POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

Indeks 363189
ISSN 0867-6747

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 42 — ZESZYT 3

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1996

Kwartalnik jest wydawany z funduszu Komitetu Badań Naukowych i dofinansowany przez instytucje:

— Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych
— Politechnika Gdańska, Wydział Elektryczny
— Politechnika Szczecińska, Instytut Elektroniki i Informatyki
— Instytut Łączności w Warszawie
— Spółka „Telekomunikacja Polska S.A."

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 42 — ZESZYT 3

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1996

RADA REDAKCYJNA

Przewodniczący
† prof. dr inż. ADAM SMOLIŃSKI
członek rzeczywisty PAN

Członkowie
prof. dr hab. inż. DANIEL JÓZEF BEM — czł. koresp. PAN, prof. dr hab. inż. MICHAŁ BIAŁKO — czł. koresp. PAN, prof. dr hab. inż. STEFAN HAHN — czł. koresp. PAN, prof. dr inż. ANDRZEJ HAŁAS, prof. dr inż. ZDZISŁAW KACHLICKI, prof. dr hab. inż. BOHDAN MROZIEWICZ, prof. dr inż. JERZY OSIOWSKI, prof. dr inż. WITOLD ROSIŃSKI — czł. rzecz. PAN, prof. dr inż. STANISŁAW SŁAWIŃSKI, prof. dr hab. inż. STEFAN WĘGRZYN — czł. rzecz. PAN, prof. dr hab. inż. WIESŁAW WOLIŃSKI — czł. koresp. PAN, prof. dr inż. ANDRZEJ ZIELIŃSKI, prof. dr inż. MARIAN ZIENTALSKI

REDAKCJA
Redaktor Naczelny
prof. dr hab. inż. WIESŁAW WOLIŃSKI

Zastępca Redaktora Naczelnego
doc. dr inż. KRYSTYN PLEWKO

Sekretarz Odpowiedzialny
mgr KRYSTYNA LELAKOWSKA

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 (0-22) 660 77 37; 25 29 18+aut. sekr.

Telefony domowe: Redaktora Naczelnego: 12 17 65
Zast. Red. Naczelnego: 26 83 41
Sekretarza Odpowiedzialnego: 25 29 18

WYDAWNICTWO NAUKOWE PWN
Warszawa, ul. Miodowa 10

Ark. wyd. 7,00 Ark. druk. 5,75	Podpisano do druku w grudniu 1996 r.
Papier offsetowy kl. III 70 g. B-1	Druk ukończono w grudniu 1996 r.

Skład: Agnieszka Chmielewska Warszawa ul. Husarii 12
Druk i oprawa: Drukarnia Braci Grodzickich, Żabieniec, ul. Przelotowa 7

Szanowni Autorzy

„Kwartalnik Elektroniki i Telekomunikcaji — Electronics and Telecommunication Quarterly" jest kontynuatorem tradycji powstałego 41 lat temu kwartalnika pt. „Rozprawy Elektrotechniczne". „Kwartalnik Elektroniki i Telekomunikacji — czasopismo Komitetu Elektroniki i Telekomunikacji Polskiej Akademii Nauk — wydawany jest przez Wydawnictwo Naukowe PWN.

Kwartalnik jest czasopismem naukowym, na którego łamach są publikowane artykuły i komunikaty prezentujące wyniki oryginalnych oraz przeglądowych prac teoretycznych i technicznych związanych z szeroko rozumianymi dziedzinami współczesnej elektroniki i telekomunikacji.

Wszystkie publikowane w Kwartalniku artykuły są recenzowane przez wybitnych krajowych specjalistów, co zapewnia, że publikacje te są uznawane jako autorski dorobek naukowy. Opublikowanie w Kwartalniku wyniku prac naukowych, badawczych i technicznych zrealizowanych w ramach „GRANTów" Komitetu Badań Naukowych, spełnia więc jeden z wymogów stawianych tym pracom.

Czasopismo ukazuje się w ponad półtysięcznym nakładzie i dociera do wszystkich zajmujących się elektroniką i telekomunikacją krajowych ośrodków naukowych oraz technicznyh i do wielu instytucji zagranicznych. Jest ponadto prenumerowane przez liczne grono specjalistów.

Każdy z Autorów otrzymuje bezpłatnie 20 egzemplarzy nadbitek swojego artykułu, co ułatwia przesłanie tej publikacji do indywidualnie wybranych przez Autora osób i instytucji w kraju lub za granicą. Ułatwia to dodatkowo fakt, że w Kwartalniku mogą być publikowane artykuły również 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

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1996. 42, z. 3, ss. 273–285

Non-uniform and non-stationary mutation in numerical optimization using genetic algorithms

SŁAWOMIR KOZIEŁ

Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska

Received 1996.05.29

Authorized 1996.08.19

The paper gives a general approach to mutation operator in classical genetic algorithms with binary representation of chromosomes. Some variations of mutation operator are presented, among others non-uniform and non-stationary one. The behaviour of the algorithm working with these operators in some simple test cases of one- and multimodal functions is shown. The problem of a convergence to global optimum and an importance of the mutation operator in the performance of the algorithm is discussed.

Key words: genetic algorithms, mutation operator.

1. INTRODUCTION

The main goal of this work is to show in which way the modified mutation operator can improve the performance of the genetic algorithm in problems of numerical optimization. Second one is to demonstrate that mutation operator may be as important as other recombination operators in a performance of genetic algorithms. It is well known, that in early works [6] mutation was treated as a background (secondary) operator, which guarantees only that every point in a search space could be reached.

Now, especially because of developing genetic algorithms working with floating point representation of chromosomes ([1], [9]), understanding of importance of mutation grows [12]. On the other hand, classical variations of genetic algorithms with binary representation of chromosomes have some drawbacks ([2], [9]), e.g. poor performance in multidimensional, high-precision numerical problems. In this work we are interested in improving of fine local tuning of solution using modified variations of mutation operator. Of course, it is not a new idea [10], but we propose more general approach to this problem. In next parts of this paper we will briefly discuss, in turn: non-uniform mutation operator, non-stationary one, dependent on fitness one, and some modifications and combinations of them.

2. NON-UNIFORM MUTATION

Let us consider global parameter of probability of mutation p_m. In the case of uniform mutation, the probability of mutation is the same for each bit of the chromosome. This makes impossible to perform an efficient fine local tuning. The reason is, that, in grown up population, if one of the most significant bits of chromosome undergoes mutation then the value of parameters leave the area of the search space, which corresponds to the best solution (in sequel, we assume that chromosome consists of subsequences of bits, which represent parameters of optimized function). Although, the fine local tuning has not been the goal of mutation operator, we can improve this operator to make possible to achieve it. It is evident, that for fixed global probability of mutation p_m and fixed length of chromosome, expected number of muted bits is stable. The main assumption of our non-uniform mutation operator is invariance of the expected number of mutated bits in every subsequence of chromosome (and in all chromosome, of course), despite of the fact, that probabilities of mutation of some bits can differ. This assumption allows to control the rate of non-uniformity of mutation and, independently, the global rate of mutation. Let $M=l\cdot p_m$ be the expected number of muted bits in chromosome, where l is a length of chromosome. Furthermore, let us consider, that we optimize the function of n variables, and each of them is represented by subsequence of l_j bits. Each bit of chromosome has its own probability of mutation p_{ji}, where $j=1,\ldots,n$ is an index of parameter and $i=l,\ldots,l_h-1$ is an index of a bit in the representation of parameter j ($i=0$ corresponds to less significant bit). The next assumption is, that probabilities of mutation of bits in each subsequence are in accordance with geometric series, as below:

$$p_{ji}=p_j^0\cdot(q)^i \quad \text{for} \quad i=1,\ldots,l_j, \tag{1}$$

where q is an effective coefficient of non-uniformity ($q\leqslant 1$), given by

$$q=W_0-(W_0-W)\cdot\left(\frac{t}{T}\right)^r \tag{2}$$

and: W_0, W — initial and final coefficient of non-uniformity,
t, T — current and maximal number of generation,
r — auxiliary parameter ($3\leqslant p\leqslant 5$);

We have made use of geometric series, because it corresponds to binary representation of chromosome the ratio of probabilities of mutation of two bits in subsequence is equal to q^i and is proportional to the ratio of "*significance*" of two bits inchromosome, where i is the distance of these bits in subsequence. Statement (2) is very useful to control the value of q coefficient as a function of age of population. This power function has an interesting property — it maintains q close to its initial value and only for t near by maximal number of generation T changes it quickly to its final value.

