalny okres próbkowania powinien być ograniczony:

T_{min} – ze względu na maksymalne możliwe nachylenie sygnału wejściowego

T_{max} – okres próbkowania nie może być zbyt długi żeby modulator mógł dobrze kodować szybkie zmiany sygnału wejściowego pojawiające się po okresie zmian niewielkich. Krócej przyjęcie zbyt długiego okresu T_{max} powoduje wzrost tzw. czasu reakcji (czasu martwego) i może prowadzić do całkowitego pomijania kodowania krótkich i szybkich zmian w sygnale wejściowym (znikają one w sygnale odtworzonym).

Przyjęto A = .25, B = 4, T_{min} = .25T, T_{max} = T.

Z zaproponowanego algorytmu wynika, że modulator T.A. Hawkes'a, P.A. Simonpieri'ego pracuje tylko z dwoma różnymi okresami próbkowania. W pracy porównano dwie pary obrazów („boy's portait" i „air view"), z których jeden wydrukowano nie stosując żadnego kodowania a drugi odtworzono po wcześniejszym zakodowaniu go metodą proponowaną przez Autorów [1]. Jak stwierdzają 3 krotnej kompresji liczby bitów niezbędnych do zakodowania tych obrazów, nie towarzyszy zauważalne pogorszenie jakości obserwowanych obrazów.

Na przełomie lat 70 i 80-tych **Dubnowski i Crochiere** [2] badali korzyści ze stosowania zmiennej szybkości bitowej podczas kodowania sygnału mowy. Jako pierwsi wykazali analitycznie poprawę właściwości szumowych, ale jedynie podczas modulacji przypadkowych drgań niestacjonarnych. Jako praktyczny przykład przedstawili propozycję systemu ADPCM ze zmienną bitową szybkością kodowania. Z kolei **LoCiero i Prezas** [3] badali korzyści z zastosowania zmiennego próbkowania w modulacjach ADM.

Szeroko udokumentowane rozważania dotyczące aplikacji zmiennego próbkowania w modulacji różnicowej można znaleźć w artykule **Ch.K. UN'a i D.H. CHO**:, *„Hybrid Delta Modulation with Variable-Rate Sampling"* [4] opublikowanego w 1982 r. Autorzy przypominają, że w grupie modulacji różnicowych ADM (CFDM, CVSD, HIDM [5] i HCDM) jednobitowych ze stałą częstotliwością próbkowania, najlepsze właściwości wykazuje modulacja hybrydowa HCDM będąca zręcznym połączeniem modulacji CFDM i CVSD w związku, z czym prezentuje zalety obu tych rozwiązań [5]. Celem poprawy właściwości szumowych lub zmniejszenia średniej szybkości bitowej tej modulacji Autorzy proponują poza adaptacją kroku Δ zastosowanie w niej zmiennej częstotliwości próbkowania.

Algorytm zmian kroku kwantyzacji Δ nie zmienia się w stosunku do tego stosowanego w HCDM [5]. Natomiast algorytm zmian częstotliwości próbkowania można scharakteryzować następująco:

- Przyjęto, że będą stosowane cztery różne częstotliwości próbkowania f_p
- f_p zmienia się zgodnie ze zmianami uśrednionego (scałkowanego) nachylenia stromości sygnału wejściowego oznaczanego jako E_i. Do ustalenia tej stromości wykorzystuje się kompandor sylabiczny modulatora HCDM. Ustalono przy tym 3 poziomy wartości E_i odpowiadające 3-em wartościom uśrednionego nachylenia stromości i gdy:

$E_i < E_1$ to: $f_{s1} = 6$ kHz
$E_1 \leq E_i < E_2$ to: $f_{s2} = 12$ kHz
$E_2 \leq E_i < E_3$ to: $f_{s3} = 16$ kHz
$E_i > E_3$ to: $f_{s4} = 24$ kHz

Autorzy badali również system z trzema f_s (8, 16, 24 kHz). Jak widać wszystkie wybrane f_s stanowią całkowitą podwielokrotność 48 kHz.

- Przyjęto, że szybkość nadawania z modulatora jest stała a $F_p = 16$ kHz. Oznacza to konieczność użycia odpowiedniego bufora transmisyjnego, który wyrównuje strumień bitów. Problemem był dobór długości bufora celem uniknięcia jego przeładowania lub całkowitego opróżnienia. Dla obliczenia długości bufora przyjęto klasyczne warunki pracy polegające na tym, że:
  - gdy bufor jest pełny, częstotliwość próbkowania kodera musi być mniejsza lub równa aktualnej częstotliwości transmitowanego sygnału
  - gdy bufor jest pusty częstotliwość próbkowania powinna być większa lub równa częstotliwości sygnału transmitowanego na zewnątrz

W proponowanym systemie dla uniknięcia przeładowania lub całkowitego opustoszenia bufora przyjęto następujące warunki:

- gdy $l_i \geq (9/10)B$ to $f_{pi} \leq F_p$
- gdy $l_i < (1/10)B$ to $f_{pi} > F_p$

gdzie:

B — długość bufora, F_p — częstotliwość transmisyjna (tut. 16 kHz), l_i-liczba bitów w buforze transmisyjnym (wyznaczono L = 3kB)

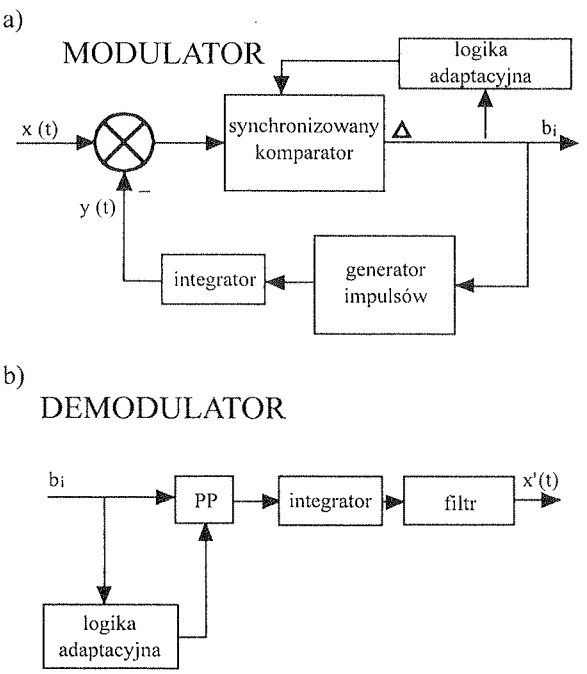
Przedstawione powyżej warunki prawidłowej pracy bufora mają priorytet nad warunkami doboru częstotliwości próbkowania jeśliby taki konflikt wystąpił.

W pracy [5] przedstawiono algorytm działania proponowanego systemu kodowania i wyniki symulacji komputerowych. Jak stwierdza Un na podstawie symulacyjnych badań porównawczych[2] użycie zmiennej częstotliwości próbkowania prowadzi do poprawy maksymalnego stosunku S/N o 3 do 4dB względem modulacji HCDM i CVSD a ponadto znacznie poszerza zakres dynamiki (o ok.35 dB) względem modulacji CVSD. Zaobserwowano również ok. 1-2 dB zysku przy stosowaniu algorytmu z 4 częstotliwościami w stosunku do algorytmu z 3 częstotliwościami próbkującymi. Wpływ długości bufora i jego sterowania jest również omówiony przez Autorów. W artykule [5] Autorzy wykazują, również, że wielkość zysku z użycia zmiennej f_p zależy od wielkości zmian mocy średniej procesu wejściowego tzn. od stopnia jego niestacjonarności. Dla czysto stacjonarnego procesu wejściowego nie ma, więc teoretycznie żadnego zysku ze zastosowania zmiennej częstotliwości próbkowania, jeśli chodzi o właściwości szumowe [5].

W ostatnich latach pojawiło się kilka prac dotyczących modulacji z nierównomiernym próbkowaniem zwanych NSDM (Nonuniform Sampling Delta Modulation) **zespołu**

[2] Porównania dokonano dla sygnału wejściowego składającego się z próbek mowy żeńskiej i męskiej o łącznej długości ok. 10 sek. Użyto 8 biegunowego dolnopasmowego filtru Butterwortha celem ograniczenia pasma przetwarzanego sygnału do wartości 3,4 kHz i do filtracji sygnału zdekodowanego

pod kierownictwem Y.S. Zhu. Jak wskazują wyniki prezentowane przez Zhu i in. [6, 7] zastosowanie nierównomiernego próbkowania w modulacjach różnicowych (NSDM i ANSDM) znacznie poprawia efektywność kodowania (zmniejsza średnią przepływność bitową- $P_{b\,śr}$ zachowując wystarczającą jakość przetwarzania lub poprawia jakość zachowując średnią przepływność bitową). Dotyczy to w szczególności modulacji ANSDM (Adaptive Nonuniform Sampling Delta Modulation). Zasady kodowania i algorytmy modulacji NSDM i ANSDM zostały przedstawione w opracowaniach [6, 7, 8, 9]. Najistotniejszym jest, iż w modulacjach ANSDM każdy bit niesie informację dla predyktora, nie tylko o znaku czy wielkości przyrostu sygnału, lecz także o czasie, po jakim nastąpi kolejne próbkowanie.



Rys.1. Schemat funkcjonalny modulatora (a) i demodulatora (b) modulacji NSDM wg propozycji Zhu

2. ZALEŻNOŚCI ANALITYCZNE OPISUJĄCE MODULATORY NSDM

Celem badań przedstawionych w niniejszej pracy była ocena podstawowych właściwości modulacji różnicowej wykorzystującej zmienną częstotliwość próbkowania jako metodę poprawy jakości przetwarzania np. zakresu dynamiki DR modulatora. W kolejnych rozdziałach wyznaczano zależności opisujące wybrane parametry wyjściowe modulatora NSDM w funkcji zmian częstotliwości próbkowania. Przedstawiono sposoby doboru kroku kwantyzacji w modulacji NSDM. Dokonano analizy wzajemnych relacji pomiędzy parametrami modulatora NSDM i przedstawiono je w postaci graficznej.

Już pierwsze badania [10] modulacji NSDM wskazały na duże znaczenie właściwego doboru jej parametrów wewnętrznych, zarówno na wartość SQNR jak i średniej przepływności bitowej $P_{b\ śr}$. Dlatego w pracy zaprezentowano diagramy wskazujące na poprawną kolejność i sposób doboru parametrów modulatorów różnicowych NSDM.

W pracy korzystano z opracowań opisujących badania analityczne i symulacyjne modulatorów różnicowych ze stałą częstotliwością próbkowania [11,12, 13].

Jednak w odróżnieniu od tych opracowań dla metody NSDM należało wyznaczyć zależności wyrażające poszczególne parametry wyjściowe w funkcji zmian częstotliwości próbkowania f_p (przy stałym kroku Δ) a nie w funkcji kroku kwantyzacji Δ przy stałej f_p. Dlatego tylko niektóre ze sprawdzonych wcześniej rozwiązań mogą być użyteczne przy analizie właściwości metody NSDM.

W pracy Abate'a [11] w oparciu o symulacje komputerowe wykazano, że minimum mocy szumów kwantyzacyjnych $(N_g + N_o)$ w modulacji delta prosta (LDM), dla modelu losowego sygnału mowy (szczegółowo opisanego w pracy O'Neal'a [13] uzyskuje się, gdy:

$$\alpha \cong ln2B \tag{1}$$

gdzie:

$$\alpha = \frac{x'_o}{(dx/dt)_{rms}} = (\Delta \cdot f_p)/\sqrt{D} \quad \text{— względny współczynnik przeciążenia stromości} \tag{1'}$$

określa stosunek szybkości zmian sygnału aproksymującego do szybkości zmian sygnału wejściowego,

$B = f_p/2f_o$ — współczynnik nadmiarowości próbkowania, Δ — krok kwantyzacji [V],

f_o — maksymalna częstotliwość w widmie sygnału wejściowego (częstotliwość odcięcia),

f_p — częstotliwość próbkowania, D — wariancja pochodnej x'(t) sygnału wejściowego,

$\sqrt{D} = \chi\sqrt{S}$ — odchylenie standardowe pochodnej x'(t) sygnału wejściowego,

χ — stała określająca rodzaj sygnału wejściowego,

S — wariancja (średnia moc) sygnału wejściowego,

$\sqrt{S}$ — odchylenie standardowe sygnału wejściowego.

Analizując dane zebrane w Tabeli 1 można stwierdzić, że minimum mocy szumów kwantyzacyjnych zależy przede wszystkim od częstotliwości próbkowania a ponadto od rodzaju i wielkości sygnału wejściowego x(t). Aby to minimum uzyskać, należy tak dobrać krok kwantyzacji Δ aby spełniony był warunek Abate'a (1).

2.1. CHARAKTERYSTYKA ZAGADNIEŃ ROZWIĄZYWANYCH W NINIEJSZYM OPRACOWANIU

Z warunku Abate'a (1) wynika również, że każda zmiana f_p czy poziomu mocy S sygnału wejściowego wymaga zmiany wartości kroku Δ dla uzyskania minimum

szumów kwantyzacyjnych. Widać, więc, że żaden rodzaj modulacji różnicowej z próbkowaniem równomiernym, w którym krok kwantyzacji jest stały nie zapewnia minimalizacji szumów kwantyzacji (oczywiście biorąc pod uwagę, że sygnał przetwarzany może posiadać różne poziomy mocy S). Jeśli jednak zmianie ulega zarówno f_p jak i S oraz kierunek tych zmian jest właściwy to na podstawie równania (1) może okazać się, że ten sam krok pozwoli zapewnić niezmienną wartość minimum szumów kwantyzacji dla różnych par f_p i S. Ujmując zagadnienie czysto matematycznie należy sprawdzić czy istnieje rzeczywiste rozwiązanie układu równań (2) względem kroku Δ dla krańcowych wartości mocy S sygnału wejściowego x(t):

$$\Delta \cdot f_{p\max}/\chi\sqrt{S_{\max}} = \ln(f_{p\ max}/f_o) \Rightarrow \Delta \cdot B_{\max}/2f_0\chi\sqrt{S_{\max}} = \ln(2B_{\max}) \qquad (2')$$

$$\Delta \cdot f_{p\min}/\chi\sqrt{S_{\min}} = \ln(f_{p\ min}/f_o) \Rightarrow \Delta \cdot B_{\min}/2f_0\chi\sqrt{S_{\min}} = \ln(2B_{\min}) \qquad (2'')$$

gdzie $B_{\max} = f_{p\ max}/2f_o$ — nadmiarowość częstotliwości maksymalnej próbkowania $f_{p\ max}$

$B_{\min} = f_{p\ min}/2f_o$ — nadmiarowość częstotliwości minimalnej próbkowania $f_{p\ min}$

Powstają pytania:

- czy istnieje taki krok kwantyzacji Δ_{opt}, który spełnia równania (2' i 2'')?. Czyli inaczej czy dla modulacji NSDM istnieje krok kwantyzacji Δ, przy którym, zmieniając częstotliwość próbkowania f_p, można uzyskać minimum szumów kwantyzacyjnych zarówno przy minimalnym jak i maksymalnym poziomie mocy S sygnału wejściowego x(t). Jeśli bowiem krok taki zostanie znaleziony to można będzie z kolei wykazać, że jeśli istnieje on dla krańcowych wartości mocy ($S_{\max}$, $S_{\min}$) to również będzie on optymalny dla wszystkich mocy pośrednich, jeśli tylko algorytm zmian częstotliwości próbkowania pozwoli je „dostrajać" do każdego z tych pośrednich poziomów mocy.

- jeśli istnieje to, jakie są ograniczenia, co do jego wielkości?

Układ równań (2) i (2') można rozwiązać względem kilku zmiennych. W przypadku modulacji NSDM interesuje nas rozwiązanie względem granicznych częstotliwości próbkowania $f_{p\ max}$ (B_{max}) lub $f_{p\ min}$ (B_{min}), które zapewnią „optymalne" przetwarzanie dla wymaganej dynamiki.

Jako parametry zadane przyjmuje się:

— dynamikę sygnału przetwarzanego DR lub $\sqrt{S_{\max}}$ i $\sqrt{S_{\min}}$,

— częstotliwość maksymalną próbkowania $f_{p\ max}$ i/lub częstotliwość minimalną $f_{p\ min}$,

— maksymalną częstotliwość w widmie sygnału wejściowego (częstotliwość odcięcia) f_o,

Dalej przedstawiono rozwiązanie układu równań 2 równocześnie dla dwóch przypadków:

Tabela 1

Zależności opisujące szumy kwantyzacyjne w modulacji LDM[3] oraz podstawowe parametry wejściowego procesu losowego o widmie scałkowanym

Parametr (funkcja)	Wartość parametru dla sygnału losowego o widmie scałkowanym
$F(\omega)$ — jednostronna moc widmowa, gęstość widma [V²/Hz] sygnału wejściowego	$\frac{S}{\omega_3 arctan\left(\frac{\omega_m}{\omega_3}\right)}\left[\frac{1}{1+\left(\frac{\omega}{\omega_3}\right)^2}\right]$
$D=\int_0^{\omega_m}\omega^2 F(\omega)\,d\omega$ — średnia moc (wariancja) pochodnej sygnału wejściowego [V²/s²]	$\omega_m^2\left[\frac{\omega_3/\omega_m}{arctan\left(\frac{\omega_m}{\omega_3}\right)}-\left(\frac{\omega_3}{\omega_m}\right)^2\right]S=\varkappa\cdot S$
$\alpha\equiv\frac{kf_s}{\sqrt{D}}$ — względny współczynnik przeciążenia stromości	$\frac{kB}{\pi\sqrt{S}}\frac{1}{\sqrt{\frac{\omega_3/\omega_m}{arctan\left(\frac{\omega_m}{\omega_3}\right)}-\left(\frac{\omega_3}{\omega_m}\right)^2}}$
$N_G=\frac{\pi^2}{6}\left(\frac{D}{\omega_m^2}\right)\frac{\alpha^2}{B^3}$ dla $\alpha<8$ (NG-szumy granulacyjne)	$\frac{\pi^2\alpha^2}{48B^3}\frac{D}{\omega_m^2}$
$N_G=\frac{\pi^2}{48}\left(\frac{D}{\omega_m^2}\right)\frac{\alpha^3}{B^3}$ dla $\alpha>8$ (NG-szumy granulacyjne)	$\frac{\pi^2\alpha^3}{48B^3}\frac{D}{\omega_m^2}$
$N_0=\frac{8\pi^2}{27}\left(\frac{D}{\omega_m^2}\right)e^{-3\alpha}(3\alpha+1)$ N_o — szumy przeciążenia stromości	$\frac{8\pi^2}{27}\frac{D}{\omega_m^2}e^{-3\alpha}(3\alpha+1)$
Minimum $N_Q=\frac{\pi^2}{6}\left(\frac{D}{\omega_m^2}\right)\left[\frac{(\ln B)^2+2{,}06\ln B+1{,}17}{B^3}\right]$ N_Q — szumy kwantyzacyjne	$\frac{\pi^2}{6}\frac{D}{\omega_m^2}\left[\frac{(\ln B)^2+2{,}06\ln B+1{,}17}{B^3}\right]$
Dla losowego sygnału telewizyjnego: $\frac{\omega_3}{\omega_m}=0.011$ i $\Delta_{opt}=0.026\frac{\ln(2B)}{B}\sqrt{S}$	
Dla losowego sygnału mowy: $\frac{\omega_3}{\omega_m}=0.23$ i $\Delta_{opt}=0.037\frac{\ln(2B)}{B}\sqrt{S}$	

a) poszukiwana jest częstotliwość minimalna $f_{p\,min}$, czyli (B_{min}), która zapewnia spełnienie układu równań (2) przy tym samym kroku kwantyzacji Δ a dana jest częstotliwość maksymalna próbkowania $f_{p\,max}$, czyli B_{max} oraz $\sqrt{S_{max}}$, $\sqrt{S_{min}}$

[3] sporządzono na podstawie pracy Abate [11]

b) poszukuje się $f_{p\,max}$ (B_{max}), która zapewnia spełnienie układu równań (2) i (2′) przy tym samym kroku kwantyzacji Δ i zadanych $f_{p\,min}$, czyli B_{min} oraz $\sqrt{S_{max}}$, $\sqrt{S_{min}}$.

2.2. WYZNACZANIE GRANICZNYCH CZĘSTOTLIWOŚCI PRÓBKOWANIA

Z układu równań (2) znajdujemy warunek, który musi być spełniony, aby zarówno przy minimalnym jak i maksymalnym poziomie mocy sygnału wejściowego S ten sam krok kwantyzacyjny Δ zapewniał uzyskanie minimum szumów kwantyzacyjnych czyli spełniał warunek Abate'a (1). Czyli poszukiwane są:

$$B_{min} = ?\ (\text{przy ustalonym } B_{max}) \quad \text{lub} \quad B_{max} = ?\ (\text{przy ustalonym } B_{min})$$

Na podstawie 2 i 2′ otrzymujemy kolejno:

$$\frac{1}{B_{min}} B_{max} \frac{\sqrt{S_{min}}}{\sqrt{S_{max}}} = \frac{1}{\ln(2B_{min})} \ln(2B_{max})$$

i dalej:

$$\frac{\sqrt{S_{max}}}{\sqrt{S_{min}}} \frac{\ln(2B_{max})}{B_{max}} B_{min} = \ln(2B_{min}) \quad \text{lub} \quad \frac{\sqrt{S_{min}}}{\sqrt{S_{max}}} \frac{\ln(2B_{min})}{B_{min}} B_{max} = \ln(2B_{max})$$

$$bB_{min} = \ln(2B_{min}) \quad \text{lub} \quad cB_{max} = \ln(2B_{max}) \qquad (3)\ \text{i}\ (3')$$

Pojawia się, więc konieczność rozwiązania równania transcendentnego o ogólnej postaci:

$$ln(2x) = \alpha x \qquad (3'')$$

Rozwiązaniem równania (3″) jest funkcja W Lamberta [14] znana również jako funkcja omega i oznaczana W(x) lub W(k, x). Funkcja W Lamberta spełnia warunek:

$$W(x) * \exp(W(x)) = x \qquad (4)$$

Podobnie jak równanie y exp(y) = x posiada nieskończoną liczbę rozwiązań y dla każdej niezerowej wartości x tak i funkcja Lambert W posiada nieskończoną liczbę zbiorów rozwiązań (gałęzi). Gałąź podstawowa (k = 0) oznaczana jest jako W(x) a wszystkie pozostałe jako W(k,x).

Ponieważ w rozpatrywanym zagadnieniu (3 i 3′) musi zachodzić $B \geq 1$ (prawo Shannona-Kotielnikowa) okazuje się, że rozwiązaniem równania (3″) jest gałąź funkcji omega oznaczana W(–1, x). Funkcja ta posiada rozwiązania rzeczywiste dla $x \in [0, 1/e]$

gdzie:
e-podstawa logarytmu naturalnego
Ostatecznie, więc:

$$B_{min} = -\frac{W(-1,\ -b/2)}{b} \quad \text{oraz} \quad B_{max} = -\frac{W(-1,\ -c/2)}{c} \qquad (5) \text{ i } (5')$$

gdzie:

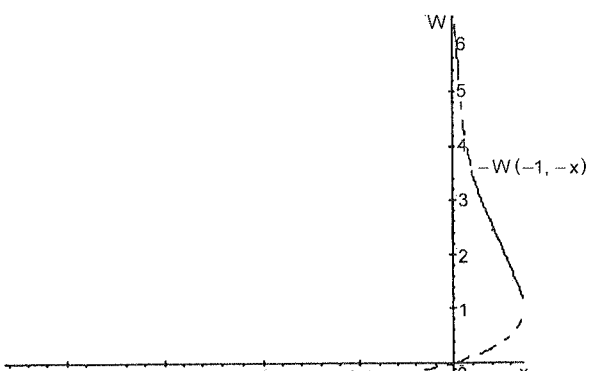
$$b = \frac{\ln(2B_{max})}{B_{max}} \frac{\sqrt{S_{max}}}{\sqrt{S_{min}}} = \frac{\ln(2B_{max})}{B_{max}} DR \text{ oraz } c = \frac{\ln(2B_{min})}{B_{min}} \frac{\sqrt{S_{min}}}{\sqrt{S_{max}}} = \frac{\ln(2B_{min})}{DR^* B_{min}}$$

(6) i (6')

$$\frac{\sqrt{S_{max}}}{\sqrt{S_{min}}} \stackrel{def}{=} DR$$

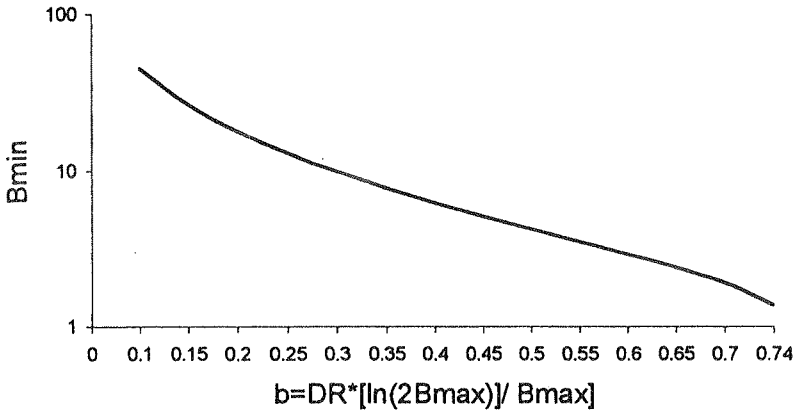
Wartość kroku kwantyzacyjnego Δ oblicza się z wzoru 2' lub 2'' znając $f_{p\,min}$ lub, $f_{p\,max}$ z (5) lub (5').

Z przedstawionych rozwiązań wynika, że wyznaczenie częstotliwości $f_{p\,min}$ lub $f_{p\,max}$ sprowadza się do obliczania wartości funkcji Lambert $-W(-1, -b/2)$. Funkcja ta posiada rozwiązania rzeczywiste dla $b/2 \vee c/2 \in [0,\ 1/e]$. Została ona w postaci ogólnej przedstawiona na rys.2.



Rys. 2. Przebieg funkcji [-W(-x)] — linia przerywana i funkcji [-W(-1, -x)] — linia ciągła

Rys. 3. przedstawia graficzną postać rozwiązania równania (5) w funkcji parametru b. Równanie to ma rozwiązania rzeczywiste dla $0 < b < 2/e$.



Rys. 3. Nadmiarowość próbkowania B_{min} (wymagana, aby zapewnić maksymalizację SQNR w całym zakresie dynamiki DR) w funkcji parametru b

Ponieważ funkcja Lambert W(–1, –x) określona jest w granicach od 0 do 1/e wobec tego:

$$b < 2/e \quad \text{lub} \quad c < 2/e \qquad (7) \text{ i } (7')$$

rozwijając b i c uzyskuje się następujące zależności:

$$\text{gdy DR}\frac{\ln(2B_{max})}{B_{max}} < \frac{2}{e} \text{ to } B_{min} > \frac{e}{2} \text{ lub gdy } \frac{1}{DR}\frac{\ln(2B_{min})}{B_{min}} < \frac{2}{e} \text{ to } B_{max} > \frac{e}{2} \qquad (8) \text{ i } (8')$$

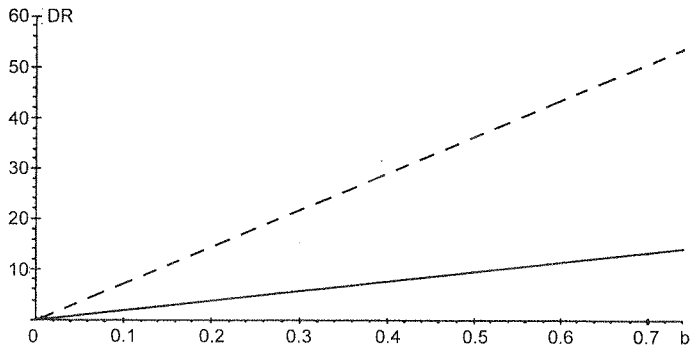
Wzór (8) pozwala wyznaczać teoretyczną maksymalną wartość zakresu dynamiki DR (w funkcji B_{max}), w którym modulator NSDM może przetwarzać z ustaloną jakością (SQNR ≈ const). Ponieważ w granicy nierówności (8) funkcja W(–1, b/2) przyjmuje wartość 1 a B_{min} wartość *e/2* to spełnione jest prawo Shannona-Kotielnikowa.

Znajomość zakresu zmiennej niezależnej b (6) pozwala wyznaczyć szczegółowe zależności pomiędzy parametrami wchodzącymi w jej skład oraz ich wartości dopuszczalne. Dalej przedstawione zostaną wykresy sporządzone na podstawie zależności (1÷8) a pozwalające scharakteryzować możliwości i podstawowe właściwości modulatorów NSDM.

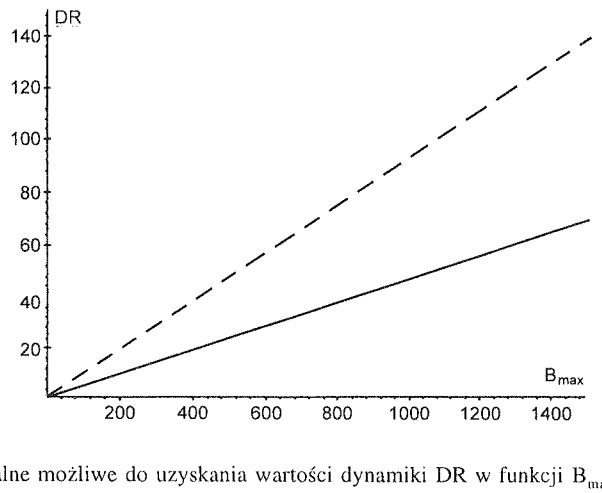
Dla przykładu wykres na rys. 4 wskazuje na liniową zależność dynamiki DR w funkcji parametru b przy B_{max} = const, co wynika z równania (6).

Nachylenie krzywych na rys. 4 potwierdza intuicyjnie wyczuwalną zależność, że większe wartości dynamiki DR wymagają zwiększenia maksymalnej częstotliwości próbkowania tu reprezentowanej przez B_{max}.

Na podstawie nierówności (8) można z kolei sporządzić wykres DR = f(B_{max}) — rys. 5.



Rys.4. Maksymalna możliwa do uzyskania dynamika DR przy B_{max} = const (linia przerywana B_{max} = 500, linia ciągła B_{max} = 100) w funkcji parametru b



Rys. 5. Maksymalne możliwe do uzyskania wartości dynamiki DR w funkcji B_{max} (Bmin = const). Linia przerywana dla b = 2/e, linia ciągła dla b = 1/e

Nierówność (8) i rys.5 potwierdzają, że optymalną wartością parametru *b* jest jego wartość graniczna $b_{opt} = 2/e$ (nie da się uzyskać lepszej relacji między DR a B_{max} niż wskazuje nachylenie linii przerywanej, b = 2/e na rys. 5). Czyli przy zadanym DR nie jest uzasadnione stosowanie większych częstotliwości granicznych, B_{max} niż wynika to z nierówności (8).

2.3. DOBÓR WIELKOŚCI KROKU KWANTYZACYJNEGO Δ

Dla modulacji LDM po przekształceniu zależności (1′) dla losowego sygnału mowy [11, 13] można wykazać, że optymalny (w tym przypadku taki, przy którym szumy osiągają minimum) krok Δ wynosi:

$$\Delta_{mowy} \cong 0.37 \frac{\ln 2B}{B} \sqrt{S} \tag{9}$$

Znak przybliżenia we wzorze (9) wynika z zaokrąglenia wartości stałej χ określającej rodzaj sygnału wejściowego. Stała χ jest w przybliżeniu równa 0,37 gdy sygnał wejściowy poddawany modulacji to proces losowy o widmie szumu białego scałkowanego i spełniającego zależność: $\frac{\omega_{3dB}}{\omega} = 0.23$.

Wzór Abate'a (1′) opisujący regułę minimalizacji szumów kwantyzacji zawiera m.in. iloczyn kroku i częstotliwości próbkowania. Minimalizacja szumów kwantyzacyjnych może więc być dokonana dla różnych kroków Δ przy różnych częstotliwościach próbkowania. W takiej sytuacji wybór kroku optymalnego zależy od innych czynników. Przy konieczności przetwarzania przez modulator sygnałów o różnych poziomach mocy czynnikiem takim może być właśnie minimalna wartość poziomu mocy lub żądana wartość dynamiki.

W efekcie tego rozumowania przyjęto, że dla modulacji NSDM krok kwantyzacji Δ powinno się wyznaczać przy mocy minimalnej S_{min}, jeśli tylko potrafimy ją wyznaczyć lub mamy daną (np. wzory 10, 10′, 10″). Jeśli bowiem zapewni się żądaną jakość (SQNR) dla mocy o poziomie minimalnym już przy minimalnej częstotliwości to spełnienie zależności (1÷8) pozwala przy tym samym kroku Δ uzyskać minimum szumów kwantyzacyjnych aż do założonego maksymalnego poziomu mocy a to dzięki wzrostowi częstotliwości próbkowania. Górny poziom mocy, która może być przetwarzana z przyjętą jakością będzie oczywiście zależał od maksymalnej częstotliwości próbkowania f_{pmax} a w rozwiązaniach praktycznych również od parametrów zapewniających skuteczne zmiany tej częstotliwości. Skuteczność tych zmian zależy od przyjętego algorytmu adaptacji.

Dla modulacji NSDM, w której znana jest wartość minimalnego poziomu mocy sygnału wejściowego S_{min}, krok kwantyzacji Δ_{mowy} może być obliczony na podstawie zależności (2″):

$$\Delta_{mowy} \cong 0.37 \frac{\ln 2B_{min}}{B_{min}} \sqrt{S_{min}} \tag{10}$$

lub zależności (2′):

$$\Delta_{wmin} = \Delta_{mowy} / \sqrt{S_{min}} \cong 0.37b \tag{10'}$$

i po przyjęciu[4]: b = 2/e i w konsekwencji B_{min} = e/2 uzyskuje się:

[4] $B_{min} \stackrel{def}{=} f_{pmin}/2f_o = -\frac{W(-1, -b/2)}{b} = e/2 \cong 1.36$ przy b = 2/e (maksymalne wartości DR uzyskuje się utrzymując $b = \frac{\ln(2B_{max})}{B_{max}} \frac{\sqrt{S_{max}}}{\sqrt{S_{min}}} = \frac{\ln(2B_{max})}{B_{max}} DR = 2/e$)

$$\Delta_{wmin} = \Delta/\sqrt{S_{min}} \cong 0.27 \text{ dla } b = 2/e \tag{11}$$

Czyli w przypadku, gdy znana jest wartość $\sqrt{S_{min}}$ i utrzymywane jest $b = 2/e$ krok kwantyzacji Δ_{mowy} należy wyznaczać z wzoru (11).

Gdy zadana jest żądana dynamika DR niezbędną (utrzymywane jest $b = 2/e$) nadmiarowość próbkowania B_{max} wyznacza się na podstawie zależności (6), która po przekształceniu przyjmuje postać równania (12):

$$\ln(2B_{max}) = \frac{2}{eDR} B_{max} \tag{12}$$

Oznaczając:

$$\frac{2}{eDR} \cong 0.735/DR = d \tag{13}$$

rozwiązaniem równania (12) podobnie jak poprzednio jest funkcja Lambert W:

$$B_{max} = -\frac{W(-1, -d/2)}{d} \tag{14}$$

Wartość kroku kwantyzacji Δ optymalną ze względu na jakość przetwarzania (SQNR), w całym zakresie od minimalnego do maksymalnego poziomu mocy, można wyznaczać za pomocą innych danych niż występujące w zależności (10). Na podstawie (6), (6'), (8), (8') i (9) uzyskuje się inne postacie wzoru (10) opisujące optymalną wartość kroku kwantyzacji w zależności od danych, którymi dysponuje projektant modulatora:

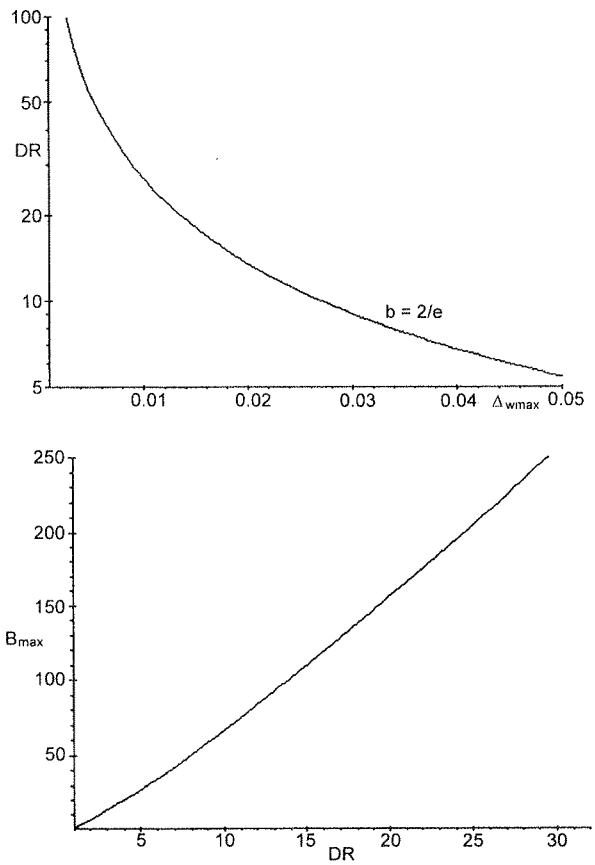
$$\Delta_{mowy} \cong 0.37 \frac{\ln 2B_{max}}{B_{max}} \sqrt{S_{max}} \tag{15}$$

W przypadku, gdy znana jest maksymalna wartość skuteczna sygnału wejściowego $\sqrt{S_{max}}$, oraz założona jest wartość B_{min} i zadana jest wartość DR względną wartość kroku kwantyzacji Δ_{wmax} (zapewniającą maksymalizację SQNR w zadanym zakresie DR) można wyznaczyć z równania:

$$\Delta_{wmax} = \Delta_{mowy}/\sqrt{S_{max}} \cong 0.37 \frac{\ln 2B_{max}}{B_{max}} \tag{15'}$$

Przy zadanej wartości parametru DR względny krok kwantyzacji można obliczyć na podstawie wzoru (17), który powstaje po przekształceniu zależności (15') wykorzystując równanie (12):

$$\Delta_{wmax} = \Delta_{mowy}/\sqrt{S_{max}} \cong \frac{0.27}{DR} \quad \text{dla } b = 2/e \tag{16}$$



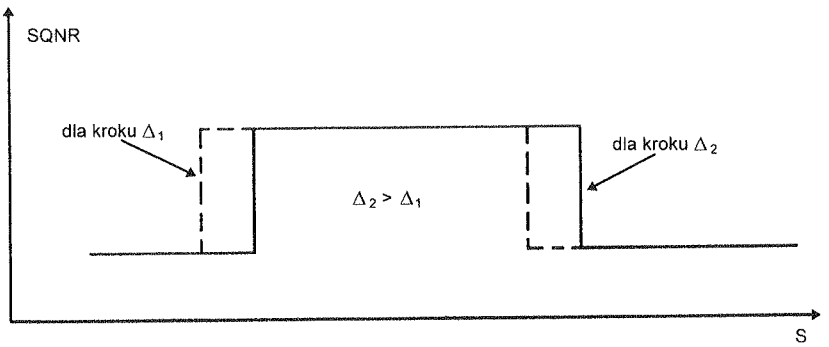
Rys. 6. (a)-wymagana względna wartość kroku kwantyzacji (Δ_{wmax}) w funkcji żądanej dynamiki DR, (b)-wymagana nadmiarowość próbkowania B_{max} dla zapewnienia żądanej dynamiki DR i utrzymania $b = 2/e$

Żądaną wartość B_{max}, która zapewnia założoną wartość dynamiki DR uzyskuje się w sposób identyczny jak przedstawiono to w równaniach (11÷14). Rys.6 ilustruje poszukiwane zależności. Na podstawie zadanej dynamiki DR uzyskuje się z jednej strony krok kwantyzacji Δ_{wmax} (rys. 6a) a z drugiej wymaganą wartość B_{max}.

Oczywiście bezwzględna wartość kroku kwantyzacji Δ uzyskana po przekształceniu wzoru (11) definiującego Δ_{wmax} i wzoru (10′) definiującego Δ_{wmin} jest taka sama. Ostatecznie spełnione jest $\Delta = \sqrt{S_{max}}\,\Delta_{wmax} = \sqrt{S_{min}}\,\Delta_{wmin}$.

Rys. 7 ilustruje wpływ wielkości kroku kwantyzacji Δ na położenie dolnego i górnego krańca zakresu dynamiki DR sygnału wejściowego, w którym dzięki adaptacji częstotliwości próbkowania modulator NSDM utrzymuje stałą maksymalną wartość SQNR. Na rys. 7 uwidoczniono ponadto, że wartość kroku kwantyzacji nie ma bezpośredniego wpływu na wielkość SQNR. Podkreślić również należy, iż do wyznaczenia wartości kroku kwantyzacji w modulacji NSDM nie wystarcza znajomość

danych w zależnościach (1) i (1′), potrzebna jest także wartość skuteczna minimalnego poziomu sygnału wejściowego $\sqrt{S_{\min}}$ lub wartość maksymalnego poziomu sygnału wejściowego $\sqrt{S_{\max}}$ i dynamika DR.



Rys.7. Ilustracja wpływu wielkości kroku kwantyzacji Δ na położenie krańców zakresu dynamiki DR sygnału wejściowego, w którym SQNR osiąga maksimum (B_{min}, B_{max} = const, czyli i DR = const)

2.3.1. Graniczne wartości kroku kwantyzacji

Na podstawie zależności (6), (6′), (8), (8′) i (9) uzyskuje się:

$$b = \frac{\Delta_{mowy}}{0.37\sqrt{S_{\min}}} < 2/e \quad \text{oraz} \quad c = \frac{\Delta_{mowy}}{0.37\sqrt{S_{\max}}} < 2/e \qquad (17) \text{ i } (17')$$

i stąd:

$$0 < \Delta_{mowy} \leq 0.27\sqrt{S_{\min}} \qquad (18)$$

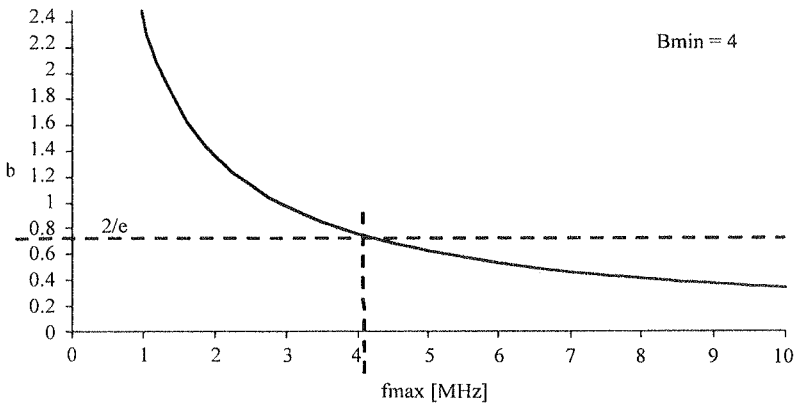
Nierówność (18) ma znaczenie praktyczne podczas projektowania modulatorów NSDM.