As we mentioned earlier, expected number of muted bits (M) must be preserved, thus probabilities p_j^0 must satisfy the following equation

$$M = l \cdot p_m = \sum_{j=1}^{n} \sum_{i=0}^{l_j-1} p_{ji} = \sum_{j=1}^{n} \sum_{i=0}^{l_j-1} (p_j^0 \cdot (q)^i) \tag{3}$$

however $l = \sum_{j=1}^{n} l_j$, thus

$$\sum_{j=1}^{n} l_j \cdot p_m = \sum_{j=1}^{n} \left(p_j^0 \cdot \sum_{i=0}^{l_j-1} q^i \right) = \sum_{j=1}^{n} \left(p_j^0 \cdot \frac{1-q^{l_j}}{1-q} \right. \tag{4}$$

$$\text{if } q \neq 1 \text{ and } \sum_{j=1}^{n} l_j \cdot p_m = \sum_{j=1}^{n} l_j \cdot p_j^0 \text{ otherwise,} \tag{5}$$

Equations (4) and (5) indicate, that

$$p_j^0 = l_j \cdot p_m \cdot \frac{1-q}{1-q^{l_j}} \text{ if } q \neq 1, \text{ and } p_j^0 = p_m \text{ if } q = 1 \tag{6}$$

for $j = 1, \ldots, n$.

Parameters W_0 and W control non-uniformity of mutation operator and it is clear, that for $W_0 = W = 1$ it changes to uniform mutation (we can say, that uniform mutation is a particular case of this operator). For $W_0 = 1$ and $W < 1$ initial mutation (i.e. for $t \ll T$) is nearly uniform and changes to the non-uniform for $t \to T$ with the rate adjusted by the value of p. For example, let us assume the values of parameters as follows $W_0 = 1$, $W = 0.8$, $p_m = 0.005$, $l_j = 20$, $t = T$. In this case we have $p_j^0 \approx 0.5$, $p_{j0} = 0.5$, $p_{j19} = 0.0000015$ but expected number of muted bits for this subsequence is stable. Described operator has an important property: it allows to search better the closest neighbourhood of found solution and prohibits of leaving this area because of small probability of mutation of most significant bits. Generally, the non-uniform mutation improves the fine local tuning, this is shown in the next parts of the paper.

3. NON-STATIONARY MUTATION

Non-stationary mutation is the simple modification of uniform mutation, where only global probability p_m is controlled. Some variations of such operator have been already described in literature [9] (and usually called non-uniform mutation). In this work, it is treated as an auxiliary operator in our global approach to mutation. Global probability of mutation p_m as a function of age of population is given below

$$p_m(t) = p_{m0} - (p_{m0} - p_m) \cdot \left(\frac{t}{T} \right)^k \tag{7}$$

where: p_{m0}, p_m — initial and final probability of mutation,
t, T — current and maximal number of generation,
k — auxiliary parameter ($1 \leqslant k \leqslant 5$).

Statement (7) is similar to equal (2) and as we mentioned in previous section, it is useful to control the value of probability of mutation. It is clear, that such operator

can improve the convergence of the algorithm, because, if $p_m \ll p_{m0}$, then for $t \to T$ the probability of modifications of optimized function's parameters decreases. However, non-stationary mutation is qualitatively different from non-uniform mutation — it doesn't ensure improving of fine local tuning, what we show later.

4. MUTATION DEPENDENT ON FITNESS

Described above non-uniform and non-stationary mutation operators make probability of mutation dependent on age of population and significance of bits in chromosome only. It seems reasonable to make this probability dependent also on the fitness of individuals in the population (exactly, in mating pool, i.e. intermediate population selected by sampling mechanism — we use the classical method — a scaled routine [4]). An intuition is, that smaller probability of mutation for the best individuals allows to preserve them in the population and large probability of mutation for worse individuals improves exploration of the search space. On the other hand, big rate of mutation can cause a destruction of some "*good*" schemata in these worse individuals. Thus, it is important to introduce an applicable function describing our mutation operator. Let us introduce notation as follows:

$\bar{f}$ — average fitness of mating pool,

$f=\frac{1}{2}(f_i+f_j)$ — average fitness of couple of parents chosen for crossover operation,

p_m — global probability of mutation;

We have taken the function, which describes a probability of mutation for each couple of parents, given below

$$p(x)=p_m \cdot \frac{C}{1+e^{\lambda(x-b)}}, \tag{8}$$

where: $$b=\frac{\ln(C-1)}{\lambda}, \tag{9}$$

$$x=\frac{f-\bar{f}}{\bar{f}}, \tag{10}$$

C, λ — auxiliary coefficients;

Above function has two important properties. First, we want the probability of mutatoin for $f=\bar{f}$ to be unchanged. It is clear, that for $f=\bar{f}$ we have $x=0$ and $p(x)=p_m$ for every value of C. This property is ensured by the choice of parameter b in accordance with equal (9). Second, we want to impose an upper limit for the multiplier of mutation probability and, of course, the lower limit should be set to zero. The function (8) satisfies these requirements, where value of C is an upper limit for the multiplier. In figure 1 the plots of $p(x)$ for some values of C and λ, are shown.

We use linear scaling of fitness function in the algorithm, so typical values of x are smaller than 0.1 and because of this we must use $\lambda > 10$ if we want to introduce

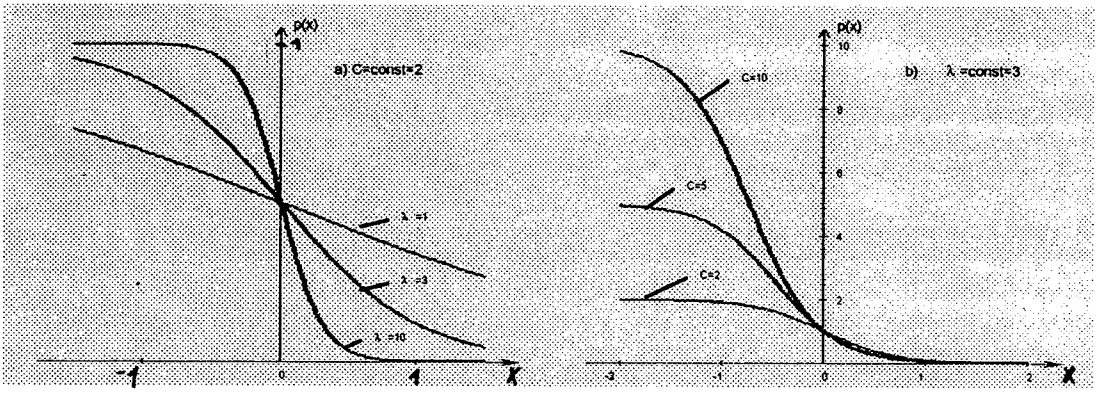


Fig. 1. Plots of function $p(x)$ for some values of coefficients C and λ

a sufficient distinction of the mutation probabilities in the population (as a function of fitness). In the next part of the paper we show the influence of this mutation operator on the performance of the algorithm.

5. OTHER IDEAS

It is possible to apply all the described mutation operators simultaneously? (We mean, that simultaneous using of different mutation operators is equivalent to superposition of these operators, e.g. if non-stationary mutation makes global probability p_m dependent on age of population t — say $p_1(t)$ — and another operator introduces dependence on fitness f — say, $p_2(p_m f)$ — so their combination is $p(t,f)=p_2(p_1(t),f)$; we do not consider the case, that each operator affects on some part of the population). It is possible to apply non-stationary mutation with non-uniform mutation together, and non-stationary mutation wit mutation dependent on fitness. However, we cannot use non-uniform and dependent on fitness mutation simultaneously, because, as we showed in section IV, for the operator dependent on fitness, the probability of mutation for the worst individuals (in mating pool) drastically increases, so the influence of non-uniform mutation becomes small. In our experiments we have found, that the algorithm with non-uniform and dependent on fitness mutation (simultaneously) performs better than algorithm with uniform mutation, but worse than algorithm with non-uniform operator only. We have introduced two modifications of the mutation operator dependent on fitness. First one is that parameter λ (see equation (8)) is now dependent on age of population:

$$\lambda(t)=\lambda_0-(\lambda_0-\lambda)\cdot\left(\frac{t}{T}\right)^d \tag{12}$$

where: λ_0, λ — initial and final values of parameter (λ is close to 0),
t, T — current and maximal number of generation,
d — auxiliary parameter ($0.1 \leqslant d \leqslant 1.0$);

This modification causes that an influence of the mutation dependent on fitness decreases, when $t \to T$. Let us denote that dependent on fitness mutation with λ described by (12) and non-uniform mutation working together is called *modified dependent on fitness mutation operator*. Second modification is a combination of dependent on fitness mutation and non-uniform one. This is a non-uniform mutation with q parameter given

$$q = (2 - C_{mul}) \cdot \left(W_0 - (W_0 - W) \cdot \left(\frac{t}{T} \right)^r \right) + C_{mul} - 1 \tag{13}$$

where $C_{mul} = \dfrac{C}{1 + e^{\lambda(x-b)}}$ (14)

and W_0, W, t, T, r are the same as in non-uniform mutation,
C, λ, x, b are the same as in dependent on fitness mutation ($C=2$).

In figure 2 the parameter q as a function of x is shown. It is clear that for $x=0$ (average fitness of couple of parents is equal to average fitness of population) we have $C_{mul}=1$ and q given by (13) is equal to q given by (2). Similarly, for $x<0$, we have $C_{mul} \to 2$ and $q \to 1$ (uniform mutation) and, for $x>0$, $C_{mul} \to 0$, and value of q becomes close to $q-(1-q)$ given by (2) (increase of non-uniformity). In other words, our considerations about global probability of mutation in dependent on fitness mutation operator are quite the same, but with respect to the value of q in modified operator. The value of λ can be constant or vary in accordance with (12). Let us note that non-uniform mutation operator with modifications given by (13) and (14) is called *modified non-uniform mutation operator*. We show that in some cases the algorithm with combined and modified mutation operators behaves better than the algorithm with only one operator.

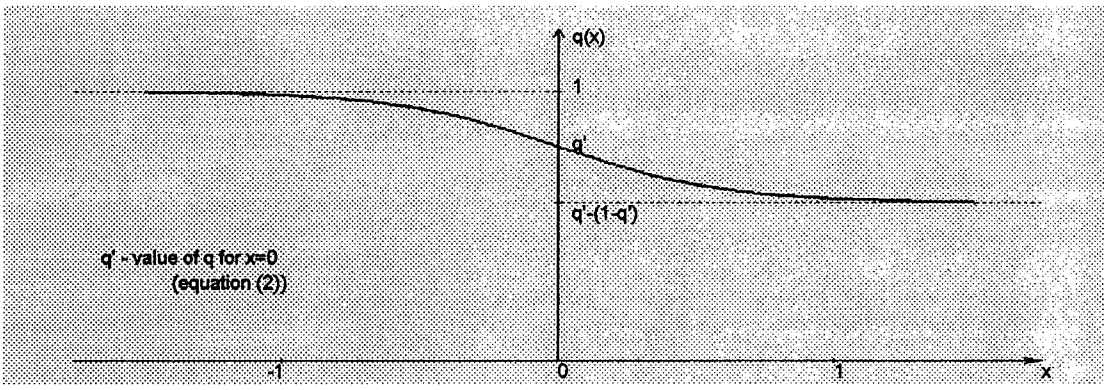


Fig. 2. Plot of typical function $q(x)$ for $C=2$, $\lambda=5$ for fixed number of generation t

6. EXPERIMENTAL RESULTS

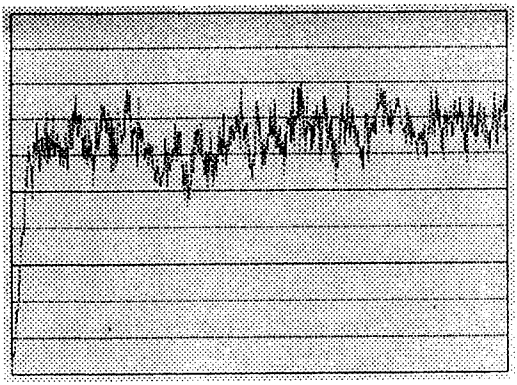
In order to test our modified mutation operators we have performed many experiments with different functions. We found out, that characteristic properties of operators are quite similar in each case. In this work, as we mentioned earlier, we are

mainly interested in demonstration of differences between introduced mutation operators, especially in improvements of performance of algorithm working with our operartors in comparison with classical mutation operator. Thus, we propose the simple function of n variables given by (11) as a useful test-bed to show above differences (the task is to maximize this function). Let us note, that it is easy to adjust a shape of this function and a number of local extrema.

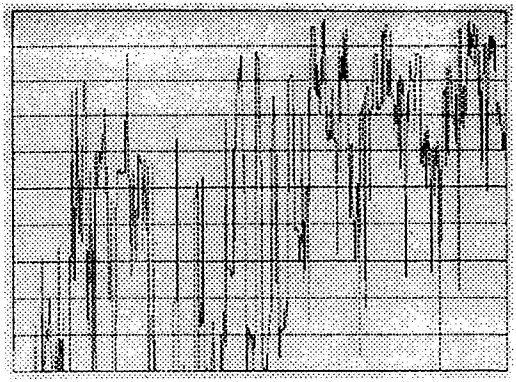
$$f(x_1,\ldots,x_n)=\prod_{i=1}^{n}(a\cdot|\sin(\pi\cdot x_i)|^{s1}+(1-a)\cdot|4x_i(1-x_i)\cdot\sin(w\pi\cdot x_i)|^{s2}) \tag{11}$$

The functions is determined by values of the coefficients: $0\leqslant a<1$, $s1>0$, $s2\geqslant 0$, w (odd natural number). The domain of the parameters is $0\leqslant x_i\leqslant 1$ for $i=1,\ldots,n$. The value of global maximum is 1 (for $x_i=0.5$). The number of local extrema is equal to about w^n. We use classical version of genetic algorithm, with simple crossover, scaled routine as sampling mechanism and linear scaling of fitness function [4]. Each parameter of test function (11) is represented by a subsequence of 20 bits; thus, the precision is six digits after decimal point. For all experiments, we have used the populations of 50 individuals, and probability of crossover was $p_c=0.3$. We have performed many experiments with described above mutation operators, for various sets of test function coefficient and various parameters of mutation operators. The behaviour of algorithm was qualitatively similar for all these cases. Figures 3 – 8 shows typical, representative runs of genetic algorithm working with some variations of mutation operator, for the following values of coefficients: $n=5$, $a=0.5$, $s1=2$, $s2=2$, $w=7$; the runs were made for 1000 generations. Let us note, that for values given above, the test function has more than 10^4 local extrema. The part a) of each figure shows average fitness of population (range of vertical axis is [0, 1]), part b) — maximal fitness in population (range of vertical axis [0.9, 1.0]).