2.4. ANALIZA WYBRANYCH WŁAŚCIWOŚCI MODULACJI NSDM W UJĘCIU GRAFICZNYM

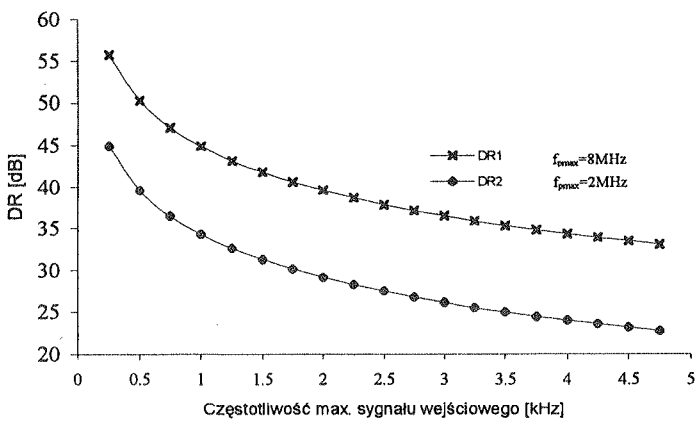
Dobór parametrów modulatora na podstawie analizy szumów kwantyzacyjnych, choć często, nie uwzględniającej szeregu czynników rzeczywiście występujących w procesie modulacji, stanowi podstawę pozwalającą wyznaczać możliwości danego rodzaju modulacji i określać zakres zmian parametrów podczas szczegółowych symulacji.

Na rys. 8 horyzontalna linia przerywana wytyczona dla granicznej wartości czynnika $b = 2/e$ pozwala wyznaczyć wartość $f_{p\,max}$, dla której modulator NSDM zapewnia już minimalizację szumów kwantyzacyjnych w całym zakresie założonej dynamiki (tutaj 36 dB). Wykres na rys. 8 podobnie jak poprzednie, wskazuje na to, że aby zapewnić

określoną dynamikę przetwarzania (tutaj 36dB), należy przyjąć odpowiednio dużą wartość $f_{p\,max}$. Wykres ten przedstawia w efekcie, jaka musi być minimalna częstotliwość $f_{p\,max}$, aby móc uzyskać założoną dynamikę (36 dB). Widać, iż dopiero częstotliwościom powyżej ok. 4 MHz odpowiadają wartości parametru *b*, dla których równanie (5) posiada rozwiązanie rzeczywiste.



Rys. 8. Wymagana wartość parametru b przy założonej dynamice (36 dB) w funkcji maksymalnej częstotliwości próbkowania $f_{p\,max}$



Rys. 9. Możliwa do uzyskania dynamika przetwarzania w funkcji maksymalnej częstotliwości f_0 sygnału wejściowego, przy stałej wartości $f_{p\,max}$ równej 8MHz dla *DR*1 i 2MHz dla *DR*2

Związek pomiędzy maksymalną częstotliwością graniczną sygnału wejściowego, a maksymalną częstotliwością próbkowania w sposób pośredni pokazany na rys. 9 jest oczywisty [Tabela 1] i udowodniony w przypadku wszystkich pozostałych modulacji różnicowych. Dwie krzywe odpowiadające odpowiednio $f_{p\,max}$ = 8MHz i $f_{p\,max}$ = 2MHz wskazują na to, że przy zadanej $f_{p\,max}$ dla zwiększenia dynamiki można obniżyć

maksymalną częstotliwość sygnału wejściowego, albo zmniejszyć wymagania, co do dynamiki przetwarzania i uzyskać możliwość przetwarzania sygnału wejściowego o większej częstotliwości granicznej.

3. METODYKA DOBORU PARAMETRÓW MODULATORA NSDM

Analiza przeprowadzona w rozdziałach 2÷2.4 obejmuje wyprowadzenie wszystkich niezbędnych wzorów opisujących parametry modulatora NSDM oraz pozwala ustalić kolejność obliczeń, której należy przestrzegać, aby możliwe było wyznaczenie wszystkich niezbędnych parametrów modulatora NSDM zależy od tego, które parametry są zadane. Metoda ustalania kolejności wynika z następujących spostrzeżeń:

- do wyznaczenia $f_{p\,min}$ wystarcza znajomość SQNR (Tabela 1-wiersz 8) i S_{min},
- przy wyznaczaniu kroku kwantyzacji Δ korzysta się z wzorów (1) i (1′) lub bezpośrednio z zależności (11)-gdy znana jest wartość S_{min} lub zależności (17)-gdy znana jest wartość S_{max}. W tym drugim przypadku niezbędna jest również znajomość wymaganego zakresu dynamiki DR.
- do wyznaczenia $f_{p\,max}$ należy znać S_{min}, S_{max} (lub DR) oraz $f_{p\,min}$- wzory (5') i (6')
- do wyznaczenia S_{max} należy znać S_{min}, $f_{p\,min}$, oraz $f_{p\,max}$ — przekształcone wzory (2) i (2′).

Na rys. 10 przedstawiono kilka przykładów poprawnej kolejności wyznaczania poszczególnych parametrów modulatora NSDM w zależności od danych początkowych.

3.1 EFEKTYWNOŚĆ UZYSKANIA OKREŚLONEGO ZAKRESU DYNAMIKI DR W MODULACJACH CFDM I NSDM

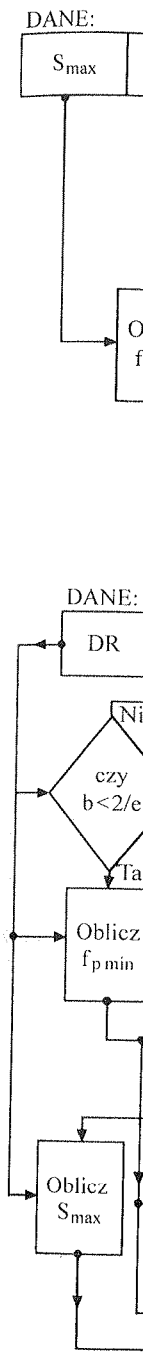
Za jedną z metod oceny porównawczej efektywności modulacji adaptacyjnych można przyjąć czułość dynamiki DR danej modulacji na zmiany parametru zapewniającego adaptację krzywej aproksymującej przebieg modulowany (przy tej samej jakości mierzonej np. stosunkiem SQNR). W przypadku modulacji CFDM tym parametrem jest krok kwantyzacji Δ natomiast dla modulacji NSDM jest to częstotliwość próbkowania f_p.

Korzystając z wzorów (2′) i (2″) (ich spełnienie zapewnia maksymalizację SQNR) można łatwo wykazać [11], że dla modulacji CFDM (stała częstotliwość próbkowania, zmienny krok próbkowania) mamy:

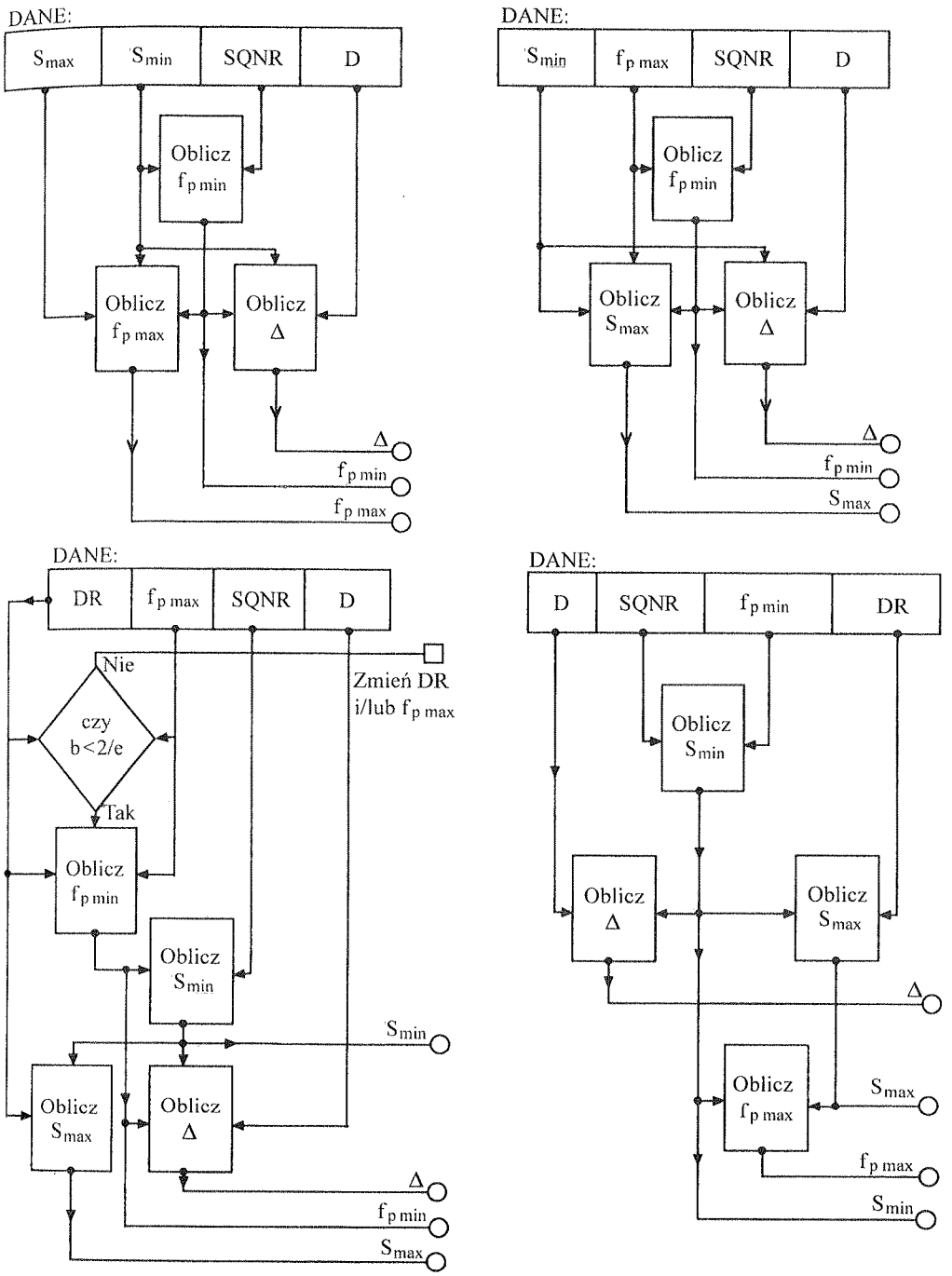
$$\frac{\Delta_{max}}{\Delta_{min}} = K_{\Delta} = DR$$

Dla modulacji NSDM (stały krok kwantyzacji, zmienna częstotliwość próbkowania) można z kolei zdefiniować:

$$f_{max}/f_{min} = K_f$$



Rys. 10. Ko



Rys. 10. Kolejność wyznaczania parametrów modulatora NSDM dla różnych danych początkowych

Czyli dla modulacji NSDM poszukuje się: $DR = f(B_{max}, B_{min})$. Również na podstawie wzorów (2′) i (2″) i zakładając $B_{min} = e/2$ uzyskuje się:

$$DR \frac{\ln(2B_{max})}{B_{max}} = \frac{2}{e} \tag{19}$$

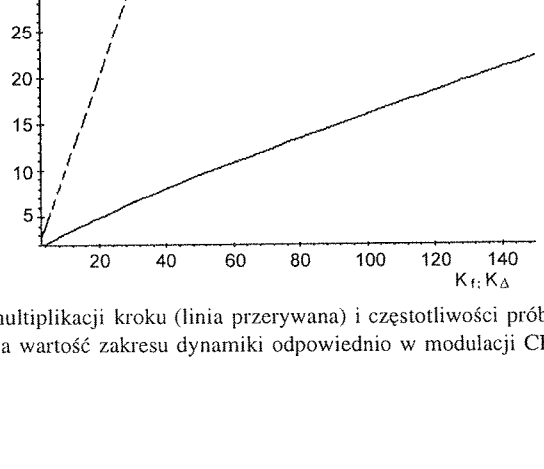
a więc zależność w postaci uwikłanej. Oznaczając: $B_{max} = K_f B_{min} = eK_f/2$ rozwiązanie równania (19) względem K_f prowadzi do postaci:

$$DR = \frac{K_f}{ln\left(\frac{e \cdot K_f}{2}\right)} \tag{20}$$

Oba rozwiązania w postaci graficznej przedstawiono na rys. 11.

Z rys. 11 widać, że system modulacji ze zmiennym krokiem kwantyzacji (CFDM, CVSD) pod względem czułości zakresu dynamiki DR na zmiany tego kroku jest systemem znacznie efektywniejszy niż modulacja ze zmienną częstotliwością próbkowania (NSDM).

Uzyskane zależności (19), (20) nie mówią nic na temat średniej przepływności bitowej $P_{bśr}$, ($P_{bśr}$ określa liczbę bitów niezbędnych do przetworzenia danego sygnału) modulacji NSDM. Wprawdzie założona jakość (SQNR) dla obu porównywanych modulacji jest taka sama, ale może się okazać, że wynikowa średnia przepływność $P_{bśr}$ metody NSDM wyraźnie się różni od niezbędnej f_p w metodzie CFDM. Jednak badania symulacyjne [15] modulacji NSDM wykazały, że na wartość SQNR w tej modulacji podobnie jak w CFDM decydujący wpływ ma przepływność bitowa $P_{bśr}$ (w przypadku NSDM średnia przepływność bitowa) natomiast pozostałe parametry wpływają w znacznie mniejszym stopniu.



Rys. 11. Wpływ multiplikacji kroku (linia przerywana) i częstotliwości próbkowania (linia ciągła) na wartość zakresu dynamiki odpowiednio w modulacji CFDM i NSDM

4. WNIOSKI

Głównym efektem pracy jest wykazanie na drodze analitycznej, że przy stałym kroku kwantyzacji Δ możliwe jest utrzymanie stałej wartości SQNR dla różnych poziomów mocy S sygnału wejściowego poprzez stosowanie dla nich różnych częstotliwości próbkowania f_p. Dlatego można stwierdzić, że system modulacji NSDM zapewnia, kompandowanie tzn. należy do systemów adaptacyjnych pozwalających na zwiększenie dynamiki przetwarzanego sygnału w stosunku do modulacji LDM. Wartości par S i f_p, dla których SQNR = const przy różnych poziomach mocy sygnału wejściowego można wyliczyć z funkcji W Lamberta [14]. Zadaniem przyjmowanego w modulacji NSDM algorytmu adaptacji f_p jest zapewnienie, aby poszczególnym poziomom mocy S odpowiadały optymalne częstotliwości próbkowania f_p.

Z warunków brzegowych funkcji Lambert W wynika ważna właściwość modulacji NSDM, polegająca na tym, że modulacja ta osiąga najlepsze parametry wyjściowe gdy spełniona jest zależność (19).

Ważnym spostrzeżeniem wynikającym z tej samej zależności jest relacja pomiędzy żądaną dynamiką, którą ma zapewnić modulacja NSDM i wymaganą do tego celu maksymalną częstotliwością próbkowania.

Porównanie efektywności metod adaptacji kroku kwantyzacji i częstotliwości próbkowania (rys. 11) pokazuje, że metodę NSDM należy stosować jedynie w przypadkach silnej niestacjonarności procesu wejściowego poddawanego modulacji. Natomiast na podstawie dotychczasowych badań eksperymentalnych i w świetle szybkiego rozwoju technik DSP wydaje się być w pełni uzasadnione prowadzenie dalszych badań nad modulacjami ANSDM, łączącymi adaptację kroku kwantyzacji i częstotliwości próbkowania.

Zaprezentowana przez Autora metoda wyznaczania podstawowych parametrów modulacji NSDM S_{max}, S_{min}, $f_{p\,min}$, $f_{p\,max}$ oraz kroku kwantyzacji Δ jest logiczną konsekwencją rozważań opisanych w rozdz. 2, ale stanowi również podsumowanie analiz przedstawionych w wielu opracowaniach dotyczących modulacji różnicowych [6, 7, 8, 10, 15, 16, 17].

BIBLIOGRAFIA

1. T.A. Hawkes, P.A. Simonpieri: *Signal coding Using Asynchronous Delta Modulation, IEEE Transaction on Communication*. Vol. COM–22, No. 3, March 1974, pp. 346–348
2. J. J. Dubnowski, R. E. Crochiere: *Variable Rate Coding of Speech*. BSTJ, vol. 58, No. 3, March 1979
3. J. L. LoCiero, P. P. Prezas: *Doubly adaptive delta modulation for bandwidth compression of speech signals*. Proc. Nat. Telecommun. Conf., 80 CH 1539–6, vol. 2, 1980, pp. 36.5.1–36.5.5
4. C. K. Un, H. S. Lee, J. S. Song: *Hybrid companding delta modulation. IEEE Trans.* Commun., vol. COM–29., pp 1337–1344, Sept. 1981
5. C. K. Un, D. H. Cho: *Hybrid Companding Delta Modulation with Variable-Rate Sampling*. IEEE Trans. On Comm., vol. COM–30, No. 4, April 1982
6. Y. S. Zhu, W. Leung, C. M. Wong: *A digital audio processing system based on nonuniform sampling delta modulation*. IEEE Trans. on Consumer Electronics, vol. 42, No. 1, Feb. 1996

7. Y. S. Zhu, W. Leung, C. M. Wong: *Adaptive Non-Uniform Sampling Delta Modulation for Audio/Image Processing*. IEEE Trans. on Consumer Electronics, vol. 42, No. 4, Nov. 1996
8. R. Golański, A. Bogusz: *Zastosowanie nierównomiernego próbkowania w modulacji różnicowej*. Mat. Konf.: Krajowe Sympozjum Telekomunikacji '97, Tom B, Bydgoszcz 1997
9. R. Golański, J. Kołodziej, P. Świrek: *„Badanie symulacyjne modulacji różnicowych ze zmiennym krokiem kwantyzacji i nierównomiernym próbkowaniem"*. Mat. Konf.: KST '99, Tom B, Bydgoszcz 1999
10. R. Golański: *Propozycja doboru parametrów adaptacji modulatora różnicowego z nierównomiernym próbkowaniem*. Mat. Konf.: KST '98, Tom B, Bydgoszcz 1998
11. J. E. Abate: *Linear and Adaptive Delta Modulation*. Proceedings of the IEEE, vol.55, No.3, March, 1967
12. N. S. Jayant: *Adaptive Delta Modulation with one-bit Memory*. Bell System Technical Journal, vol.49, No 3, March 1970, pp 321–342
13. J. B. O'Neal, Jr: *Delta Modulation Quantizing Noise Analytical and Computer Simulation Results for Gaussian and Television Input Signals*. BSTJ, January 1966
14. Maple Manuals. *Maple V Release 3*, Waterlo Maple Software, 1981–1994
15. R. Golański, R. Dmoch, J. Kochanowski, K. Szklarski: *Propozycja znajdowania optymalnych parametrów modulacji różnicowej NSDM w oparciu o badania symulacyjne*. Mat. Konf.: KST '2000, Tom B, Bydgoszcz, wrzesień 2000
16. W. Pogribny, I. Różnakowski, I. Zieliński, Z. Nowakowski: *Graniczna częstotliwość dyskretyzacji eliminująca przeciążenie delta-kodera*. Mat. Konf. KST'98, T.B, Bydgoszcz wrzesień 1998
17. *Wybrane systemy i układy scalone w telekomunikacji cyfrowej*. Pod red. R. Golańskiego, Wydawnictwa AGH, Kraków 1995

R. GOLAŃSKI

NONUNIFORM SAMPLING AS A COMPANDING METHOD IN DIFFERENTIAL MODULATIONS

Summary

Fundamental works of Hawkes, Un, Zhu and Abate, are shortly presented. Some results in the paper are based upon previously obtained formulas for DM and ADM delta modulators. The staircase in the delta modulation system can be adapted to the amplitude variations of a input signal by varying sampling frequency (Nonuniform Sampling Delta Modulation method). In the article signal to noise formula for NSDM modulator having a random inputs is shown. The calculation method of the minimum ($f_{p\,min}$) and maximum ($f_{p\,max}$) sampling frequency based on changing of the *slope overload factor* is presented in this paper. Design of the parameters ($f_{p\,min}$, $f_{p\,max}$ and quantization step size Δ) for NSDM delta modulator is derived. Formula of maximization SQNR for different power levels of the input signal by means of various sampling frequency in the NSDM modulation is the main effect of discussion.The paper presents analytical and graphical dependences between the most important parameters of the NSDM. These dependences help to establish a method of procedure during the choice of the parmeters of the NSDM modulator. This relationships could be the base to design of a non-uniform sampling modulators. Some example of design of the parameters is presented.

Keywords: differential modulations, non-uniform sampling, Lambert W function, S/N ratio, dynamic range, NSDM modulator design.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2001, 47, z. 2, ss. 183–207

Zastosowanie algorytmu kratowego typu LS w adaptacyjnym podpasmowym systemie usuwania echa akustycznego

JACEK FALKIEWICZ

Instytut Systemów Elektronicznych, Politechnika Warszawska
ul. Nowowiejska 15/19, 00-665 Warszawa

Otrzymano 2001.02.23
Autoryzowano 2001.03.09

W pracy omówiono problem adaptacyjnego usuwania echa akustycznego. Przedstawiono dwie podstawowe klasy algorytmów adaptacyjnych: LMS oraz RLS i porównano je pod kątem zastosowania w systemach usuwania echa. Wykazano, że algorytm kratowy LSL stanowi rozsądny kompromis przy wyborze filtru adaptacyjnego, od którego wymaga się z jednej strony dużej i niezależnej od sygnału wejściowego szybkości zbieżności, z drugiej zaś liniowo zależnej od rzędu filtru złożoności obliczeniowej oraz stabilności numerycznej. Pokazano, że w porównaniu z algorytmem LSL pracującym w pełnym pasmie, zastosowanie filtru LSL w podpasmowym systemie usuwania echa akustycznego prowadzi do zmniejszenia złożoności obliczeniowej całego algorytmu filtracji adaptacyjnej oraz do poprawy skuteczności tłumienia echa w początkowym okresie pracy systemu.

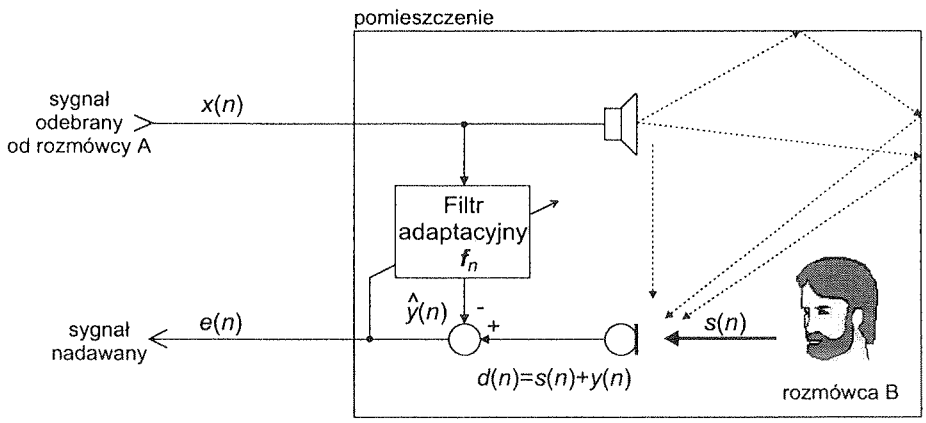
Działanie algorytmów zweryfikowano symulacyjnie. Badania symulacyjne przeprowadzono z wykorzystaniem zarówno sygnałów syntetycznych, jak i rzeczywistych. Potwierdziły one dobre właściwości podpasmowego adaptacyjnego systemu usuwania echa akustycznego z filtrami LSL, który do analizy i syntezy sygnałów podpasmowych wykorzystuje bank filtrów z nadpróbkowaniem.

Słowa kluczowe: usuwanie echa akustycznego, algorytmy adaptacyjne, telefoniczne systemy głośnomówiące

1. WPROWADZENIE

Konstrukcja zapewniających swobodną konwersację systemów telefonii *hands-free* oraz telekonferencyjnych terminali audio, które nie wymagają warunków studyjnych, wiąże się ściśle z problemem usuwania echa akustycznego. Powstaje ono w wyniku odbić fal akustycznych w pomieszczeniu oraz sprzężenia akustycznego między mikrofonem i głośnikiem. Echo powoduje nieprzyjemne wrażenia słuchowe u rozmówców,

polegające na percepcji zniekształconych i opóźnionych w czasie ich własnych wypowiedzi. W szczególnie niesprzyjających warunkach może ono doprowadzić do sprzężenia akustycznego objawiającego się tzw. wyciem. Nowoczesną metodą walki z echem jest technika filtracji adaptacyjnej. Polega ona na rekursywnej identyfikacji odpowiedzi impulsowej środowiska akustycznego, a następnie na wytworzeniu możliwie wiernej repliki sygnału echa, którą odejmuje się od sygnału nadawanego. Sposób powstawania echa akustycznego w systemie telefonii *hands-free* oraz koncepcję jego adaptacyjnego usuwania przedstawiono na rys. 1.



Rys. 1. Ilustracja powstawania echa akustycznego w systemie telefonii *hands-free* i koncepcja jego adaptacyjnego usuwania

Sygnał odebrany $x(n)$, pochodzący od rozmówcy A, dociera do głośnika i jest emitowany w pomieszczeniu. Dociera on do rozmówcy B oraz — po przekształceniu w otoczeniu akustycznym — do mikrofonu, jako sygnał $y(n)$, który jest sygnałem echa wtórnie transmitowanym do rozmówcy A. Aby uniknąć tego niepożądanego efektu, należy wytworzyć replikę $\hat{y}(n)$ sygnału echa i odjąć ją od sygnału $d(n)$, pochodzącego z wyjścia mikrofonu. Sygnał $e(n)$, będący różnicą między sygnałem $d(n)$ i estymowanym sygnałem echa $\hat{y}$(n), nazwiemy *resztkowym sygnałem echa*, który przenika do rozmówcy znajdującego się po drugiej stronie łącza.

Wytworzenie repliki echa jest realizowane za pomocą filtru adaptacyjnego, którego zadaniem jest estymacja zmiennej w czasie odpowiedzi impulsowej pomieszczenia. Jeżeli przez $h_n(0),\ldots,h_n(L-1)$ oznaczymy pierwsze L próbek odpowiedzi impulsowej, które należy wyestymować, aby zapewnić jej identyfikację z wymaganą dokładnością, to zadanie filtru adaptacyjnego sprowadza się do estymacji wektora $\boldsymbol{h}_n = [h_n(0),\ldots, h_n(L-1)]^T$ w każdej chwili $n = 1, 2,\ldots$ Replikę $\hat{y}(n)$ sygnału echa uzyskuje się w wyniku filtracji sygnału odebranego $x(n)$ za pomocą transwersalnego filtru SOI o strojonym adaptacyjnie wektorze współczynników $\boldsymbol{f}_n = [f_n(0),\ldots,f_n(L-1)]^T$, będącym jednocześnie estymatą wektora $\boldsymbol{h}_n$.

Proces identyfikacji jest prowadzony w czasie rzeczywistym, nie może mieć zatem zbyt dużej złożoności obliczeniowej. Wysokie wymagania co do wartości tłumienia echa, postawione systemom jego usuwania w normach telekomunikacyjnych (G.167 ITU-T), implikują natomiast konieczność uzyskania dostatecznie dokładnej estymaty odpowiedzi impulsowej pomieszczenia. Wymienione wymagania są w praktyce trudne do spełnienia z uwagi na warunki w jakich przychodzi realizować zadanie identyfikacji nieznanego układu liniowego, który modeluje środowisko akustyczne. Analiza tych warunków pozwala wyróżnić dwie przyczyny komplikacji procesu identyfikacji. Są to:

- charakterystyka sygnału mowy propagowanego w systemie *hands-free*;
- właściwości środowiska akustycznego.

Sygnał mowy charakteryzuje się dużą dynamiką, silną korelacją oraz niestacjonarnością. Wszytkie wymienione cechy wpływają niekorzystnie na szybkość i dokadność prowadzonej przez filtr adaptacyjny identyfikacji odpowiedzi impulsowej lub, co jest równoważne, transmitancji pomieszczenia. Powodują one spowolnienie zbieżności oraz potencjalną utratę stabilności algorytmów działania filtrów adaptacyjnych. Dodatkowo należy tu rozważyć sygnał $s(n)$ (rys. 1). W przypadku, gdy rozmówca A słucha, a rozmówca B mówi, jest on użytecznym sygnałem mowy nadawanym do rozmówcy A, natomiast w sytuacji odwrotnej, jest szumem środowiska po stronie rozmówcy B. Można sobie wyobrazić sytuację jednoczesnego nadawania (ang. *double talk*). Wtedy sygnał $s(n)$ jest zakłócającym sygnałem mowy, który, ze względu na swoje wcześniej wymienione właściwości oraz poziom mocy, może nawet uniemożliwić przeprowadzenie procesu identyfikacji.

Druga przyczyna jest związana z właściwościami środowiska akustycznego. Czas pogłosu typowych pomieszczeń biurowych wynosi kilkaset milisekund, zaś w dużych salach konferencyjnych przekracza jedną sekundę. Chcąc uzyskać dostatecznie dokładną estymację transmitancji pomieszczenia, należy zastosować filtr adaptacyjny, którego czas trwania odpowiedzi impulsowej będzie tego samego rzędu, co czas pogłosu pomieszczenia. Zakładając, że pasmo transmisyjne dla sygnału mowy ma szerokość 4 kHz (częstotliwość próbkowania 8 kHz), można łatwo obliczyć, że należy zastosować filtr o rzędzie równym kilka tysięcy. Strojenie filtru adaptacyjnego o tak dużej liczbie współczynników, nawet za pomocą algorytmów o niewielkiej złożoności obliczeniowej, wymaga zastosowania procesorów sygnałowych o znacznej mocy obliczeniowej.

Ważną właściwością środowiska akustycznego, mającą istotny wpływ na działanie systemu usuwania echa, jest bardzo duża czułość odpowiedzi impulsowej pomieszczenia na zmiany przestrzenne w jego wnętrzu, takie jak ruch ludzi czy zmiany w ustawieniu mebli [6]. Z właściwości tej wynika konieczność identyfikacji zmiennej w czasie odpowiedzi impulsowej, czyli, innymi słowy, identyfikacji liniowego układu niestacjonarnego.

Omówione powyżej trudności określają specyfikę problemu adaptacyjnego usuwania echa akustycznego i pozwalają na postawienie wymagań algorytmom adaptacyjnym, które można do usuwania echa potencjalnie wykorzystać. I tak, od algorytmów tych oczekuje się:

1. jak największej szybkości zbieżności wektora estymat $\boldsymbol{f}_n$ do wektora $\boldsymbol{h}_n$, niezależnej od charakterystyk sygnału wejściowego filtru adaptacyjnego;
2. małego poziomu mocy sygnału resztkowego echa $e(n)$;
3. dobrej zdolności śledzenia niestacjonarności identyfikowanego systemu (ang. *tracking capability*);
4. niewielkiej złożoności obliczeniowej;
5. odporności procesu adaptacji na szum środowiska $s(n)$ o znacznym poziomie mocy.

Wydaje się intuicyjnie oczywistym, że wymagania te są wzajemnie sprzeczne i nie istnieje jeden dobry algorytm adaptacyjny, który spełniłby je wszystkie w zadowalającym stopniu. W części 2 pracy przedstawiono dwie podstawowe klasy algorytmów adaptacyjnych (LMS i RLS) oraz przeprowadzono ich porównanie pod kątem zastosowania w adaptacyjnym systemie usuwania echa akustycznego (punkt 2.2). Przedstawiono także algorytm kratowy LSL (punkt 2.3) będący — według autora — rozsądnym kompromisem, pozwalającym spełnić w pewnym stopniu wymagania postawione algorytmom usuwania echa. W części 3 omówiono technikę filtracji w podpasmach, która pozwala zredukować złożoność obliczeniową algorytmu LSL oraz zwiększyć tłumienie echa w początkowym okresie jego pracy. Część 4 zawiera wyniki badań symulacyjnych podpasmowego adaptacyjnego systemu usuwania echa akustycznego z filtrami LSL, który do analizy i syntezy sygnałów podpasmowych wykorzystuje bank filtrów z nadpróbkowaniem. Badania te przeprowadzono używając zarówno sygnałów syntetycznych, jak i rzeczywistych.

2. ALGORYTMY ADAPTACYJNE

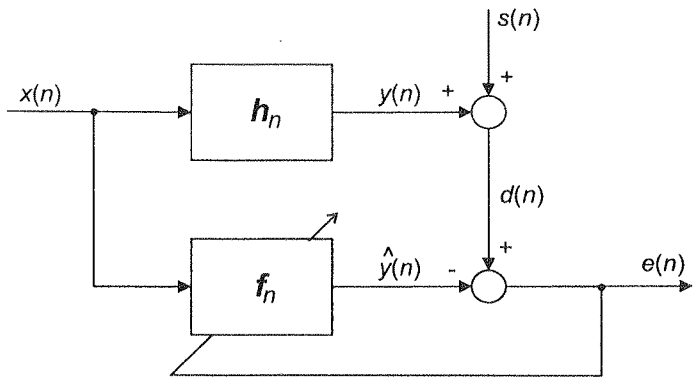
Narzędziami rekursywnej identyfikacji odpowiedzi impulsowej nieznanego układu, stanowiącej podstawową zasadę działania systemów usuwania echa akustycznego, są algorytmy adaptacyjne. Prezentacja dwóch podstawowych klas algorytmów adaptacyjnych, tj. LMS (*Least Mean Square*) i RLS (*Recursive Least Squares*) zostanie ograniczona jedynie do przytoczenia ich rekursji oraz podania podstawowych parametrów pozwalających je porównać. Pominięte zostaną wyprowadzenia tych algorytmów, dostępne w literaturze poświęconej filtracji adaptacyjnej [9, 14].

2.1. ALGORYTMY LMS i RLS

Na rys. 2 przedstawiono schemat adaptacyjnego systemu identyfikacji nieznanego układu niestacjonarnego. Zadaniem filtru adaptacyjnego w tym systemie jest takie przetworzenie sygnału $x(n)$, aby *błąd estymacji* $e(n)$, będący różnicą między *sygnałem wzorcowym* $d(n)$ a sygnałem wyjściowym filtru $\hat{y}(n)$ był minimalny w sensie pewnego kryterium. Dokonując minimalizacji funkcji kosztu o następującej postaci:

$$J_n = e^2(n) = d(n) - \hat{y}(n) = [d(n) - \boldsymbol{f}_n \boldsymbol{x}_n^T]^2, \tag{1}$$

gdzie $\boldsymbol{x}_n = [x(n), x(n-1),\ldots, x(n-L+1)]^T$ jest wektorem danych wejściowych (L — rząd filtru adaptacyjnego SOI[1]), otrzymujemy algorytm adaptacyjny LMS.



Rys. 2. Schemat adaptacyjnego systemu identyfikacji nieznanego układu niestacjonarnego

Tabela 1

Algorytm LMS

Warunki początkowe
$\boldsymbol{f}_0 = 0$
Dla kolejnych chwil czasu n obliczamy
$e(n) = d(n) - \boldsymbol{f}_n^T \boldsymbol{x}_n$
$\boldsymbol{f}_{n+1} = \boldsymbol{f}_n + \alpha e(n) \boldsymbol{x}_n$

Z równania (1) wynika, że w każdej chwili czasu n, wektor współczynników filtru $\boldsymbol{f}_n$ jest tak dobierany, aby minimalizować chwilową moc błędu estymacji w tejże chwili.

W tabeli 1 zamieszczono rekursje algorytmu LMS wraz z warunkami początkowymi. Pojawiająca się w drugiej rekursji stała α jest nazywana *krokiem adaptacji*. Wpływ doboru jej wartości na szybkość zbieżności algorytmu, poziom sygnału błędu estymacji w stanie ustalonym oraz zdolność śledzenia niestacjonarności omówiony zostanie dokładnie wpunkcie 2.2.

Drugi z rozważanych algorytmów — RLS, otrzymujemy w wyniku minimalizacji funkcji kosztu określonej wzorem:

[1] Pod pojęciem rzędu filtru adaptacyjnego SOI rozumiemy liczbę jego sukcesywnie aktualizowanych współczynników.

$$J_n = \sum_{i=1}^{n} \lambda^{n-1} e^2(i), \tag{2}$$

gdzie λ jest *stałą zapominania*, przyjmującą wartości z przedziału $(0, 1]$. Wprowadzenie tego parametru powoduje wykładnicze „oknowanie" sygnału błędu $e(n)$, tzn. działanie polegające na tym, że starsze próbki sygnału błędu estymacji są brane do sumarycznej miary błędu J_n z odpowiednio mniejszą wagą. Za miarę „pamięci" algorytmu RLS można przyjąć odwrotność dopełnienia współczynnika λ do jedności, $1/(1-\lambda)$. W przypadku podstawowej wersji algorytmu RLS, stosowanej z reguły dla sygnałów stacjonarnych, przyjmuje się $\lambda = 1$. Dostajemy wtedy rekursywne rozwiązanie klasycznego, dobrze znanego *zagadnienia najmniejszych kwadratów*, a „pamięć" algorytmu jest wtedy nieskończona. Przyjmując $\lambda \neq 1$, otrzymujemy tzw. *algorytm RLS z wykładniczą stałą zapominania*. Algorytm ten zamieszczono w tabeli 2.

Tabela 2

Algorytm RLS

Warunki początkowe:

$$\boldsymbol{P}_0 = \gamma \boldsymbol{I} \qquad \gamma \gg 1$$

$$\boldsymbol{f}_0 = \boldsymbol{0}$$

Dla kolejnych chwil czasu n obliczamy

$$e(n|n-1) = d(n) - \boldsymbol{f}_{n-1}^T \boldsymbol{x}_n$$

$$\boldsymbol{k}_n = \frac{\boldsymbol{P}_{n-1}\boldsymbol{x}_n}{\lambda + \boldsymbol{x}_n^T \boldsymbol{P}_{n-1}\boldsymbol{x}_n}$$

$$\boldsymbol{f}_n = \boldsymbol{f}_{n-1} + \boldsymbol{k}_n e(n|n-1)$$

$$\boldsymbol{P}_n = \frac{1}{\lambda}[\boldsymbol{P}_{n-1} - \boldsymbol{k}_n \boldsymbol{x}_n^T \boldsymbol{P}_{n-1}]$$

Występująca w tym algorytmie macierz $\boldsymbol{P}_n$ o wymiarze $\dim(\boldsymbol{P}_n) = L \times L$ jest estymatą w chwili n macierzy $\boldsymbol{R}^{-1}$, odwrotnej do macierzy autokorelacji $\boldsymbol{R} = \mathrm{E}[\boldsymbol{x}_n \boldsymbol{x}_n^T]$ sygnału wejściowego $x(n)$ filtru adaptacyjnego ($\mathrm{E}[\cdot]$ jest symbolem operatora wartości oczekiwanej). Symbol $e(n|n-1)$ oznacza tu błąd estymacji *a priori*, w odróżnieniu do wykorzystanego w kryterium minimalizacji (2) błędu *a posteriori* $e(n)$. Błąd estymacji *a posteriori* oblicza się wykorzystując aktualny wektor współczynników filtru $\boldsymbol{f}_n$, zaś błąd *a priori* — korzystając z wektora współczynników filtru $\boldsymbol{f}_{n-1}$ z chwili poprzedniej. Wektor $\boldsymbol{k}_n$ nazywany jest *wektorem wzmocnienia algorytmu*.

2.2. PORÓWNANIE ALGORYTMÓW LMS I RLS POD KĄTEM ZASTOSOWANIA W SYSTEMACH USUWANIA ECHA AKUSTYCZNEGO

Przypomnijmy raz jeszcze krótko wymagania stawiane algorytmom adaptacyjnym pretendującym do zastosowania w systemach usuwania echa akustycznego: (1) duża szybkość zbieżności; (2) mały poziom mocy sygnału $e(n)$; (3) dobra zdolność śledzenia niestacjonarności identyfikowanego systemu; (4) mała złożoność obliczeniowa; (5) odporność procesu tłumienia echa na szum środowiska $s(n)$. Poniżej przeanalizujemy w jakim stopniu algorytmy LMS i RLS spełniają te wymagania.