The values of parameters of mutation operators were selected experimentally. The subject of our future work will be investigating of influence of these parameters on the performance of the algorithm for different optimization problems. Our recent experiments indicates, that the same set of parameters gives similar results for various optimized functions (i.e. operators demonstrate similar properties in each case).

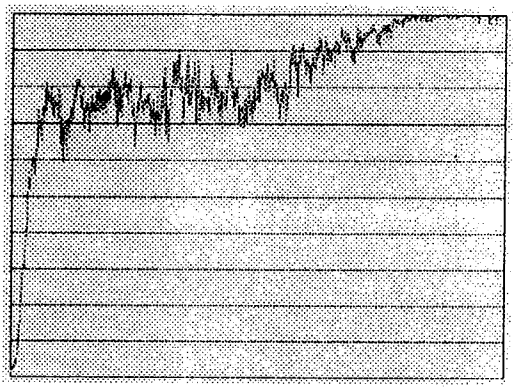


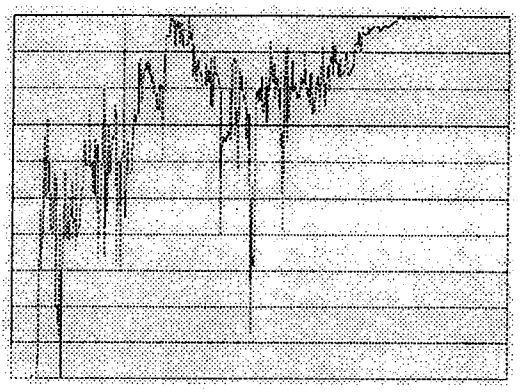
a) average fitness



b) maximal fitness

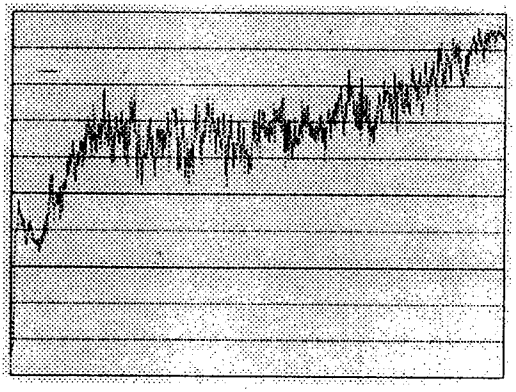
Fig. 3. Typical performance of algorithm with uniform mutation operator

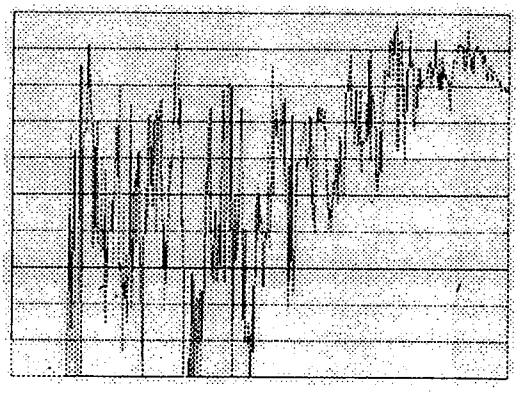




a) average fitness b) maximal fitness

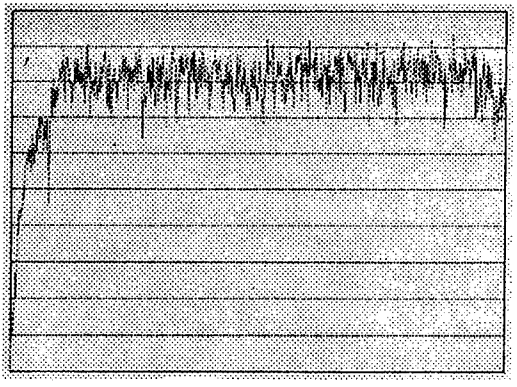
Fig. 4. Typical performance of algorithm with non-uniform mutation operator

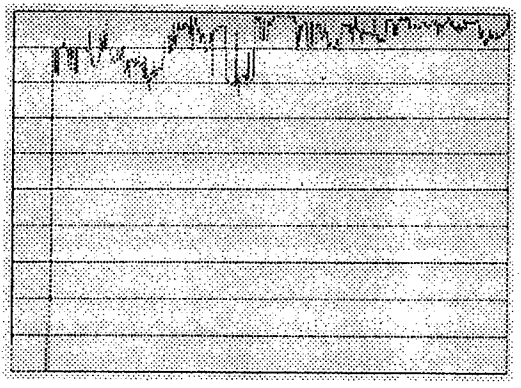




a) average fitness b) maximal fitness

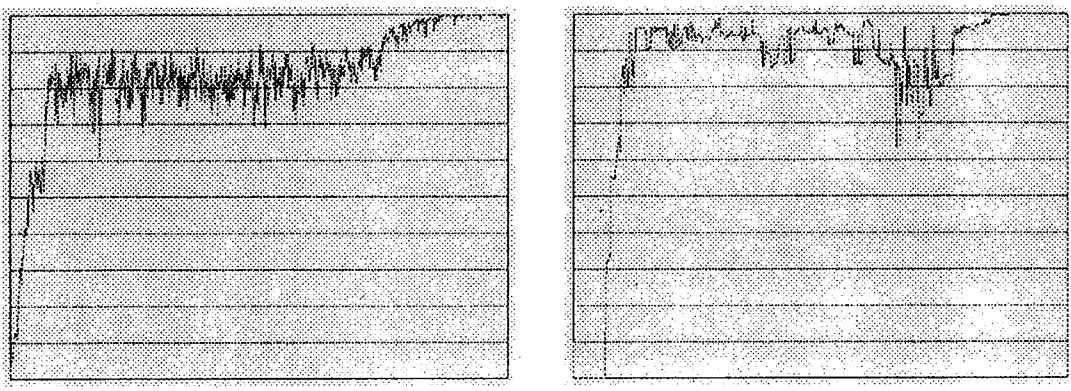
Fig. 5. Typical performance of algorithm with non-stationary mutation operator





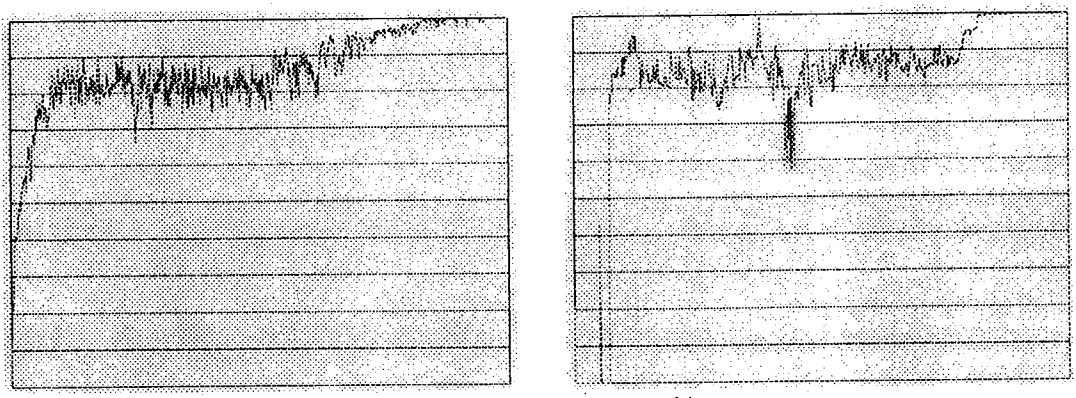
a) average fitness b) maximal fitness

Fig. 6. Typical performance of algorithm with dependent on fitness mutation operator



a) average fitness b) maximal fitness

Fig. 7. Typical performance of algorithm with modified dependent on fitness mutation operator



a) average fitness b) maximal fitness

Fig. 8. Typical performance of algorithm with modified non-uniform mutation operator