1. Szybkość zbieżności algorytmu LMS silnie zależy od charakterystyk widmowych sygnału wejściowego $x(n)$, które są ściśle powiązane z rozrzutem wartości własnych macierzy autokorelacji $\boldsymbol{R}$ tego sygnału [9]:

$$\chi(\boldsymbol{R}) = \frac{\lambda_{max}}{\lambda_{min}} \leq \frac{S_{max}}{S_{min}}, \tag{3}$$

gdzie λ_{max} i λ_{min} są odpowiednio największą i najmniejszą wartością własną macierzy $\boldsymbol{R}$, zaś S_{max} i S_{min} — największą i najmniejszą wartością widma mocy tego sygnału. Dla sygnałów silnie skorelowanych, a takim jest sygnał mowy, o dużym rozrzucie $\chi(\boldsymbol{R})$ wartości własnych macierzy $\boldsymbol{R}$, szybkość zbieżności algorytmu LMS jest mała. Powodem tego jest jednakowa wartość kroku adaptacji α dla wszystkich współczynników filtru LMS. Zwiększenie wartości parametru α zwiększa wprawdzie szybkość zbieżności, ale jest ono ograniczone warunkiem stabilności algorytmu LMS [15]:

$$0 < \alpha < \frac{2}{\lambda_{max}}. \tag{4}$$

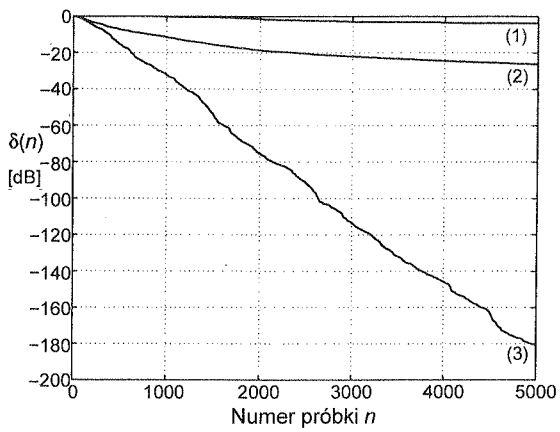
Z drugiej strony tzw. *stała zbieżności* τ_{LMS} algorytmu LMS (definiowana jako czas, po którym błąd estymacji najwolniej zbieżnego współczynnika filtru adaptacyjnego maleje e-krotnie [15]) jest odwrotnie proporcjonalna do kroku adaptacji oraz najmniejszej wartości własnej macierzy $\boldsymbol{R}$:

$$\tau_{LMS} = \frac{1}{\alpha \lambda_{min}}. \tag{5}$$

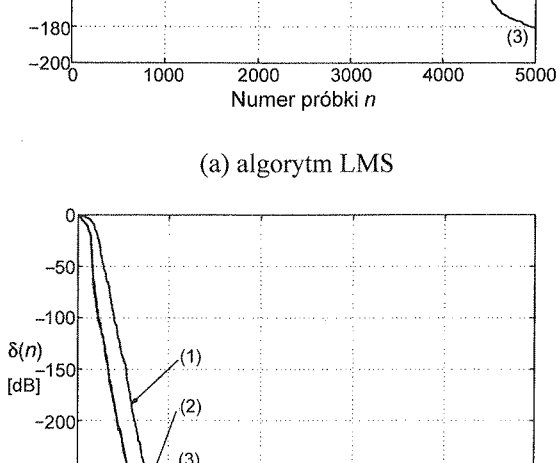
W rezultacie stała zbieżności algorytmu LMS nie może być mniejsza od wartości $\lambda_{max}/2\lambda_{min}$.

Szybkość zbieżności algorytmu RLS jest natomiast niezależna od widma mocy sygnału wejściowego, dzięki indywidualnej regulacji zbieżności poszczególnych współczynników filtru. Regulacja ta odbywa się w wyniku mnożenia błędu *a priori* $e(n|n-1)$ w głównej rekursji algorytmu RLS dla wektora $\boldsymbol{f}_n$ przez wektor wzmocnienia $\boldsymbol{k}_n$. Wektor ten z kolei zależy od macierzy $\boldsymbol{P}_n$, adaptującej się do macierzy $\boldsymbol{R}^{-1}$. Tym samym do

regulacji zbieżności filtru adaptacyjnego, pracującego według algorytmu RLS, jest wykorzystywana informacja o strukturze korelacyjnej sygnału wejściowego $x(n)$. W efekcie zapewnia to znacznie szybszą zbieżność algorytmu RLS w porównaniu z algorytmem LMS.



(a) algorytm LMS



(b) algorytm RLS

Rys. 3. Błąd względny $\delta(n)$ estymacji odpowiedzi impulsowej systemu dla różnych typów sygnałów wejściowych: (3) szum biały, $\chi(\boldsymbol{R}) = 1$; (2) proces AR(4), $\chi(\boldsymbol{R}) \simeq 630$; (1) sygnał mowy

Omawiane właściwości potwierdzają rezultaty przeprowadzonych badań symulacyjnych. Badano szybkość zbieżności filtrów adaptacyjnych LMS i RLS rzędu $L = 128$, których zadaniem była rekursywna identyfikacja odpowiedzi impulsowej układu stacjonarnego o skończonej długości równej L. Filtry i układ pobudzano sygnałami o różnych

właściwościach widmowych: szumem białym ($\chi(\boldsymbol{R}) = 1$), procesem AR rzędu 4 ($\chi(\boldsymbol{R}) = 630$) oraz rzeczywistym sygnałem mowy. W celu uniezależnienia wyników od wpływu szumu zakłócającego proces identyfikacji, przyjęto $s(n) = 0$, zatem współczynnik sygnał–szum SNR, zdefiniowany jako stosunek mocy sygnału wyjściowego $y(n)$ identyfikowanego systemu do mocy sygnału zakłócającego $s(n)$, wynosił $\sigma_y^2/\sigma_s^2 = \infty$. Błąd względny $\delta(n)$ estymacji odpowiedzi impulsowej obliczano według wzoru:

$$\delta(n) = 10\log\frac{\|\boldsymbol{h}-\boldsymbol{f}_n\|^2}{\|\boldsymbol{h}\|^2}, \tag{6}$$

gdzie $\boldsymbol{h}$ jest wektorem próbek odpowiedzi impulsowej identyfikowanego systemu, zaś $\boldsymbol{f}_n$ estymowanym wektorem próbek odpowiedzi impulsowej w chwili n. Z porównania krzywych zamieszczonych na rys. 3a i 3b wynika, że szybkość zbieżności algorytmu LMS przy pobudzeniu szumem białym jest znacznie mniejsza niż dla algorytmu RLS i gwałtownie maleje dla sygnałów skorelowanych (procesu AR(4) i sygnału mowy). Szybkość zbieżności algorytmu RLS nie zależy od sygnału wejściowego. Nieznaczne przesunięcie krzywej względnego błędu estymacji w przypadku pobudzania sygnałem mowy wynika z faktu, iż użyta do symulacji fraza rozpoczyna się trwającym kilkaset próbek okresem ciszy. Tak więc, ze względu na szybkość zbieżności, algorytm RLS jest znacznie lepszy od algorytmu LMS.

2. Parametrem określającym w sposób pośredni poziom mocy sygnału resztkowego echa $e(n)$ jest *niedopasowanie M* (ang. *misadjustment*). Jest ono zdefiniowane wzorem:

$$M = \frac{J_\infty - J_{min}}{J_{min}}, \tag{7}$$

gdzie J_{min} jest *minimalnym błędem średniokwadratowym* optymalnego rozwiązania wienerowskiego (tzn. rozwiązania zadania optymalnej filtracji dla kryterium MMSE, przy założeniu funkcji kosztu o postaci $J_n = \mathrm{E}[e^2(n)]$), zaś J_∞ jest wartością błędu średniokwadratowego, do którego zbiegają się rozwiązania generowane przez poszczególne algorytmy adaptacyjne. Niedopasowanie algorytmów LMS i RLS w przypadku stacjonarności identyfikowanego systemu oraz stacjonarności sygnałów $x(n)$ i $d(n)$ określają wzory [2]:

$$M_{LMS} = \alpha\,\mathrm{tr}(\boldsymbol{R}), \tag{8a}$$

$$M_{RLS} = \frac{1-\lambda}{1+\lambda}L, \tag{8b}$$

gdzie $\mathrm{tr}(\boldsymbol{R})$ we wzorze (8a) oznacza ślad macierzy $\boldsymbol{R}$.

Z porównania wzorów (5) i (8a) wynika, że dobór kroku adaptacji α algorytmu LMS wiąże się z kompromisem między szybkością zbieżności algorytmu a jego

niedopasowaniem. Wraz ze wzrostem wartości kroku adaptacji rośnie szybkość zbieżności, ale towarzyszy temu również wzrost niedopasowania. Ze wzoru (8a) wynika ponadto, że algorytm LMS nigdy nie jest w stanie zapewnić minimalnego poziomu mocy sygnału $e(n)$. Natomiast w przypadku stacjonarności sygnałów $x(n)$ i $d(n)$ oraz stacjonarności identyfikowanego systemu, rozwiązania generowane przez algorytm RLS z nieskończoną „pamięcią" ($\lambda = 1$) są zbieżne do optymalnego rozwiązania wienerowskiego. Oznacza to — w tym przypadku — wyższość algorytmu RLS nad algorytmem LMS pod względem niedopasowania.

3. Zagadnienie zdolności do śledzenia przez algorytm niestacjonarności układu nie jest tak jednoznaczne, jak problem jego szybkości zbieżności czy poziomu niedopasowania w warunkach stacjonarności. Aby porównać algorytmy pod tym względem, trzeba uściślić pojęcie niestacjonarności układu w rozważanym kontekście. W pracy [11] wyróżniono trzy typy niestacjonarności: (1) szybkie skokowe zmiany odpowiedzi impulsowej systemu; (2) permanentne zmiany deterministyczne, np. liniowy dryft oraz (3) permanentne zmiany losowe tejże odpowiedzi. Z uwagi na pierwszy rodzaj niestacjonarności, zdolność śledzenia algorytmu należy rozważać w kontekście stanu przejściowego. Czas trwania tego stanu zależy od szybkości zbieżności algorytmu, a więc dla tego rodzaju niestacjonarności analiza jego zdolności do śledzenia sprowadza się do przeprowadzonej wyżej analizy szybkości zbieżności.

Zmiany transmitancji pomieszczenia spowodowane poruszaniem się w nim ludzi biorących udział w telekonferencji mają raczej znamiona niestacjonarności drugiego lub trzeciego typu. W tych przypadkach zdolność śledzenia algorytmu należy rozpatrywać w kategoriach stanu ustalonego [10], istotne jest bowiem wówczas zachowanie algorytmu w czasie trwania zmian charakterystyki układu, a nie tylko po ich ustaniu. W tym przypadku odpowiedź na pytanie, który z algorytmów lepiej sobie radzi ze śledzeniem tych zmian nie jest jednoznaczna, ponieważ zależy to od doboru odpowiednich parametrów algorytmów.

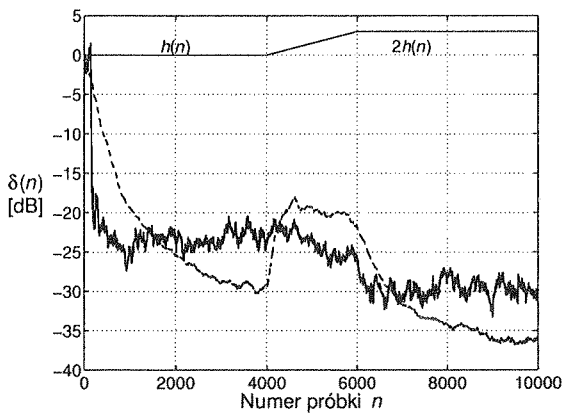
Załóżmy, że zmiany wektora współczynników odpowiedzi impulsowej systemu niestacjonarnego są modelowane procesem Markowa pierwszego rzędu, którego wariancja wynosi σ_N^2. Niedopasowanie rozwiązań generowanych zarówno przez algorytm LMS, jak i RLS wyrażają wtedy dwa składniki [2]:

$$M_{LMS} = \alpha \operatorname{tr}(\boldsymbol{R}) + \frac{L}{4\alpha}\frac{\sigma_N^2}{J_{min}}, \tag{9a}$$

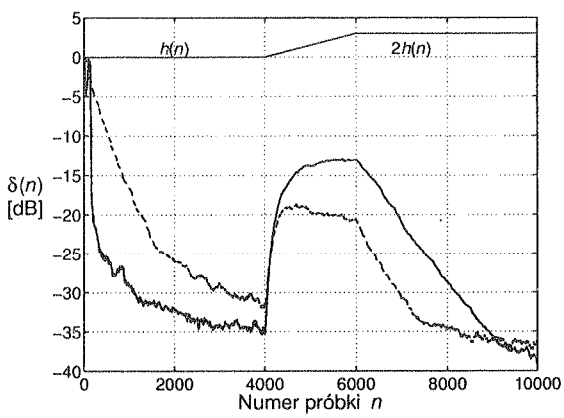
$$M_{RLS} = \frac{1-\lambda}{1+\lambda} L + \frac{1}{2(1-\lambda)} \operatorname{tr}(\boldsymbol{R}) \frac{\sigma_N^2}{J_{min}}. \tag{9b}$$

Pierwsze składniki we wzorach (9a) i (9b) są związane z błędami estymacji, które są wynikiem założeń poczynionych przy wyprowadzaniu algorytmów (zastąpienie gradientu funkcji kosztu jego estymatą w przypadku algorytmu LMS oraz „oknowanie" sygnału

błędu $e(n)$ w przypadku algorytmu RLS). Składniki te określają niedopasowanie w przypadku stacjonarnym. Drugie składniki w obu wzorach wynikają z opóźnienia w śledzeniu przez algorytmy zmian transmitancji systemu. Zarówno w przypadku algorytmu LMS, jak i RLS, zmiana stałej algorytmu (odpowiednio α lub λ), powodująca zmniejszenie jednego ze składników, powoduje natychmiast zwiększenie drugiego. Wzrost wartości kroku adaptacji α w algorytmie LMS, zapewniający lepsze śledzenie niestacjonarności, powoduje również zwiększenie błędu ekscesu $J_{ex} = J_{\infty} - J_{min}$ w czasie, gdy system jest stacjonarny. Podobnie, zmniejszenie wartości stałej zapominania λ w algorytmie RLS, które sprawia, że algorytm lepiej nadąża za zmianami odpowiedzi impulsowej systemu, powoduje jednoczesny wzrost niedopasowania w przypadku stanu stacjonarnego.



(a) $\alpha = 0,003;\ \lambda = 0,990$



(b) $\alpha = 0,003;\ \lambda = 0,999$

Rys. 4. Błąd względny $\delta(n)$ estymacji odpowiedzi impulsowej systemu niestacjonarnego przez algorytmy: LMS (linia przerywana) i RLS (linia ciągła).
Sygnał $x(n)$ — proces AR(5), $\chi(\boldsymbol{R}) = 100$, SNR = 30 dB

W celu potwierdzenia tych właściwości algorytmów LMS i RLS przeprowadzono badania symulacyjne, których wyniki są zamieszczone na rys. 4a i 4b. Na rysunkach tych pokazano krzywe względnego błędu $\delta(n)$ estymacji odpowiedzi impulsowej układu niestacjonarnego. Identyfikację tego układu realizowano za pomocą algorytmu LMS (linia przerywana) oraz algorytmu RLS (linia ciągła). Badano zachowanie się algorytmów w przypadku liniowego dryftu wartości próbek odpowiedzi impulsowej nieznanego układu, które wzrastały dwukrotnie w przedziale czasu od 4000 do 6000 próbek. Jako sygnał pobudzający wybrano proces AR(5) o rozrzucie wartości własnych macierzy korelacji $\chi(\boldsymbol{R}) = 100$. Jako sygnał zakłócający $s(n)$ przyjęto szum biały, którego moc zapewniała stosunek SNR = 30 dB. Dla algorytmu LMS dobrano doświadczalnie największą wartość kroku adaptacji, zapewniającą jego stabilną pracę. Przy takim doborze porównywano zachowanie tego algorytmu w przypadku niestacjonarności układu z zachowaniem algorytmu RLS dla różnych wartości stałej λ. Z porównania tego wynika, że dobór stałej zapominania, który zapewnia mniejszy względny błąd estymacji algorytmu RLS w porównaniu z algorytmem LMS w czasie trwania niestacjonarności, pociąga za sobą większy, niż w przypadku algorytmu LMS poziom tego błędu w przedziałach czasu, kiedy układ możemy uznać za stacjonarny (rys. 4a). Dla odpowiednio większej wartości stałej λ relacje między błędami są odwrotne (rys. 4b).

Podsumowując, należy stwierdzić, że nie ma jednoznacznej odpowiedzi na pytanie, który z algorytmów lepiej nadaje się do pracy w warunkach niestacjonarności identyfikowanego systemu, ponieważ zależy to od doboru odpowiednich parametrów roboczych algorytmów.

4. Złożoność obliczeniowa algorytmu LMS jest rzędu $O(L)$, zaś algorytmu RLS $O(L^2)$ [15]. Prostota algorytmu LMS jest jego dużym atutem i stanowi zasadniczą przyczynę powszechności stosowania go w adaptacyjnych systemach usuwania echa akustycznego. Kwadratowa zależność ilości operacji, koniecznych do wykonania w każdym kroku algorytmu RLS, od rzędu filtru, sprawia, że w przedstawionej w punkcie 2.1 wersji podstawowej nie może on być praktycznie wykorzystany w systemach usuwania echa.

5. Wymóg stabilnej pracy algorytmu w warunkach jednoczesnego nadawania, bądź w przypadku, gdy poziom mocy szumu środowiska jest porównywalny z mocą sygnału nadawanego okazuje się być zbyt wygórowany. Żaden z rozpatrywanych algorytmów nie jest w stanie zapewnić zadowalającego działania w tak niekorzystnych warunkach. Do detekcji stanu jednoczesnego nadawania i zapewnienia odpowiedniego działania w tym stanie stosowane są dodatkowe metody, które w tej pracy nie będą rozpatrywane.

WNIOSKI

Ze względu na swą prostotę, algorytm LMS jest atrakcyjnym narzędziem rekursywnej identyfikacji odpowiedzi impulsowych o długim czasie trwania. Jego podstawową wadą jest wolna zbieżność, zależna od sygnału wejściowego.

Algorytm RLS, charakteryzujący się znacznie większą od algorytmu LMS i niezależną od charakterystyk sygnału wejściowego szybkością zbieżności oraz mniejszym

błędem ekscesu w warunkach stacjonarności identyfikowanego układu, wydaje się być znacznie lepszym od algorytmu LMS. Możliwość zastosowania go w systemach usuwania echa akustycznego w wersji podstawowej wyklucza jednak bardzo duża złożoność obliczeniowa, której przy wysokich rzędach filtru (kilka tysięcy) nie są w stanie sprostać nawet obecnie najszybsze systemy pracujące w czasie rzeczywistym. Algorytmem, który najlepiej spełniałby wszystkie postawione wymagania byłby algorytm, który zapewniłby rozwiązanie takie jak algorytm RLS i miałby jednocześnie złożoność obliczeniową liniowo zależną od rzędu filtru. W kolejnym punkcie omówiono algorytm o takich właśnie właściwościach.

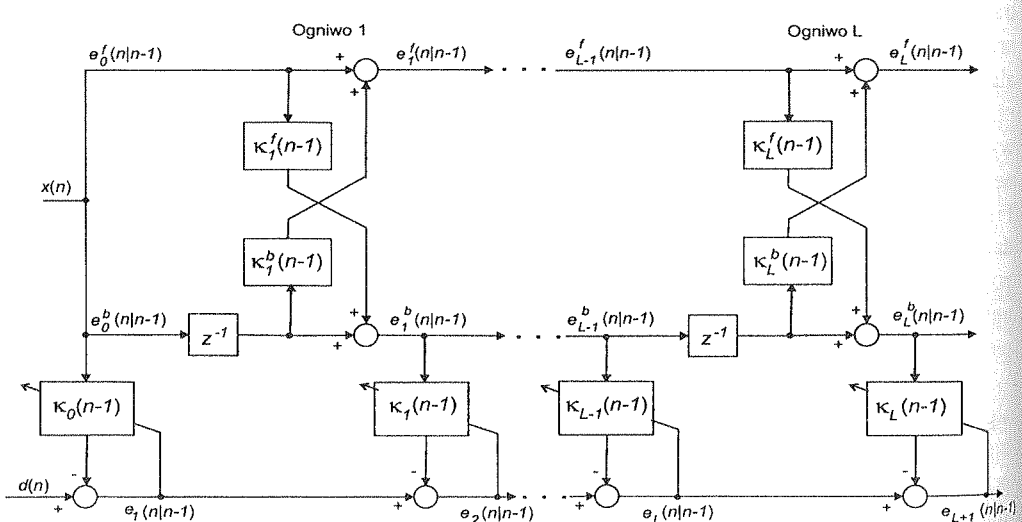
2.3. SZYBKI ALGORYTM KRATOWY LSL

Algorytm RLS, ze względu na dużą i niezależną od sygnału wejściowego szybkość zbieżności oraz mały poziom mocy sygnału błędu estymacji, jest bardzo atrakcyjnym narzędziem rekursywnej identyfikacji odpowiedzi impulsowych nieznanych układów. Jednakże znaczna złożoność obliczeniowa (rzędu ($O(L^2)$)) wyklucza jego aplikację w systemach usuwania echa akustycznego, w których wymagane jest stosowanie filtrów o rzędzie równym kilka tysięcy. Znane są jednak algorytmy, które wyznaczają rekursywne rozwiązanie zagadnienia najmniejszych kwadratów, mając jednocześnie liniowo zależną od długości filtru złożoność obliczeniową. Algorytmy te można podzielić na transwersalne i kratowe. Podstawowym algorytmem pierwszej klasy jest szybki filtr transwersalny FTF (*Fast Transversal Filter*), zaś do drugiej należy szybki algorytm kratowy LSL (*Lattice Least Squares*). Znaczące obniżenie złożoności obliczeniowej algorytmu FTF (złożoność około $7L^2$) uzyskuje się przez zastąpienie kosztownej obliczeniowo procedury uaktualniania macierzy $\boldsymbol{P}_n$, której znajomość pozwala na wyznaczenie wzmocnienia $\boldsymbol{k}_n$, operacjami wykonywanymi przez trzy filtry transwersalne, korzystające z niezmienniczości czasowego przesunięcia danych wejściowych [1]. Filtry te (predykcyjny „w przód" i „wstecz" oraz wzmocnienia) wykorzystują wektory danych wejściowych z chwili aktualnej i poprzedniej, aby wyznaczyć wektor $\boldsymbol{k}_n$ potrzebny do aktualizacji filtru głównego $\boldsymbol{f}_n$. Algorytm FTF jest niestety niestabilny numerycznie, tzn. kumulacja błędów numerycznych, która występuje przy jego implementacji w arytmetyce o skończonej precyzji prowadzi do rozbiegania się algorytmu. W literaturze można znaleźć opis różnych zabiegów, których celem jest zapewnienie algorytmowi FTF stabilności numerycznej, kosztem większego bądź mniejszego zwiększenia jego złożoności obliczeniowej [1, 4, 13]. Badania symulacyjne przeprowadzone przez autora wykazały jednak, że żadna z tych procedur nie jest w stanie zapewnić stabilnej pracy filtru FTF przy długotrwałym pobudzaniu sygnałem mowy.

Alternatywą dla stabilizowanych numerycznie wersji algorytmu FTF, w skrócie oznaczanych w literaturze przez SFTF, jest szybki algorytm kratowy LSL [3, 9]. Podobnie jak w przypadku szybkiego filtru transwersalnego, złożone obliczenia ak-

[2] Złożoność obliczeniowa jest tu rozumiana jako liczba operacji mnożenia i dzielenia konieczna do przetworzenia jednej próbki sygnału wejściowego

tualizacji odwrotności macierzy autokorelacji sygnału wejściowego są realizowane z wykorzystaniem filtrów predykcyjnych. Różnica polega na zastąpieniu transwersalnych predyktorów „w przód" i „wstecz" ich kratowymi odpowiednikami, co pociąga za sobą również inny sposób obliczania błędu estymacji $e(n)$. W strukturze kratowej (rys. 5) nie występuje jawnie wektor zawierający estymowane próbki odpowiedzi impulsowej (odpowiednik wektora $\boldsymbol{f}_n$ w strukturze transwersalnej), który można jednakże wyznaczyć za pomocą odpowiedniej transformacji liniowej wektora *współczynników regresji* $[\kappa_0,\ldots,\kappa_L]^T$. Iloczyn skalarny tego wektora z wektorem błędów predykcji „wstecz" *a priori* $[e_0^b(n|n-1),\ e_1^b(n|n-1),\ldots,e_L^b(n|n-1)]^T$, obliczany w kolejnych chwilach czasu n, stanowi estymatę sygnału wzorcowego $d(n)$. Podobnie, istnieje możliwość przekształcenia *współczynników odbicia* struktury kratowej κ_m^f i κ_m^b $(m = 1,\ldots,L)$ na odpowiednie wektory transwersalnych współczynników predykcji. W tym przypadku jest to jednak transformacja nieliniowa. Ze względu jednak na to, że w systemach usuwania echa akustycznego nie ma potrzeby, aby wartości próbek estymowanej odpowiedzi impulsowej były bezpośrednio znane, a istotny jest tylko sygnał błędu estymacji $e(n)$, transformacje te nie są konieczne i w systemach tych można zastosować strukturę *estymatora procesu łącznego* (ang. *joint process estimator*) [9, 15], opartą na algorytmie LSL. Na rys. 5 pokazano tę strukturę dla przypadku algorytmu LSL zaimplementowanego w wersji opartej na wykorzystaniu błędów predykcji i estymacji *a priori*, którego równania zamieszczono w tabeli 3. Wersja ta — w odróżnieniu od niektórych innych — charakteryzuje się bardzo dobrą stabilnością numeryczną[3]. Jej złożoność obliczeniowa jest większa niż algorytmu SFTF i wynosi około $30L$.



Rys. 5. Kratowa struktura estymatora procesu łącznego jako implementacja algorytmu LS wykorzystującego błędy predykcji i estymacji *a priori*

[3] Przez bardzo dobrą stabilność numeryczną autor rozumie stabilną pracę algorytmu w dowolnie niekorzystnych warunkach. Często w literaturze można spotkać algorytmy uznane za pracujące stabilnie, których właściwości zbadano niestety tylko w określonych, korzystnych dla nich warunkach.

Charakterystyczną cechą algorytmów kratowych jest występowanie dwóch typów rekursji. Rekursje pierwszego typu, podobnie jak w przypadku algorytmów transwersalnych, uaktualniają zmienne algorytmu po czasie, drugiego zaś typu zmienne względem rzędu filtru. Stąd też pochodzi często spotykana w literaturze anglosaskiej nazwa tych filtrów — *order-recursive adaptive filters*. W rekursjach po rzędzie filtru do obliczenia błędu predykcji predyktora rzędu $m+1$ konieczna jest znajomość błędu predyktora rzędu m. Cecha ta jest bardzo wygodna, bowiem jak wiadomo [3] dołączenie kolejnych ogniw (zwiększenie rzędu filtru) odbywa się bez konieczności przeliczania wcześniejszych współczynników struktury kratowej od nowa, co w niektórych zastosowaniach jest bardzo pożądane. Rozważany tu algorytm LSL cechuje charakterystyczna dla algorytmów klasy LS duża szybkość zbieżności (taka, jak klasycznego algorytmu RLS), a ponadto liniowo zależna od rzędu filtru złożoność obliczeniowa oraz atrakcyjna z punktu widzenia implementacji modularność. Właściwości te sprawiają, że jest on istotnym narzędziem z punktu widzenia aplikacji w systemach usuwania echa akustycznego.

Tabela 3

Rekursje algorytmu LSL

Inicjalizacja pętli czasowej: dla $m = 1,\ldots,L$

$\varepsilon^f_{m-1}(0) = \delta, \quad \varepsilon^b_{m-1}(-1) = \delta, \quad \delta$ — mała liczba dodatnia

$\kappa^f_m(0) = \kappa^b_m(0) = 0 \quad \kappa_{m-1}(0) = 0$

Inicjalizacja pętli po rzędzie: dla $n \geq 1$

$e^f_0(n|n-1) = e^b_0(n|n-1) = x(n) \quad e_0(n|n-1) = d(n) \quad \gamma_0(n-1) = 1$

Predykcja: dla $n \geq 1$ i $m = 1,\ldots,L$

$$\varepsilon^f_{m-1}(n) = \lambda\varepsilon^f_{m-1}(n-1) + \gamma_{m-1}(n-1)[e^f_{m-1}(n|n-1)]^2$$

$$\varepsilon^b_{m-1}(n-1) = \lambda\varepsilon^b_{m-1}(n-2) + \gamma_{m-1}(n-1)\,[e^b_{m-1}(n-1|n-2)]^2$$

$$e^f_m(n|n-1) = e^f_{m-1}(n|n-1) + \kappa^f_m(n-1)e^b_{m-1}(n-1|n-2)$$

$$e^b_m(n|n-1) = e^b_{m-1}(n-1|n-2) + \kappa^b_m(n-1)e^f_{m-1}(n|n-1)$$

$$\kappa^f_m(n) = \kappa^f_m(n-1) - \frac{\gamma_{m-1}(n-1)e^b_{m-1}(n-1|n-2)}{\varepsilon^b_{m-1}(n-1)}\,e^f_m(n|n-1)$$

$$\kappa^b_m(n) = \kappa^b_m(n-1) - \frac{\gamma_{m-1}(n-1)e^f_{m-1}(n|n-1)}{\varepsilon^f_{m-1}(n)}\,e^b_m(n|n-1)$$

$$\gamma_m(n-1) = \gamma_{m-1}(n-1) - \frac{\gamma^2_{m-1}(n-1)[e^b_{m-1}(n-1|n-2)]^2}{\varepsilon^b_{m-1}(n-1)}$$

Filtracja: dla $n \geq 1$ i $m = 1,\ldots,L$

$$e_m(n|n-1) = e_{m-1}(n|n-1) - \kappa_{m-1}e^b_{m-1}(n|n-1)$$

$$\kappa_{m-1}(n) = \kappa_{m-1}(n-1) + \frac{\gamma_{m-1}(n)e^b_{m-1}(n|n-1)}{\varepsilon^b_{m-1}(n)}\,e^f_m(n|n-1)$$

3. PRZETWARZANIE W PODPASMACH JAKO SPOSÓB REDUKCJI ZŁOŻONOŚCI OBLICZENIOWEJ FILTRÓW ADAPTACYJNYCH O WYSOKICH RZĘDACH

Jak podkreślono we wprowadzeniu, adaptacyjne usuwanie echa akustycznego wymaga identyfikacji odpowiedzi impulsowej pomieszczenia o bardzo dużej długości. Rząd adaptacyjnego filtru SOI realizującego proces identyfikacji może sięgać nawet kilku tysięcy. Realizacja jego strojenia przy użyciu nawet najprostszych algorytmów adaptacyjnych wymaga wykonania bardzo dużej ilości operacji arytmetycznych w czasie jednego okresu próbkowania T_S. Technika filtracji adaptacyjnej w podpasmach stanowi sposób zmniejszenia złożoności obliczeniowej algorytmów rekursywnej identyfikacji długich odpowiedzi impulsowych. W literaturze spotyka się najczęściej adaptacyjne systemy podpasmowe pracujące z filtrami NLMS. Algorytm NLMS różni się od klasycznego algorytmu LMS jedynie normalizacją wektora poprawki współczynników filtru $\Delta \boldsymbol{f}_n = \alpha e(n)\boldsymbol{x}_n$ względem estymaty chwilowej mocy sygnału $x(n)$. Za estymatę mocy chwilowej przyjmuje się kwadrat normy euklidesowej wektora $\boldsymbol{x}_n$, co prowadzi do następującej postaci podstawowej rekursji algorytmu:

$$\boldsymbol{f}_{n+1} = \boldsymbol{f}_n + \frac{\alpha}{\varrho + \|\boldsymbol{x}_n\|^2} e(n)\boldsymbol{x}_n = \boldsymbol{f}_n + \frac{\alpha}{\varrho + \boldsymbol{x}_n^T \boldsymbol{x}_n} e(n)\boldsymbol{x}_n. \tag{10}$$

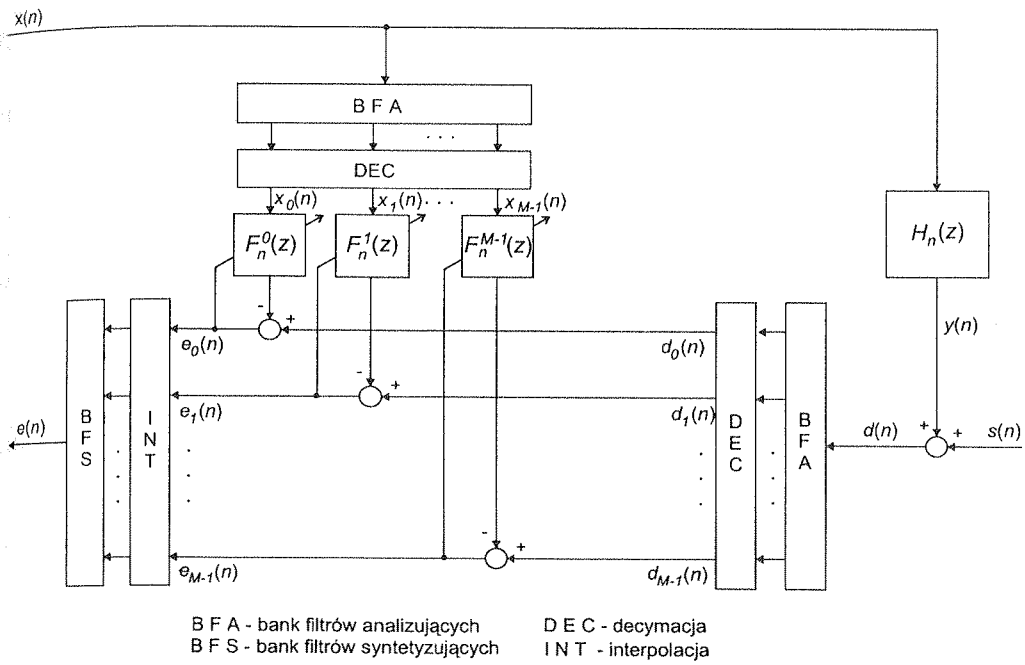
Stała ϱ jest małą liczbą zapewniającą, że mianownik będzie różny od zera w przypadku gdy wektor danych $\boldsymbol{x}_n = \boldsymbol{0}$ (co jest równoznaczne z brakiem pobudzenia filtru NLMS).

Autorowi niniejszej pracy znany jest tylko jeden artykuł, gdzie zaproponowano zastosowanie w systemie podpasmowym algorytmu klasy LS [8]. Omówiono tam system podpasmowy z filtrami FTF, które poddawano asynchronicznej reinicjalizacji w celu zapewnienia stabilnej pracy. Zastosowanie filtrów LSL w systemie podpasmowym prowadzi do uzyskania nie tylko mniejszej złożoności obliczeniowej przy zachowaniu immanentnej dla tego algorytmu stabilności numerycznej, ale również do złagodzenia efektu wzrostu błędu estymacji generowanego przez algorytm LSL w fazie „wchodzenia sygnału w filtr". Szerzej zjawisko to zostanie omówione w części 4, w której przedstawione będą wyniki badań symulacyjnych.

3.1. KONCEPCJA FILTRACJI ADAPTACYJNEJ W PODPASMACH

Struktura *adaptacyjnego filtru podpasmowego* (SAF — *Subband Adaptive Filter*) została przedstawiona na rys. 6. Sygnały $x(n)$ i $d(n)$ poddawane są dekompozycji na sygnały podpasmowe. Na proces dekompozycji składają się dwa etapy. Pierwszy — analizy, który przeprowadzany jest przez *bank filtrów analizujących* BFA oraz drugi — *decymacji* DEC. W każdym z M kanałów umieszczony jest filtr adaptacyjny $F^i(z)$, $i = 0,\ldots,M-1$ o rzędzie znacznie mniejszym od długości L identyfikowanej odpowiedzi impulsowej. Każdy z filtrów adaptacyjnych generuje własny błąd estymacji. Zatem, aby

uzyskać sumaryczny sygnał błędu estymacji (ang. *full band estimation error*), należy dokonać syntezy podpasmowych sygnałów błędu. Do tego celu konieczny jest blok interpolacji i bank syntezy sygnałów podpasmowych.



Rys. 6. Struktura podpasmowego filtru adaptacyjnego (SAF) zastosowana do usuwania echa akustycznego

Znacząca redukcja złożoności obliczeniowej systemu podpasmowego wynika z obniżenia częstotliwości próbkowania w poszczególnych kanałach. Przy zastosowaniu *krytycznego próbkowania* (współczynnik decymacji N jest równy liczbie podpasm M), zarówno rzędy filtrów adaptacyjnych w podpasmach, jak i częstotliwość ich aktualizacji maleje M-krotnie. W idealnym przypadku zatem, jeśli zaniedbamy obliczenia konieczne do przeprowadzenia analizy i syntezy sygnałów podpasmowych, złożoność obliczeniowa algorytmu filtracji podpasmowej maleje M-krotnie w stosunku do algorytmu pracującego w dziedzinie czasu w całym pasmie. Jeśli współczynnik decymacji będzie mniejszy od liczby podpasm ($N<M$), co oznacza zastosowanie *nadpróbkowania* sygnałów podpasmowych, to złożoność obliczeniowa nadal maleje proporcjonalnie do liczby podpasm, lecz ze współczynnikiem proporcjonalności mniejszym od jedności. Dalsze obniżenie zysku obliczeniowego wynika z konieczności filtracji analizującej i syntetyzującej sygnały podpasmowe.

Technika filtracji adaptacyjnej w podpasmach posiada niestety dwie zasadnicze wady:

1. wprowadzanie przez bloki analizy i syntezy sygnałów podpasmowych opóźnienia, które w systemie usuwania echa akustycznego nie może przekroczyć pewnego progu (próg ten jest związany z takim opóźnieniem, które jest już zauważalne przez użytkownika);
2. zmniejszenie szybkości zbieżności algorytmu wynikające ze zjawiska *aliasingu*, które jest spowodowane niedoskonałością filtrów wchodzących w skład banków analizy sygnałów podpasmowych.

Obydwie niedogodności są ze sobą powiązane. Filtry zapewniające duży poziom tłumienia w pasmie zaporowym i charakteryzujące się stromym zboczem przejściowym charakterystyki amplitudowej (cechy te osłabiają niekorzystny wpływ aliasingu), muszą mieć dostatecznie wysoki rząd. Filtry analizujące wysokich rzędów powodują z kolei duże opóźnienia i, co nie jest bez znaczenia, zwiększają złożoność obliczeniową całego algorytmu. Samo zjawisko aliasingu jest na tyle poważne, że wymaga stosowania specjalnych dodatkowych sposobów zmniejszania jego niekorzystnego oddziaływania na pracę całej struktury podpasmowego filtru adaptacyjnego. Zarówno ze studiów literaturowych, jak i z doświadczenia autora wynika, że najlepszym sposobem walki z aliasingiem w systemach adaptacyjnej filtracji podpasmowej jest zastosowanie do analizy/syntezy sygnałów podpasmowych banku filtrów z nadpróbkowaniem. Należy dodać, że proponowane są także inne metody redukcji aliasingu w tego typu systemach, jak np. zastosowanie skrośnych filtrów adaptacyjnych [5], czy wykorzystanie filtrów NOI jako banku analizy/syntezy sygnałów podpasmowych [12].

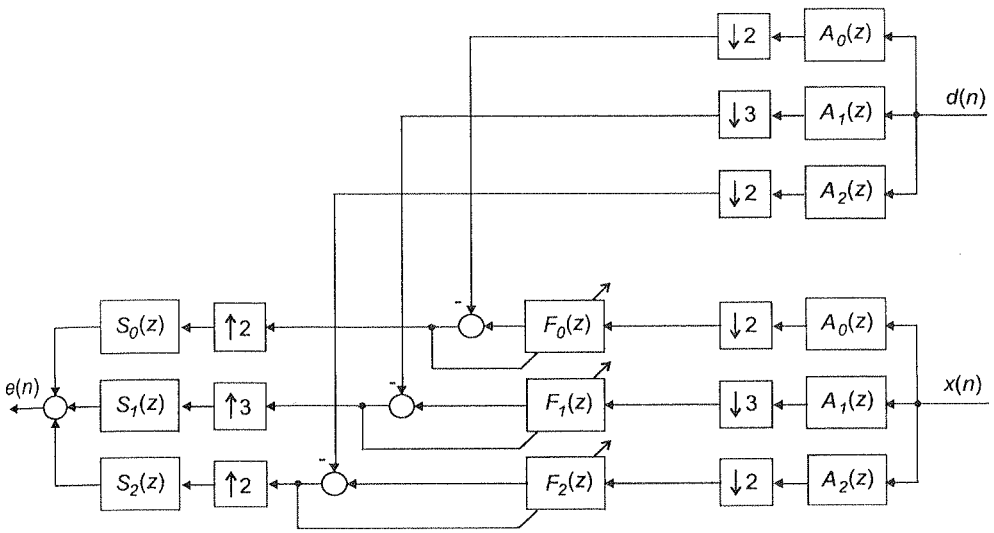
3.2. ZASTOSOWANIE ALGORYTMU LSL W PODPASMOWYM SYSTEMIE FILTRACJI ADAPTACYJNEJ Z NADPRÓBKOWANIEM

Jak podkreślono wcześniej, w podpasmowych systemach filtracji adaptacyjnej najczęściej wykorzystuje się algorytm NLMS. System taki ma nie tylko mniejszą złożoność obliczeniową, ale również większą szybkość zbieżności w porównaniu z algorytmem NLMS pracującym w pełnym pasmie. Korzyści wynikające z zastosowania filtrów adaptacyjnych typu LSL w systemie podpasmowym są natomiast następujące:

1. podobnie, jak w przypadku zastosowania filtrów NLMS, maleje złożoność obliczeniowa algorytmu podpasmowego w porównaniu z algorytmem LSL pracującym w pełnym pasmie;
2. wartość tłumienia echa w początkowym okresie pracy algorytmu jest większa niż wartość tłumienia uzyskiwana przy zastosowaniu algorytmu pracującego w pełnym pasmie;
3. szybkość zbieżności podpasmowego systemu filtracji adaptacyjnej z filtrami LSL jest większa niż dla algorytmów NLMS — klasycznego i podpasmowego.

Na rys. 7 przedstawiono schemat trójkanałowego systemu podpasmowej filtracji adaptacyjnej zaproponowanego przez Hartenecka w pracy [7], w którym do analizy i syntezy sygnałów podpasmowych wykorzystano bank filtrów z nadpróbkowaniem.

Oryginalność podejścia Hartenecka do zagadnienia projektowania banku filtrów z nadpróbkowaniem polega na zastosowaniu różnych współczynników decymacji w poszczególnych kanałach. Prowadzi to do uzyskania minimalnego narzutu obliczeniowego związanego z nadpróbkowaniem.