Table 1 includes the values of f_{max}, $\bar{f}_{max}$ and S for the best of ten independent runs of the algorithm, for the values of test function coefficients: $n=5$, $a=0.5$, $s1=2$, $s2=2$, $w=7$; where:

f_{max} — fitness of the best individual,

$\bar{f}_{max}$ — the highest value of average fitness for population,

$S=\sigma/\bar{f}$ — a measure of convergence of the algorithm, where σ — standard deviation of fitness in population, $\bar{f}$ — average fitness of population. The table 1 includes an average value of S for last 10 generations of the algorithm. The runs of the algorithm were performed for 2000 generations. For comparison, we show the results for algorithm working with uniform mutation operator, too. Let us note, that parameters $\bar{f}_{max}$ and S are very important as a measure of convergence of the algorithm. We can find out, that all modified operators outperforms classical uniform operator with respect to every criterion: precision of find of optimum and convergence of algorithm.

Table 1

Results of the best of ten independent runs of algorithm for various mutation operators (function coefficients: $n=5$, $a=0.5$, $s1=2$, $s2=2$, $w=7$)

Operator		Parameter		
		f_{max}	$\bar{f}_{max}$	S
I	uniform mutation	0.998629	0.844167	0.354053
II	non-uniform mutation	1.000000	1.000000	<0.000001
III	non-stationary mutation	0.999340	0.988485	0.079382
IV	dependent on fitness mutation	0.999661	0.943628	0.360735
V	non-uniform+non-stationary mutation	1.000000	1.000000	0.000003
VI	dependent on fitness+non-stationary mut.	0.999872	0.993301	0.102303
VII	modified dependent on fitness mutation	1.00000	1.00000	0.0000015
VIII	modified non-uniform mutation	1.00000	1.000000	0.0000011
IX	modified dep. on fitn. mut.+non-stationary	1.000000	1.000000	<0.000001
X	modified non-uniform mut.+non-stationary	1.000000	1.000000	0.000015

Parameters of mutation operators: I: $p_m=0.01$; II: $W_0=0.95$, $W=0.5$, $r=4$, $p_m=0.01$; III. $p_{m0}=0.01$, $p_m=0.0005$, $k=4$, IV: $C=5$, $\lambda=50$, $p_m=0.01$; V: $W_0=0.95$, $W=0.5$, $r=4$, $p_{m0}=0.01$, $p_m=0.0005$, $k=4$; VI: $C=5$, $\lambda=100$, $p_{m0}=0.01$, $p_m=0.0005$, $k=4$; VII: $C=5$, $\lambda_0=100$, $\lambda=0$, $d=0.5$, $p_{m0}=0.01$, $W_0=1.0$, $W=0.5$, $r=4$; VIII: $C=2$, $\lambda_0=50$, $\lambda=0$, $d=0.2$, $p_{m0}=0.01$, $W_0=0.95$, $W=0.5$, $r=4$; IX: $C=5$, $\lambda_0=100$, $\lambda=0$, $d=0.5$, $p_{m0}=0.01$, $p_m=0.0005$, $k=4$, $W_0=1.0$, $W=0.5$, $r=4$;X: $C=2$, $\lambda_0=50$, $\lambda=0$, $d=0.5$, $p_{m0}=0.01$, $p_m=0.0005$, $W_0=1.0$, $W=0.5$, $r=4$;.

Table 2

The best results of ten independent runs of algorithm for various mutation operators (function coefficients: $n=10$, $a=0.5$, $s1=0$, $s2=2$, $w=13$)

Operator		Parameter		
		f_{max}	$\bar{f}_{max}$	S
I	uniform mutation	0.926258	0.525083	0.7214
II	non-uniform mutation	1.000000	0.999998	0.0027
III	non-stationary mutation	0.996511	0.987103	0.1115
IV	dependent on fitness mutation	0.998444	0.923170	0.4728
V	non-uniform+non-stationary mutation	1.000000	1.000000	0.0079
VI	dependent on fitness+non-stationary mut.	0.998629	0.983042	0.1186
VII	modified dependent on fitness mutation	1.000000	0.999999	0.0026
VIII	modified non-uniform mutation	0.999999	0.999996	0.0011
IX	modified dep. on fitn. mut.+non-stationary	1.000000	1.000000	0.0049
X	modified non-uniform mut.+non-stationary	1.000000	1.000000	0.0025

Parameters of mutation operators: I: $p_m=0.01$; II: $W_0=0.9$, $W=0.5$, $r=4$, $p_m=0.01$; III. $p_{m0}=0.01$, $p_m=0.0005$, $k=3$, IV: $C=5$, $\lambda=50$, $p_m=0.01$; V: $W_0=0.9$, $W=0.5$, $r=6$, $p_{m0}=0.01$, $p_m=0.0005$, $k=4$; VI: $C=5$, $\lambda=100$, $p_{m0}=0.01$, $p_m=0.0005$, $k=4$; VII: $C=5$, $\lambda_0=100$, $\lambda=0$, $d=0.5$, $p_{m0}=0.01$, $W_0=1.0$, $W=0.5$, $r=3$; VIII: $C=2$, $\lambda_0=50$, $\lambda=0$, $d=0.2$, $p_{m0}=0.01$, $W_0=0.95$, $W=0.5$, $r=5$; IX: $C=5$, $\lambda_0=100$, $\lambda=0$, $d=0.5$, $p_{m0}=0.01$, $p_m=0.0005$, $k=4$, $W_0=1.0$, $W=0.5$, $r=4$;X: $C=2$, $\lambda_0=50$, $\lambda=0$, $d=0.5$, $p_{m0}=0.01$, $p_m=0.0005$, $W_0=1.0$, $W=0.5$, $r=5$;.

In the table 2 the results for the best of ten independent runs of the algorithm are shown, for the following values of test function coefficients: $n=10$, $a=0.5$, $s1=2$, $s2=2$, $w=13$ (number of local extrema more than 10^{11}); the runs for this function were performed for 10000 generations.

Table 3 includes the best resutls of ten independent runs of algorithm for values of test function coefficients: $n=5$, $a=0.998$, $s2=2$, $w=7$ (and factor $|\sin(\pi \cdot x_i)|^{s1}$ is replaced by constant value 1); the runs were performed for 2000 generations. In this case, differences between values of local extrema are quite small — the value of global minimum is about 0.99, and global maximum is 1.00. It indicates, that finding of optimum can be as difficult as in previous cases — the algorithm can easily falls into local optimum; however, it takes place due to another reasons — sampling mechanism is too weak to distinguish the best individuals. A performance of the algorithm for such cases can be improved using the power scaling of the objective function [9]; however, but we have shown, that our mutation operators are able to provide good results without additional modifications of the algorithm.