Rys. 7. Schemat trójkanałowego adaptacyjnego filtru podpasmowego z nadpróbkowaniem

Charakterystyki amplitudowe filtrów analizujących wchodzących w skład układu pokazanego na rys. 7 spełniają następujące warunki. Pasma przepustowe filtrów $A_0(z)$ i $A_2(z)$ nie zawierają pulsacji unormowanej równej $\pi/2$. Szerokość ich pasm przepustowych jest wówczas mniejsza od połowy szerokości całego pasma, co przy zastosowaniu współczynnika decymacji równego 2 (dla kanałów 0 i 2) prowadzi do nadpróbkowania sygnałów podpasmowych w tych kanałach. Pasmo przepustowe filtru $A_1(z)$ zawiera się natomiast w przedziale $(\pi/3;\ 2\pi/3)$ pulsacji unormowanych, zatem jego szerokość jest mniejsza od 1/3 szerokości całego pasma. Można więc zastosować współczynnik decymacji równy 3, unikając jednocześnie niepożądanego zjawiska aliasingu.

Poniżej przeprowadzimy analizę złożoności obliczeniowej podpasmowego filtru adaptacyjnego z filtrami typu LSL, wykorzystującego do analizy i syntezy sygnałów podpasmowych bank filtrów z nadpróbkowaniem. Przez L_a i L_s oznaczymy rzędy filtrów analizującego i odpowiednio syntetyzującego sygnały podpasmowe, a przez L_f rząd filtru adaptacyjnego pracującego w pełnym pasmie, wymagany do estymacji odpowiedzi impulsowej pomieszczenia z akceptowalną dokładnością. Jak wspomniano w p. 2.3 złożoność obliczeniowa algorytmu LSL pracującego w pełnym pasmie wynosi $C_f = 30L_f$. Z kolei można pokazać, że złożoność obliczeniowa algorytmu podpasmowego wyraża się następującym wzorem:

$$C_s = 2 \cdot 3 \cdot L_a + 3 \cdot L_s + 2 \cdot \left(\frac{30L_f}{2 \cdot 2}\right) + \frac{30L_f}{3 \cdot 3}. \tag{11a}$$

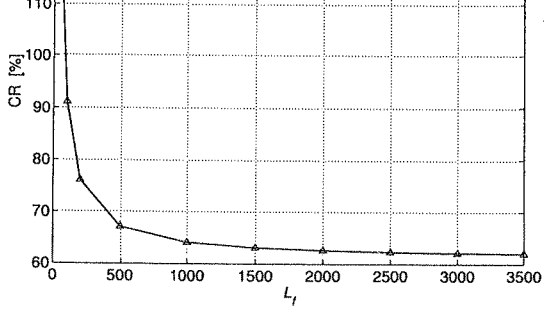
Przy założeniu, że rzędy filtrów analizującego i syntetyzującego są takie same i wynoszą $L(L_a = L_s = L)$, złożoność obliczeniową algorytmu podpasmowego można obliczyć korzystając z formuły:

$$C_s = 9L + \frac{11}{18} \cdot 30L_f. \tag{11b}$$

Współczynnik złożoności obliczeniowej CR (ang. *complexity ratio*), definiowany jako stosunek złożoności algorytmu podpasmowego do złożoności algorytmu pracującego w pełnym pasmie, jest zatem opisany wzorem:

$$\mathrm{CR} = \frac{C_s}{C_f} \cdot 100\% = \left(\frac{11}{18} + \frac{3}{10} \cdot \frac{L}{L_f}\right) \cdot 100\%. \tag{12}$$

Na rys. 8 pokazano wykres zależności współczynnika CR od długości estymowanej odpowiedzi impulsowej L_f, przy założeniu, że rząd filtrów analizującego i syntetyzującego jest stały i wynosi $L = 100$. Z wykresu tego wynika, że zastosowanie struktury podpasmowej jest już opłacalne przy $L_f = 100$. Dla długich odpowiedzi (rzędu kilka tysięcy), jak ma to miejsce w przypadku estymacji odpowiedzi impulsowej pomieszczenia w systemie usuwania echa akustycznego, współczynnik złożoności obliczeniowej przy wykorzystaniu banku trójkanałowego dochodzi do 62%. Oznacza to zmniejszenie nakładu obliczeniowego o 38%. Stosując do analizy i syntezy sygnałów podpasmowych banki filtrów o większej ilości kanałów, uzyskuje się jeszcze większy zysk obliczeniowy, jednak wraz ze wzrostem liczby podpasm poprawa ta jest niweczona przez obniżenie szybkości zbieżności spowodowane obecnością składowych aliasingowych.



Rys. 8. Zależność współczynnika złożoności obliczeniowej CR od długości estymowanej odpowiedzi impulsowej przy stałej długości filtrów analizy/syntezy

4. WYNIKI BADAŃ SYMULACYJNYCH

Omówiony wyżej system filtracji adaptacyjnej z filtrami LSL, wykorzystujący do analizy i syntezy sygnałów podpasmowych bank filtrów z nadpróbkowaniem, będziemy oznaczać SUB–LSL. System ten poddano porównawczym badaniom symulacyjnym, w których użyto zarówno sygnałów syntetycznych, jak i rzeczywistego sygnału mowy. Właściwości algorytmu SUB-LSL porównano z właściwościami algorytmów NLMS i LSL pracujących w pełnym pasmie oraz podpasmowego algorytmu NLMS — w skrócie SUB–NLMS.

4.1. WYNIKI UZYSKANE PRZY WYKORZYSTANIU SYGNAŁÓW SYNTETYCZNYCH

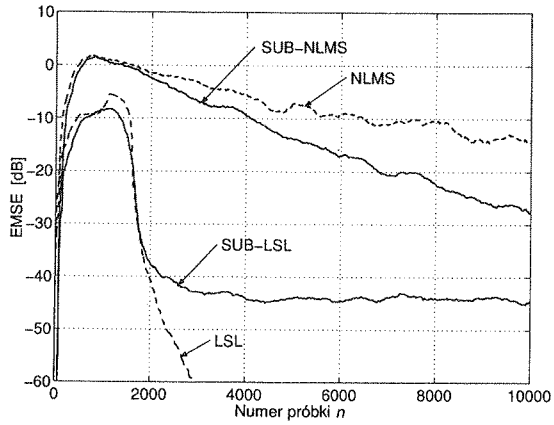
Do badania podpasmowych oraz pracujących w pełnym pasmie algorytmów NLMS i LSL użyto sygnału AR(5) o rozrzucie wartości własnych macierzy autokorelacji $\chi(\boldsymbol{R}) = 100$, zaś jako identyfikowaną odpowiedź impulsową wykorzystano 1000 pierwszych próbek rzeczywistej odpowiedzi impulsowej pomieszczenia biurowego. W eksperymencie wykorzystano filtry analizujące i syntetyzujące o rzędzie $L_a = L_s = 73$. Parametrem, który poddano ocenie był empiryczny błąd średniokwadratowy — EMSE, który jest określony ogólnym wzorem:

$$\mathrm{EMSE}(n) = \frac{1}{K}\sum_{k=0}^{K-1}[e^{(k)}(n)]^2, \tag{13}$$

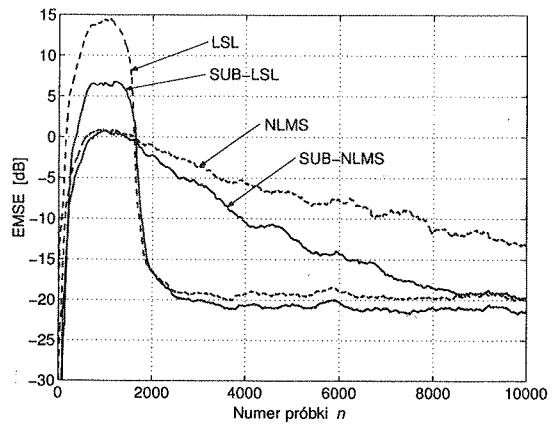
gdzie k oznacza numer realizacji sygnału $e(n)$, zaś K liczbę wszystkich realizacji podlegających uśrednianiu. W przeprowadzonych badaniach przyjęto $K = 1$, tzn. nie dokonywano uśredniania.

Na rys. 9 i 10 pokazano trajektorie empirycznego błędu średniokwadratowego EMSE uzyskane przy użyciu do procesu identyfikacji odpowiedzi impulsowej nieznanego układu stacjonarnego algorytmów: NLMS, LSL, SUB-NLMS oraz SUB-LSL. Trajektorie przedstawione na rys. 9 uzyskane zostały przy braku szumu zakłócającego (SNR = ∞), zaś trajektorie z rys. 10 przy stosunku sygnał–szum równym 30 dB. Stała adaptacji α dla algorytmów typu NLMS oraz współczynnik zapominania λ dla algorytmów typu LSL miały w obydwu eksperymentach taką samą wartość i wynosiły odpowiednio: $\alpha = 1$, $\lambda = 0{,}998$. W celu wygładzenia trajektorii EMSE, sygnał błędu średniokwadratowego poddawano filtracji za pomocą filtru SOI o stałej odpowiedzi impulsowej $g(n) = 0{,}005$ dla $n = 0,\ldots,499$.

Jak wynika z rys. 9, zastosowanie techniki podpasmowej w odniesieniu do algorytmu NLMS owocuje znaczącym przyspieszeniem zbieżności. W przypadku algorytmu LSL nie obserwujemy tego zjawiska. Co więcej, algorytm podpasmowy generuje większy błąd ekscesu. Jest to skutkiem zjawiska aliasingu, błąd ekscesu zależy bowiem od poziomu tłumienia filtrów wchodzących w skład banków analizy i syntezy. Zaletami



Rys. 9. Trajektorie empirycznego błędu średniokwadratowego EMSE estymacji odpowiedzi impulsowej uzyskane przy zastosowaniu podpasmowych (linie ciągłe) i pracujących w pełnym pasmie (linie przerywane) algorytmów NLMS i LSL (SNR = ∞)



Rys. 10. Trajektorie empirycznego błędu średniokwadratowego EMSE estymacji odpowiedzi impulsowej uzyskane przy zastosowaniu podpasmowych (linie ciągłe) i pracujących w pełnym pasmie (linie przerywane) algorytmów NLMS i LSL (SNR = 30 dB)

podpasmowej wersji algorytmu LSL są jednak niewątpliwie mniejsza złożoność obliczeniowa oraz mniejszy poziom błędu EMSE osiągany w początkowym okresie pracy algorytmu. W obecności szumu zakłócającego druga zaleta jest znacznie łatwiej zauważalna. Na rys. 10 można zobaczyć, że poziom błędu EMSE dla algorytmu SUB-LSL jest w początkowej fazie jego pracy o około 5 dB mniejszy niż dla pełnopasmowego algorytmu LSL. Okazuje się również, że w obecności sygnału zakłócającego, poziomy błędu EMSE uzyskiwane przez algorytmy typu LSL w stanie ustalonym[4] mają bardzo

[4] Stan ustalony oznacza tu okres pracy algorytmu po osiągnięciu zbieżności, przy założeniu stacjonarności sygnałów $x(n)$ i $d(n)$ oraz stacjonarności identyfikowanego systemu.

podobne wartości. Zastowanie algorytmu LSL w systemie podpasmowym prowadzi zatem do obniżenia jego złożoności obliczeniowej oraz do poprawy jego działania w początkowym okresie pracy.

4.2. WYNIKI UZYSKANE PRZY WYKORZYSTANIU RZECZYWISTEGO SYGNAŁU MOWY

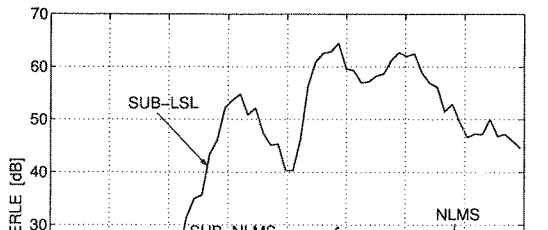
Badania symulacyjne przeprowadzono również z wykorzystaniem rzeczywistego sygnału mowy oraz zmierzonej odpowiedzi impulsowej pomieszczenia biurowego. W eksperymencie użyto dwusekundowej frazy (16 000 próbek) składającej się z dwóch angielskich słów: *substitute* i *convenience*, wypowiedzianych męskim głosem. Odpowiedź impulsowa pomieszczenia biurowego o wymiarach: szerokość 3,5 m, długość 5,2 m, wysokość 2,8 m, miała czas trwania równy 350 ms, co przy częstotliwości próbkowania 8 kHz daje 2800 próbek. W przypadku użycia algorytmu podpasmowego filtry adaptacyjne w poszczególnych kanałach miały następujące rzędy: $L_{f0} = L_{f2} = 1500$, $L_{f1} = 1000$. Stała adaptacji w przypadku algorytmów typu NLMS wynosiła $\alpha = 0{,}7$, zaś współczynnik zapominania algorytmu SUB-LSL miał wartość $\lambda = 0{,}995$.

Do oceny algorytmów usuwania echa akustycznego pracujących z rzeczywistym sygnałem mowy jako sygnałem wejściowym wykorzystano powszechnie stosowany do tego celu parametr oznaczany akronimem ERLE (ang. *Echo Return Loss Enhancement*). Wskaźnik ten jest obliczany zwykle nie dla każdej chwili czasu, lecz dla bloków trwających 32 ms:

$$\mathrm{ERLE}(n) = 10 \log_{10} \frac{\sum_{i=0}^{P-1} y^2(nP+i)}{\sum_{i=0}^{P-1} e^2(nP+i)}, \qquad P = 256. \tag{14}$$

Na rys. 11 przedstawiono trajektorie tłumienia echa ERLE, uzyskane przy wykorzystaniu omawianych w pracy algorytmów. Przedstawione na rysunku wyniki dowodzą niezbicie wyższości algorytmu SUB-LSL zarówno nad podpasmowym algorytmem SUB-NLMS, jak i klasycznym pracującym w pełnym pasmie algorytmem NLMS. Algorytm SUB-LSL zapewnia już po 750 ms pracy poziom tłumienie echa równy 45 dB, który jest wymagany przez standard G.167, podczas gdy tłumienie uzyskiwane przez dwa pozostałe badane w tej serii eksperymentów algorytmy nie osiągają tej granicy podczas całego doświadczenia. Nawet w newralgicznym początkowym okresie pracy algorytm SUB-LSL osiąga tłumienie echa wyższe niż pozostałe algorytmy.

Opisane powyżej doświadczenia symulacyjne dowodzą, że algorytm filtracji podpasmowej z filtrami LSL pracującymi w poszczególnych kanałach może być z powodzeniem stosowany w adaptacyjnych systemach usuwania echa akustycznego.



Rys. 11. Trajektorie tłumienia echa ERLE uzyskane przy zastosowaniu algorytmów podpasmowych (linie ciągłe) oraz algorytmu NLMS (linia przerywana), które pobudzano rzeczywistym sygnałem mowy

5. PODSUMOWANIE

W artykule przedstawiono podpasmowy algorytm filtracji adaptacyjnej wykorzystujący filtry kratowe typu LSL. Zaprezentowano dwie podstawowe klasy algorytmów adaptacyjnych — LMS i RLS, a następnie dokonano ich krytycznego porównania pod kątem zastosowania w adaptacyjnym systemie usuwania echa akustycznego. Przeprowadzona analiza porównawcza tych dwóch klas algorytmów pozwoliła na sformułowanie wymagań, jakie powinien spełniać algorytm zastosowany w systemie usuwania echa. Omówiono algorytm kratowy LSL, który zdaniem autora w dużej mierze spełnia te wymagania. Wykazano, że zastosowanie algorytmu LSL w systemie podpasmowym dodatkowo poprawia niektóre właściwości systemu usuwania echa, istotne z praktycznego punktu widzenia. Stosując algorytm podpasmowy SUB-LSL, uzyskuje się większą szybkość zbieżności oraz wyższe tłumienie echa niż ma to miejsce przy zastosowaniu algorytmów NLMS i SUB-NLMS. W porównaniu z algorytmem LSL pracującym w pełnym pasmie, algorytm SUB-LSL ma mniejszą złożoność obliczeniową oraz uzyskuje większy poziom tłumienia w fazie „wchodzenia sygnału w filtr". Wyżej wymienione cechy algorytmu SUB-LSL potwierdziły badania symulacyjne, w których wykorzystano zarówno sygnały syntetyczne, jak i rzeczywisty sygnał mowy.

BIBLIOGRAFIA

1. J. M. Ciofii, T. Kailath: *Fast, recursive-least-squares transversal filters for adaptive filtring.* IEEE Trans. on Speech and Audio Processing, vol. ASSP-32, no 5, pp. 304–337
2. E. Eleftheriou, D. D. Falconer: *Tracking properties and steady-state performance of RLS adaptive filter.* IEEE Trans. on Speech and Audio Processing, vol. ASSP-34, no 2, pp. 1097–1109
3. B. Friedlander: *Lattice filters for adaptive processing algorithms.* Proceedings of IEEE, vol. 70, no 8, pp. 829–867

4. A. Gilloire, T. Petillon: *Signal processing V: theories and application. A comparison of NLMS and fast RLS algorithms for the identification of time-varying systems with noisy outputs. Application to acoustic echo cancellation.* Elsevier Amsterdam, 1990, pp. 417–420
5. A. Gilloire, M. Vetterli: *Adaptive filtring in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation.* IEEE Trans. on Signal Processing, vol. 40, no 8, pp. 1862–1875
6. E. Hänsler: *The hands-free telephone problem.* ISCAS-92, San Diego (USA), 1992, pp. 1914–1917
7. M. Harteneck: *Real valued filter banks for subband adaptive filters.* PhD thesis, University of Strathclyde, april 1998
8. B. Hätty: *Recursive least squares algorithms using multirate systems for cancellation of acoustical echoes.* ICASSP' 90, Albuquerque (New Mexico), 1990, pp. 1145–1148
9. S. Haykin: *Adaptive filter theory.* New York, Prentice-Hall, 1991
10. S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, P. C. Wei: *Adaptive tracking of linear time-variant systems by extended RLS algorithms.* IEEE Trans. on Signal Processing, vol. 45, no 5, pp. 1118–1128
11. O. Macchi: *Signal processing V: theories and application. A general methodology for comparison of adaptive filtering algorithms in a nonstationary context.* Elsevier Amsterdam, 1990, pp. 189–192
12. P. A. Naylor, O. Tanrikulu, A. G. Constantinides: *Subband adaptive filtering for acoustic echo control using allpass polyphase IIR filterbanks.* IEEE Trans. on Speech and Audio Processing, vol. 6, no 2, pp. 143–155
13. Z. Ren, I. Hartmann, H. Schütze: *A fast exact FTF adaptive algorithm.* Annals of Telecommunication, vol. 49, no 7–8, pp. 398–406
14. L. Rutkowski: *Filtry adaptacyjne i adaptacyjne przetwarzanie sygnałów.* Warszawa, WNT, 1994
15. J. Szabatin: *Optymalne przetwarzanie sygnałów.* Niepublikowany konspekt wykładu, 1995

J. FALKIEWICZ

APPLICATION OF LATTICE LS-TYPE ALGORITHM IN SUBBAND ADAPTIVE ECHO CANCELLATION SYSTEM

Summary

In this paper the problem of adaptive echo cancellation is presented. Classical adaptive algorithms (LMS and RLS) are discussed and compared from the point of view of their application in acoustic echo cancellation systems. The key requirements for acoustic echo cancellation algorithms are formulated. On the one hand fast and signal independent convergence is necessary, on the other hand low computational complexity and numerical stability are needed. It has been shown that LSL algorithm meet this requirements well. In comparison with fullband LSL algorithm, LSL in subband scheme leads to decrease of computational complexity and to improvement of echo suppression at the work beginning. The proposed algorithm has been verified via computer simulations using both synthetic and real signals. Simulation results confirm the good performance of the subband adaptive filtering system with LSL filters in subbands.

Keywords: acoustic echo cancellation, adaptive filtering, hands-free telephony

Prawd

W
poziom
nieskorel
mo-norm
operacyjn
tak jak
niezależn
potwierdz
zakłóceń

Słowa klu

Jednym z
jęcie decyzji
wykrywanego
a procedura d
detektora. Por
magnetycznyc
dyskretną zm
powiednio:
$d = 0$ — obec
$d = 1$ — obec
Przyjęcie każd
tabela 1, w k

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2001, 47, z. 2, ss. 209–219

Prawdopodobieństwo fałszywego alarmu detektora log-t w obecności skorelowanych zakłóceń typu K

ANDRZEJ JAKUBIAK, ANDRZEJ ZIELIŃSKI

Instytut Telekomunikacji, Politechnika Warszawska
ul. Nowowiejska 15/19, 00-665 Warszawa

Otrzymano 2001.02.05
Autoryzowano 2001.03.30

W artykule przedstawiono strukturę tak zwanego detektora log-t, który utrzymuje stały poziom prawdopodobieństwa fałszywego alarmu w szerokim zakresie zmian parametrów nieskorelowanych zakłóceń biernych o niegaussowskich modelach probabilistycznych: logarytmo-normalnym, Weibulla i typu K. Metodą symulacji Monte Carlo oszacowano charakterystykę operacyjną tego detektora w przypadku, kiedy próbki zakłóceń typu K są silnie skorelowane, tak jak ma to często miejsce w rzeczywistości. Uzyskane wyniki wskazują na praktyczną niezależność prawdopodobieństwa fałszywego alarmu od stopnia skorelowania zakłóceń, co potwierdza dużą przydatność struktury log-t w detekcji SPFA w obecności radiolokacyjnych zakłóceń biernych.

Słowa kluczowe: detekcja, radiolokacja, zakłócenia bierne.

1. WPROWADZENIE

Jednym z podstawowych zadań, stawianych systemom radiolokacyjnym, jest podjęcie decyzji o obecności lub braku sygnału użytecznego, czyli o obecności lub braku wykrywanego obiektu. Decyzję taką podejmuje urządzenie, nazywane detektorem, a procedura decyzyjna jest oparta na kryterium, od którego zależy struktura oraz jakość detektora. Ponieważ do apretury anteny radaru dociera również wiele sygnałów elektromagnetycznych, które są sygnałami zakłócającymi, to decyzję detektora można opisać dyskretną zmienną losową d, przyjmującą wartości binarne $\{0,1\}$, oznaczające odpowiednio:
$d = 0$ — obecność samych zakłóceń (przyjęcie hipotezy H_0),
$d = 1$ — obecność sygnału użytecznego (przyjęcie hipotezy H_1).
Przyjęcie każdej z dwu hipotez jest obarczone ryzykiem popełnienia błędu. Obrazuje to tabela 1, w której przedstawiono kombinację decyzji, uwarunkowanych rzeczywistą

Tabela 1

Zależności operacyjne detektora dwudecyzyjnego

		Rzeczywiste zdarzenie	
		$\mathbf{H_0}$	$\mathbf{H_1}$
Decyzja detektora	0	Poprawne niewykrycie sygnału użytecznego	Błędne niewykrycie sygnału użytecznego
	1	Błędne wykrycie sygnału użytecznego — fałszywy alarm	Poprawne wykrycie sygnału użytecznego — poprawna detekcja

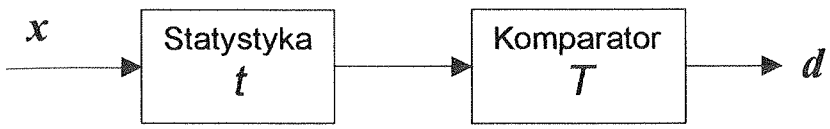
sytuacją. W dwóch przypadkach decyzje są podejmowane prawidłowo, w dwóch błędnie. Podjęcie decyzji $d = 1$ w sytuacji, gdy w rzeczywistości występują same zakłócenia (prawdziwa jest hipoteza H_0), nazywane jest błędem fałszywego alarmu (FA). Podjęcie takiej samej decyzji, kiedy prawdziwa jest hipoteza H_1, nosi nazwę poprawnej detekcji (PD). Warunkowe prawdopodobieństwa podjęcia decyzji $d = 1$ noszą nazwę odpowiednio prawdopodobieństwa fałszywego alarmu (PFA) oraz prawdopodobieństwa poprawnej detekcji (PPD) i tworzą zestaw, określany mianem charakterystyki operacyjnej detektora [12]. Charakterystyka operacyjna opisuje całkowicie probabilistyczne cechy detektora, ponieważ pozostałe warunkowe prawdopodobieństwa podejmowania decyzji $d = 0$ są uzupełnieniem PFA oraz PD do wartości 1.

Projektowanie optymalnych detektorów polega najogólniej rzecz biorąc na takim doborze procedury decyzyjnej, żeby uzyskać maksymalną wartość PD przy minimalnej wartości PFA. Jest to możliwe do spełnienia w przypadku tak zwanej detekcji parametrycznej, kiedy przy znajomości modeli probabilistycznych zakłóceń oraz sygnału użytecznego można zastosować odpowiednie reguły decyzyjne (np. maksymalnej wiarygodności, Bayesa, Neymana-Pearsona) [4]. W praktyce informacja o przychodzącym sygnale jest ograniczona, w związku z czym stosuje się suboptymalne metody detekcji, oparte na przykład na procedurach nieparametrycznych, „distribution free'' lub „robust'' [10].

W technice radiolokacyjnej szczególnie ważną klasę stanowią detektory, które niezależnie od zmian parametrów czy też statystyk zakłóceń utrzymują stały poziom prawdopodobieństwa fałszywego alarmu (SPFA) [1,11]. Stosowane we współczesnych radarach detektory SPFA umożliwiają w znacznym stopniu zautomatyzowanie procesu wykrywania obiektów użytecznych, zapewniając właściwe i stabilne poziomy prawdopodobieństw podjęcia błędnych decyzji. W kolejnych rozdziałach artykułu została przedstawiona struktura detektora progowego, który w szerokim zakresie zmian parametrów zakłóceń biernych (clutter) utrzymuje stałe prawdopodobieństwo fałszywego alarmu. Ponieważ dotychczasowe analizy dotyczyły sytuacji, w której próbki zakłóceń były nieskorelowane, wyznaczono metodami symulacji Monte Carlo zależność PFA od stopnia skorelowania zakłóceń biernych o modelu probabilistycznym typu K.

2. DETEKTOR LOG-T

Opisany po raz pierwszy w pracy [2] tak zwany detektor log-t jest progowym detektorem, którego schemat blokowy przedstawiono na rysunku 1.



Rys. 1. Schemat blokowy detektora log-t

Na podstawie ciągu N próbek sygnału wejściowego $\boldsymbol{x} = \{x_1, x_2, ..., x_i, ..., x_N\}$ zostaje wyznaczona wartość statystyki t, zgodnie z zależnością:

$$t = \frac{\ln x_0 - \frac{1}{N}\sum_{i=1}^{N} \ln x_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\ln x_i - \frac{1}{N}\sum_{i=1}^{N} \ln x_i\right]^2}} \tag{1}$$

gdzie x_0 — aktualnie badana próbka sygnału (jedna z ciągu wejściowego $\boldsymbol{x}$).

W układzie komparatora wartość statystyki t jest porównywana z wartością progu T. Decyzja d jest podejmowana zgodnie z regułą:

$$d = \begin{cases} 0 \ \ dla \ \ t < T \\ 1 \ \ dla \ \ t \geq T \end{cases} \tag{2}$$

Można łatwo wykazać, że jeśli każda z próbek x_i zostanie zastąpiona przez wielkość αx_i^{β}, gdzie α i β są parametrami, to wartość statystyki t nie ulegnie zmianie. Dlatego w przypadku, kiedy sygnał x jest zakłóceniem modelowanym rozkładem prawdopodobieństwa o dwu parametrach — skali i kształtu, to statystyka detektora log-t nie zależy od tych parametrów i PFA detektora również od nich nie zależy [2].

Przedstawiona wyżej właściwość detektora log-t została wykorzystana do detekcji SPFA sygnałów użytecznych, znajdujących się na tle niegaussowskich zakłóceń biernych, modelowanych rozkładami log-normal i Weibulla [2]. Obydwa należą do klasy rozkładów dwuparametrycznych, z których jeden jest parametrem skali a drugi kształtu. Prawdopodobieństwo fałszywego alarmu detektora log-t w obecności zakłóceń log-normalnych i Weibulla zależy tylko od wartości progu T oraz liczby N próbek sygnału, natomiast zmiany wartości parametrów modeli probabilistycznych nie mają wpływu na poziom PFA.

Oprócz dwóch wymienionych niegaussowskich modeli radiolokacyjnych zakłóceń biernych istnieje jeszcze wiele innych, z których najczęściej stosuje się model oparty na

rozkładzie typu K [5]. Jest to bardzo uniwersalny model, który dobrze oddaje dynamikę zakłóceń („kolczastość"), jednocześnie nie przerysowując „ogonów" rozkładu. Model ten stosuje się zarówno do opisu różnego rodzaju zakłóceń meteorologicznych (szczególnie odbić od chmur lub ściany deszczu) jak i zakłóceń od powierzchni ziemi [7]. Funkcja gęstości prawdopodobieństwa amplitudy x próbek zakłóceń typu K jest wyrażona zależnością:

$$P_{TK}(X) = \frac{2c}{\Gamma(\nu)}\left(\frac{cX}{2}\right)^{\nu} K_{\nu-1}(cX), \qquad X \geq 0 \tag{3}$$

gdzie c i ν — parametry rozkładu,

$\Gamma(\,)$ — funkcja gamma,

$K_{\nu-1}(\,)$ — modyfikowana funkcja Bessela (funkcja Mc Donalda) rzędu $\nu - 1$.

Parametr c jest parametrem skali, w związku z tym PFA detektora log-t nie zależy od zmian wartości c. Drugi parametru rozkładu — ν — nie jest niestety parametrem kształtu, chociaż odgrywa podobną rolę. W związku z tym PFA detektora log-t zależy nie tylko od wartości progu T oraz liczby próbek N, ale również w jakimś stopniu od zmian parametru ν.

Przy założeniu braku korelacji między próbkami sygnału oraz przy liczbie N dążącej do nieskończoności, prawdopodobieństwo fałszywego alarmu detektora log-t w obecności zakłóceń typu K można wyznaczyć z dystrybuanty zmiennej losowej z:

$$z = \frac{\ln x - m}{\sigma} \tag{4}$$

gdzie x — zmienna losowa o rozkładzie danym zależnością (3),

m — wartość średnia zmiennej ln x,

σ — odchylenie standardowe zmiennej ln x.

PFA wyraża się wzorem [6]:

$$PFA = \frac{2}{\Gamma(\nu)} \exp\left\{\nu\left[\frac{T}{4}\zeta(2,\ \nu) + \frac{T\pi^2}{24} + \frac{1}{2}\Psi(\nu) - \frac{\gamma}{2}\right]\right\} \times K_{\nu}\left\{2\exp\left[\frac{T}{4}\zeta(2,\ \nu) + \frac{T\pi^2}{24} + \frac{1}{2}\Psi(\nu) - \frac{\gamma}{2}\right]\right\} \tag{5}$$

gdzie: $\zeta(\cdot,\cdot)$ — uogólniona funkcja zeta,

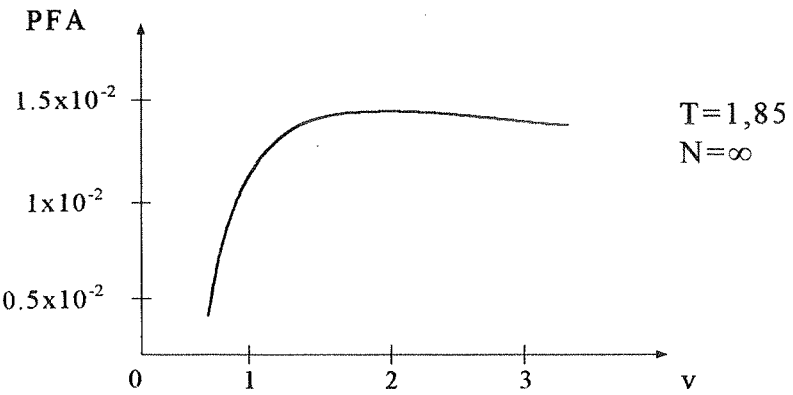
$\Psi(\cdot)$ — funkcja psi,

γ — stała Eulera ($\gamma = 0{,}577...$),

pozostałe oznaczenia jak we wzorze (3).

Przebieg zależności (5) w funkcji parametru ν, dla ustalonej wartości progu T, pokazano na rysunku 2.

Dla innych wartości progu T przebieg krzywej PFA w funkcji parametru ν jest podobny, ulegając jedynie przesunięciu wzdłuż osi pionowej. Również oszacowanie



Rys. 2. PFA detektora log-t w obecności nieskorelowanych zakłóceń typu K

metodami symulacyjnymi przebiegu PFA dla skończonych wartości liczby N przyniosło zbliżone rezultaty [6]. Na tej podstawie można przyjąć, że przy nieskorelowanych zakłóceniach typu K detektor log-t prawie utrzymuje stały poziom prawdopodobieństwa fałszywego alarmu dla wartości parametru v większych od 1. Dla uzyskania pełnego „obrazu" PFA tego detektora w obecności niegaussowskich zakłóceń biernych przeprowadzono badania symulacyjne w warunkach skorelowania próbek zakłóceń typu K. Warunki i wyniki tej symulacji zostały przedstawione w następnym rozdziale.

3. BADANIA SYMULACYJNE

Występujące w rzeczywistości radiolokacyjne zakłócenia bierne są na ogół silnie skorelowane, zarówno przestrzennie jak i czasowo. Zgodnie z wcześniejszymi rozważaniami sama statystyka detektora log-t nie zależy od parametrów kształtu i skali rozkładów prawdopodobieństwa. W ten sposób zostaje wyeliminowany wpływ tego typu parametrów na poziom prawdopodobieństwa fałszywego alarmu, zatem nawet w obecności skorelowanych zakłóceń o modelu Weibulla i logarytmo-normalnym detektor log-t jest SPFA. W przypadku skorelowanych zakłóceń typu K detektor ten również eliminuje wpływ parametru skali c, natomiast uzyskanie zależności analitycznej na PFA (nawet przy założeniu $N \to \infty$), pozwalającej na określenie wpływu drugiego parametru, jest praktycznie niemożliwe. W tej sytuacji zostały przeprowadzone badania metodą symulacji Monte Carlo, których podstawowym celem było określenie w jaki sposób stopień korelacji między próbkami zakłóceń zmienia zależność PFA od parametru v rozkładu typu K. Stopień skorelowania określono zgodnie z zależnością:

$$R_{i,j} = \frac{E[x_i x_j] - E[x_i]E[x_j]}{\sqrt{D^2(x_i)D^2(x_j)}}, \qquad i,j = 1, 2, ..., N \tag{6}$$

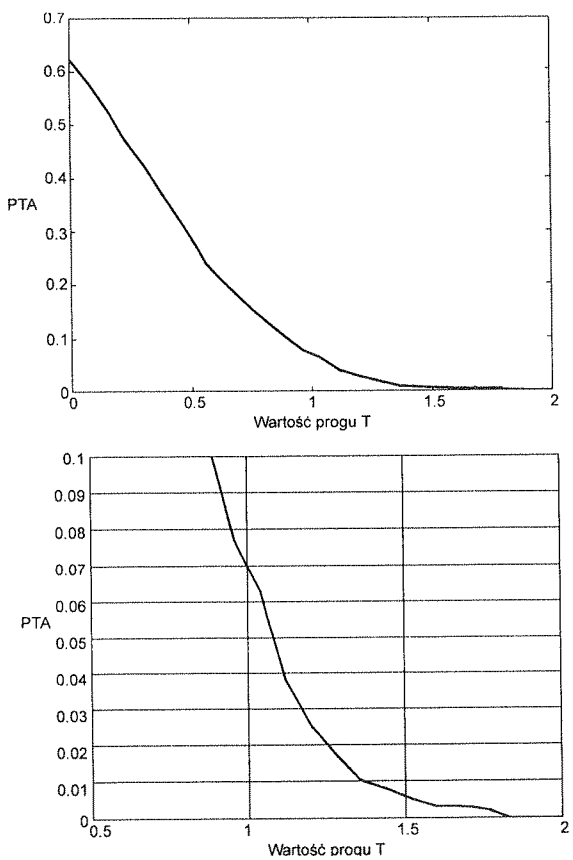
gdzie $E[\ ]$ oznacza uśrednianie probabilistyczne a $D^2(x)$ wariancję zmiennej x. Dla $|i-j| = 1$ $R_{i,j}$ jest nazywane współczynnikiem korelacji, oznaczanym jako ϱ.

Eksperymenty symulacyjne wykonano za pomocą komputera klasy PC, wykorzystując pakiet MATLAB [9] jako narzędzie programowe. Do wygenerowania ciągu N skorelowanych próbek zakłóceń typu K wykorzystano algorytm, opisany w [8]. Badania zostały przeprowadzone w dwóch grupach:

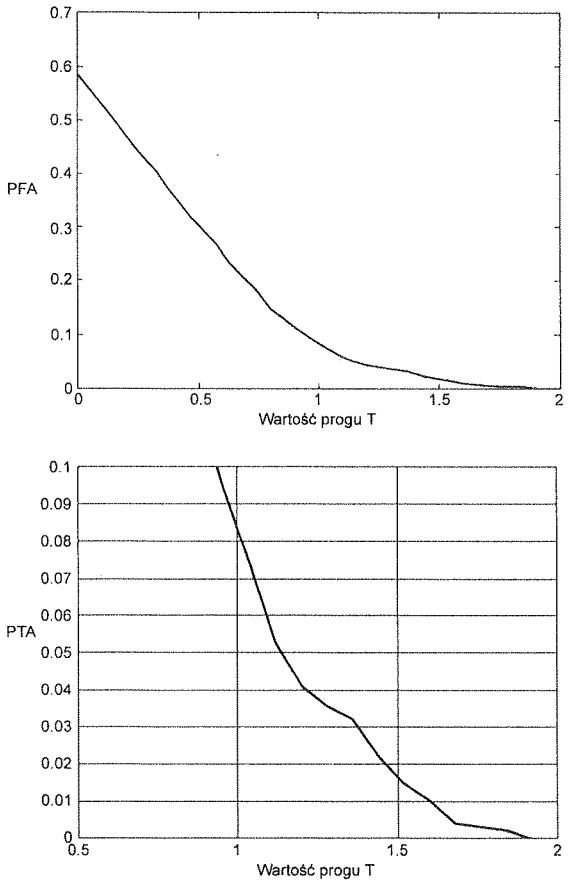
1. Dla ustalonych wartości parametru ν szacowano wartość PFA przy różnych poziomach progu T, oddzielnie dla dwóch wartości współczynnika korelacji — $\varrho = 0{,}2$ oraz $\varrho = 0{,}8$.
2. Dla ustalonych wartości PFA szacowano zmiany wartości progu T w zależności od zmian parametru ν oraz liczby próbek N, oddzielnie dla dwóch wartości współczynnika korelacji — $\varrho = 0{,}2$ oraz $\varrho = 0{,}8$.

Dla określonego zestawu parametrów każdy eksperyment powtarzano 10 000 razy, co umożliwiło oszacowanie PFA z dokładnością około 10^{-3} [3].

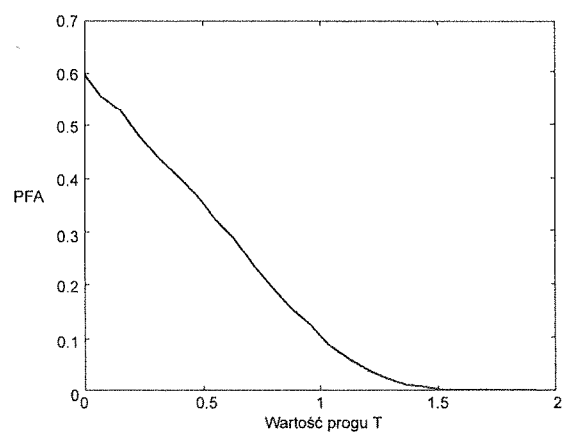
Wybrane wyniki badań symulacyjnych pierwszej grupy zostały przedstawione na rysunkach 3–6. Każdy z nich zawiera dwa wykresy — górny przedstawia pełny

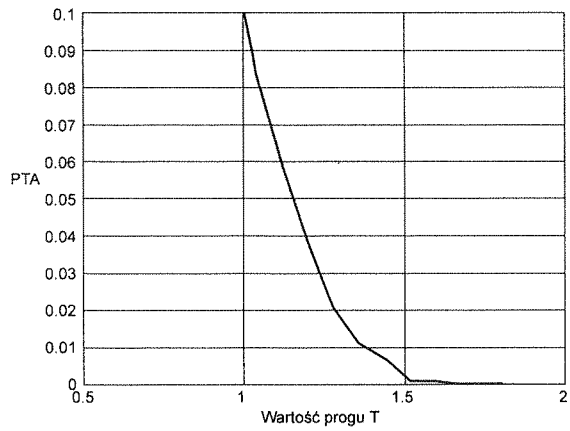


Rys. 3. Zależność PFA od progu detekcji T dla $\varrho = 0.2$ oraz $\nu = 1{,}0$

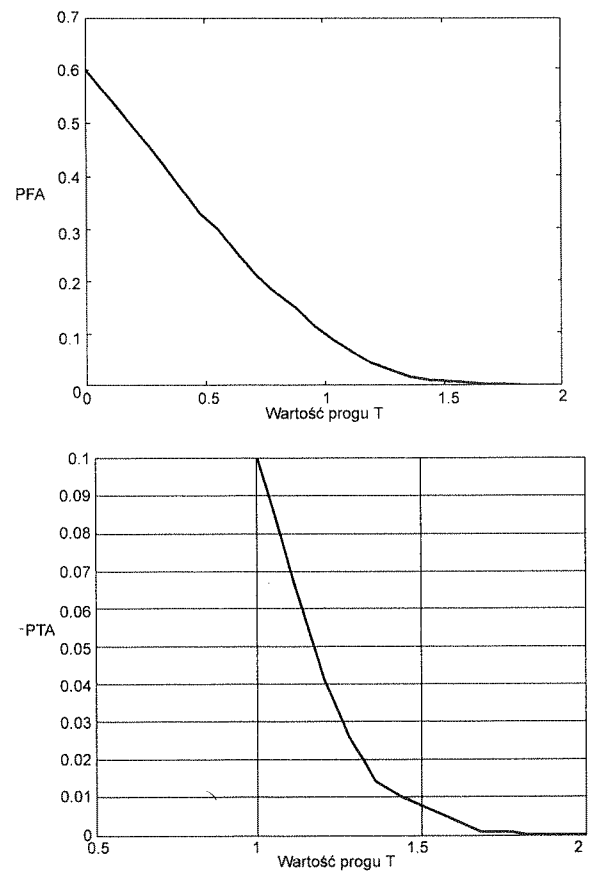


Rys. 4. Zależność PFA od progu detekcji T dla $\varrho = 0.2$ oraz $\nu = 2{,}0$





Rys. 5. Zależność PFA od progu detekcji T dla $\varrho = 0.8$ oraz $\nu = 1{,}0$

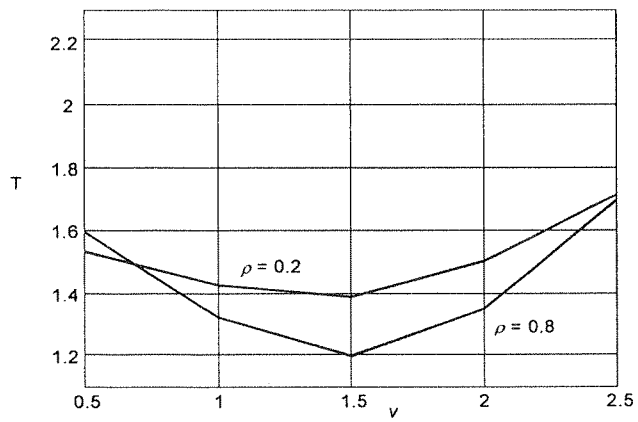


Rys. 6. Zależność PFA od progu detekcji T dla $\varrho = 0.8$ oraz $\nu = 2{,}0$

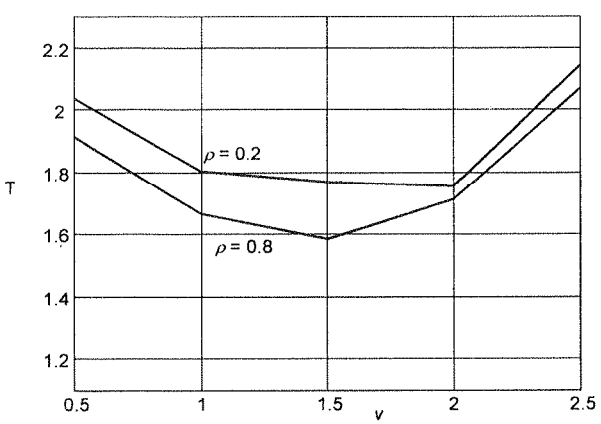
przebieg PFA w funkcji progu T, drugi, dolny wykres, jest powiększonym fragmentem pierwszego, co umożliwia precyzyjne określenie żądanych parametrów.

Wyniki badań z zakresu drugiej grupy zobrazowano na rysunkach 7 i 8. Zależność wartości progu od zmian parametru v, przy stabilizacji wartość PFA, przedstawiono jednocześnie dla niskiego i wysokiego stopnia korelacji. W ten sposób wyznaczono na płaszczyźnie zakres możliwych zmian progu T dla dwóch typowych wartości PFA: 10^{-2} oraz 10^{-3}.

Wszystkie zaprezentowane wyniki uzyskano przy liczności próby $N = 100$. Jest to typowa liczba próbek, możliwa do uzyskania przy pojedynczym oświetleniu tak zwanej elementarnej komórki rozróżnialności przestrzennej radaru, charakteryzowanej przez szerokość kątową listka głównego charakterystyki promieniowania anteny i długość impulsu sondującego.



Rys. 7. Zależność progu T od parametru v oraz stopnia korelacji przy PFA = 10^{-2}



Rys. 8. Zależność progu T od parametru v oraz stopnia korelacji przy PFA = 10^{-3}

Na podstawie przeprowadzonych symulacji można stwierdzić, że zmiany stopnia korelacji między próbkami zakłóceń typu K nie mają istotnego wpływu na poziom PFA, przy odpowiednim doborze wartości progu. Dzieje się tak w dużym zakresie zmian wartości parametru v rozkładu prawdopodobieństwa amplitud zakłóceń.

4. PODSUMOWANIE

Przedstawiona analiza oraz wyniki przeprowadzonych badań symulacyjnych pozwalają na sformułowanie wniosku, że detektor log-t jest detektorem SPFA w szerokim zakresie zmian parametrów rozkładów amplitud niegaussowskich zakłóceń biernych. Właściwość ta zostaje zachowana również przy znacznym stopniu skorelowania próbek zakłóceń. Uzyskane wyniki umożliwiają dobór odpowiedniej wartości progu detektora, tak, żeby utrzymać założony maksymalny poziom prawdopodobieństwa fałszywego alarmu.

Na koniec warto podkreślić, że przy zadanym poziomie PFA wartości progów dla różnych modeli zakłóceń są różne. Dlatego ostateczna postać detektora SPFA, opartego na statystyce log-t, powinna zawierać element adaptacyjny, dopasowujący wartość progu do aktualnego modelu probabilistycznego zakłóceń. Wymaga to zastosowania dodatkowego urządzenia, klasyfikującego rzeczywiste próbki zakłóceń ze względu na ich realny rozkład prawdopodobieństwa.

BIBLIOGRAFIA

1. J.D. Echard: *Estimation of radar detection and false alarm probabilities*. IEEE Trans. on AES, vol. AES-27, No. 2, March 1991, pp. 255–260.
2. G.B. Goldstein: *False-alarm Regulation in Log-normal and Weibull Clutter*. IEEE Trans. Aerospace Electr. Systems, vol. AES-9, No 5, 1973, pp. 84–92.
3. G.J. Hahn: *Sample size for Monte Carlo simulation*. IEEE Trans. on SM&C, Nov. 1972, pp. 678–680.
4. C.W. Helstrom: *Elements of Signal Detection and Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
5. E. Jakeman, P.N. Pusey: *A Model for Non-Rayleigh Sea Echo*. IEEE Trans. AP, vol. AP-24, 1976, pp. 806–814.
6. A. Jakubiak: *Signal Detection in Non-Gaussian Clutter*. IEEE Trans. AES, vol. AES-27, No 5, Sept. 1991, pp. 758–760.
7. A. Jakubiak: *Metody klasyfikacji radiolokacyjnych zakłóceń biernych*. Prace Naukowe — Elektronika, z. 126, Warszawa, Oficyna Wyd. Politechniki Warszawskiej, 2000.
8. A. Jakubiak: *Metody i algorytmy symulacji radiolokacyjnych zakłóceń biernych*. Kwartalnik Elektroniki i Telekomunikacji, t. 46, z. 2, 2000, ss. 237–253.
9. *MATLAB High Performance Numeric Computation and Visualisation Software*. The Math Works Inc., Natick, Mass., 1993.
10. W.L. Root: *An introduction to the theory of the detection of signal in noise*. Proc. of the IEEE, vol. 58, No. 5, May 1970, pp. 610–622.
11. M.I. Skolnik: *Radar Handbook*. New York, McGraw-Hill, 1990.
12. C.W. Therrien: *Decision, estimation and classification*. New York, Wiley&Sons, 1989.

A. JAKUBIAK, A. ZIELIŃSKI

FALSE ALARM PROBABILITY OF LOG-T DETECTOR IN CORRELATED K-DISTRIBUTED CLUTTER

Summary

The log-t detector is described in this paper. The CFAR properties of this detector in non-Gaussian radar clutter like log-normal, Weibull and K-distributed are discussed. The operating characteristic for correlated K-distributed clutter are evaluated by using the Monte Carlo simulation. The results show that the false alarm probability practically does not depend on v parameter of clutter distribution.

Keywords: detection, radar, clutter

W
nych, k
struktur
pojęciu
dokonar
niegaus
K. Osz
decyzji
ści od
urządze
warunki

Słowa k

Występu
przypadkowy
się na drodz
wykrywania
średniego za
nych celów.
Stosowa
i tłumienia

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2001, 47, z. 2, ss. 221–232

Klasyfikator Kullbacka-Leiblera w obecności skorelowanych zakłóceń biernych

ANDRZEJ JAKUBIAK, MARIAN MUCZKO

Instytut Telekomunikacji, Politechnika Warszawska
ul. Nowowiejska 15/19, 00-665 Warszawa

Otrzymano 2001.01.21
Autoryzowano 2000.03.24

W artykule przedstawiono problematykę klasyfikacji radiolokacyjnych zakłóceń biernych, której celem jest rozpoznanie modelu probabilistycznego tych zakłóceń. Ze znanych struktur, na podstawie przyjętych założeń i warunków, wybrano klasyfikator oparty na pojęciu średniej miary Kullbacka-Leiblera. Wykorzystując technikę symulacji Monte Carlo dokonano kompleksowej oceny tego klasyfikatora w przypadku występowania skorelowanych niegaussowskich zakłóceń biernych, o modelach logarytmo-normalnym, Weibulla bądź typu K. Oszacowano warunkowe prawdopodobieństwa podejmowania poprawnych i błędnych decyzji w funkcji zmian parametrów zakłóceń i stopnia ich skorelowania, a także w zależności od liczności próby. Uzyskane wyniki pozwalają na ocenę badanego systemu, jako urządzenia o wysokiej wiarygodności podejmowanych decyzji i spełniającego postawione warunki, przede wszystkim aplikacyjne.

Słowa kluczowe: klasyfikacja, radiolokacja, zakłócenia bierne.

1. WPROWADZENIE

Występujące w radiolokacji dość powszechnie zakłócenia bierne powstają na skutek przypadkowych odbić fali elektromagnetycznej od naturalnych przeszkód, znajdujących się na drodze jej transmisji. Taka sytuacja w poważnym stopniu ogranicza możliwości wykrywania i obserwacji, prowadzonej za pomocą radaru, a nawet przy braku bezpośredniego zagrożenia utrudnia bądź uniemożliwia poprawną identyfikację obserwowanych celów.

Stosowane współcześnie techniki automatycznej detekcji sygnałów użytecznych i tłumienia zakłóceń powodują utratę wiele użytecznych informacji o otaczającej

przestrzeni elektromagnetycznej i dodatkowo uniemożliwiają operatorowi rozpoznanie rodzaju eliminowanych zakłóceń — obraz na wskaźniku syntetycznym zawiera głównie specjalne symbole a ingerencja człowieka jest ograniczona do minimum. W tej sytuacji niezwykle ważnym jest kontynuowanie badań nad coraz lepszą identyfikacją i klasyfikacją radiolokacyjnych zakłóceń biernych, tak, aby automatyczne podejmowanie decyzji w obecności tych zakłóceń było obarczone jak najmniejszym błędem.

W przypadku radiolokacyjnych zakłóceń biernych można wyróżnić dwa podstawowe cele klasyfikacji:

1. Rozpoznanie źródła powstawania zakłóceń. W tym przypadku klasami są wymienione w podrozdziale 1.1 typy zakłóceń, takie jak zakłócenia atmosferyczne, morskie, lądowe, pochodzące od stad ptaków, insektów itp.
2. Przyporządkowanie modelu matematycznego przychodzącemu zakłóceniu. W tym przypadku klasami mogą być modele probabilistyczne bądź inne, których cechy dają się ująć w odpowiednie reguły analityczne.

Pierwszy z wymienionych celów ma praktyczne znaczenie dla użytkownika stacji radiolokacyjnej, który może na podstawie podjętej decyzji klasyfikacyjnej uruchamiać odpowiednie procedury, eliminujące lub ograniczające wpływ zakłóceń. Tego typu klasyfikacji została poświęcona monografia [29], a szereg szczegółowych rozwiązań urządzeń klasyfikacyjnych znaleźć można w pracach [10], [11], [12], [21], [23].

Drugi cel jest szczególnie użyteczny przy projektowaniu torów odbiorczych radaru. Współczesny, automatyczny detektor działa w oparciu o „wyrafinowane" metody decyzyjne, dostosowując je najczęściej w sposób adaptacyjny do aktualnej postaci przetwarzanego sygnału wejściowego [2], [5], [7], [22]. Projektowanie takich zaawansowanych systemów, wykrywających sygnały użyteczne o poziomie niższym od poziomu zakłóceń, jest praktycznie niemożliwe bez znajomości modeli matematycznych tych zakłóceń. Zagadnienia te stanowią niezwykle ważny obszar badań nad sygnałami radiolokacyjnymi. W efekcie większość współcześnie konstruowanych detektorów, pracujących w obecności zakłóceń biernych, zawiera blok funkcjonalny, klasyfikujący te zakłócenia ze względu na ich hipotetyczny model matematyczny [5], [7], [9], [17], [19], [34]. Przyjmując jako cel klasyfikacji rozpoznanie, jaki model matematyczny najlepiej opisuje przychodzące zakłócenia bierne, należy określić klasy oraz reguły decyzyjne, które mogą realizować postawiony cel. Ze względu na losowy charakter zakłóceń klasy tworzy się w sposób niejako „naturalny", w oparciu o cechy probabilistyczne procesów stochastycznych, modelujących zakłócenia bierne. Z wielu możliwych cech, przydatnych do określenia klas zakłóceń, dwie są najczęściej wykorzystywane, tj. zależności korelacyjne (funkcja autokorelacji lub widmo mocy) oraz brzegowy rozkład prawdopodobieństwa amplitudy zakłóceń [3], [4], [6], [8], [10], [12], [14], [20], [21]. W związku z wymienionymi wyżej zastosowaniami w procedurach detekcji, do dalszych rozważań zdefiniowano trzy następujące klasy:

1. klasa zakłóceń opisanych modelem logarytmo-normalnym, nazywana w skrócie LN,
2. klasa zakłóceń opisana modelem Weibulla, nazywana w skrócie WB,
3. klasa zakłóceń opisana modelem typu K, nazywana w skrócie TK.

Klasy te odpowiadają współcześnie stosowanym tak zwanym niegaussowskim modelom

radiolokacyjnych zakłóceń biernych, które opisują właściwości tych zakłóceń w sposób najbliższy rzeczywistości [16], [32].

Głównym zadaniem klasyfikatora jest podjęcie jednoznacznej decyzji, do której z wyżej wymienionych trzech klas należy analizowane zakłócenie. Ze względu na to, że cechą wspólną obiektów, należących do tej samej klasy, jest rozkład prawdopodobieństwa, to w konsekwencji omawiana procedura klasyfikacji prowadzi do rozpoznania postaci funkcji gęstości prawdopodobieństwa próbek zakłóceń. Właśnie ta procedura będzie dalej nazywana klasyfikacją radiolokacyjnych zakłóceń biernych.

Urządzenie, które podejmuje decyzje o przynależności ciągu obserwowanych próbek sygnału $\boldsymbol{x} = \{x_1, x_2, ..., x_N\}$ do jednej z określonych klas, nosi nazwę klasyfikatora, a zasada, na podstawie której jest dokonywana klasyfikacja, nazywa się regułą decyzyjną. Określenie reguły decyzyjnej determinuje strukturę klasyfikatora.

Parametrem służącym do oceny klasyfikatora jest średnie ryzyko $\bar{r}$, którego miarą jest [28], [29], [31]:

$$\bar{r} = 1 - \sum_{l=1}^{L} P(x_l, d_l), \tag{1}$$

gdzie L oznacza liczbę klas a $P(\boldsymbol{x}_l, d_l)$ jest prawdopodobieństwem łącznym wystąpienia sygnału $\boldsymbol{x}$ z l-tej klasy oraz podjęcia decyzji d_l o przynależności sygnału $\boldsymbol{x}$ do l-tej klasy.

Reguła decyzyjna minimalizująca średnie ryzyko $\bar{r}$ nazywa się regułą optymalną [28], [29], [31]. Łatwo zauważyć, że minimalizację średniego ryzyka $\bar{r}$ można zastąpić maksymalizacją sumy w wyrażeniu (1), która jest średnim prawdopodobieństwem podjęcia poprawnej decyzji klasyfikacyjnej, nazywanym dalej średnim prawdopodobieństwem poprawnej klasyfikacji $\overline{PPK}$. Korzystając z definicji prawdopodobieństwa warunkowego [25] średnie prawdopodobieństwo poprawnej klasyfikacji można wyrazić jako:

$$\overline{PPK} = \sum_{l=1}^{L} P(x_l)\, P(d_l|x_l) \tag{2}$$

gdzie $P(\boldsymbol{x}_l)$ jest prawdopodobieństwem a priori wystąpienia obiektu $\boldsymbol{x}$ z klasy l , natomiast $P(d_l/\boldsymbol{x}_l)$ jest prawdopodobieństwem podjęcia decyzji d_l pod warunkiem, że obiekt $\boldsymbol{x}$ należy do klasy l. To ostatnie prawdopodobieństwo będzie dalej nazywane warunkowym prawdopodobieństwem poprawnej klasyfikacji $PPK|_l$.

Niestety, dość często prawdopodobieństwa a priori nie są znane. W tej sytuacji klasyfikator jest oceniany właśnie na podstawie warunkowych prawdopodobieństw poprawnej klasyfikacji. Optymalizacja struktury klasyfikatora sprowadza się wówczas do znalezienia reguły decyzyjnej, dla której wszystkie $PPK|_l$ są maksymalne.

Sprecyzowany wcześniej cel klasyfikacji oraz zdefiniowane klasy wyznaczają zasadnicze obszary poszukiwań reguł decyzyjnych, określających strukturę klasyfikatora. Należą do nich między innymi testy służące do weryfikacji hipotez statystycznych, reguły decyzyjne statystycznej teorii decyzji oraz miary porównawcze teorii informacji

[4], [8], [25], [28], [31], [33]. Coraz częściej podejmowane są próby zastosowania metod, które można określić mianem niestandardowych, jak na przykład użycie sieci neuronowych bądź elementów teorii chaosu [3], [11], [13], [18], [30].

Specyfika zastosowań radiolokacyjnych pozwala na sformułowanie dodatkowych warunków, które powinien spełniać klasyfikator zakłóceń biernych. Zawęża to możliwości wyboru reguł decyzyjnych, precyzując jednocześnie kryteria ich przyjęcia. Warunki te są następujące:

1. Prawdopodobieństwa a priori wystąpienia każdej klasy zakłóceń są nieznane.
2. Klasyfikacja powinna być możliwa nawet przy stosunkowo małej liczbie próbek sygnału (50–100).
3. Klasyfikator powinien być mało wrażliwy na zmiany parametrów zakłóceń.
4. Struktura klasyfikatora powinna być fizycznie realizowalna, a decyzje powinny być podejmowane w czasie rzeczywistym.
5. Próbki zakłóceń mogą być silnie skorelowane.

Ostatni z powyższych warunków jest szczególnie krytyczny, ponieważ większość znanych testów decyzyjnych, możliwych do zastosowania jako reguła decyzyjna klasyfikatora, zakłada niezależność statystyczną badanych próbek sygnału, natomiast radiolokacyjne zakłócenia bierne charakteryzują się znaczną korelacją przestrzenno-czasową [6], [9], [16], [26].

W monografii [16] dokonano obszernej analizy wielu reguł decyzyjnych, identyfikujących statystyki sygnałów o charakterze stochastycznym, pod kątem zastosowań w radiolokacji. Przeprowadzono także testy symulacyjne, pozwalające na ocenę wybranych struktur klasyfikatorów. Dwa z nich — oparty na średniej informacji Kullbacka-Leiblera oraz na sieci neuronowej Kohonena — spełniają w największym stopniu postawione warunki, przy czym pierwszy został zbadany przy założeniu braku korelacji między próbkami zakłóceń. Wysoka wiarygodność podejmowanych decyzji oraz teoretyczna możliwość zastosowania klasyfikatora Kullbacka-Leiblera w przypadku zakłóceń skorelowanych, stanowiły imperatyw do podjęcia dalszych badań nad tą strukturą. Wyniki tych badań zostały przedstawione w dalszej części niniejszej pracy.

2. STRUKTURA KLASYFIKATORA KULLBACKA-LEIBLERA

W pracy [1] H. Akaike wskazał na możliwość wykorzystania zmodyfikowanej miary Kullbacka, znanej jako średnia informacja Kullbacka-Leiblera, do oceny stopnia dopasowania rozkładu prawdopodobieństwa próbek sygnału do analizowanego modelu. Średnia informacja Kullbacka-Leiblera jest zdefiniowana wzorem:

$$I(p; p_l) = C(p; p) - C(p; p_l), \tag{3}$$

gdzie p i p_l są funkcjami łącznej gęstości prawdopodobieństwa losowych próbek $\{x_i; i = 1, ..., N\}$ zmiennej X, a $C(p; p_l)$ wyraża się zależnością:

$$C(p;\, p_l) = \int_X \cdots \int_X p(x_1, \ldots, x_N) \log p_l(x_1, \ldots, x_N)\, dx_1 \ldots dx_N. \tag{4}$$

Jeżeli próbki $\{x_i\}$ są wzajemnie niezależne statystycznie, $p(...)$ jest nieznanym rozkładem prawdopodobieństwa tych próbek a $p_l(...)$ reprezentuje jeden z hipotetycznych rozkładów (z grupy L analizowanych modeli), to naturalny estymator wielkości $C(p;\, p_l)$ wynosi [1]:

$$\hat{C}(p;\, p_l) = \frac{1}{N} \sum_{i=1}^{N} \log[p_l(x_i)]. \tag{5}$$

Estymator ten należy do klasy estymatorów wiarygodności i nie zależy od nieznanej z definicji postaci rozkładu $p(...)$. W miejsce parametrów rozkładu $p_l(...)$ należy podstawić ich estymatory, otrzymane metodą największej wiarygodności. Średnia informacja Kullbacka-Leiblera $I(p;\, p_l)$ jest nieujemna i przyjmuje wartość zero w przypadku równości rozkładów p oraz p_l. Ponieważ maksymalizacja $C(p;\, p_l)$ minimalizuje wartość $I(p;\, p_l)$ [1], to wybór tego modelu, dla którego wartość estymatora wyznaczona z zależności (5) jest największa, może stanowić proste kryterium decyzyjne klasyfikatora.

Klasyfikator stosowany do analizy radiolokacyjnych zakłóceń biernych, działający w oparciu o średnią informację Kullbacka-Leiblera, został po raz pierwszy opisany w pracy [19], a ograniczone w swoim zakresie wyniki jego testowania opublikowano w pracach [14], [16]. Ograniczenie to dotyczy przede wszystkim zagadnienia korelacji próbek zakłóceń — badania były prowadzone przy założeniu niezależności stochastycznej tychże próbek, co jest uwarunkowane stosowaniem estymatorów największej wiarygodności parametrów rozpatrywanych modeli probabilistycznych. Jednak wykorzystanie estymatorów otrzymanych metodą momentów daje teoretyczne podstawy do klasyfikacji zakłóceń skorelowanych, przy czym ze względu na obciążenie tych estymatorów również estymator (5) będzie obciążony i należy spodziewać się gorszych wyników klasyfikacji.

Przyjmując trzy wymienione we wstępie modele zakłóceń można wyznaczyć estymatory średniej informacji Kullbacka-Leiblera. Na podstawie postaci funkcji gęstości prawdopodobieństwa, podanych w [hab.], estymatory te wynoszą odpowiednio:

$$\hat{C}(p;\, p_{LN}) = \frac{1}{N} \sum_{i=1}^{N} \log\left\{ \frac{1}{\sqrt{2\pi}\, \hat{s} x_i} \exp\left[-\frac{\ln^2\left(\frac{x_i}{\hat{d}}\right)}{2\hat{s}^2} \right] \right\}, \tag{6}$$

$$\hat{C}(p;\, p_{WB}) = \frac{1}{N} \sum_{i=1}^{N} \log\left\{ \frac{\hat{a}}{\hat{b}} \left(\frac{x_i}{\hat{b}}\right)^{\hat{a}-1} \exp\left[-\left(\frac{x_i}{\hat{b}}\right) \right] \right\}, \tag{7}$$

$$\hat{C}(p;\, p_{TK}) = \frac{1}{N} \sum_{i=1}^{N} \log\left[\frac{2\hat{c}}{\Gamma(\hat{\nu})} \left(\frac{\hat{c} x_i}{2}\right)^{\hat{\nu}} K_{\hat{\nu}-1}(\hat{c} x_i) \right], \tag{8}$$

gdzie $\Gamma(\cdot)$ jest funkcją gamma, $K_m(\cdot)$ — modyfikowaną funkcją Bessela rzędu m, natomiast $\hat{d}$, $\hat{s}$, a, $\hat{b}$, $\hat{c}$, $\hat{v}$ są estymatorami parametrów rozkładów LN, WB i TK. Wyznaczone metodą momentów estymatory parametrów dane są zależnościami [16]:

$$\hat{d} = \frac{\hat{m}_X^2}{\sqrt{\hat{E}\{X^2\}}}, \tag{9}$$

$$\hat{s} = \sqrt{\ln \frac{\hat{E}\{X^2\}}{\hat{m}_X^2}}, \tag{10}$$

$$\hat{a} = \frac{\pi}{\hat{\sigma}\sqrt{6}}, \tag{11}$$

$$\hat{b} = \exp\left(\hat{m}_X + \frac{\gamma\hat{\sigma}\sqrt{6}}{\pi}\right), \tag{12}$$

$$\hat{v} = \frac{2\hat{E}^2\{X^2\}}{\hat{E}\{X^4\} - 2\hat{E}^2\{X^2\}}, \tag{13}$$

$$\hat{c} = 2\sqrt{\frac{\hat{v}}{\hat{E}\{X^2\}}}, \tag{14}$$

gdzie γ jest stałą Eulera, a $\hat{m}_X$, $\hat{E}\{X^2\}$, $\hat{E}\{X^4\}$ oraz $\hat{\sigma}$ wyznacza się z ciągu próbek zakłóceń $\{x_i;\ i = 1, ..., N\}$ odpowiednio:

$$\hat{m}_X = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{15}$$

$$\hat{E}\{X^2\} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \tag{16}$$

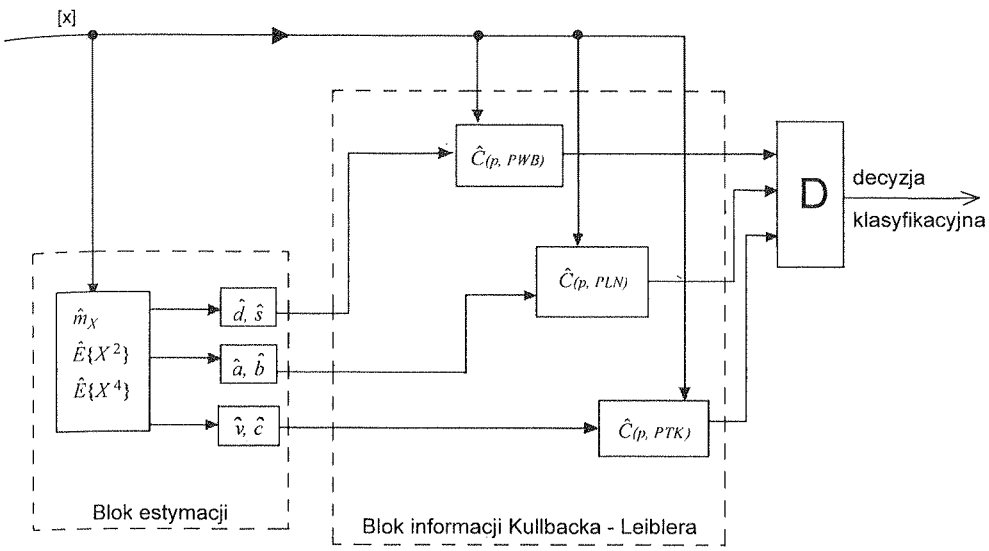
$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{m}_X)^2} \tag{17}$$

$$\hat{E}\{X^4\} = \frac{1}{N}\sum_{i=1}^{N} x_i^4. \tag{18}$$

Schemat funkcjonalny klasyfikatora Kullbacka- Leiblera przedstawia rysunek 1.

Poszczególne etapy algorytmu działania klasyfikatora Kullbacka-Leiblera są następujące:

1. Na podstawie próbek zakłóceń $\{x_i\}$ oblicza się wartości estymatorów parametrów wszystkich rozpatrywanych modeli zakłóceń, według zależności 9–14.



Rys. 1. Schemat funkcjonalny klasyfikatora Kullbacka-Leiblera

2. Dla każdego z rozpatrywanych modeli zakłóceń wyznacza się z próbek $\{x_i\}$ wartość estymatora $\hat{C}$, podstawiając do zależności 6–8 wartości estymatorów parametrów, wyznaczone w pierwszym kroku.
3. Wybiera się ten model zakłóceń, dla którego wartość estymatora $\hat{C}$ jest największa.

3. WYNIKI SYMULACJI MONTE CARLO

Klasyfikator Kullbacka-Leiblera, działający zgodnie z przedstawionym algorytmem, został szczegółowo przetestowany za pomocą metody Monte Carlo, ponieważ przy założeniu korelacji zakłóceń analityczna ocena jego parametrów jest praktycznie niemożliwa. W eksperymentach symulacyjnych wykorzystano generatory skorelowanych próbek zakłóceń, opisane w [15], natomiast same procedury utworzono w oparciu o środowisko programowe MATLAB [24].

Stopień skorelowania między próbkami określono współczynnikiem korelacji ϱ, zgodnie z zależnością:

$$\varrho = R_{i,i+1} = \frac{E[x_i x_{i+1}] - E[x_i]E[x_{i+1}]}{\sqrt{D^2(x_i)D^2(x_{i+1})}}, \qquad i = 1, 2, ..., N-1 \tag{19}$$

gdzie $E[\ ]$ oznacza uśrednianie probabilistyczne a $D^2(x)$ wariancję zmiennej x.

Ze względu na brak możliwości wyznaczenia prawdopodobieństwa a priori wystąpienia danego modelu zakłóceń, oszacowano warunkowe PPK, oddzielnie dla każdego z rozpatrywanych modeli.

W tabelach 1÷3 przedstawiono wyniki badań symulacyjnych warunkowych prawdopodobieństw poprawnej klasyfikacji klasyfikatora Kullbacka-Leiblera dla różnych wartości współczynnika korelacji zakłóceń o modelach odpowiednio: logarytmo-normalnym, Weibulla i typu K. We wszystkich przypadkach przyjęto typowe dla danych modeli wartości parametrów s, a, v [16], natomiast pozostałe, będące parametrami skali, są równe 1. Liczność próby $N = 100$ jest typową ilością próbek, uzyskiwanych przy

Tabela 1

Parametry klasyfikatora Kullbacka-Leiblera w obecności zakłóceń logarytmo-normalnych (s = 1,3)

Współczynnik korelacji ϱ	$PPK_{\|LN}$	Prawdopodobieństwa błędnych decyzji		
		Weibeull	Typu K	Brak
0,10	0,98	0,00	0,01	0,00
0,50	0,89	0,02	0,09	0,00
0,70	0,78	0,02	0,19	0,01
0,90	0,73	0,03	0,23	0,01

Tabela 2

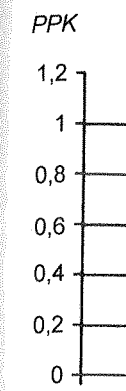
Parametry klasyfikatora Kullbacka-Leiblera w obecności zakłóceń Weibulla (a = 1)

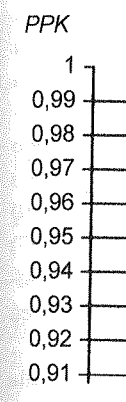
Współczynnik korelacji ϱ	$PPK_{\|WB}$	Prawdopodobieństwa błędnych decyzji		
		Log-normalny	Typu K	Brak
0,10	0,99	0,01	0,00	0,00
0,50	0,99	0,01	0,00	0,00
0,70	0,98	0,02	0,00	0,00
0,90	0,94	0,04	0,02	0,00

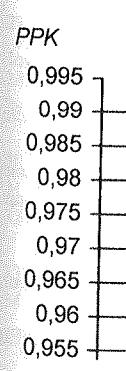
Tabela 3

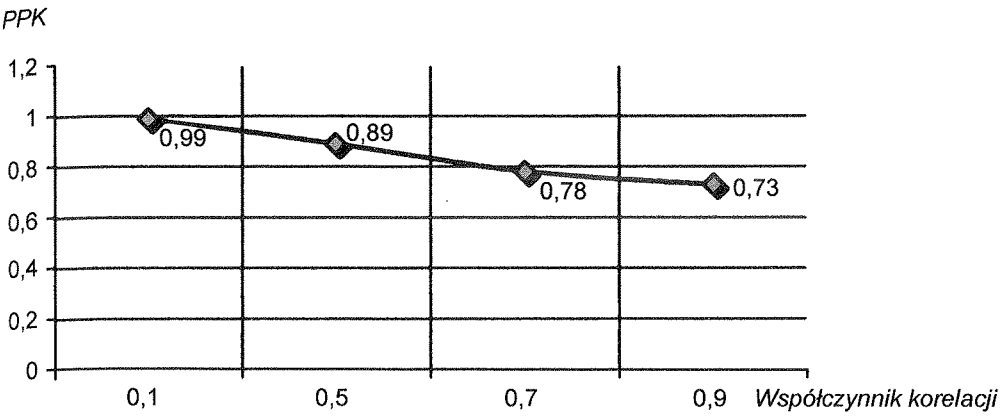
Parametry klasyfikatora Kullbacka-Leiblera w obecności zakłóceń typu K (v = 1,5)

Współczynnik korelacji ϱ	$PPK_{\|TK}$	Prawdopodobieństwa błędnych decyzji		
		Weibull	Log-normalny	Brak
0,10	0,97	0,03	0,00	0,00
0,50	0,99	0,01	0,00	0,00
0,70	0,99	0,00	0,00	0,01
0,90	0,99	0,00	0,00	0,01

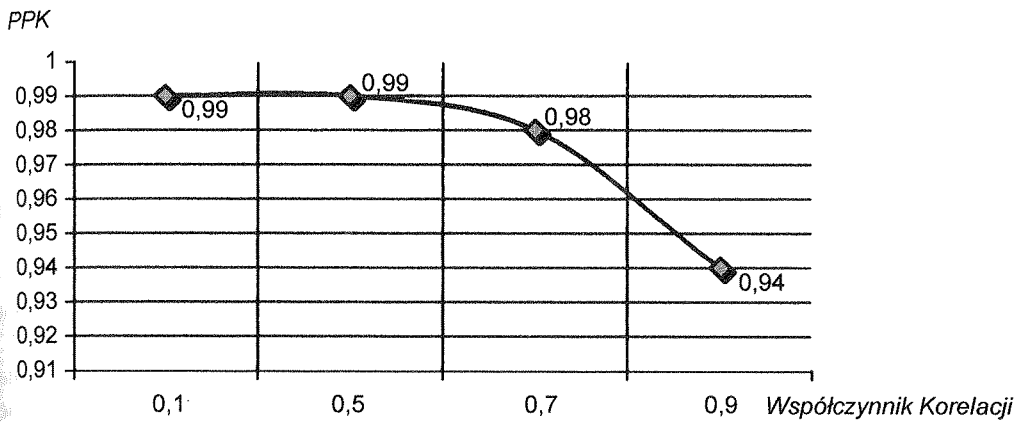




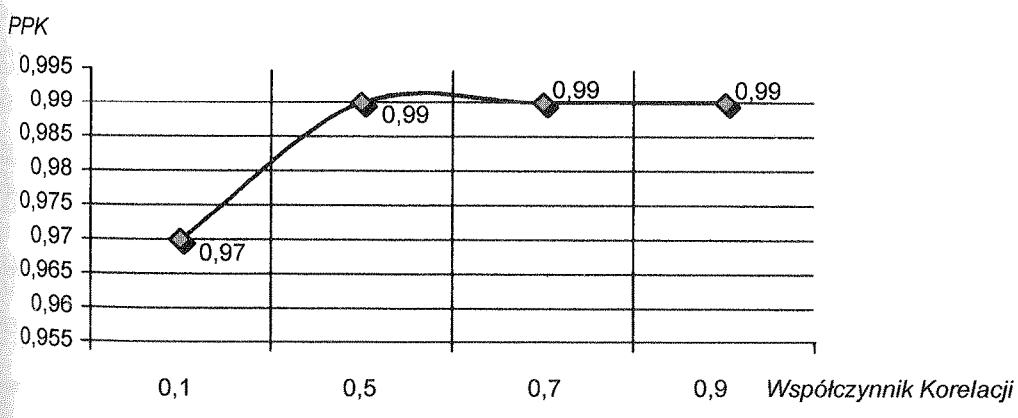




Rys. 2. Wpływ stopnia korelacji na prawdopodobieństwo poprawnej klasyfikacji w obecności zakłóceń logarytmo-normalnych



Rys. 3. Wpływ stopnia korelacji na prawdopodobieństwo poprawnej klasyfikacji w obecności zakłóceń Weibulla



Rys. 4. Wpływ stopnia korelacji na prawdopodobieństwo poprawnej klasyfikacji w obecności zakłóceń typu K

pojedynczym „oświetleniu" tak zwanej komórki elementarnej [22]. W tabelach umieszczone zostały także wyniki oszacowania prawdopodobieństwa błędnych decyzji, to znaczy rozpoznania innego modelu, niż na wejściu klasyfikatora, bądź nie przyjęcie żadnego z modeli (oznaczone jako brak).

Warunkowe prawdopodobieństwa poprawnej klasyfikacji w zależności od stopnia skorelowania próbek zakłóceń zostały przedstawione w formie wykresów na rysunkach 2÷4.

4. WNIOSKI

Przedstawione wyniki badań symulacyjnych pozwalają na sformułowanie ogólnego wniosku o możliwości zastosowania klasyfikatora Kullbacka-Leiblera do identyfikacji modeli probabilistycznych zakłóceń nawet w przypadku ich znacznej korelacji. Warunkowe prawdopodobieństwa poprawnej klasyfikacji są szczególnie wysokie w obecności zakłóceń Weibulla lub typu K. W przypadku zakłóceń logarytmo-normalnych można zaobserwować pewien spadek PPK przy korelacji powyżej 0,5 połączony ze wzrostem prawdopodobieństwa błędnej decyzji, wskazującej na model typu K.

Przyjmując hipotetycznie, że prawdopodobieństwo a priori wystąpienia każdego z rozpatrywanych modeli jest jednakowe i wynosi 1/3, można wyznaczyć średnie ryzyko badanego klasyfikatora. Dla wysokiego stopnia skorelowania próbek zakłóceń ($\varrho = 0{,}9$) średnie ryzyko przyjmuje wartość $\bar{r} = 0{,}11$; przy nieco niższej korelacji ($\varrho = 0{,}7$) $\bar{r} = 0{,}08$. Utrzymanie średniego ryzyka na poziomie 10% świadczy o dużej wiarygodności podejmowanych decyzji przez klasyfikator Kullbacka-Leiblera w obecności zakłóceń skorelowanych.

Na koniec należy podkreślić, że przeprowadzono także badania klasyfikatora dla innych typowych wartości parametrów rozkładów zakłóceń, uzyskując wyniki zbliżone do przedstawionych w pracy.

BIBLIOGRAFIA

1. H. Akaike: *A new look at the statistical model identification.* IEEE Trans. Automatic Control, vol. AC-19, No 6, Dec. 1974, pp. 716–723.
2. J.H. Blythe, D.E. Rice, W.L. Altwood: *The CFE clutter model with application to automatic detection.* The Marconi Rewiev, vol. XXXII, No. 174, 1969.
3. C. Bouvier i in.: *Radar Clutter Classification Using Autoregressive Modelling, K-distribution and Neural Network.* Proc. IEEE Int. Conf. Acoustics, Speech & Signal Proc., ICASSP'95, Detroid, 1995, pp. 1820–1823.
4. S.C. Choi: *Statistical methods of discrimination & classification.* Pergamon Press, New York, 1986.
5. E. Conte i in.: *Asymptotically optimum radar detector in non-Rayleigh clutter.* IEE Proceedings, vol. 134, Pt. F, No 7, 1987, pp. 667–672.
6. W.A. Gardner: *A unifying view of second-order measures of quality for signal classification.* IEEE Trans. on Com., vol. Com-28, No. 6 , June 1980, pp. 807–816.
7. L.M. Garth, H.V. Poor: *Detection of non-Gaussian signals: a paradigm for modern statistical signal processing.* Proceedings of the IEEE, vol. 82, No 7, 1994, pp. 1061–1095.

8. A.K. Gupta: *On classification rule for multiple measurements*. Comp. & Maths. with Appls., vol. 12A, No 2, 1986, pp. 301–308.
9. S. Haykin: *Radar clutter attractor: implications for physics, signal processing and control*. IEE Proc. Radar, Sonar Navig., vol. 146, No 4, 1999, pp. 177–187.
10. S. Haykin: S.B. Kesler, B.W. Currie: *An experimental classification of radar clutter*. Proceedings of the IEEE, vol. 67, No 2, Feb. 1979, pp. 332–333.
11. S. Haykin, C. Deng: *Classification of Radar Clutter Using Neural Networks*. IEEE Trans. Neural Networks, vol. 2, No 6, Nov. 1991, pp. 589–600.
12. S. Haykin i in.: *Classification of Radar Clutter in an Air Traffic Control Environment*. Proc. IEEE, vol. 79, No 6, June 1991, pp. 742–772.
13. J. Hertz, A. Krogh, R. Palmer: *Introduction to the theory of neural computation*. New York, Addiso Wesley, 1991.
14. A. Jakubiak: *A new algorithm for statistical model identification*. 32 Int. Wissenschaftliches. Kolloquium, Ilmenau, 1987, Mat. Konf., pp. 53–55
15. A. Jakubiak: *Metody i algorytmy symulacji radiolokacyjnych zakłóceń biernych*. Kwartalnik Elektronika i Telekomunikacja, 2000, 46, z. 2, ss. 237–253.
16. A. Jakubiak: *Metody klasyfikacji radiolokacyjnych zakłóceń biernych*. Prace Naukowe — Elektronika, z. 126, Warszawa, Oficyna Wyd. Politechniki Warszawskiej, 2000.
17. A. Jakubiak: *An Adaptive Detection in Non-Gaussian Clutter*. Signal Processing IV, Elsevier Science Publishers B.V., 1988, pp. 351–353.
18. A. Jakubiak i in: *Radar clutter classification using Kohonen neural network*. Proc. RADAR'97 Conf., Edinghburrg, UK, 1997, pp. 185–188.
19. A. Jakubiak: *Metoda detekcji i klasyfikacji odbieranych zakłóceń niegaussowskich*. Rozprawa doktorska, politechnika Warszawska, 1981.
20. A. Jakubiak, B. Wala: *Comparison of Two Algorithms for Statistical Model Identification*. Proc. Intern. AMSE Conf. Signal & Systems, Brighton, 1989, vol. 1A, pp. 93-100.
21. S.B. Kesler, S. Haykin: *The maximum entropy method applied to the spectral analysis for radar clutter*. IEEE Trans. on IT, vol. IT-24, No. 2, March 1978, pp. 269–272.
22. J. Kroszczyński i in.: *Technika współczesnej radiolokacji*. Warszawa, WkiŁ, 1975.
23. M. Ogata i in.: *Discrimination of Radar Clutter by Texture Analysis*. IEE Proc. , vol. 133, Pt. F, No 3, June 1986, pp. 257-263.
24. *MATLAB High Performance Numeric Computation and Visualisation Software*. The Math Works Inc., Natick, Mass., 1993.
25. Z. Pawłowski: *Statystyka matematyczna*. Warszawa, PWN, 1976.
26. J.R. Riley: *Radar cross section of insects*. Proceedings of the IEEE, vol. 73, No. 2, Feb. 1985, pp. 228-232.
27. F.W. Smith, J.A. Malin: *Models for radar scatterer density in terrain images*. IEEE Trans. on AES, vol. AES-22, No. 5, Sept. 1986, pp. 642-647.
28. W. Sobczak, W. Malina: *Metody selekcji informacji*. Warszawa, WNT, 1978.
29. W. Stehwien: *Radar clutter classification*. Ph.D. disscrtation, McMaster University, Hamilton, Ontario, Canada, 1989.
30. R. Tadeusiewicz: *Sieci neuronowe*. Warszawa, Akademicka Oficyna Wydawnicza RM, 1993.
31. C.W. Therrien: *Decision, estimation and classification*. New York, Wiley & Sons, 1989.
32. P. Thomas, S. Haykin: *Stochastic modelling of radar returns*. IEE Proceedings, vol. 133, Pt. F, No 5, August 1986, pp. 476–481.
33. H.L. Trees Van: *Detection, Estimation, and Modulation Theory*. New York, Wiley&Sons, 1971.
34. B.C.Y. Wong, I.F. Blake: *Detction in multivariate Non-Gaussian noise*. IEEE Trans. on Communication, vol. 42, No 2/3/4, 1994, pp. 1672-1683.