Table 3

The best resutls of ten independent runs of algorithm for various mutation operators (function coefficients: $n=5$, $a=0.998$, $s2=2$, $w=7$)

Operator		Parameter		
		f_{max}	$\bar{f}_{max}$	$S(\cdot 10^{-6})$
I	uniform mutation	0.999992	0.999350	991.20
II	non-uniform mutation	1.000000	1.000000	20.31
III	non-stationary mutation	0.999996	0.999954	181.70
IV	dependent on fitness mutation	0.999999	0.999676	1606.00
V	non-uniform + non-stationary mutation	1.000000	1.000000	<0.1
VI	dependent on fitness + non-stationary mut.	1.000000	0.999996	151.70
VII	modified dependent on fitness mutation	1.00000	1.00000	<0.1
VIII	modified non-uniform mutation	1.00000	1.000000	<0.1
IX	modified dep. on fitn. mut. + non-stationary	1.000000	1.000000	0.85
X	modified non-uniform mut. + non-stationary	1.000000	1.000000	20.02

Parameters of mutation operators are the same as given under table 1.

Presented results allow to draw some conclusions. Let us denote the version II, III and IV (see tables) as the single mutation operators, and the versions V – X as the multiple mutation operators. First of all, introduced mutation operators behave better than uniform one. Dependent on fitness mutation improves maximum found value of fitness (we call it f_{max}), but does not improve the convergence of the algorithm. Non-stationary mutation gives a moderate growth of both f_{max} and the convergence. The best of single operators is the non-unifiorm mutation. As we saw, this operator ensured a remarkable improvement of f_{max} (usually with accuracy of six digits after

decimal point) and the same growth of convergence of algorithm. Combination of single operators (II, IV) and non-stationary mutation (V, VI) gives some additional improvements. It is important to distinguish a performance of modified operators (VII and VIII); these seem to be the best of all. Let us note, that dependent on fitness mutation gives considerable improvement of f_{max} very early (with respect to the age of population), thus, in other words, it "prepares" the population for non-uniform operator, which operates in final stage of performance of the algorithm. Both VII'th and VIII'th version of mutation operator realise above idea. The operators IX and X are multiple ones, where IX and X are just versions VII and VIII combined with the non-stationary mutation. We are able to compare only convergence of the algorithm working with these operators, because all of them usually find global optimum with highest precision. In some cases, version IX and X give worse convergence than VII and VIII. The reason is, as we think, that introducing of non-stationary mutation to modified operators (VII, VIII) makes average value of probability of mutation too small to give changes of genes sufficient to effective searching of a solution domain.

It seems that non-uniform mutation operator is the most important one of all described in this paper. It is clear, that non-uniform mutation performs best, both as single operator and as multipleo ne. Others, non-stationary and dependent on fitness mutation we can treat as the auxiliary operators. At times they play an important role, e.g. if maximal number of generations T is small and differences between extrema are small too, algorithm with only non-uniform mutation can easily fail into local optimum and, as we have found out, other operators are able to preserve it.

CONCLUSIONS

The results obtained in the work show, that modified mutation operators can improve the performance of genetic algorithm in the field of numerical optimization. Furthermore, appropriate mutation operator is able to replace other mechanisms, like some methods of scaling of fitness function [9] or special sampling mechanisms [4], in the sense of ability of algorithm to find an optimal solution without those operators. The best of presented operators ensure, that almost all population converges into global optimum with high precision. In this paper, of course, we have presented results for some simple test functions only, but our main goal was to show the differences between performance of algorithm working with various mutation operators. The goal of our future work is to test presented mutation operators for other functions (among others for classical set of de Jong's functions [3]) and to perform more comprehensive studies about an influence of values of operators parameters on the performance of the algorithm. The algorithm working with some of modified operators (non-uniform and non-stationary ones) have been already applied to several difficult problems. One of them was fitting parameters to the model of non-uniformly doped MOSFET transistor [8]. This problem is equivalent to optimizing the multimodal function with very "narrow" local optima. Results are very encouraging. Presented mutation operators have been also applied to testing the novel method of numerical optimization with constraints [7], where as a test-bed served some functions known from literature ([5], [11]). Our algorithm was able to find solutions very close to global optima.

It seems, that except for time performance of the algorithm, a binary representation of chromosomes can be as good as floating point one in numerical optimizatoin problems. We think, that mutation operator is quite noticeable as a way of improvement of genetic algorithms performance.

REFERENCES

[1] H.J. Antonisse: *A new Interpretation of Schemat Notation that Overturns the Binary Encoding Constraint,* in [12], pp. 86–91

[2] L. Davis (Editor): *Handbook of Genetic Algorithms.* Van Nostrand Reinhold, New York, 1991

[3] K.A. De Jong: *An Analysis of the behaviour of a class of genetic adaptive systems,* (Doctoral dissertation, University of Michigan). Dissertation Abstract International, 36(10), 5140B (University Microfilms No 76-9381)

[4] D.E. Goldberg: *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Publishing Company, Inc., 1989

[5] W. Hock, K. Schittkowski: *Test Example for Nonlinear Programming Codes.* Lecture Notes in Economics and Mathematical Systems, Vol. 187, Springer-Verlag, 1981

[6] J. Holland: *Adaptation in natural and Artificial Systems.* University of Michigan Press, Ann Arbor, 1975

[7] S. Kozieł: *Algorytmy ewolucyjne w zagadnieniach optymalizacji numerycznej z ograniczeniami na zbiorach wypukłych,* submitted for publication, 1996

[8] S. Kozieł, Kordalski, W.: *Applications of Genetic Algorithms to Fitting Parameters to an Unified Model of the NOn-uniformly Doped MOSFET,* to appear, Proceedings of the XIX KKTOiUE, 1996

[9] Z. Michalewicz: *Genetic Algorithms+Data Structures=Evolution Programs.* Springer-Verlag, New York, 2nd edition, 1994

[10] Z. Michalewicz, T.D. Logan, S. Swaminathan: *Evolutionary Operators for Continous Convex Parameters Spaces.* Proceedings of the Third Annual Conference on EP, World Scientific, 1994, pp. 84–97

[11] Z. Michalewicz, G. Nazhiyath, M. Michalewicz: *A Note on Usefulness of Geometrical Crossover for Numerical Optimizatio Problems.* Proceedings of the Fifth Annual Conference on Evolutionary Programming, 1996

[12] J. Schaffer (Editor): *Proceedings of the Third International Conference on Genetic Algorithms.* Morgan Kaufmann Publishers, Los Altos, CA, 1989

S. KOZIEŁ

NIEJEDNORODNE I NIESTACJONARNE OPERATORY MUTACJI W OPTYMALIZACJI NUMERYCZNEJ PRZY UŻYCIU ALGORYTMÓW GENETYCZNYCH

Streszczenie

Artykuł stanowi ogólne podejście do operatora mutacji w klasycznych algorytmach genetycznych z reprezentacją binarną ciągów kodowych. Zaprezentowano kilka wariantów operatora mutacji, między innymi operatory niejednorodne i niestacjonarne. Pokazano działanie algorytmu pracującego z wprowadzonymi operatorami na przykładzie kilku prostych funkcji testowych. Omówiono problem zbieżnosci algorytmu oraz dodatkowe funkcje jakie może pełni operator mutacji w algorytmach genetycznych.

Słowa kluczowe: algorytmy genetyczne, operator mutacji.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1996. 42, z. 3, ss. 287–294

Experimental determination of the band bending the surface of semiconductors

BOŻENA SEROCZYŃSKA-WOJAS

Instytut Fizyki, Politechnika Warszawska

JÓZEF WOJAS

Prof. emeryt., Politechnika Warszawska

Received 1996.06.05

Authorized 1996.08.23

The coefficient D determining the relative band bending at the surface of the semiconductor was defined. By using of this coefficient, the rough estimations of the degree of space-charge-region disturbance are possible.

Key words: surface of semiconductors, band deformations, disturbance degree.