A. JAKUBIAK, M. MUCZKO

THE KULLBACK-LEIBLER CLASSIFIER IN CORRELATED CLUTTER

Summary

The problem of radar clutter statistics classification is described in this paper. The structure of classifier based on Kullback-Leibler mean information was choosen from many others, according to the assumed conditions. The conditional probabilities of correct and wrong decisions in presens of correlated non-gaussian clutter were evaluated by using the Monte Carlo method. The results show that the classifier decision practically does not depend on clutter correlation level.

Keywords: classification, radar, clutter

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2001, 47, z. 2, ss. 233–253

Układy mnożące przez stały współczynnik implementowane w układach programowalnych FPGA

KAZIMIERZ WIATR, ERNEST JAMRO

Katedra Elektroniki, Akademia Górniczo-Hutnicza
Al. Mickiewicza 30, 30-059 Kraków
e-mail: wiatr@uci.agh.edu.pl

Otrzymano 2000.11.21
Autoryzowano 2001.01.24

Poniższy artykuł przedstawia różne architektury równoległe układów mnożących o stałym współczynniku mnożenia, implementowanych w układach programowalnych FPGA. W pierwszej części artykułu zostały opisane układy mnożące bezmnożne MM (ang. *Multiplierless Multiplication*). Układy MM wykorzystują reprezentację kanoniczną cyfry ze znakiem CSD (ang. *Canonic Sign Digit*) lub/i dzielnie wspólnej podstruktury SS (ang. *Sub-structure Sharing*). Opisany został również nowy, zoptymalizowany pod kątem generowanego układu MM algorytm konwersji z kodu uzupełnień do dwóch do reprezentacji CSD. Druga część artykułu została poświęcona układom mnożącym wykorzystującym pamięć typu LUT (ang. *Look-Up Table*) i nazywanym w skrócie LM (ang. *LUT based Multiplication*). W konsekwencji opisano wykorzystywanie różnych modułów pamięci oraz znajdowanie optymalnej kombinacji pamięć — układ dodający. Dla układów mnożących LM rozważona została również redukcja szerokości magistrali adresowej dla każdej komórki pamięci jak również możliwość dzielenia wspólnej pamięci dla komórek pamięci o tej samej zawartości. W ostatniej części artykułu podano wyniki implementacji dla układów firmy Xilinx serii XC4000 oraz Virtex.

Słowa kluczowe: układy dedykowane, procesory specjalizowane, układy programowalne, arytmetyka rozproszona

1. WPROWADZENIE

Mnożenie jest jedną z podstawowych operacji w cyfrowym przetwarzaniu danych. Niestety dla niektórych aplikacji, układ procesora ogólnego przeznaczenia jak również układ procesora DSP nie pozwalają sprostać postawionym wymaganiom czasowym, dlatego w tej sytuacji należy zastosować układy dedykowane ASIC (ang. *Application Specific Integrated Circuit*) lub układy programowalne FPGA (ang. *Field Programmab-*

le Gate Array). Architektura układu mnożącego dla układów ASIC i FPGA wydaje się być taka sama — implementacja układu mnożącego przebiega w podobny sposób i wykorzystywany jest podobny algorytm, jak np. układ mnożarki równoległej typu macierzowego (ang. *parrallel-array*) [1] czy też z wykorzystaniem drzewka Wallace [2]. Jednakże najbardziej znaczącą różnicą pomiędzy układami FPGA a układami ASIC jest to że układy FPGA mogą być łatwo i szybko przeprogramowane. To z kolei umożliwia zmianę współczynnika mnożenia dla układów FPGA albo przez zmianę wartości jednego z argumentu mnożenia dla układu mnożącego o zmiennym współczynniku mnożenia VCM (ang. *Variable Coefficient Multiplier*) lub też zamianę architektury układu mnożącego o stałym współczynniku mnożenia KCM (ang. *Constant Coeficient Multiplier*), dla którego współczynnik mnożenia jest określony przez strukturę układu. Układ mnożący KCM w porównaniu z układem VCM zajmuje dużo mniej powierzchni (26–33% układu VCM dla FPGA [3, 4]) i dlatego jest układem zalecanym pod warunkiem, że współczynnik mnożenia jest relatywnie stały [5] podczas procesu obliczeniowego.

W układach FPGA logika jest implementowana poprzez wykorzystanie pamięci typu LUT (ang. *Look-Up Table*), dlatego naturalną architekturą implementacji mnożenia KCM wydaje się być mnożenie z wykorzystaniem pamięci LUT w układach LM (ang. *LUT based Multiplication*). W celu zmniejszenia rozmiaru pamięci LUT, duża pamięć LUT jest dzielona na mniejsze pamięci w połączeniu z odpowiednimi układami dodającymi [3, 4, 6]. Z drugiej strony w układach ASIC, KCM jest implementowany z wykorzystaniem układów mnożących bezmnożnych MM (ang. *Multiplierless Multiplication*) wykorzystując operacje dodawania i odejmowania [7]. W konsekwencji architektury KCM implementowane w układach ASIC i FPGA są inne. Układy FPGA nowszej generacji posiadają wbudowany układ propagacji przeniesienia dla układu dodającego z przeniesieniem skrośnym (ang. *ripple-carry adder*) dzięki czemu układ dodający zajmuje dwa razy mniejszą powierzchnię układu scalonego i jego szybkość uległa znaczącej poprawie [8]. Jednakże ulepszenie układu dodającego nie zostało uwzględnione w efektywności architektur MM i LM dla układów FPGA i w konsekwencji architektura MM nie jest zalecana dla układów FPGA. Dlatego celem tej publikacji jest porównanie wyników implementacji LM i MM w układach FPGA z uwzględnieniem ich wzajemnej relacji czasowo-powierzchniowej.

W pierwszej części artykułu opisane zostały układy MM wykorzystujące reprezentację kanoniczną cyfry ze znakiem CSD (ang. *Canonic Sign Digit*) i dzielenie wspólnej podstruktury SS (ang. *Sub-structure Sharing*). Zaprezentowany został również nowy, zmodyfikowany algorytm konwersji z kodu uzupełnień do dwóch do kodu CSD, który uwzględnia budowę układu dodającego w układach FPGA. W dalszej części artykułu szczególnie rozpatrywana jest architektura LM. Układy FPGA posiadają różne moduły pamięci, np. układy Virtex firmy Xilinx posiadają pamięci 16×1, 32×1, 256×1, itd.), dlatego znalezienie optymalnej kombinacji tych pamięci i układów dodających jest skomplikowanym zadaniem, któremu poświęcono dużo uwagi. Ponadto, niektóre komórki pamięci mogą posiadać skróconą szerokość magistrali adresowej lub też dwie (a nawet więcej) komórki pamięci posiadające tę samą zawartość, przez co mogą być

zastąpione tylko przez jedną, wspólną komórkę pamięci. W tym miejscu należy zwrócić uwagę, że każdy etap referowanej pracy jest potwierdzony rzeczywistymi wynikami badań, weryfikując w ten sposób prowadzone rozważania i stawiane tezy dotyczące różnych architektur.

2. UKŁADY MNOŻĄCE BEZMNOŻNE (MM)

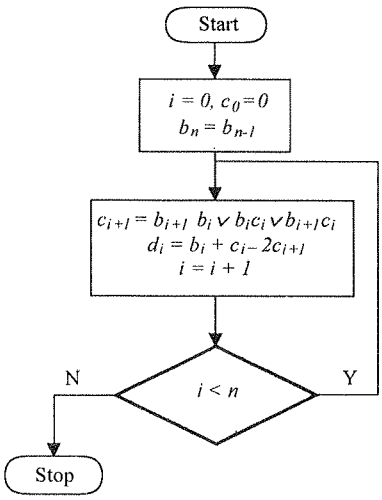
Układ mnożący o stałym współczynniku mnożenia KCM jest najczęściej implementowany w postaci bezmnożnej MM (ang. *Multiplierless Multiplication*) czyli przy wykorzystaniu tylko układów dodających i odejmujących oraz przesunięcia bitowego, których struktura jest określona przez wartość współczynnika mnożenia, a konkretnie jego binarnej reprezentacji BR. Dla przykładu: $A \cdot B$, gdzie $B = 14 = 1110_2$ może być zaimplementowane jako $(A \ll 1) + (A \ll 2) + (A \ll 3)$, gdzie „$A \ll n$" oznacza przesunięcie bitowe liczby A o n bitów w lewo. Należy podkreślić, że koszt (zajmowany obszar układu scalonego) układu mnożącego zależy bezpośrednio od wartości współczynnika mnożącego — liczba jedynek w jego binarnej reprezentacji powinna być jak najmniejsza. W związku z tym wiele algorytmów zostało tak zorganizowanych, aby sprostać temu wymogowi w celu zmniejszenia kosztów implementacji układu mnożącego, np. dla filtrów FIR [9]. Jednakże w tym artykule wartość współczynnika mnożenia jest narzuconym parametrem wejściowym, którego wartość nie może być zmieniona.

2.1. POSTAĆ KANONICZNA CYFRY ZE ZNAKIEM (CSD)

W reprezentacji CSD redukcję obszaru zajmowanego przez układ mnożący uzyskuje się poprzez redukcję jedynek w reprezentacji liczby [10]. W reprezentacji CSD w porównaniu z BR (reprezentacja binarna) wprowadzona została dodatkowa cyfra $\overline{1}$ mająca wartość -1 i symbolizująca operację odejmowania. Użycie reprezentacji CSD powoduje, że mnożenie może być przeprowadzone z użyciem mniejszej liczby operacji dodawania (odejmowania) w porównaniu z reprezentacją binarną. W poprzednim przykładzie dla BR: mnożenie $A \cdot B$, gdzie $B = 14 = 1110_2 = 100\overline{1}0_{CSD}$ wymagało 2 operacji dodawania, natomiast dla CSD wymagana jest tylko jedna operacja odejmowania: $(A \ll 4) - (A \ll 1)$. Przeciętnie reprezentacja CSD zawiera o 33% mniej jedynek niż BR [10], co pociąga oszczędności rzędu 33%.

Dodatkowa cyfra $\overline{1}$ umożliwia zakodowanie cyfry binarnej na wiele różnych sposobów. Przykładowo liczba $11 = 1011_2$ może być zakodowana w następujący sposób: $1011_{CSD} = 8+2+1$, $110\overline{1}_{CSD} = 8+4-1$, $10\overline{1}0\overline{1}_{CSD} = 16-4-1$. Sposób kodowania liczby w kodzie CSD decyduje o koszcie układu mnożącego, dlatego konwersja do kodu CSD zostanie przedstawiona bardziej szczegółowo.

Standardowa konwersja z reprezentacji binarnej (lub uzupełnień do dwóch — TC) liczby $B = b_{n-1}, b_{n-2}, \ldots, b_0$ do reprezentacji CSD $D = d_{n-1}, d_{n-2}, \ldots, d_0$ przebiega według algorytmu przedstawionego na rysunku 1 [7].



Rys. 1. Algorytm konwersji do reprezentacji CSD

Podstawą działania powyższego algorytmu jest założenie, że w kodzie CSD nie mogą wystąpić obok siebie dwie (lub więcej) jedynki. W przeciwnym przypadku na pozycję najmniej znaczącej jedynki wpisywany jest symbol $\overline{1}$. Dzieje się to według zależności $011..11 = 100..0\,\overline{1}$. Jak widać w rezultacie działania tego algorytmu liczba operacji odejmowania lub dodawania ulega zmniejszeniu lub pozostaje niezmieniona. Warto w tym miejscu podkreślić, że powyższy algorytm konwersji do CSD traktuje operację dodawania i odejmowania jako operacje równoważne, tzn. zajmujące taką samą powierzchnię układu scalonego. Dla przykładu, liczba $3 = 11_2$ podlega konwersji na $10\,\overline{1}$ co nie zmniejsza liczby operacji, a wprowadza odejmowanie zamiast dodawanie.

W przypadku układów FPGA operacje dodawania i odejmowania nie są sobie równoważne. Aby to wykazać przybliżona zostanie struktura układu dodającego i odejmującego implementowanego w układach FPGA. Operacja dodawania $S = A + B$ wykorzystuje następujące operacje:

$$s_i = a_i \; xor \; b_i \; xor \; c_i \tag{1}$$

$$if \; a_i = b_i \; then \; c_i = a_i \; else \; c_i = c_{i-1}. \tag{2}$$

Podobnie operacja odejmowania $S = A - B$ może być wyrażona jako:

$$s_i = a_i \; xor \; not \; b_i \; xor \; c_i \tag{3}$$

$$if \; a_i = not \; b_i \; then \; c_i \; a_i \; else \; c_i = c_{i-1}. \tag{4}$$

Dla układów FPGA (np. uklady serii XC4000 firmy Xilinx) operacje (1) i (3) są implementowane w czterowejściowych pamięciach LUT, które mogą zaimplementować

dowolną funkcję logiczną czterech argumentów. Operacje (2) i (4) są implementowane w dedykowanym układzie logiki przeniesienia [8], dzięki czemu sygnał przeniesienia propagowany jest 5–50 razy szybciej niż w przypadku standardowych operacji oraz nie zajmuje dodatkowej powierzchni układu. Porównując operacje (1–4) można zobaczyć, że operacja odejmowania wymaga tylko dodatkowej operacji negacji bitów b_i, dlatego założenie równoważności (tego samego kosztu) operacji odejmowania i dodawania wydaje się być spełnione, co zresztą jest potwierdzone w ogólnym przypadku dla układów FPGA. Jednakże dla dodawania, dla którego jeden z argumentów jest przesunięty bitowo w lewo, bity mniej znaczące (LSB) argumentu drugiego są bezpośrednio kopiowane na wyjście, dzięki czemu zmniejsza się koszt układu dodającego. Niestety w przypadku operacji odejmowania bity LSB odjemnika nie mogą być kopiowane na wyjście ponieważ bity te muszą być zanegowane i musi zostać dodana jedynka do najmniej znaczącego bitu ($c_0 = 1$ w przypadku odejmowania, patrz operacje (3, 4)). W konsekwencji operacje dodawania i odejmowania nie są równoważne. Na przykład dla operacji mnożenia przez $3 = 11_2 = 10\overline{1}_{CSD}$ (dla BR) LSB może być skopiowany bezpośrednio na wyjście, natomiast dla CSD nie jest to możliwe. Dla układów ASIC, dodatkowe operacje negacji i dodanie jedynki do LSB mogą być zaimplementowane na półsumatorach, których koszt jest znacznie niższy. Niestety w przypadku układów FPGA koszt pełnego sumatora i półsumatora jest w przybliżeniu taki sam.

2.2. ZMODYFIKOWANY ALGORYTM KONWERSJI DO CSD

Standardowy algorytm konwersji reprezentacji BR do kodu CSD zakłada, że operacje dodawania i odejmowania są sobie równoważne, czyli nie uwzględnia wcześniejszych rozważań. W konsekwencji algorytm konwersji został zmodyfikowany tak, aby konwersja do symbolu $\overline{1}$ (symbolizującego operacje odejmowania) została przeprowadzona tylko wtedy, gdy liczba operacji (liczba niezerowych symboli w reprezentacji CSD) ulega zmniejszeniu.

W celu opisania zmodyfikowanego algorytmu konwersji do CSD (nazywanego dalej MCSD) zostanie wprowadzona nowa funkcja $Q(i, j)$, gdzie $j \geq i$. Niech b_j reprezentuje j-ty bit reprezentacji binarnej (lub uzupełnień do dwóch) współczynnika mnożenia $B (M = A \cdot B)$. Funkcja $Q(i, j)$ niech będzie zdefiniowana iteracyjnie:

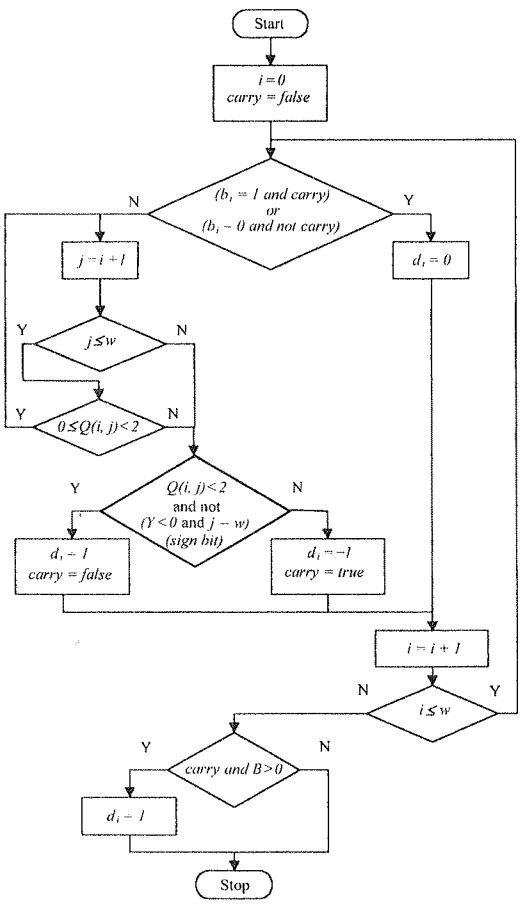
$$Q(i, i) = 0$$

$$Q(i, j+1) = \begin{cases} Q(i, j)+1 & gdy \quad b_{j+1} = 1 \\ Q(i, j)-1 & gdy \quad b_{j+1} = 0 \end{cases} \tag{5}$$

Funkcja $Q(i, j)$ jest inkrementowana jeżeli binarny symbol b_{j+1} jest jedynką i dekrementowana jeżeli jest zerem.

Zmodyfikowany algorytm konwersji (MCSD) jest pokazany na rysunku 2. Konwersja do symbolu $\overline{1}$ ma miejsce tylko wtedy gdy ogólna liczba operacji ulegnie zmniej-

szeniu. W konsekwencji liczba kolejnych jedynek w BR musi być równa co najmniej 3. Jeżeli natomiast zero przerywa sąsiadujące jedynki wtedy liczba kolejnych jedynek (po odrzuceniu zera) powinna być zwiększona o 2, lub też równoważnie, licznik jedynek powinien być zmniejszony o 1, jak to ma miejsce w przypadku funkcji $Q(i, j)$. Proces liczenia jedynek (funkcja $Q(i, j)$) powinien być zakończony w momencie kiedy funkcja $Q(i, j)$ jest mniejsza od zera — wstawiony ma być symbol 0 lub 1, lub też jeżeli jest równa 2 — wstawiony symbol $\overline{1}$.



Rys. 2. Zmodyfikowany algorytm konwersji do CSD (gdzie: $b_i - i$-ty bit BR, $Q(i, j)$ — funkcja zdefiniowana w równaniu (5), w — szerokość współczynnika mnożenia-1 czyli liczba bitów, na których należy zapisać współczynnik mnożenia-1)

Przykładowe wyniki dla standardowego i zmodyfikowanego algorytmu konwersji podane są w Tabeli 1. Dla liczby *3* i *11* standartowy algorytm wprowadził symbol $\overline{1}$, mimo tego, że liczba operacji nie uległa zmianie, a jedynie operacja dodawania jest

zastąpiona operacją odejmowania. Zmodyfikowany algorytm w tym przypadku nie wprowadza żadnych zmian co jest w zgodzie z przyjętym założeniem wprowadzania zmian w kodzie tylko w wypadku redukcji liczby operacji. W przypadku liczby 7 zastosowanie znaku $\overline{1}$ zmniejsza liczbę operacji dlatego oba algorytmy konwersji wprowadzają ten symbol. W przypadku liczby 23 zmodyfikowany algorytm wprowadza tylko jeden symbol $\overline{1}$ natomiast algorytm standardowy aż dwa symbole $\overline{1}$.

Tabela 1

Przykład wyników dla standardowego i zmodyfikowanego algorytmu konwersji do CSD

Współczynnik	BR	CSD	MCSD
3	11	$10\overline{1}$	11
7	111	$100\overline{1}$	$100\overline{1}$
11	1011	$10\overline{1}0\overline{1}$	1011
23	10111	$10\overline{1}00\overline{1}$	$1100\overline{1}$

2.3. DZIELENIE WSPÓLNEJ PODSTRUKTURY (SS)

Inną metodą redukcji obszaru zajmowanego przez układ mnożący jest dzielnie wspólnej podstruktury SS (ang. *Sub-structure Sharing*) [11]. Na przykład, mnożenie przez liczbę $27 = 11011_2$ może być zrealizowane z użyciem dodatkowej zmiennej pomocniczej *tmp*, tak jak to pokazano w poniższym równaniu, dla którego liczba operacji dodawania została zmniejszona z 3 do 2.

$$\begin{aligned} tmp &= a + (a \ll 1) \\ 27 \cdot a &= tmp + (tmp \ll 3) \end{aligned} \tag{6}$$

Operacja SS może być również przeprowadzona na reprezentacji CSD, dlatego kombinacja SS i CSD została również uwzględniona podczas implementacji algorytmu optymalizacji generowanego układu. Z drugiej jednak strony, CSD może zakłócać optymalizację SS, dlatego algorytm SS został zaimplementowany zarówno na BR jak i CSD, a wybrany został lepszy rezultat.

2.4. WYNIKI DOŚWIADCZALNE

Porównanie opisanych w tym rozdziale technik jest trudne ze względu na to, że wynik i wybór optymalnego algorytmu w dużym stopniu zależy od wartości współczynnika mnożenia. Jednakże można wyciągnąć ogólne wnioski dotyczące średniego i maksymalnego kosztu układu oraz liczby wystąpień najlepszego algorytmu.

Wyniki doświadczalne zostaną przedstawione dla 8-bitowego wejścia bez-znaku (najbardziej popularny format danych dla przetwarzania obrazu) oraz współczynników mnożenia o szerokości $K = 3-12$ bitów. Wyniki podane są w Tabeli 2.

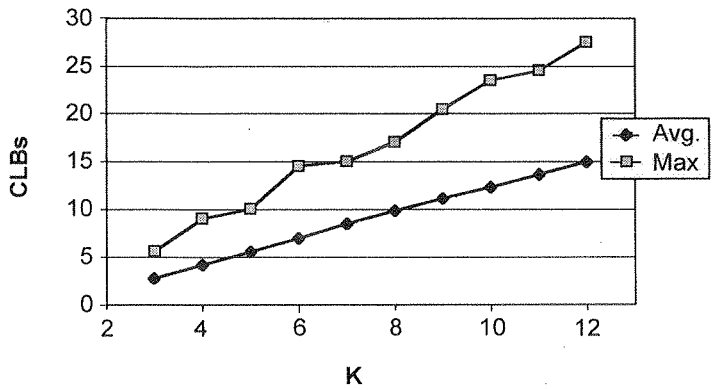
Tabela 2

Średnia i maksymalna liczba komórek CLB dla XC4000 oraz odpowiadająca im liczba operacji dodawania (odejmowania) L dla najlepszego algorytmu (wybieranego osobno dla każdego współczynnika mnożenia). K – maksymalna szerokość współczynnika mnożenia (współczynnik mnożenia w zakresie $[1, 2^K-1]$). Najlepszy algorytm — liczba współczynników mnożenia dla których podany algorytm daje najlepszy rezultat, jeżeli rezultat (koszt w CLB a nie liczba operacji) dwóch algorytmów jest taki sam to wybierany jest algorytm wcześniejszy (w tabeli bardziej na lewo)

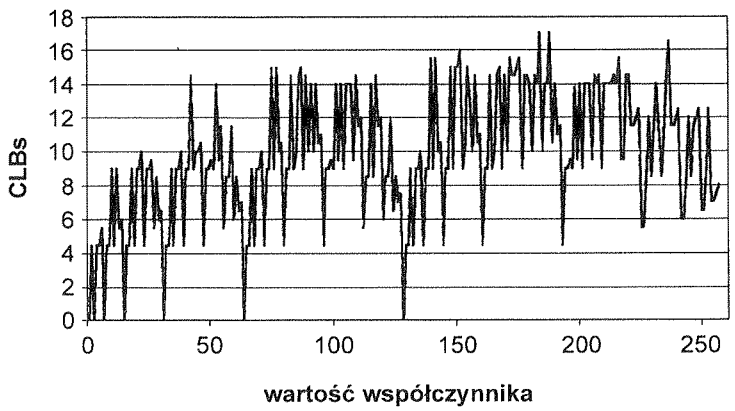
K	Średnio		Max.			Najlepszy algorytm			
	CLB	L	CLB	L	Współ.	BR	CSD	SS	CSD-SS
3	2,71	0,57	5,5	1	7	6	1	0	0
4	4,13	0,87	9	2	11	12	3	0	0
5	5,52	1,16	10	2	23	22	9	0	0
6	6,92	1,44	14,5	3	43	39	17	4	0
7	8,44	1,75	15	3	75	68	43	16	0
8	9,8	2,03	17	3	183	114	94	44	3
9	11,1	2,31	20,5	4	309	188	193	107	15
10	12,3	2,57	23,5	4	747	300	407	254	62
11	13,6	2,83	24,5	4	1463	478	797	579	193
12	14,9	3,08	27,5	4	3381	746	1510	1285	554

Analizując wyniki w Tabeli 2 można zauważyć, że algorytmy optymalizacji (CSD, SS, CSD-SS) są bardziej skuteczne dla szerszych współczynników mnożenia. Średnia liczba operacji dla BR jest w przybliżeniu równa *K/2-1*, co dla $K = 12$ daje *5*, w porównaniu z *3,08* z zastosowaniem technik optymalizacji. Średni koszt układu mnożącego wzrasta w przybliżeniu liniowo ze wzrostem szerokości współczynnika mnożenia K, jak to jest pokazane na wykresie (rys. 3 — linia Avg). Podobny jest przebieg maksymalnego kosztu (na wykresie linia Max), chociaż dla tego przebiegu odchylenia są znacznie większe, szczególne dla K, dla którego następuje wzrost maksymalnej liczby operacji L (dla $K = 4, 6, 9$). Warto zwrócić uwagę, że występuje wzrost kosztu maksymalnego pomimo, że liczba operacji pozostaje stała. W konsekwencji, wybór najlepszego algorytmu nie zależy tylko od liczby operacji dodawania (odejmowania), ale musi zostać uwzględniony całkowity koszt układu mnożącego.

Koszt układu mnożącego silnie zależy od wartości współczynnika mnożącego co jest pokazane na wykresie (rys. 4). Można zauważyć, że najbardziej kosztowny układ mnożący występuje dla współczynnika o wartości 60–75% pełnego zakresu binarnego.



Rys. 3. Średnia i maksymalna powierzchnia zajmowana przez 8-bitowy układ mnożący dla różnej szerokości współczynnika mnożącego K



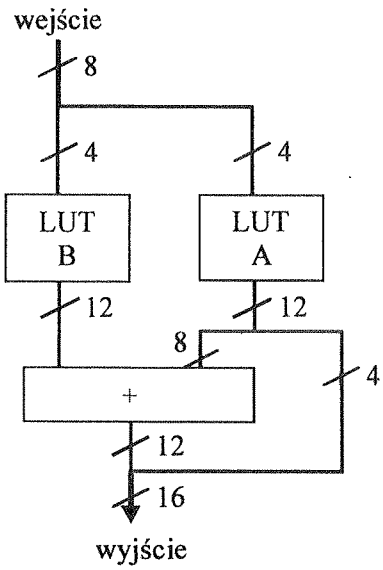
Rys. 4. Koszt układu mnożącego o wejściu 8-bitowym dla różnych wartości współczynników mnożenia. Układ był implementowany w XC4000

3. UKŁADY MNOŻĄCE Z WYKORZYSTANIEM PAMIĘCI LUT (LM)

3.1. ZASADA DZIAŁANIA

Obliczenie dowolnej ograniczonej wartościowo funkcji możliwe jest z wykorzystaniem pamięci typu LUT, na której wejście adresowe podane są wartości argumentów, a na wyjściu otrzymuje się wynik funkcji. Teoretycznie zastosowanie pamięci LUT daje najszybszą implementację funkcji ponieważ nie jest wymagane zastosowanie żadnych

układów arytmetycznych. Niestety, zastosowanie w praktyce pojedynczej pamięci LUT dla mnożenia jest możliwe tylko dla małych wartości argumentu, ponieważ wielkość pamięci wzrasta wykładniczo wraz z szerokością argumentu. Na przykład, dla L-bitowego argumentu i K- bitowego współczynnika mnożenia, wymagana jest pamięć o rozmiarze $(L+K)\cdot 2^L$. Dlatego praktycznym rozwiązaniem jest zastosowanie kombinacji małych pamięci LUT i układów dodających, poprzez rozbicie argumentu, użycie małych pamięci LUT i następnie odpowiednie dodanie wyjść pamięci LUT [3, 6, 12]. Przykład takiego układu mnożącego jest pokazany na rysunku 5, gdzie operacja mnożenia $Y = A \cdot B$ jest przeprowadzona według zależności $Y = 2^4 \cdot B \cdot A_{bity\ 7-4} + B \cdot A_{bity\ 3-0}$.



Rys. 5. Układ mnożący LM dla szerokości argumentu L = 8 i szerokości współczynnika mnożącego K = 8

Zawartość pamięci LUT dla mnożenia $Y = A \cdot B$ może być obliczona bezpośrednio z mnożenia tak jak to jest pokazane na przykładzie w Tabeli 3.

Warto zwrócić uwagę, że zawartość komórek pamięci LUT zależy tylko od linii adresowych o wadze mniejszej lub równej wadze komórki wyjściowej. Dlatego w przedstawionym przykładzie w Tabeli 3, komórka pamięci y_0 zależy tylko od linii adresowej a_0; y_1 zależy od a_0 i a_1; itd. W rezultacie szerokość linii adresowej wzrasta z wagą bitu wyjściowego pamięci LUT. Maksymalna szerokość magistrali adresowej jest jednak ograniczona od góry przez szerokość n magistrali adresowej pamięci LUT. W rezultacie szerokość magistrali adresowej komórki y_i w ogólnym przypadku wynosi $MAX(i+1, n)$. W konsekwencji $n-1$ najmniej znaczących bitów pamięci wymaga mniejszej szerokości magistrali adresowej, a przez to i mniejszych rozmiarów pamięci, co powoduje znaczącą redukcję kosztów pamięci i będzie oznaczane dalej jako LAWR (ang. *LSBs Address Width Reduction*).

Tabela 3

Zawartość komórek pamięci LUT ($y_5 - y_0$) dla różnych wartości argumentu podanego na magistralę adresową; współczynnik mnożenia wynosi 19; szerokość adresu — szerokość magistrali adresowej dla poszczególnych komórek pamięci

Adres	wartość	y_5	y_4	y_3	y_2	y_1	y_0
0	0	0	0	0	0	0	0
1	19	0	1	0	0	1	1
2	38	1	0	0	1	1	0
3	57	1	1	1	0	0	1
Szerokość adresu		1	1	2	2	2	1

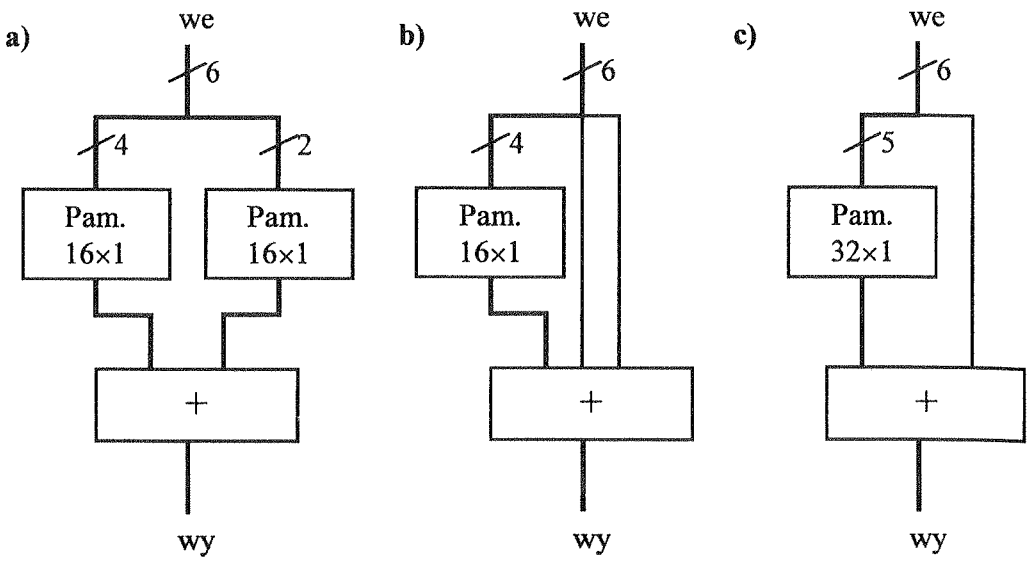
Dodatkową redukcję szerokość magistrali adresowej można zaobserwować dla komórek pamięci, dla których zawartość pamięci nie zależy od pewnej linii adresowej. Osiągnięte dzięki tej metodzie oszczędności będą dalej w skrócie nazywane jako DAWR (ang. *Don't care Address Width Reduction*). DAWR jest z reguły obserwowany dla najbardziej znaczących bitów pamięci LUT, co potwierdza przykład w Tabeli 3, dla którego DAWR występuje dla komórek pamięci y_5 i y_4. Dodatkowe oszczędności można osiągnąć dzięki dzieleniu wspólnych komórek pamięci MS (ang. *Memory Sharing*), jeżeli te komórki mają taką samą zawartość. Na przykładzie Tabeli 3 można zauważyć, że komórki pamięci y_0 i y_4 są takie same, dlatego tylko jedna z nich jest wystarczająca do zastosowania. Warto w tym miejscu podkreślić, że redukcję szerokości magistrali adresowej DAWR oraz dzielenie wspólnej pamięci MS nie można uogólnić i zależą one silnie od wartości współczynnika mnożenia i szerokości magistrali adresowej. Ponadto, wymagany jest zaawansowany algorytm przeszukujący i implementujący wspomniane techniki. W przedstawionych rozważaniach zastosowano algorytm porównujący wszystkie możliwe kombinacje komórek i linii adresowych.

3.2. IMPLEMENTACJA UKŁADU LM W UKŁADACH FPGA

W układzie mnożącym rozbicie dużej pamięci LUT na mniejsze powinno być przeprowadzone z uwzględnieniem relacji koszt–rozmiar pamięci oraz koszt–rozmiar układów dodających. Układy serii XC4000 firmy Xilinx mają wbudowane pamięci 16×1 oraz 32×1, jednakże koszt pamięci 32×1 jest równy 1 CLB i jest dwa razy większy niż dla pamięci 16×1 ($^1/_2$ CLB). Koszt układu dodającego jest równy $^1/_2$ CLB/bit. Ponadto, istnieje dodatkowa wirtualna pamięć 2×1, która może być zaimplementowana jako bezpośrednie połączenie (z ominięciem pamięci — zerowy koszt) do układu dodającego.

W konsekwencji, znalezienie optymalnej kombinacji różnych modułów pamięci oraz układów dodających jest skomplikowanym zadaniem, które zależy od rozmiaru

danej wejściowej i narzuconej wartości współczynnika mnożenia. Na podstawie przeprowadzonych badań doświadczalnych, dla danej wejściowej dużo większej niż 4 preferowanym rozmiarem pamięci jest pamięć 16×1. Jeżeli jednak szerokość argumentu nie dzieli się przez 4, mogą być użyte różne bloki pamięci.

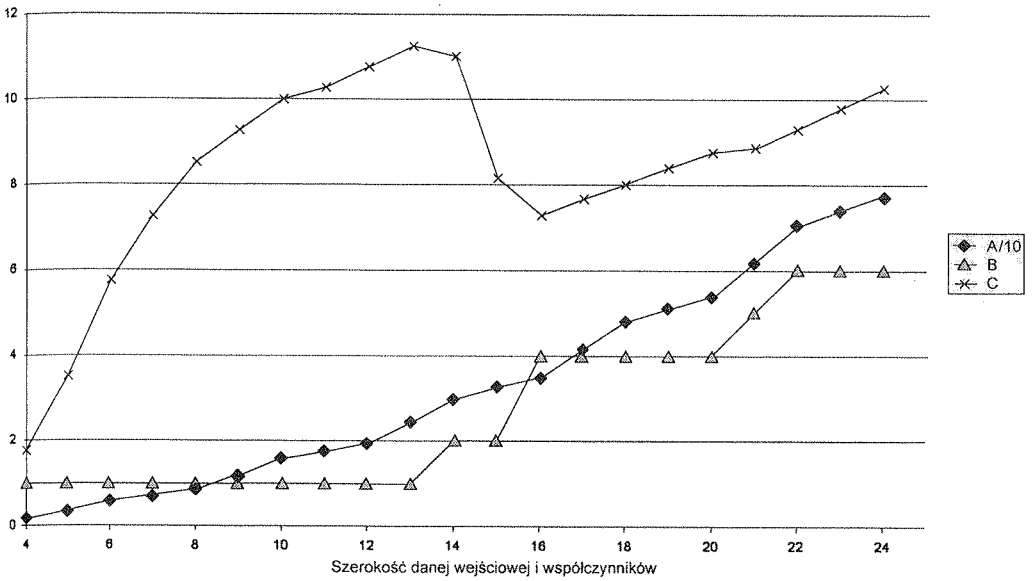


Rys. 6. Różne korzystne metody implementacji układu mnożącego LM dla szerokości danej wejściowej równej 6

Rysunek 6 przedstawia kilka dobrych rozwiązań układu mnożącego; koszt układów mnożących implementowanych w układach XC4000 wyniesie dla poszczególnych rozwiązań odpowiednio: a) $11^1/_2$ CLB, b) 12 CLB, c) 12 CLB. Jak widać różnica kosztów dla poszczególnych rozwiązań jest niewielka. Uogólniając, jednym z najlepszych rozwiązań jest układ, który używa tylko pamięci 16×1. Jeżeli pozostaje tylko jedna linia adresowa, która jest niepodzielna przez 4, należy użyć bezpośredniego połączenia do układu dodającego.

Układy serii Virtex firmy Xilinx oferują jeszcze bogatszy wybór pamięci. Oprócz pamięci 16×1 i 32×1 wprowadzono w układach Virtex dodatkowe duże bloki pamięci BSR (ang. *Block Select RAM*) o rozmiarze 4kb i różnej szerokości magistrali danych: 4k×1, 2k×2, 1k×4, 512×8, 256×16 [13]. Powierzchnia krzemu zajmowana przez pojedynczy BSR jest porównywalna do powierzchni zajmowanej przez 16 Virtex CLB (64 RAM 16×1). Jednakże rzeczywisty koszt tych pamięci może zależeć od wolnych zasobów układu FPGA, np. układ ma już zajęte wszystkie CLB, ale ma jeszcze wolne zasoby BSR, dlatego rzeczywistą zależność kosztów tych pamięci należy dostosować do aktualnego projektu. Wykres na rysunku 7 przedstawia ekwiwalentny koszt pamięci BSR (dodatkowy koszt układu po zastąpieniu pamięci BSR pamięciami 16×1 i 32×1

i podzieleniu przez liczbę użytych BSR) dla różnych szerokości wejścia i współczynnika mnożenia. Na podstawie tego wykresu można odczytać, że pojedyncza pamięć BSR może być zastąpiona przez średnio 8–11 VCLB (około 32–44 RAM 16×1).



Rys. 7. Powierzchnia układu LM dla różnej szerokości magistrali wejściowej i współczynnika mnożenia K: a) powierzchnia (podawana dla Virtex CLB (1 VCLB = 2 XC4000 CLB) w skali 1:10 dla LM z użyciem tylko pamięci 16×1 i 32×1; b) liczba użytych bloków BSR 256×16; c) ekwiwalentny koszt (podawany dla VCLB) jednego bloku BSR w porównaniu z rozwiązaniem przy użyciu tylko pamięci 16×1 i 32×1

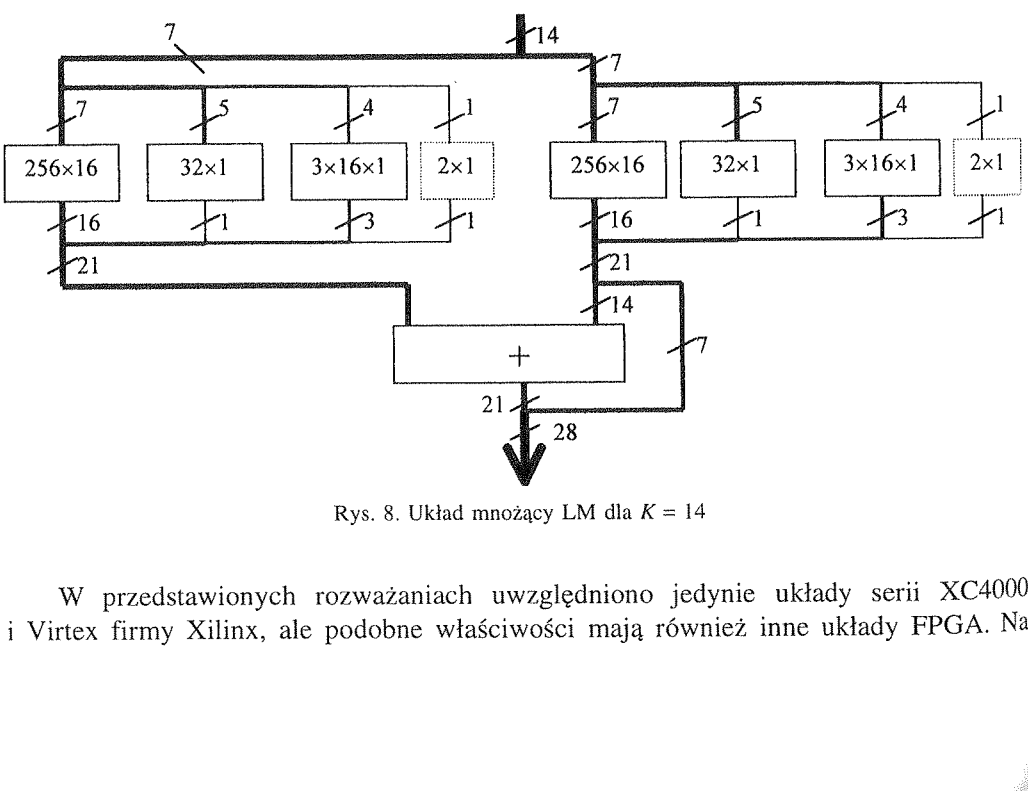
Bloki pamięci BSR są relatywnie duże i dlatego trudno jest znaleźć architekturę układu mnożącego, dla której są one całkowicie wykorzystane, a stopień wykorzystania bloków pamięci bardzo wpływa na ich ekwiwalentny koszt. Konsekwentnie, dla małej szerokości magistrali wejściowej i współczynnika mnożenia K ($K<8$) ekwiwalentny koszt BSR jest raczej mały (rys. 7), co nie zachęca do użycia tych pamięci. Dla $K \leq 13$ ekwiwalentny koszt BSR wzrasta. Natomiast dla $K>13$, dwa lub więcej bloki BSR są użyte co wymaga użycia dodatkowych układów arytmetycznych oraz ponownie pojawia się problem pełnego wykorzystania pamięci BSR (występuje efekt ziarnistości dużych bloków BSR), przez co zmniejsza się ekwiwalentny koszt pamięci BSR. Dla $K = 16$, liczba użytych bloków BSR wzrasta gwałtownie, jako że bloki BSR są grupowane w pary stanowiące w sumie dużą pamięć 256×32, co również zmniejsza ekwiwalentny koszt BSR. Dalsze zwiększanie K powoduje wzrost ekwiwalentnego kosztu, wydaje się jednak, że wartość maksymalna 11 nie zostanie przekroczona.

Wykres A na rysunku 7 pokazuje koszt LM przy użyciu tylko pamięci 16×1 i 32×1. Można zauważyć gwałtowny wzrost kosztu układu mnożącego dla $K = 9, 13,$

17, 23, kiedy to K przekracza liczbę podzielną przez 4 — liczbę wejść do pamięci 16×1.

W przedstawionych rozważaniach w celu znalezienia optymalnej architektury układu mnożącego wykorzystano technikę przeszukiwania wszystkich możliwych kombinacji (ang. *exhausted search*). Zostały rozpatrzone wszystkie możliwe i korzystne kombinacje pamięci BSR (32×1 i 16×1) oraz układów dodających i został wybrany najlepszy układ. W celu wizualizacji rozważanych architektur na rysunku 8 przedstawiono przykład układu mnożącego, w którym użyto kombinacje różnych pamięci RAM i układu dodającego. W przykładzie tym (jak i we wszystkich rozważaniach w tym podrozdziale) nie uwzględniono możliwej optymalizacji MS i DAWR, ze względu na ich ścisłe powiązanie z wartością współczynnika mnożenia. Dlatego po uwzględnieniu tych optymalizacji schemat na rysunku 8 może ulec dalszej komplikacji.

Warto podkreślić, że opisane tutaj wyniki zostały uzyskane dzięki specjalnie zaprojektowanemu w tym celu pakietowi AuToCon (ang. *Automated Tool for Convolution processor design*) napisanemu w języku C++ i VHDL, który automatycznie generuje optymalną architekturę układu mnożącego. Dokładne opisanie działania tego narzędzia przekracza ramy tego artykułu. Program AuToCon nie zakłada jakichkolwiek związków pomiędzy kosztami modułów pamięci i układów dodających, dlatego moduły pamięci mogą być dowolnie zdefiniowane i program może być użyty dla różnych rodzin układów FPGA, a także dla projektów ASIC. Parametrami wejściowymi jest koszt układów dodających oraz rozmiar i koszt modułów pamięci. Wyjściem programu jest plik tekstowy VHDL, który po syntezie może być zaimplementowany do dowolnego układu FPGA.



Rys. 8. Układ mnożący LM dla $K = 14$

W przedstawionych rozważaniach uwzględniono jedynie układy serii XC4000 i Virtex firmy Xilinx, ale podobne właściwości mają również inne układy FPGA. Na

przykład, układy Apex 20k firmy Altera [14] mają podobne relacje kosztów jak w przypadku układów Virtex. Apex posiada czterowejściową pamięci LUT 16×1 i dedykowany układ generacji przeniesienia w każdym elemencie logicznym LE (ang. *Logic Element*) oraz posiada duże bloki pamięci (128×16, 256×8, 512×4, 1024×2 i 2048×1) w ESB (ang. *Embedded System Block*). Rozmiar pamięci dla układów Apex jest o połowę mniejszy niż w przypadku BSR układów Virtex, jakkolwiek w przeważającej większości 4 kb BSR nie jest w pełni wykorzystany i dlatego równie znaczącą różnicą jest brak pamięci 32×1 dla układów Apex.

Należy zwrócić uwagę, że liczba użytych bloków BSR zależy od relacji kosztów. Na przykład, dla 16-bitowego układu mnożącego ($K = 16$) liczba BSR wzrasta z malejącym kosztem pamięci BSR, jak to jest pokazane w Tabeli 4.

Tabela 4

Wpływ kosztu bloku pamięci BSR na architekturę LM dla $K = 16$; dla kosztów: 16×1 0,25 VCLB oraz sumator = 0,25 VCLB/bit

Koszt BSR [VCLB]	liczba BSRs	Koszt LUT RAM [VCLB]	Koszt Sumatora [VCLB]	Ekw. koszt BSR [VCLB]
≥7,75	0	19	16	–
≥6,5	2	9,5	10	7,75
≤6,25	4	0	6	7,25

Bloki pamięci BSR podczas implementacji bardzo często nie są w pełni wykorzystywane (patrz przykładowy schemat z rysunku 8, w którym jedna linia adresowa pamięci BSR nie została użyta), co może sugerować o ich nieoptymalnych rozmiarach. Pamięci BSR: 4k×1, 2k×2, 1k×4 oraz 512×8 nigdy nie zostały użyte, a wykorzystywane były tylko pamięci 256×16. W konsekwencji korzystne wydaje się wprowadzenie pamięci o szerszej magistrali danych, np. 128×32. Uogólniając, można stwierdzić, że optymalna wielkość pamięci, aby została w pełni wykorzystana powinna być następująca:

$$W_D = W_C + W_A - W_L \tag{7}$$

gdzie: W_D *— szerokość magistrali danych;* W_C *— szerokość współczynnika konwolucji;* W_A *— szerokość magistrali adresowej;* W_L *— szerokość magistrali adresowej następnej mniej kosztownej pamięci, dla układu Virtex* $W_L = 5$.

Podstawiając do równania (7) $W_C = 13$, $W_A = 8$, $W_L = 5$ otrzymujemy $W_D = 16$, co odpowiada pamięci 256×16, i dlatego maksymalny ekwiwalentny koszt pamięci BSR jest obserwowany na rysunku 7 dla szerokości współczynnika mnożenia $K = 13$.

Dodatkowe oszczędności można uzyskać w przypadku, gdy binarny zakres wejścia nie jest w pełni wykorzystywany. Na przykład, dla zakresu danej wejściowej 0-127

(zakres binarny) oraz dla zakresu 0-99 (zakres dziesiętny) i współczynnika mnożenia równego 81, wynik implementacji układu mnożącego wynosi odpowiednio: 14,5 CLB oraz 13,5 CLB (dla XC4000).

Ponadto zoptymalizować można arytmetykę liczb ujemnych. Generalnie obszar optymalizacji może być podzielony na cztery zakresy:

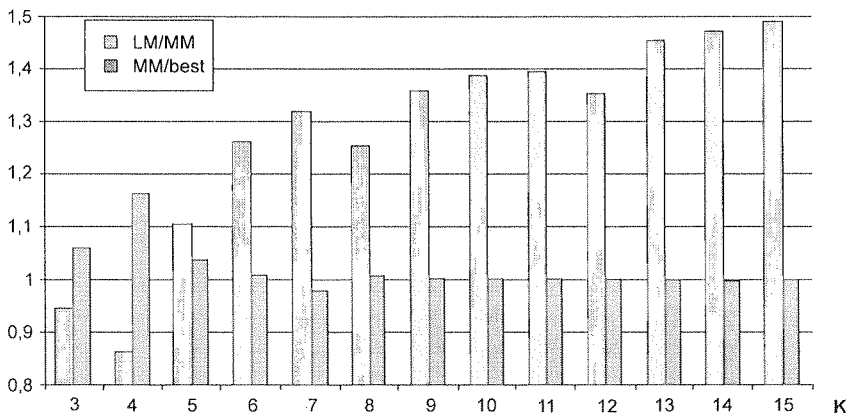
- Współczynnik mnożenia i dana wejściowa jest dodatnia – nie uzyskuje się żadnej optymalizacji.
- Współczynnik jest dodatni, a dana wejściowa jest zapisana w kodzie uzupełnień do dwóch. W tym przypadku tylko najbardziej znaczący blok pamięci operuje na liczbach ujemnych i dodatnich, pozostałe bloki tylko na liczbach dodatnich. W tym przypadku również nie można uzyskać żadnej optymalizacji.
- Ujemny współczynnik mnożenia i dodatnia dana wejściowa. Wszystkie pamięci LUT operują na danych ujemnych (w kodzie uzupełnień do dwóch — TC), dlatego dodawanie może być zastąpione odejmowaniem i dzięki temu wszystkie pamięci będą pracowały tylko na dodatnich liczbach (współczynnik mnożenia jest zanegowany, a przez to dodatni). W konsekwencji bit znaku nie będzie musiał być implementowany i wszystkie pamięci LUT będą o jeden bit węższe. Niestety podwójne odejmowanie ($s = -a - b$) nie może być zaimplementowane w układach FPGA i musi być odłożone do następnego poziomu dodającego ($s = -(a+b)$), co powoduje, że wynik dodawania jest zanegowany ($s = -a - b - c - \ldots = -(a + b + c + \ldots)$), a także na koniec konieczne jest dokonanie dodatkowej operacji negacji co jest raczej kosztowne. Dlatego najlepszym rozwiązaniem jest użycie odejmowania dla wszystkich pamięci LUT z wyjątkiem ostatniej najmniej znaczącej. Dzięki temu przełamany jest łańcuch podwójnego odejmowania. Wybór padł na LSB LUT ponieważ w tym wypadku możliwe jest kopiowanie bitów LSB przy przesunięciu bitowym dwóch argumentów jak to było opisane w podrozdziale 2.1.
- Ujemny współczynnik mnożenia i dana wejściowa w formacie uzupełnień do dwóch. W tym przypadku wszystkie pamięci oprócz MSB LUT operują na danych ujemnych i dlatego powinny być zaimplementowane tak jak w poprzednim zakresie. MSB LUT operuje jednak na danych w kodzie uzupełnień do dwóch dlatego jego wyjście może być dodawane do wyniku. Jednakże ta operacja dodawania może zakłócić etap bezpośredniego kopiowania LSB dla pozostałych operacji odejmowania. Dlatego korzystne wydaje się również dla tej pamięci zastosowanie operacji odejmowania, czyli układ działa podobnie jak dla poprzedniego zakresu.

4. PORÓWNANIE UKŁADÓW MNOŻĄCYCH

4.1. POWIERZCHNIA UKŁADU

W przedstawionych rozważaniach omówiono dwie różne techniki implementacji KCM: LM oraz MM. Dlatego powstaje pytanie, która z nich jest bardziej efektywna.

Wykres na rysunku 9 przedstawia ich wzajemne zależności uwzględniając wartości średnie zajmowanej powierzchni. Analizując wykres na rysunku 9 można zauważyć, że dla XC4000 układ LM jest preferowany dla $K \leq 4$, a dla $K \geq 5$ najlepszy rezultat jest z reguły osiągnięty dla układu MM. Warto jednak zwrócić uwagę, że wybór pomiędzy układami LM i MM zależy od wybranego współczynnika mnożenia i wykres z rysunku 9 pokazuje tylko statystyczną relację pomiędzy tymi układami. Dlatego obie architektury powinny być przeanalizowane i lepsza z nich wybrana dla konkretnego przypadku. Jednakże, analizując stosunek MM/best można dojść do wniosku, że dla $K > 5$ zysk z użycia układu LM jest minimalny i jego znaczenie spada wraz ze wzrostem K.



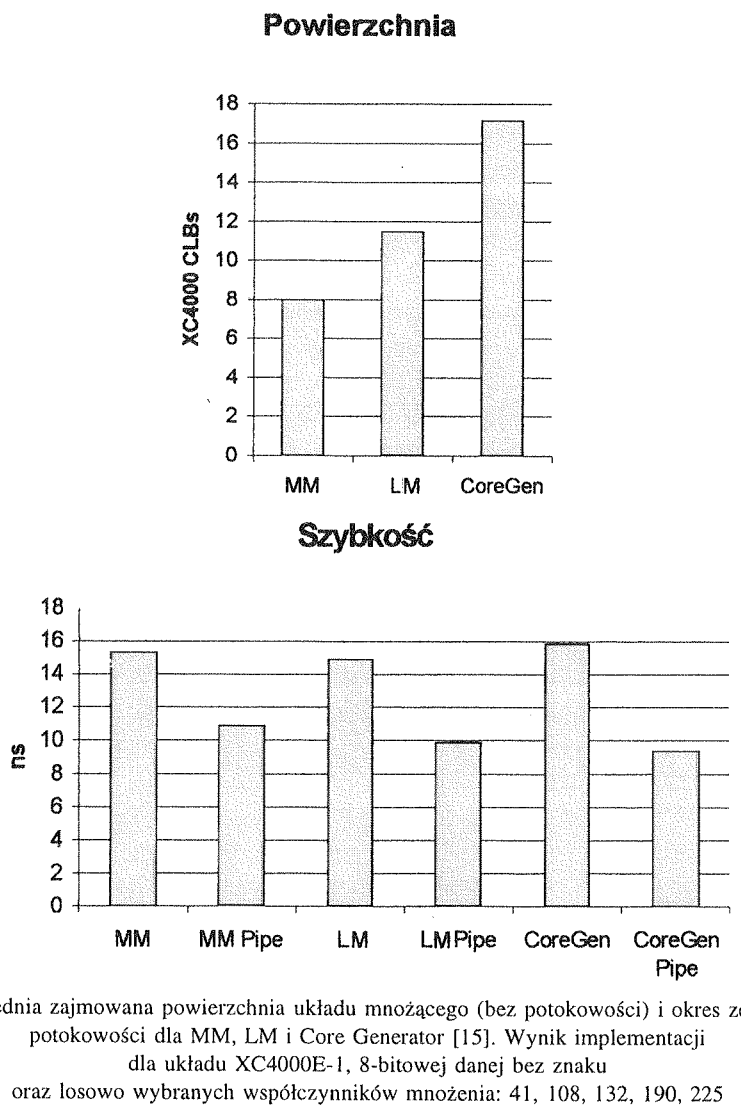
Rys. 9. Relacja pomiędzy stosunkiem średnich powierzchni układu XC4000 zajętych przez LM/MM oraz MM/best. Wynik dla różnej szerokości K danej wejściowej (zakres danej wejściowej $0 \div 2^K - 1$) i współczynnika mnożenia (zakres: $1 \div 2^K - 1$)

Z wykresu na rysunku 9 i Tabeli 2 można wysnuć bardziej ogólny wniosek: dla układu MM optymalizacja typu CSD i SS jest coraz bardziej skuteczna wraz ze wzrostem K.

Kolejnym aspektem wzajemnych porównań jest badanie relacji ilościowych jak dużą redukcję powierzchni można osiągnąć dzięki zastosowaniu technik optymalizacji MS i DAWR dla LM. Według wyników doświadczalnych uzyskuje się średnio zmniejszenie powierzchni układu o $5 \div 20\%$ w zależności od szerokości wejścia K, przy czym optymalizacja ta jest bardziej skuteczna dla małych pamięci LUT.

4.2. CZĘSTOTLIWOŚĆ PRACY

W poprzednim rozdziale rozważana była tylko powierzchnia zajmowanego układu. W rzeczywistym układzie należy jednak rozważyć zarówno powierzchnię jak i częstot-



Rys. 10. Średnia zajmowana powierzchnia układu mnożącego (bez potokowości) i okres zegara z/bez potokowości dla MM, LM i Core Generator [15]. Wynik implementacji dla układu XC4000E-1, 8-bitowej danej bez znaku oraz losowo wybranych współczynników mnożenia: 41, 108, 132, 190, 225

liwość pracy układu (w szczególności stosunek wartości powierzchni do częstotliwości pracy układu). W konsekwencji, aby zwiększyć przepustowość układu mnożącego należy zastosować architekturę potokową tym bardziej, że po każdym elemencie logicznym układy FPGA zawierają wbudowane rejestry typu D. Dlatego w pierwszym przybliżeniu zastosowanie architektury potokowej nie wymaga dodatkowej powierzchni układu FPGA. Jednakże, niektóre sygnały nie wymagają logiki, dlatego często rejestry potokowe muszą być użyte bez powiązanej logiki, zgodnie z zasadą tnij-wstaw (ang. *cut-set*) [7]. W konsekwencji dla całkowicie potokowej architektury (rejestry występują po każdej logice), powierzchnia zajmowana przez układ jest określona raczej przez liczbę rejestrów niż liczbę użytej logiki (układów dodających i pamięci). Dodatkowa

powierzchnia zajmowana przez układ potokowy znajduje się w zakresie 0–50% i redukuje się do zera dla mniejszej liczby etapów potokowości gdy np. rejestry są implementowane co dwa poziomy logiki. Jednakże redukcja liczby etapów potokowych powoduje zmniejszenie częstotliwości taktowania układu. Z drugiej strony, architektura potokowa powoduje znaczne zwiększenie częstotliwości pracy układu i dlatego w większości przypadków można zaakceptować wzrost powierzchni układu potokowego, szczególnie wtedy gdy nie jest on duży. Warto podkreślić, że koszt rejestrów został również brany pod uwagę podczas poszukiwania optymalnej architektury układu mnożącego. Dla przykładu, technika optymalizacji SS ma tendencję do implementacji dużej liczy rejestrów w przeciwieństwie do techniki CSD i dlatego metoda CSD jest skuteczniejsza dla architektury potokowej.

Wykresy na rysunku 10 przedstawiają średnią powierzchnię i okres zegara dla układów MM i LM. Widać, że układy MM są bardziej efektywne niż LM i charakteryzują się tylko nieznacznym spadkiem częstotliwości pracy spowodowanej propagacją przeniesienia w dedykowanym układzie dodającym. Spadek szybkości w niektórych przypadkach może powodować, że przy porównywalnych powierzchniach układu bardziej optymalną architekturą jest układ LM. Wykres na rysunku 10 przedstawia również wynik implementacji dla komercyjnego programu Core Generator [15]. Uwidacznia on także, że rozwiązanie układów MM i LM opracowane w ramach referowanych w tym artykule badań przewyższa to rozwiązanie komercyjne.

5. WNIOSKI

W prezentowanych rozważaniach przedstawiono dwie różne architektury implementacji mnożenia: LM i MM. Wyniki doświadczalne pokazują, że dla małych szerokości wejścia i współczynnika mnożenia układy LM są z reguły lepsze, ale ze wzrostem tych szerokości układy MM stają się coraz bardziej atrakcyjne i przewyższają układy LM. Dzieje się tak dzięki większej efektywności optymalizacji CSD i SS.

Ponadto, podano ulepszony algorytm konwersji z reprezentacji binarnej do kodu CSD. Algorytm ten uwzględnia fakt, że koszt odejmowania jest wyższy od kosztu dodawania ponieważ dla odejmowania nie zachodzi kopiowanie najmniej znaczących bitów przesuniętego w prawo odjemnika. W konsekwencji odejmowanie (symbol $\bar{1}$ w kodzie CSD) jest implementowane tylko wtedy gdy dzięki temu zmniejszy się ogólna liczba operacji. Oprócz tego, dla architektury LM opisane zostało użycie różnych bloków pamięci, co wymaga użycia zaawansowanych technik i wyszukiwania optymalnego rozwiązania. Na zakończenie zostały podane relacje pomiędzy powierzchnią, a szybkością omawianych architektur. Warto w tym miejscu podkreślić, że osiągnięte w ramach prezentowanych badań rozwiązanie przewyższa rozwiązanie komercyjne.

Przedstawiono także różne metody i rezultaty implementacji dla układu mnożącego przez stały współczynnik. Warto podkreślić, że największy zakres pracy, który został wykonany w ramach prowadzonych badań, został włożony w opracowanie zautomatyzowanego oprogramowania narzędziowego AuToCon do syntezy tych układów. Dzięki

temu, bez udziału człowieka i bez jego rozległej wiedzy na temat budowy układów mnożących w układach FPGA, generowany jest kod VHDL układu mnożącego, który po syntezie może być implementowany do rzeczywistego układu FPGA. Zbudowane oprogramowanie wspomagające AuToCon nie zakłada jakichkolwiek wstępnych relacji kosztów i dlatego może być użyte dla dowolnego układu FPGA, a nawet układów typu ASIC. Parametrami wejściowymi programu AuToCon są: zakres danej wejściowej, wartość współczynnika mnożenia, koszt układu dodającego i rejestrów oraz rozmiar i koszt pamięci. W konsekwencji dzięki opracowanemu oprogramowaniu AuToCon, nie tylko zredukowany został koszt układu mnożącego przez stały współczynnik, ale również znacząco został skrócony czas realizacji projektu.

BIBLIOGRAFIA

1. S. Waser: *High-speed monolithic multipliers for real-time digital signal processing.* IEEE Computer Magazine, Vol. 11, No. 10, pp. 19–29, 1978
2. C. S. Wallace: *A suggestion for a fast multiplier.* IEEE Trans. On Electron. Comput., Vol. EC-13, pp. 14–17, 1964
3. K. Chapman: *Constant Coefficient Multipliers for the XC4000E.* Xilinx Application Note, XAPP 054 December 1996
4. R. Petersen, B. L. Hutchings: *An Assessment of the Suitability of FPGA-Based Systems for Use in Digital Signal Processing.* In 5th International Workshop on Field Programmable Logic and Applications, Oxford England, pp. 293–302, August 1995
5. M. J. Wirthlin, B. L. Hutchings: *Improving Functional Density Through Run-Time Constant Propagation.* ACM/SIGDA International Symposium on Field Programmable Gate Arrays, pp. 86–92, 1997
6. K. Chapman: *Fast Integer Multiplier fit in FPGA's.* EDN 1993 Design Idea Winner, EDN May 12th 1994
7. P. Pirsch: *Architectures for Digital Signal Processing.* Wiley 1998
8. Xilinx Co.: *Using the Dedicated Carry Logic in XC4000E.* Xilinx Application Note XAPP 013 July 4, 1996
9. H. Samueli: *An imroved search algorithm for the design of multiplierless FIR filters with power-of--two coefficients.* IEEE Transactions on Circuits and Systems, Vol. 36, pp. 1044–1047, July 1989
10. H. Garner: *Number Systems and Arithmetic.* Advances in Computing, vol. 6, pp. 131–194, 1965
11. R. I. Hartley: *Subexpression Sharing in Filters Using Canonic Signed Digit Multipliers.* IEEE Transactions on Circuits and Systems II — Analog and Digital Signal Processing, vol. 43, no. 10, Oct. 1996
12. A. R. Omondi: *Computer Arithmetic Systems. Algorithms Architecture and Implementations.* Prentice Hall 1994
13. Xilinx Co.: *The Programmable Logic Data Book 1999*
14. Altera Co.: *Apex 20K Programmable Logic Device Family, Data Sheet.* ver. 2.05, Nov. 1999
15. Xilinx Co.: *Core Generator.* Foundation 2.1i Software Packet, 1999

K. WIATR, E. JAMRO

CONSTANT COEFFICIENT MULTIPLICATION IN FPGA STRUCTURES

Summary

This paper investigates different architectures implementing bit-parallel constant coefficient multiplication in FPGA structures. At first the multiplierless multiplication (MM) architectures employing Canonic Sign Digit (CSD) and sub-structure sharing methods are addressed, and a novel algorithm for the conversion from two's complement to CSD representation is presented. In the second part of this paper the Look up table based Multiplication (LM) is investigated. Correspondingly, the usage of different memory modules and finding the optimal combination of the memory and adders are considered. The LM architecture considers also reduction of the address width for each memory cell and the possibility of memory sub-structure sharing (the search for the same memory cells is implemented). Finally the implementation results for Xilinx XC4000 and Virtex families are presented. As a result, the MM generally surpasses the LM architecture, however the actual choice between these two architectures is coefficient and input parameters dependent.

Keywords: dedicated circuits, specialised processors, programmable logic devices, distributed arithmetic

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2001, 47, z. 2, ss. 255–271

Szerokopasmowy układ formowania wiązki antenowej dookólnego systemu namiaru źródeł promieniowania

BRONISŁAW STEC, ZDZISŁAW CHUDY, LESZEK KACHEL

Wydział Elektroniki, Wojskowa Akademia Techniczna
00-809 Warszawa, ul. Kaliskiego 2

Otrzymano 2001.01.29
Autoryzowano 2001.04.11

Systemy namiaru źródeł promieniowania elektromagnetycznego używane są do wykrywania i pomiaru parametrów sygnałów promieniowanych przez różne urządzenia. Wykorzystują one często monoimpulsowe sposoby pracy. Zbudowane są z dookólnej sieci antenowej oraz układu formowania wiązki antenowej. W artykule przedstawiono układ formowania złożony z sześciu sprzęgaczy kwadraturowych i dwóch prowadnic falowych. Ponadto zamieszczono omówienie metody i wynik analizy teoretycznej wraz z wynikami pomiarów modelu pracującego w paśmie 2÷4GHz.

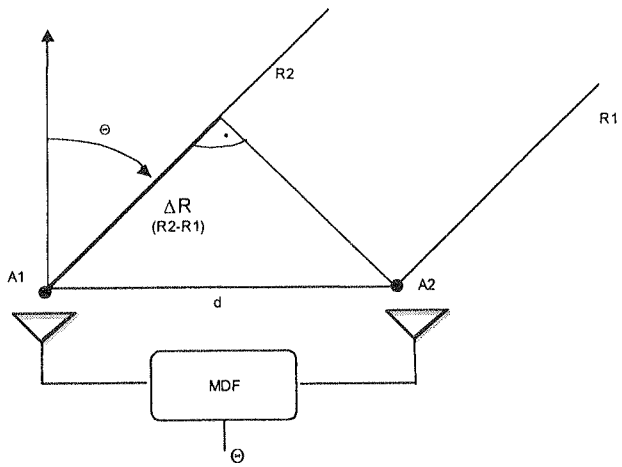
Słowa kluczowe: namiar, układ formowania, sprzęgacz kierunkowy, promieniowanie elektromagnetyczne

1. NAMIAR ŹRÓDEŁ PROMIENIUJĄCYCH

W rozpoznaniu elektronicznym, w zależności od jego zadań wykorzystywany jest pomiar wielu parametrów sygnałów. Najczęściej określana jest częstotliwość, parametry czasowe, a następnie rodzaj i lokalizacja źródła. Dla określenia parametrów sygnałów radioelektronicznych, trwających niekiedy dziesiątki lub setki nanosekund potrzebne są odbiorniki szerokopasmowe. Do grupy tej zalicza się odbiorniki detektorowe, natychmiastowego pomiaru częstotliwości oraz superheterodynowe nie przestrajane odbiorniki macierzowe. W obecnej chwili najczęściej wykorzystywanym jest odbiornik natychmiastowego pomiaru częstotliwości NPCz. Gwarantuje on pomiar parametrów w systemie monoimpulsowym.

Przeprowadzone analizy i testy potwierdzają, że wykonanie dokładnego i szybkiego namiaru źródła promieniowania elektromagnetycznego pracującego losowo lub krótko-

trwale, możliwe jest jedynie przy zastosowaniu metod monoimpulsowych. Namiar może być realizowany metodą fazową, amplitudową lub amplitudowo-fazową. Wyróżniane metody fazowe opierają się na wykorzystaniu różnicy faz φ sygnałów odbieranych przez minimum dwie anteny odpowiednio usytuowane (rys. 1). Występująca różnica dróg ($\Delta R = R_2 - R_1$) sygnałów odbieranych przez anteny powoduje uzależnienie określanej wartości kąta θ jako namiaru.



Rys. 1. Zasada funkcjonowania monoimpulsowo-fazowej metody namiaru

Różnica faz sygnałów φ wywołana różnicą dróg ΔR oraz namiar θ można ocenić korzystając z zależności:

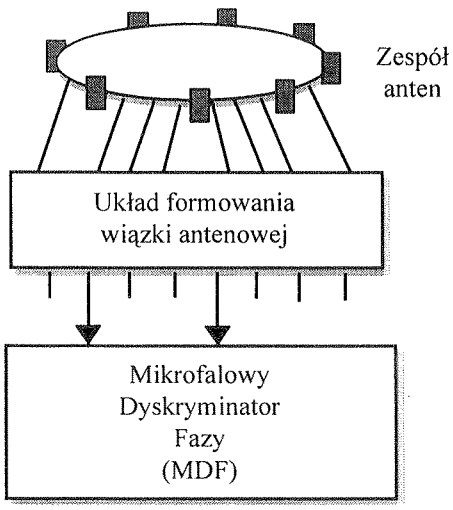
$$\varphi = 2\pi \frac{d}{\lambda} \cdot \sin\theta \qquad \theta = \arcsin\frac{\varphi\lambda}{2\pi d} \tag{1}$$

gdzie: λ — długość fali, d — baza
Dokładność namiaru wykrywanego promieniowania zależy od dokładności pomiaru różnicy faz analizowanych sygnałów:

$$\frac{d\varphi}{d\theta} = 2\pi \frac{d}{\lambda} \cdot \cos\theta \qquad \Delta\theta = \frac{\Delta\varphi}{2\pi \frac{d}{\lambda} \cdot \cos\theta} \tag{2}$$

gdzie: $\Delta\varphi$ — błąd określenia różnicy faz sygnałów.
Porównując metody monoimpulsowego pomiaru współrzędnych kątowych można stwierdzić, że dokładności metod fazowych są o rząd wyższe od dokładności uzyskiwanych metodami amplitudowymi. Stosując metodę fazowego interferometru można uzyskać dokładność rzędu 0.1°÷3°, a dalszy wzrost precyzji pomiaru może być osiągnięty poprzez stosowanie modyfikacji rozwiązań układowych.

Przykładem dookólnego systemu namiaru jest układ anten połączony z wielowrotnikiem mikrofalowym będącym układem formowania wiązki antenowej (rys.2) [1, 3, 6]. Fazy sygnałów na wyjściach układu formowania sygnałów są funkcjami różnicy dróg od wejścia do wyjścia. Jednocześnie są one zmodulowane w amplitudzie własnościami kierunkowymi anten. Analiza przesunięć fazowych sygnałów z odpowiednio dobranych wyjść pozwala na natychmiastowe określenie kierunku na źródło promieniowania znajdujących się w obserwowanym obszarze. Szybki pomiar różnicy faz umożliwia mikrofalowy dyskryminator fazy (MDF) [8].



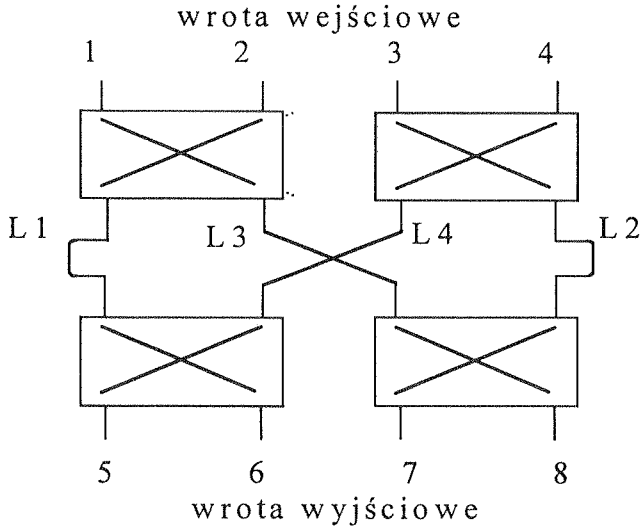
Rys.2. Schemat blokowy dookólnego systemu namiaru

Tak więc, dookólny system automatycznego wykrywania sygnałów radiolokacyjnych składa się z cylindrycznego szyku antenowego, układu formującego wiązkę antenową oraz układu MDF jako miernika fazy. Szyk antenowy zbudowany jest z 2n jednakowych anten rozłożonych regularnie na okręgu. Zaletą układu kołowego jest to, że może on być stosowany do przeszukiwania przestrzeni w zakresie 360^0 lub mniejszym bez konieczności mechanicznego obracania systemu antenowego [3, 4, 5].

2. UKŁAD FORMOWANIA WIĄZEK ANTENOWYCH

Układ formowania wiązek antenowych rozdziela sygnały wejściowe do poszczególnych wyjść, wprowadzając odpowiednie przesunięcia fazowe [3,5,6]. Przykładowe rozwiązanie sieci formowania przedstawiono na rys. 3. Moce sygnałów wejściowych są dzielone na wszystkie wyjścia równomiernie i posiadają stałe różnice faz między wyjściami. Analiza parametrów układu formującego pozwala znaleźć wrota wyjściowe, dla których różnice faz między sygnałami wyjściowymi odpowiadają kątowemu roz-

mieszczeniu anten odbiorczych dołączanych do jego wejść. Cechę tą określa odpowiedni dobór przesuwników fazy w połączeniu z wielkością sieci formowania. Oznacza to, że analizując własności fazowe sygnałów w określonych wrotach wyjściowych układu, otrzymamy informację odpowiadającą kierunkowi odbieranego sygnału.



Rys. 3. Struktura układu formowania wiązki antenowej systemu 4-antenowego

Prezentowany układ składa się z czterech sprzęgaczy 3dB/90° i czterech odcinków linii je łączących ($L_1 \div L_4$). Odcinki L_1 i L_2 oraz L_3 i L_4 są parami równe. Jednocześnie odcinki L_1 i L_2 są dłuższe o 1/8 długości fali od odcinków L_3 i L_4. Odpowiada to przesunięciu fazowemu 45° na środkowej częstotliwości pasma pracy układu. Parametry elementów składowych są tak dobrane, aby droga fazowa od wejścia do wyjścia charakteryzowała się założoną różnicą faz. Dla środkowej częstotliwości pasma pracy, zmiany faz φ między odpowiednimi wyjściami i wejściami oraz różnice faz $\Delta\varphi$ między wrotami wyjściowymi 7–5 i 7–6 dla układu z rys. 2 przedstawiono w tabeli 1.

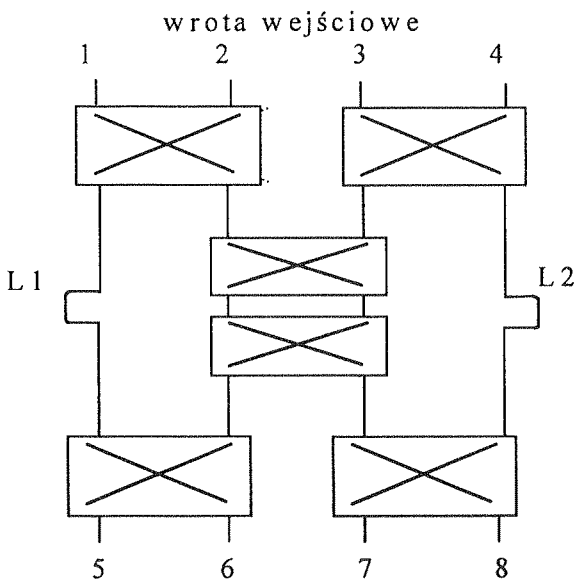
Określając różnice faz sygnałów na wyj. 7 i wyj. 5 lub na wyj. 7 i wyj. 6 zauważymy, że odpowiadają one rozłożeniu anten na kole w punktach o azymucie 45°, 135° , 225° i 315°. Różnica między parami wyjść 7 i 5 oraz 7 i 6 polega na innym kierunku obiegu koła przy założeniu tego samego połączenia anten z wejściami. Mierząc zatem różnicę faz sygnałów z odpowiednich wyjść możemy określić, z którego wejścia sygnał pochodzi. Jeżeli do wejść dołączone są anteny kierunkowe to uzyskujemy możliwość określenia namiaru źródła odbieranego promieniowania. Wyznaczając różnicę faz między wybraną parą wyjść, otrzymamy odwzorowanie przestrzennego usytuowania anten. Wymienione wartości różnic faz $\Delta\varphi$ przedstawione w tabeli 1 wynikają z parametrów i struktury układu formującego. Błąd namiaru wyrazi się różnicą wartości kąta obliczonego i rzeczywistego kąta przychodzenia sygnału Θ. W prawidłowo funkcjonują-

Tabela 1

Wartości zmian faz sygnałów

Nr wyj. / Nr wej.	5	6	7	8	$\varphi_7-\varphi_5$	$\varphi_7-\varphi_6$
1	45°	135°	90°	180°	+45°	−45° (+315°)
2	135°	225°	0°	90°	−135° (+225°)	−225° (+135°)
3	90°	0°	225°	135°	+135°	+225° (+135°)
4	180°	90°	135°	45°	−45° (+315°)	+45°

cym układzie wskazanie miernika fazy dołączonego do wyjść sieci określać będzie kątowe położenia anten, a tym samym wartości namiaru źródła. Przedstawiony układ formowania nie posiada jednak własności układu szerokopasmowego ze względu na występujące przesuwniki fazy zrealizowane na odcinkach linii długich. Na rys. 4 przedstawiono podobny układ formowania sygnałów, gdzie w roli elementu gwarantującego przesunięcie fazy występuje układ tandemowego połączenia dwóch sprzęgaczy z liniami odniesienia L_1 i L_2.



Rys.4. Układ formowania wiązki antenowej systemu cztero antenowego

Wartości faz sygnałów między poszczególnymi wrotami wyjściowymi, a wejściowymi oraz różnice faz $\Delta\varphi$ między wrotami wyjściowymi 7–5 i 7–6 dla środkowej częstotliwości pasma pracy przedstawiono w tabeli 2. Określając różnice faz sygnałów na wyj. 7 i wyj. 5 lub na wyj. 7 i wyj. 6 zauważymy, że odpowiadają one rozłożeniu anten podobnie jak w poprzednim przypadku na kole w punktach 45°, 135°, 225° i 315°. Wykorzystanie podukładu tandemowego połączenia dwóch sprzęgaczy kierunkowych w miejsce dwóch krzyżujących się linii transmisyjnych zmienia cechy układu.

Tabela 2

Zmiany faz sygnałów

Nr wyj. / Nr wej.	5	6	7	8	$\varphi_7-\varphi_5$	$\varphi_7-\varphi_6$
1	−135°	135°	−180°	90°	−45°	+45° (−315°)
2	135°	45°	−90°	−180°	135° (−225°)	225° (−135°)
3	−180°	−90°	45°	135°	−135°	+315° (−45°)
4	90°	−180°	135°	−135°	45° (−315°)	−45°

Upraszcza realizację techniczną, a jednocześnie umożliwia wykonanie w technologii NLP na wspólnym podłożu.

Rozważania analityczne przeprowadzono dla układu zawierającego bezstratne sprzęgacze zbliżeniowe 3dB o transmitancjach: do wrót bezpośrednich:

$$T_b = \frac{\sqrt{1-k^2}}{\sqrt{1-k^2}\cos\Theta + j\sin\Theta} = |T_b|e^{j\varphi_b} \tag{3}$$

do wrót sprzężonych:

$$C_S = \frac{jk\sin\Theta}{\sqrt{1-k^2}\cos\Theta + j\sin\Theta} = |C_S|e^{j\varphi_S} \tag{4}$$

gdzie: f_0 — częstotliwość środkowa odpowiadająca długości fali λ_0

Θ — elektryczna długość sprzęgacza, k — współczynnik sprzężenia sprzęgacza

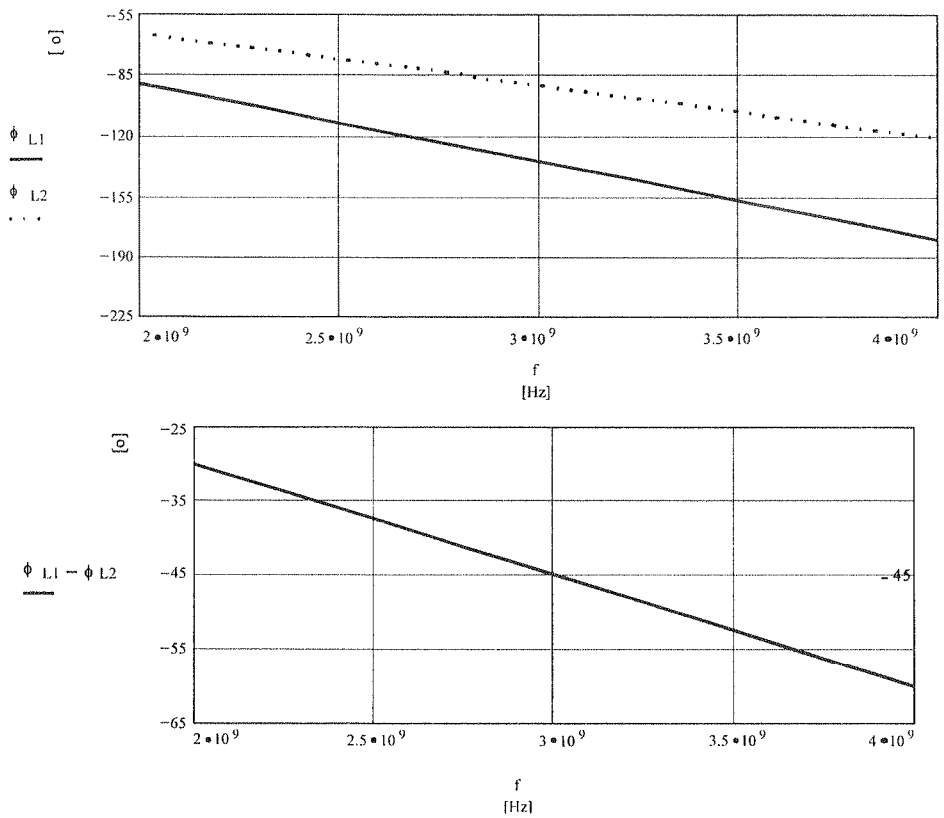
$$\Theta = \frac{\pi}{2}\cdot\frac{f}{f_0} \tag{5}$$

3. UKŁAD WZGLĘDNEJ ZMIANY FAZY

Zasadniczymi elementami sieci formującej oprócz sprzęgaczy kierunkowych są przesuwniki fazy (PF). W wersji podstawowej mogą być zrealizowane w oparciu o linie długie. Linia transmisyjna o określonej długości l powoduje opóźnienie fazowe ϕl:

$$\phi_l = \frac{2\pi}{\lambda} l \tag{6}$$

gdzie: λ — długość fali elektromagnetycznej,
Różnica faz wynikająca z różnicy długości pomiędzy dwiema liniami L_1, L_2 w funkcji częstotliwości jest proporcjonalna do różnicy ich długości. Charakterystyki fazowe oraz wnoszoną względną różnicę faz dwóch różnych linii przedstawiono na rys. 5.