1. INTRODUCTION

At a semiconductor surface, which is approximately defined as a plane between a crystal and vacuum contact, the additional energetic levels appear, refered to as the surface states. They originate as a result of the crystal lattice termination at the surface (The Tamm and Shokley states) and as a result of nonideality at the surface (defects). The defects appear as a result of the chemical contamination of the material, atoms adsorption from the surrounding, the mechanical damage of the surface etc. The surface states can be populated by the additional charge of electrons and holes, which influences the deformation of the semiconductor band model at the surface.

The layer at which the band model is disturbed is referred to as the surface layer. The physical properties of this layer can differ by far from the remaining part of the crystal, which we refer to as the bulk properties. Usually we distinguish two cases: enhancement (Fig. 1a) and inversion (Fig. 1b) at the surface layer [1; 2; 3]. If at the n-type semiconductor the acceptor states prevaile and they attach the electrons, the surface charges negatively, and potential of the surface layer rises with respect to the vacuum.

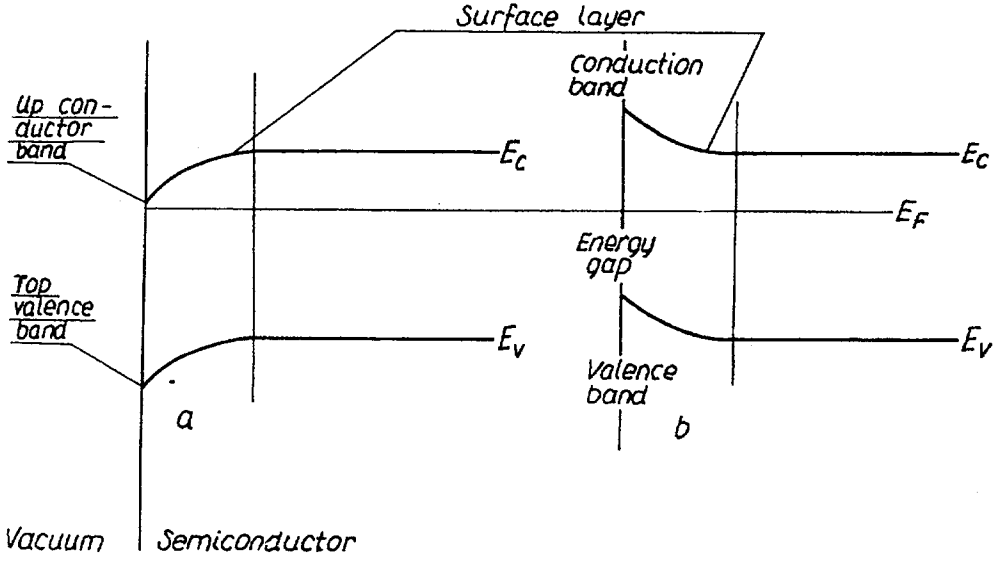


Fig. 1. Simplified diagrams of energetic band bending at a semiconductor surface, a — enhancement, b — inversion

In the case of inversion in a *p*-type semiconductor the surface should trap holes, since in this case the surface layer become to be of n-type. This mechanism is presented at Fig. 2 by the band bending [4, 5]. In the "a" case the bands appear to be bended upwards and the Fermi level approaches the top of the valence band, while its separation from the conduction band rises. Whereas the reverse is true in the "b" case.

Owing to the fact that the properties of semiconductor play an essential role, in many phenomenas (i.e. the contact voltage, photoconductivity, photovoltaic effect etc.) to the variety of studies even the approximate informations are valuable about the kind and intensity of deformation of the surface layer, which gives informations about the kind and the defect density at the surface. The need accurred in the to date experience of the autors concerning with the determination of work function and quantum efficiency in the variety of semiconductor surfaces. So the authors have defined the rough criterion of the disturbance sharpnes in the surface zone by introducing the coefficient D determining the relative band bending at the surface.

The data to the general relation (which will be presented below) as well as informations required for the drawings of models in each cases were obtained from determination the energetic distances δ of the Fermi level from the top of the valence band: a) in the bulk region δ_v and b) in the surface layer δ_s of the material under study. The energetic distances regarding to the δ_v and δ_s values, in the case of n and *p*-types are illustrated in Fig. 2. Since δ_v and δ_s usually differs in *n* and *p*-type semiconductors, in Fig. 2 the notation: δ_{vn} and δ_{sn} are used for *n*-type, and δ_{vp} and δ_{sp} — for the *p*-type.

The value δ_v is found from the Hall constant and conductivity measurements, and δ_s from the semiconductor surface studies using the photoelectron emission method. This will be discussed in the following sequence:

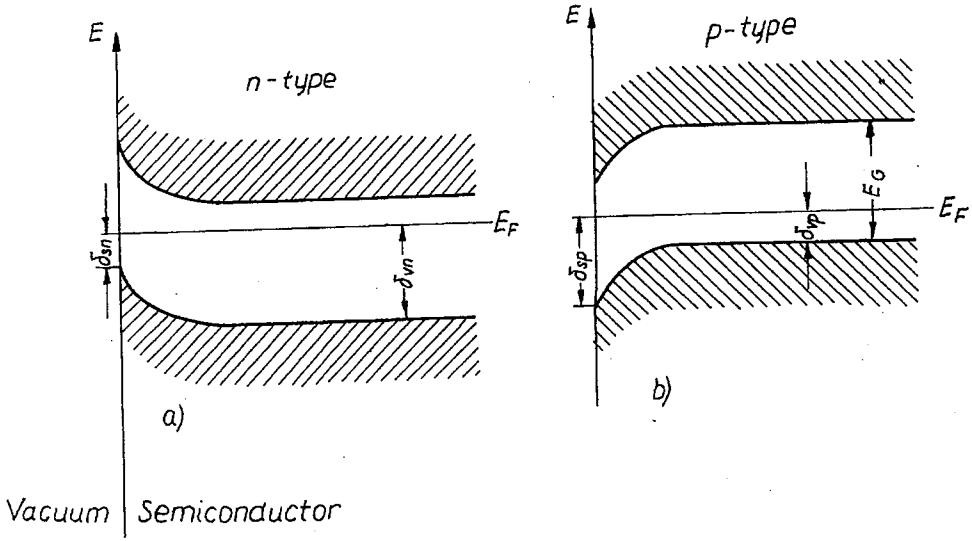


Fig. 2. Diagram of band bending in n and p-type semiconductors

2. DETERMINATION OF δ_v VALUE

Band deformation in the surface layer occurs of course in the policrystalline samples as well as in the crystals. In policrystals the disturbance degree is found usually to be greater owing to the much greater defects density. In oriented monocrystals the surface properties of the crystal depend less or more on the orientation plane. For instance it is apparent in semiconductors of III – V group [6].

The Hall constant R and conductivity values, which are used as the data to obtain informations concerning the bulk properties of a crystal (the principal information concerns the position of the Fermi level) can be direction dependent in crystals. This is connected with the band structure in the semiconductor bulk, more precisely — with the shape of equienergetic surfaces. In the case of spherical surfaces the ideal isotropy is found for the R and σ values. According to data in the literature [7] a spherical conduction band were found in some III – V semiconductors, i.e. in InSb and InAs. However GaAs, according to Ref. [8], has a more complicated shape of the conductivity band, which splits into two bands of boundaries energetically shifted in (111) direction and electron effective masses in these bands differ considerably. In germanium and silicon the equienergetic surfaces of conduction band were found to be ellipsoide, this is connected with anisotropy of the effective mass and other physical properties. The anisotropy coefficient A, equal to the rate of effective masses in directions of two principal axis of ellipse is 20 for Ge and about 5 for Si [7].