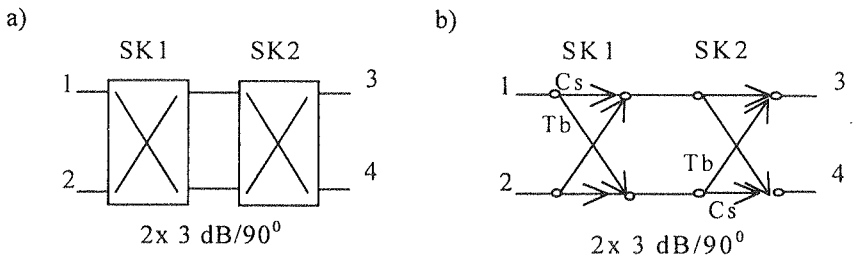


Rys.5. Własności fazowe linii transmisyjnych w funkcji częstotliwości

Przedstawiony układ na rys. 3 wykorzystuje linie długie jako przesuwniki fazy. Niestety łatwo zauważyć, że posiada on cechy układu „wąskopasmowego". Zmiany wartości

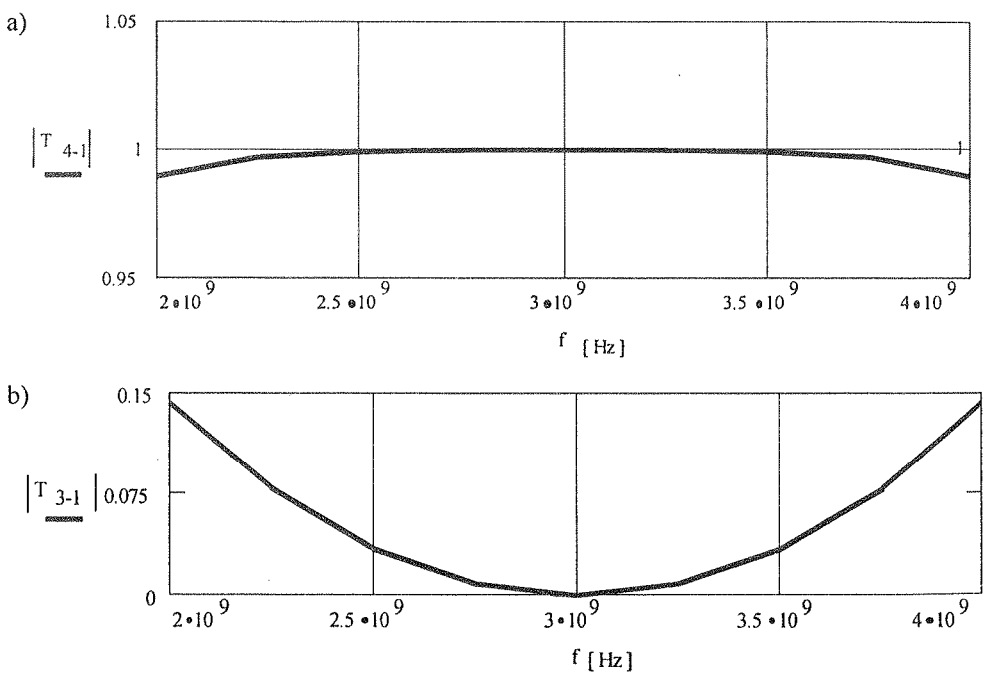
wnoszonej różnicy faz w funkcji częstotliwości (rys. 5) będą przyczyną błędów w układzie pracującym szerokopasmowo.

Przeprowadzona analiza teoretyczna oraz uzyskane wyniki doświadczeń przyczyniły się do zrealizowania modelu przesuwnika fazy z wykorzystaniem tandemowego połączenia dwóch sprzęgaczy kierunkowych. Strukturę przesuwnika przedstawiono na rys. 6 [2].



Rys.6. Tandemowy przesuwnik fazy: a) struktura układu; b) graf przepływu

Traktując powyższy czterowrotnik jako zastępcze „skrzyżowanie linii" należy oczekiwać, że jego własności w postaci wartości modułów parametrów S określających własności transmisyjne wynosić powinny odpowiednio: $|S_{31}| = |S_{42}| = 0$, $|S_{41}| = |S_{32}| = 1$. Oznacza to, że sygnał doprowadzony do wrót 1 propaguje się wyłącz-

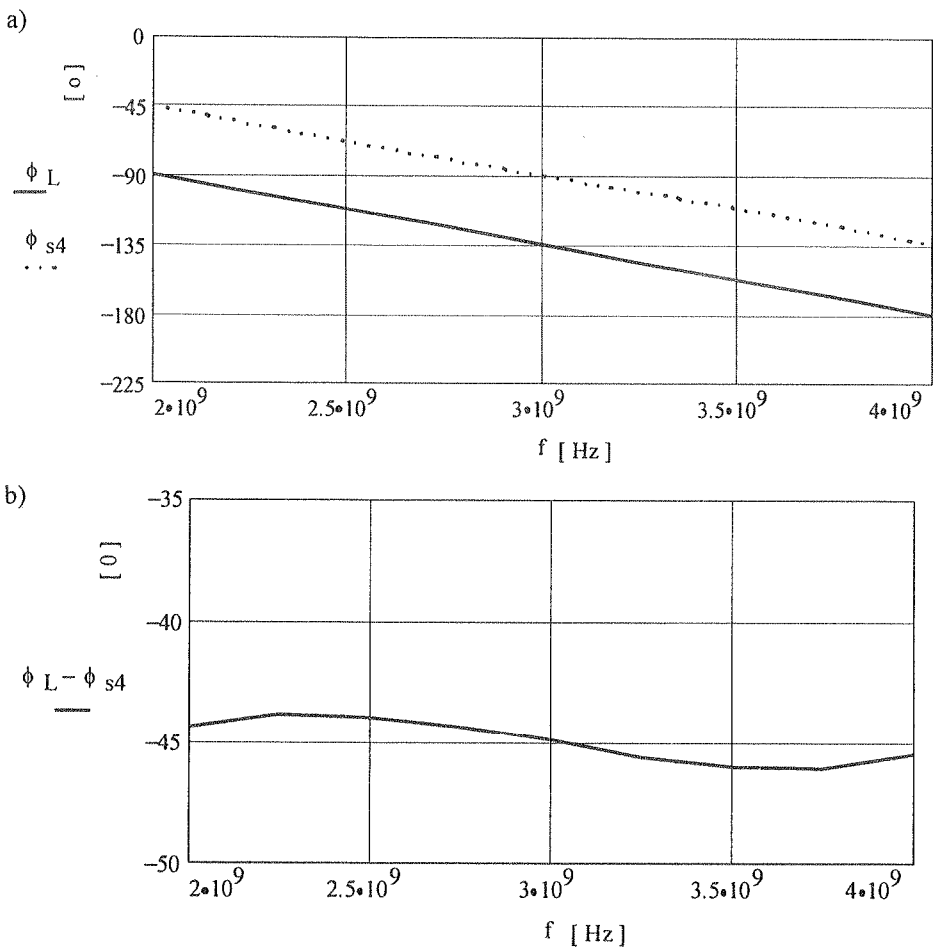


Rys. 7. Charakterystyki amplitudowo-częstotliwościowe tandemowego układu „skrzyżowania":
a) zmiany amplitudy sygnału we wrotach wyjściowych 4
b) zmiany amplitudy sygnału we wrotach wyjściowych 3

nie do wrót 4, a w przypadku wrót 2 — do wrót 3. Uproszczony graf przepływu dla tandemowego połączenia dwóch sprzęgaczy kierunkowych pełniących rolę „skrzyżowania" przedstawiono na rys. 6b. Transmitancje $T_{sk(m-n)}$ określające własności tandemu dla częstotliwości środkowej pasma pracy posiadają postać:

$$T_{sk3-1} = (C_s \cdot C_s) + (T_b \cdot T_b) = \frac{1}{2} e^{-j0} + \frac{1}{2} e^{-j\pi} = \frac{1}{2}(e^{-j0} + e^{-j\pi})$$

$$T_{sk4-1} = (C_s \cdot T_b) + (T_b \cdot C_s) = \frac{1}{2} e^{-j\frac{\pi}{2}} + \frac{1}{2} e^{-j\frac{\pi}{2}} = \frac{1}{2}(e^{-j\frac{\pi}{2}} + e^{-j\frac{\pi}{2}}) \quad (7)$$



Rys. 8. Charakterystyki fazowo-częstotliwościowe tandemowego układu „skrzyżowania":
a) faza sygnału we wrotach wyjściowych 4 — $\phi_{s4}(f)$ i linii odniesienia S_L — $\phi_L(f)$
b) różnica faz sygnału z wrót wyjściowych 4 i linii odniesienia $(\phi_L(f) - \phi_{s4}(f))$

Tak więc tandemowe „skrzyżowanie" charakteryzuje się bezstratnym przesyłaniem sygnału i wnoszeniem ściśle określonego przesunięcia fazy. Wyznaczone na podstawie obliczeń własności amplitudowo-częstotliwościowe tandemu przy założeniu obecności sygnału wejściowego we wrotach 1, przedstawiono na rys. 7. Ponieważ rozważany układ ma pełnić rolę przesuwnika fazy w układzie formowania, należy uwzględnić jego charakterystyki fazowe w stosunku do linii odniesienia. W roli odniesienia występuje linia przesyłowa, dobrana tak by uzyskać założone parametry układu formującego (tabela 2). Wyznaczone na podstawie obliczeń własności fazowe w stosunku do linii odniesienia S_L, przy założeniu obecności sygnału wejściowego we wrotach 1, przedstawiono na rys. 8.

Uzyskane charakterystyki fazowo-częstotliwościowe układu skrzyżowania pozwalają stwierdzić, że układ tandemowego połączenia sprzęgaczy z uwzględnieniem odniesienia do linii długiej zapewnia układowi formowania własności szerokopasmowe. Odchylenia fazy rzędu ±1.5 stopnia w paśmie 2÷4GHz w porównaniu z innymi źródłami błędów nie stanowią głównego źródła błędów układu namiaru.

4. CZĘSTOTLIWOŚCIOWA ANALIZA WŁASNOŚCI SZEROKOPASMOWEGO UKŁADU FORMUJĄCEGO WIĄZKI ANTENOWE

Przeprowadzenie pełnej analizy częstotliwościowej układu formowania wymaga wykonania dość złożonych obliczeń. Celem wykazania ogólnych własności układu, poniższe rozważania oparto na wykorzystaniu uproszczonego grafu przepływu układu formowania, którego strukturę przedstawiono na rys. 4. Uwzględniając zastosowanie podukładu tandemu w układzie formowania uzyskany graf przepływu sygnałów przedstawiono na rys. 9. Analiza cech układu przy założeniu bezstratności elementów, pozwala otrzymać w i-tych wrotach wyjściowych (5÷8) układu formowania sygnały opisane transmitancjami $Y_5(f)\div Y_8(f)$ (9). Celem ułatwień w opisie, wprowadzono pomocniczo opis własności transmisyjnych tandemu zgodnie z jego grafem (rys.6a) i grafem ogólnym układu formowania wiązki antenowej (rys.9).

$$T_{1s}(f) = X_3 \cdot C_s(f) + X_4 \cdot T_b(f)$$

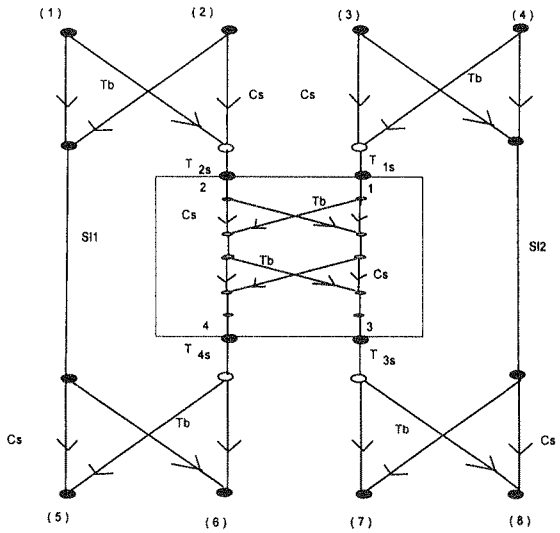
$$T_{2s}(f) = X_1 \cdot T_b(f) + X_2 \cdot C_s(f) \tag{8}$$

$$T_{3s}(f) = [T_{2s}(f) \cdot T_b(f) + T_{1s}(f) \cdot C_s(f)] \cdot C_s(f) + [T_{2s}(f) \cdot C_s(f) + T_{1s}(f) \cdot T_b(f)] \cdot T_b(f)$$

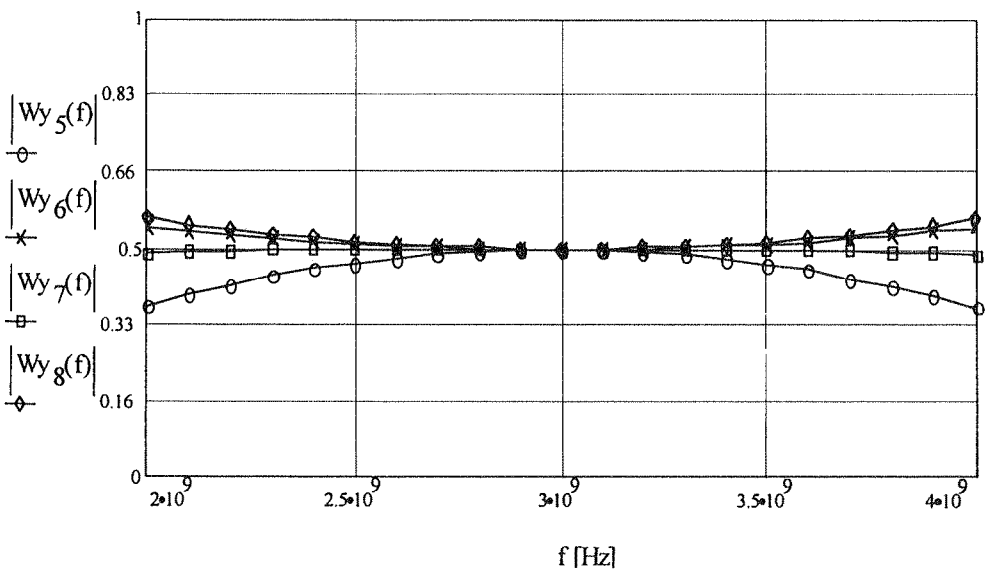
$$T_{4s}(f) = [T_{2s}(f) \cdot C_s(f) + T_{1s}(f) \cdot T_b(f)] \cdot C_s(f) + [T_{2s}(f) \cdot T_b(f) + T_{1s}(f) \cdot C_s(f)] \cdot T_b(f)$$

Transmitancje $T_{1s}(f)\div T_{4s}(f)$ określają transmisję sygnałów wejściowych X1÷X4 w obwodzie „skrzyżowania". Wyjściowe uogólnione transmitancje zespolone od wejścia do wyjścia układu opisane są zależnościami:

$$
\begin{aligned}
Y_5(f) &= [X_1 \cdot C_s(f) + X_2 \cdot T_b(f)] \cdot S_{L1}(f) \cdot C_s(f) + T_{4s}(f) \cdot T_b(f) \\
Y_6(f) &= [X_1 \cdot C_s(f) + X_2 \cdot T_b(f)] \cdot S_{L1}(f) \cdot T_b(f) + T_{4s}(f) \cdot C_s(f) \\
Y_7(f) &= T_{3s}(f) \cdot C_s(f) + [X_3 \cdot T_b(f) + X_4 \cdot C_s(f)] \cdot S_{L2}(f) \cdot T_b(f) \\
Y_8(f) &= T_{3s}(f) \cdot T_b(f) + [X_3 \cdot T_b(f) + X_4 \cdot C_s(f)] \cdot S_{L2}(f) \cdot C_s(f)
\end{aligned}
\tag{9}
$$



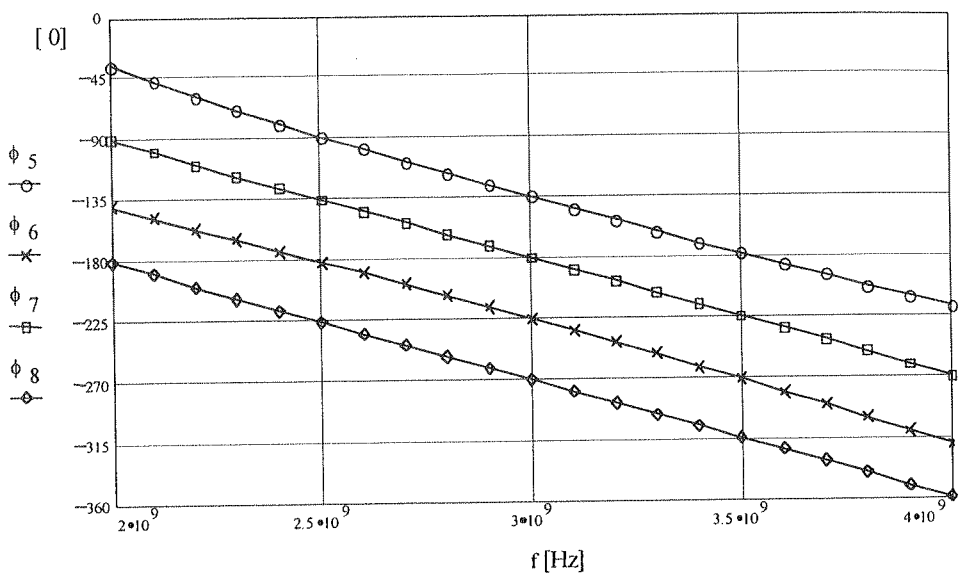
Rys. 9. Uproszczony graf przepływu układu formowania wiązki antenowej



Rys.10. Charakterystyki amplitudowe układu formowania przy pobudzeniu wrót wejścia 1

Symbole X_1, X_2, X_3, X_4 oznaczają sygnały podawane na odpowiednie wejścia 1, 2, 3 i 4. Symbole $Y_5(f)$, $Y_6(f)$, $Y_7(f)$ i $Y_8(f)$ oznaczają sygnały wychodzące z odpowiednich wyjść 5, 6, 7 i 8. Transmitancje odcinków linii łączących sprzęgacze oznaczono przez $S_{L1}(f)$ i $S_{L2}(f)$. Tak przedstawiony algorytm programu pozwala na analizę dowolnego układu mikrofalowego z czterema wejściami i czterema wyjściami. Dla innych układów zmianie ulegną jedynie parametry wewnętrzne sieci mikrofalowej, które muszą być zdefiniowane oddzielnie dla każdego z nich. W perspektywie algorytm ten umożliwi analizę systemu namiaru z dołączonymi na wejścia antenami. Szczególnie będzie przydatny gdy sygnał pojawi się na dwóch wejściach, co jest normalne gdy sygnał przyjdzie z kierunków zbliżonych do styku charakterystyk kierunkowych sąsiednich anten. Przebieg zmian amplitud na poszczególnych wyjściach przy doprowadzeniu sygnału na wejście 1 przedstawiono na rys. 10.

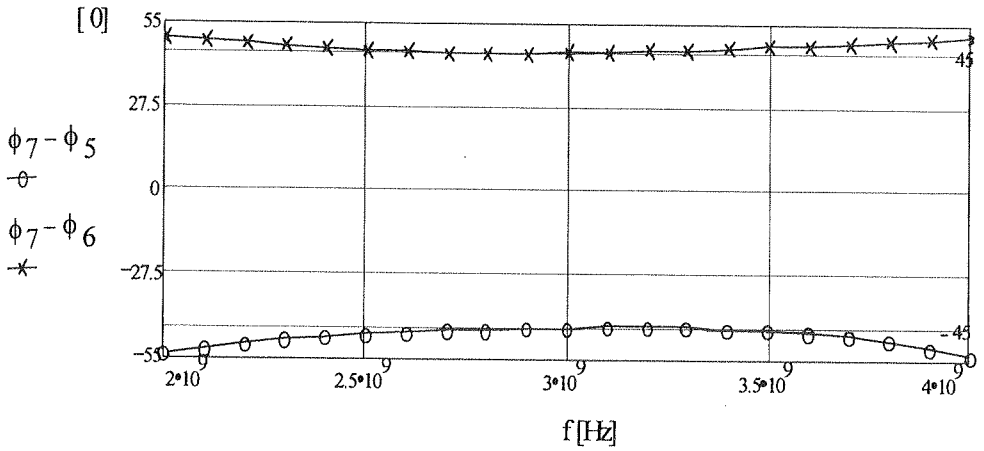
Istotniejszym parametrem układu są zmiany faz sygnałów ($\phi_5 \div \phi_8$) w poszczególnych wrotach wyjściowych, które przedstawiono na rys.11.



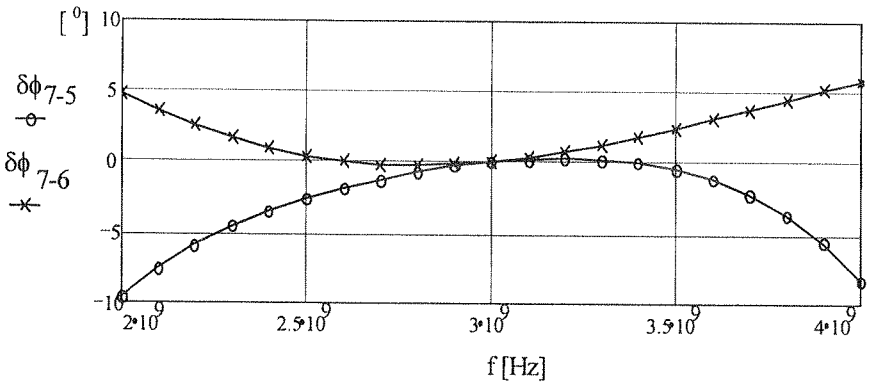
Rys. 11. Charakterystyki fazowe układu formowania przy pobudzeniu wrót wejścia 1

Idea pracy układu formowania wykorzystuje różnicę faz sygnałów w wskazanych wrotach wyjściowych. Wartości te odwzorowują kierunek odbioru. Dokładność oceny różnic faz decyduje o dokładności systemu. Na rys.12 przedstawiono przebieg zmian różnic faz sygnałów między wrotami 7 i 5 oraz 7 i 6.

Otrzymane wartości poprawnie wskazują na wartości +45° i –45°. Wartości odchyleń w funkcji częstotliwości od wymienionych wartości jako wielkość wnoszonego błędu $\delta\phi$ przedstawiono na rys.13.



Rys.12. Przebieg zmian różnicy faz na wyjściach 7–5 oraz 7–6

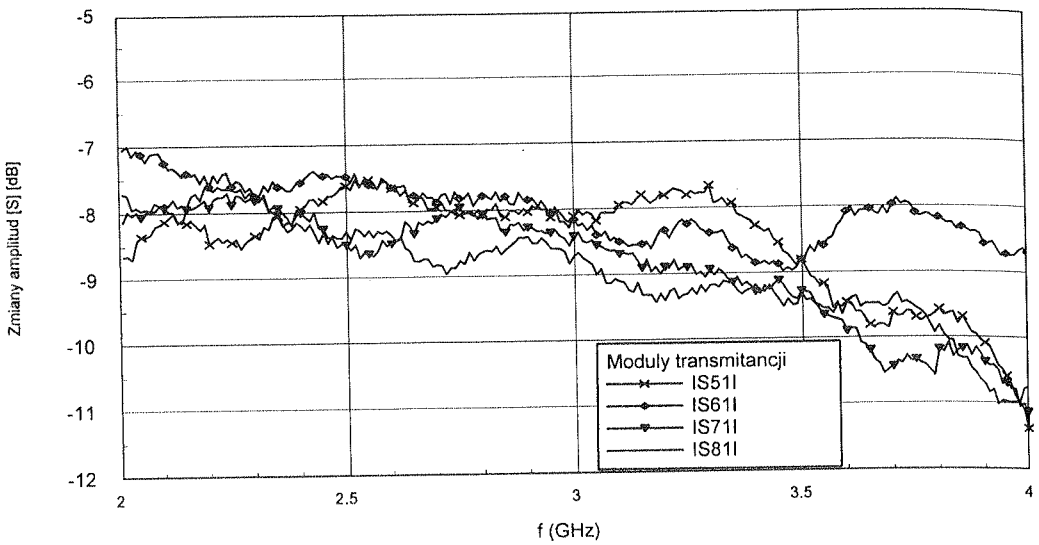


Rys.13. Wartości błędów fazowych układu formowania

Wartości odchyleń fazy rzędu $+5° \div -10°$ zależnie od kombinacji wyjść (ocenianego kierunku) na krańcach pasma pracy są wartościami wystarczającymi dla wielu zastosowań. Przykładowo urządzenia ostrzegania i wstępnego wykrywania nie wymagają znacznych dokładności pomiaru kąta. Uwzględniając prostą budowę systemu i wykorzystując uzyskane wyniki analizy, należy stwierdzić, że układ spełnia zadania układu formowania wiązki i umożliwia realizację wersji układu w technologii NLP.

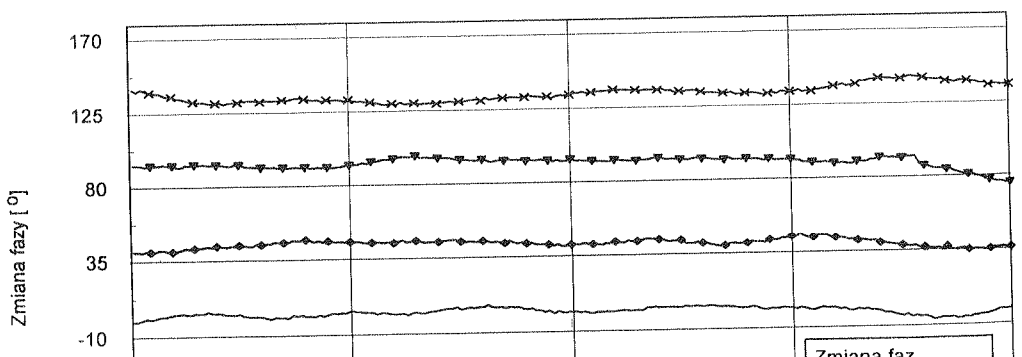
5. WYNIK BADAŃ DOŚWIADCZALNYCH

Dla potrzeb weryfikacji zbudowano układ wg rys. 4. W części środkowej zastosowano układ tandemu. W modelu zastosowane zostały sprzęgacze wykonane z NLP jako sprzęgacze o strukturze palczastej, zwane sprzęgaczami Lang'a i odcinki NLP. Materiałem podłożowym był laminat szklano-epoksydowy o względnej przenikalności dielektrycznej $\varepsilon_r = 4.6$ i grubości $h = 1.5$mm. Przykładowe przebiegi zmian mocy sygnału od wejścia 1 do poszczególnych wyjść przedstawiono na rys. 14.



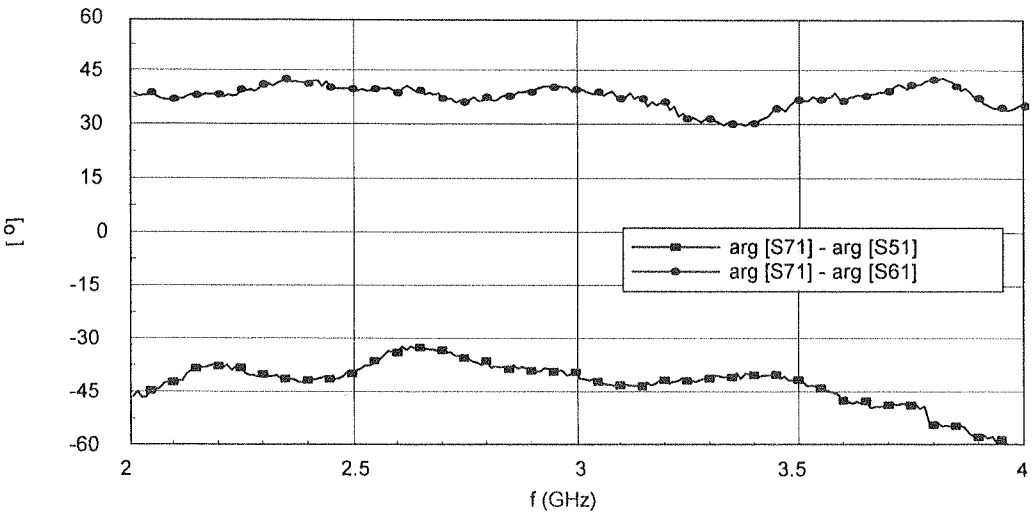
Rys. 14. Przebieg zmian amplitud sygnałów wyjściowych w układzie formowania wiązki antenowej ze „skrzyżowaniem"

Występujące odchylenia w przebiegu zmian amplitud sygnału w paśmie pracy są wynikiem istnienia odbić spowodowanych brakiem dopasowań. Spadek amplitud w górnym zakresie częstotliwości jest wynikiem wnoszonego tłumienia przez laminat wraz ze wzrostem częstotliwości. Z punktu widzenia własności układu formowania, ważnym jest zapewnienie stałej zmiany faz w szerokim paśmie pracy. Zmiany faz sygnałów do poszczególnych wrót wyjściowych przy obecności sygnału wejściowego we wrotach wejściowych 1 przedstawiono na rys.15.

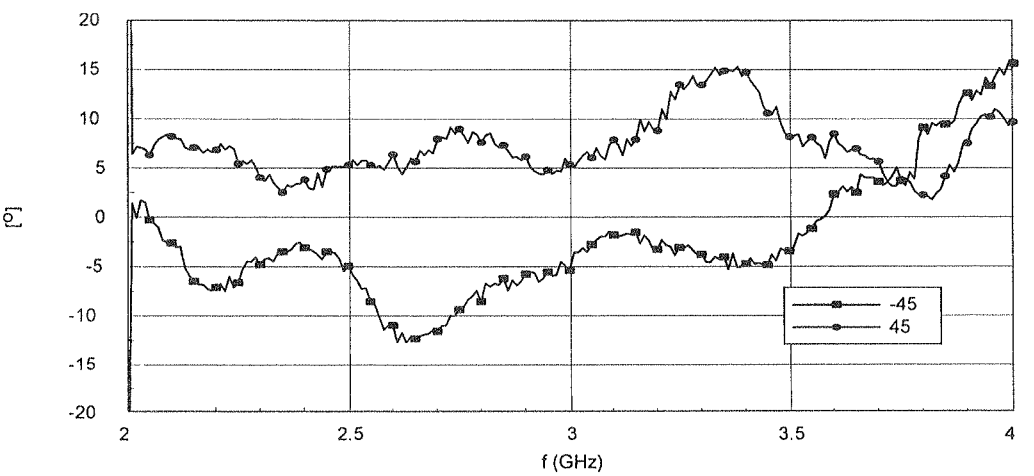


Rys. 15. Przebieg zmian faz sygnałów do wyjść układu formującego

W ocenie kierunku odbioru sygnału wykorzystuje się różnicę faz sygnałów w wrotach wyjściowych układu. Wybierając wrota wyjściowe nr 5, 6 i 7 jako wrota do oceny kierunku w badanym układzie, uzyskano przebiegi różnic faz w paśmie 2÷4GHz, które przedstawiono na rys. 16.



Rys. 16. Przebieg zmian różnic faz sygnałów we wrotach 7–5 oraz 7–6



Rys. 17. Dokładność oceny „kierunku"

Nierównomierność charakterystyk różnicy faz między wrotami wyjściowymi w przypadku jest konsekwencją istnienia odbić wewnętrznych poszczególnych podzespołów. Uzyskane wartości różnic faz pozwalają bezpośrednio przyporządkować wejścia

układu formowania wiązki dla oceny kierunku przychodzenia sygnału w zorientowanym systemie namiaru. Na podstawie uzyskanych rezultatów pomiarów określone zostały błędy wnoszonych zmian faz w stosunku do wartości oczekiwanych związanych z kierunkiem, jako wielkości bezwzględne odchyleń wnoszonych przez układ badany w stosunku do cech układu idealizowanego rozważanego teoretycznie. Dokładność oceny „kierunku" +45° i –45° w paśmie 2÷4GHz w postaci błędu bezwzględnego przedstawiono na rys. 17.

Przebieg wykreślonych charakterystyk wnoszonego błędu kątowego wskazuje, że układ formowania wiązki antenowej ze „skrzyżowaniem" zapewnia dużą dokładność ocen. Występuje równomierny charakter zmian błędu. Stałe odchylenia wartości względem zera w funkcji częstotliwości pozwalają wnioskować o obecności błędu systematycznego wywołanego odbiciami wewnątrz układu. Zaproponowany układ formowania wiązki antenowej dla 4-ro antenowego systemu namiaru kierunku ma cechy układu szerokopasmowego. Szerokość pasma pracy zależy od własności częstotliwościowych sprzęgaczy 3dB/90°. Stosując sprzęgacze wielooktawowe, można znacznie poszerzyć szerokość pasma pracy układu formowania wiązki antenowej.

6. PODSUMOWANIE

W wyniku przeprowadzonej analizy należy stwierdzić, że istnieje możliwość budowy zintegrowanego układu formowania wiązki antenowej w technologii NLP, jako układu o dużym stopniu scalenia. Wyniki pomiarów doświadczalnych wskazują na prawidłowość postawionych na wstępie założeń o szerokopasmowości układu. Przedstawiona konstrukcja pozwala na budowę całego układu na jednej płaszczyźnie, co znacznie upraszcza technologię wykonania i zmniejsza koszty budowy urządzenia. Zastosowanie anten szerokopasmowych typu szerokopasmowej anteny tubowej oraz sprzęgaczy wielooktawowych umożliwi budowę wielooktawowego prostego monoimpulsowego systemu namiaru źródła promieniowania elektromagnetycznego. Przy rozbudowanej strukturze przekazywania sygnałów, układ może być wykorzystany jako urządzenie ostrzegania lub wstępnego wykrywania sygnałów promieniowania elektromagnetycznego.

BIBLIOGRAFIA

1. B. Stec: *Wielowrotowy mikrofalowy układ transformacji amplitudowo fazowej.* Nowe Techniki Antenowe PIT, 01–03.06.93 Rozciszów.
2. B. Stec, A. Rutkowski: *A planar microwave phase discriminator.* 27 European Microwave Conference Jerusalem 1997, Volume I pp. 51–55.
3. B. Stec: *Radioelectronic Reconnaisance Monopulse Station.* International Defence Conference, Abu Dhabi,19–23 March 1995 r.
4. A. Sheleg: *A matrix-fed circular array for continuous scanning.* IEEE, vol. 56, Nov. 1968 r.
5. Z. Chudy, B. Stec, L. Kachel: *Układy wykrywania i namiaru źródeł sygnałów zakresu mikrofalowego.* III Konferencja Naukowa pt. „Systemy rozpoznania i walki radioelektronicznej", Żegiestów 1997r.

6. B. Stec: *Monoimpulsowa stacja rozpoznania radioelektronicznego*. Materiały I Konferencji Naukowo-Technicznej Systemy Rozpoznania i Walki Radioelektronicznej, Żegiestów 1995 r.
7. B. Stec, A. Rutkowski: *Zintegrowany mikrofalowy dyskryminator częstotliwości*. Materiały III Konferencji Naukowo-Technicznej pt. „Systemy rozpoznania i walki radioelektronicznej" Żegiestów 1997 r.
8. B. Stec: *Szerokopasmowy układ homodynowy do pomiaru obwodów mikrofalowych*. MECS Gdańsk 1980 r.

B. STEC, Z. CHUDY, L. KACHEL

WIDEBAND ANNTENAS BEAM FORMING CURCUIT IN CIRCLE DIRECTION FINDING SYSTEM

Summary

Receivers are main compenents of communication, detecting and warning systems. They are used to detect and measure parameters of signals radiated by different power sources. They contain antenna arrays and beam forming networks. Analysis of phase shifts between signals at corresponding outputs permits to estimate the direction of source of radiation. So constructed circuit makes it possible to measure any direction of incoming signals without movement of antennas and observes all 360° or part of space in azimuth. The antenna array can be composed of 4, 8, 16, 32 etc. antennas. Four antennas receiving system is the simplest arrangement, in which signal forming circuit can be used. The paper introduces a new two design of signal forming circuit composed of quadrate couplers and phase shifters. The method and results of theoretical analysis as well as results of measurements in the range of 2÷4GHz are also presented.

Keywords: direction finding, forming circuit, directional coupler, phase shifter, electromagnetic radiation

T. Ostrow
R. Golańs
różni
J. Falkiew
usuw
A. Jakubi
skole
A. Jakubi
biern
K. Wiatr,
gram
B. Stec, Z
syste
Informacje

T. Ostrow
R. Golańs
J. Flakiwi
A. Jakubia
A. Jakubia
K. Wiatr,
B. Stec, Z
syste
Informatic

SPIS TREŚCI

CONTENTS

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

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 powinno 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. PODZIĘKOWANIE
  8. BIBLIOGRAFIA
- Układ streszczenia w dodatkowej wersji językowej powinien być następujący;

  TYTUŁ (w języku angielskim – o ile artykuł pisany jest w języku polskim i na odwrót),
  AUTOR (inicjał imienia i nazwisko).

Obszerne jednostronne streszczenie (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.
- Wszystkie strony muszą mieć numerację ciągłą.

Sposób pisania tekstu.
Tekst powinien być pisany bez używania wyróżnień, a w szczególności nie dopuszcza się spacjowania, 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 adiustacyjnych np. podkreślenie linią przerywaną oznacza spacjowanie (rozstrzelenie), podkreślenie linią ciągłą – pogrubienie, podkreślenie wężykiem – 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 do wielkości czcionki maszyny do pisania (minimum 12 punktów). Przy podziale pracy na rozdziały i podrozdziały cyfrowe ich oznaczenia nie powinny być większe niż III 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. 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 w wytycznymi IEE (Intemational Electronical Commision) oraz ISO (Intemational Organization of Standarization).

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 satelitę. E.T.T. 1990, vol. 2, no 5, pp. 141–148
- nieperiodyczne — 2. K. Andersen: A resource allocation framework. XVI Intemational Symposium, Stockholm (Sweden), may 1991, paper A 2.4
- książki — 3. Y.P. Tvidis: Operation and modeling ofthe 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 ołowiem 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. Rysunki powinny być cytowane w tekście.

Uwagi dodatkowe.
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.

Autora obowiązuje 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 stuff will accept for publishing only original monographic and survey papers concerning widely understood electronics. Because of the fact that KWATRALNIK 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 stuff 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, address 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:
  - Title (in Polish, if article was prepared in English and vice versa)
  - Author (firs name initials and surname)
  - One page summary (in Polish, if article was prepared in English and vice versa).

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 numbered using no more than 3 levels (i.e. 4.1.1.).

Tables
Tables with their titles should be placed 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 place in brackets on the right side of the page. Units of measure, letter and graphic symbols should be printed according to requirements of IEE (International Electronical Commission) and ISO (International Organisation of Standardisation).

References

References should be placed at the and of the main text with 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 satelite. E.T.T. 1990, vol.2, no 5, pp. 141–148
2. K. Anderson: A recource allocation framework. XVI International Symposium, Stockholm (Sweden), May 1991. paper A 2.4
3. Y.P. Tvidis: Operation and modelling 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 grater than 10×15 cm. On the margin of each drawing and on the back side of each photo author name and abbreviation of the title of article should be placed. Figure's captures should be placed 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 one's expense.

Author is obliged to perform the author's correction, which should be accomplished within 3 days starting from the date of receiving of the text from the editorial stuff. Corrected text should be return to the editorial stuff personally or by mail. Correction marks should be placed on the margin of copies received from the editorial stuff 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 stuff of the publisher.

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

Subscritpion for external subscribers:

The promotional subscription price in 2001 is $ 90 including postage for institutions. At subscriber's request this journal will be air maild at additional postage to 50% dross proce to European countries nad 65% overseas.

– Foreign Trade Enterprise ARS Polona, Krakowskie Przedmieście 7, 00-068 Warszawa, Poland,

– RUCH S.A. tel. (4822) 53-28-819, 53-28-823, fax 53-28-731.

Pragnę Państwa zawiadomić, że można już zamawiać prenumeratę czasopism naukowych w Wydawnictwie Naukowym PWN. Z pewnością usprawni to obsługę prenumeratorów i pozwoli zebrać informacje o czytelnikach.

Zamówienia można składać telefonicznie pod numerami (022) 695-42-74, (022) 695-40-47, przysłać faksem pod nr (022) 695-42-70, lub listownie (wzór zamówienia jest dostępny w Dziale Publikacji Zleconych Wydawnictwa Naukowego S.A.).