In the case of such complicated energetic structure, that dependence of R and σ on direction in the crystal can not be ommited, the δ_v and δ_s values correlated in the correct way can be obtained, when the orientation of uncovered surface will be in

accordance with the direction of electron movement in the crystal during (the Hall voltage measurement). In this case the energetic models, considering the band bending at the surface, would be considerably differing even for crystallografic directions with the low indexes of Miller. Generally, for policrystals we can assume the isotropy of mentioned above physical quantities due to the strong disturbance of geometrical structure of the crystal in its entire volume, which can lead to equivalence of many directions. However even in this case it is difficult to reduce the surface influence in such a degree that R and σ values would represent the bulk properties. In the case of single type conductivity, when the forbidden gap is known, from the above measurements it is possible to determine the δ_v value, which represents the Fermi level position in relation to the top of the valence band. Here it is possible to use the relation for the Hall constant and conductivity.

The Fermi energy E^F is related to the carrier concentration, to effective masses (m_e^* and m_h^*) and to temperature.

It is troublesome to measure accurately the effective mass however the values are known at the bands boundary for several semiconductors. The effective masses are known to be independent of energy in parabolic bands.

When the two kind of conductivity occur in the sample, and the semiconductor is not degenerated, the following relation can be used to determine the Fermi level

$$n_e \cdot n_h = n_i^2,$$

where n_i is the intrinsic carrier concentration, the value of which can be found in the literature for different semiconductors.

When all mentioned above values of three concentrations are known, the Fermi level can be found by using the tables of the Fermi integrals and the relation

$$n_e = \frac{n_i^2}{n_h} = \frac{16\pi^2}{n_h}\left(2m_e^* \frac{kT}{h^2}\right)^3 \left\{F_{1/2}\left(\frac{E_F}{kT}\right)\right\}^2$$

The $F_{1/2}$ function is known as the Fermi integral

$$F_{n(a)} = \int_0^\infty \frac{x^n dx}{\exp[(x-n)+1]}.$$

The values of Fermi integrals when n have half values can be easy found in the literature [9]. It is more difficult to determine the Fermi level in semiconductors with complicated band structure, when the conduction band is not parabolic and the valence band is applied and shows departure from the classical structure. Perturbation of the band model occurs in all two or more component semiconductors, due to deffect associated with departure from stechiometry. The semiconductors of III–V group seem to be the best known. The Kane's theory, created in 1957, describes mathematically the band structure of InSb and can be used to others III–V group compounds. Dependences between effective mass, the carrier concentration and the position of the Fermi level were found to be more complicated compared to the above

mentioned (bell shaped conduction band), but the theory allows, for at last six compounds of III – V group, to use the appropriate relations and graphics to approximate E_F (works of Kane Ehrenreich [8], Cardona [11, 12] and Kołodziejczak and coworkers [13, 14]).

3. DETERMINATION OF δ_s VALUE

Authors of this work made approximation for δ_s from photoelectron emission* from excitation by the ultrafiolet monochromatic radiation [15 – 19] and using the method of spheric field retardation. From the obtained current-voltage characteristics Fig. 3 the maximal energy of electrons can be calculated. Hence, taking into account the assumption that the electrons having the largest speed came from the top of valence band, position of the top with respect to the vacuum level is estimated. It is essential that the distance from the vacuum level denoted as φ_{vac} defines the position of the valence band top in the zone of the surface perturbation, because the zone is known to penetrate up to distance of 10^{-4} cm from the surface, however the geometric depth of photoelectrons excitation in semiconductors is known to be at most 10^{-5} cm. From the characteristics obtained, the values of collector-emitter contact voltage (Fig. 3) can also be read out. The value of φ_e can be found from the relation:

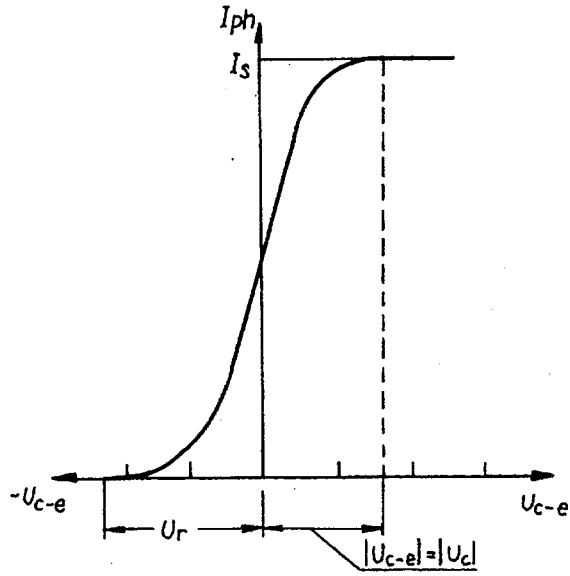


Fig. 3. Characteristic of photoemission current vs collector-emitter voltage

* Measurements of the field emission of electrons can be used as the alternative method for determination of near-surface band bending.

$$\varphi_e = \varphi_c - U_{con},$$

this is "thermionic work function" of the semiconductor emitter, which is identified as the Fermi level depth position with respect to the vacuum level. It is apparent that work function φ_c of the collector layer should be known and be unvariable during measurements. If value of φ_e is given, we can calculate δ_s from the simple relation: $\delta_s = \varphi_{vac} - \varphi_e$. When using the described above method for comparison of the band deformation for different surface types, the photoelectric measurements should be carried out by monochromatic light using in all cases the thereshold wavelenght. The results obtained by using considerably differing wave-lenghts λ could be unadequate, since the δ_s values can be function of $h\nu$ in some energy intervals of radiation.
The method of retarding spheric field allows us to obtain the results with the error not lower than ± 0.02 eV. The more exact values of δ_s can be obtained by determining the thermionic work functoin φ_e, using the method of the contact potential difference in a metal-semiconductor structure [20–21].

4. DEFINITION OF THE COEFFICIENT D CALCULATION EXAMPLES OF BAND DEFORMATION AT SURFACE FOR FEW DIFFERENT SEMICONDUCTORS

To compare the surface disturbance degree of the structure in investigated samples, the authors have defined the coefficient D, which they expressed for n-type samples as

$$D_n = \frac{\delta_{sn} - \delta_{vn}}{E_G}$$

and for p-type samples as

$$D_p = \frac{\delta_{vp} - \delta_{sp}}{E_G},$$

where E_G is the energy gap expressed in eV.
Since δ_v and δ_s are also given in eV, the coefficient D can be expressed in percentage. In the case of inversion at the surface, the D is negative, in the case of enhancement — positive. Table 1 and 2 contain the D values calculated from the author's experimental results as well as from experimental data in the literature. Different semiconductors with surfaces prepared in different way were used.

Table 1

Values of D calculated from Gobeli and Allen results [20] φ_{ph} were found by extrapolation from quantum yield curves

Semiconductor, and orientation	Conductivity type in the bulk	The kind of the surface deformation	E_G^+ [eV]	| D | [%]
GaAs (110)	*n*	inversion	1.40	43
InAs (110)	*n*	enhancement	0.36	41.5
GaSb (110)	*p*	enhancement	0.70	11.5
Ge (111)	*p*	enhancement	0.67	~7.5
Si (111)	*p*	inversion	1.09	~2

+ The value obtained in the work [20]

Table 2

Values of D calculated from the results of the retarding field method in spherical condensator

Autor	Semiconductor	The kind of sample and surface	Energy of incident radiation [eV]	E_G [eV]	| D | [%]
Haneman [22]	InSb	semiconductor cleaved in vacuum, the surface (110)	5.2	0.18	17
Wojas [17]	GaAs	the etched surface (111)	5.8	1.35	47.5
Wojas [16]	HgSe	policrystal *n*-type; the cleaved surface	5.38	0.08	45
Seroczyńska--Wojas [17]	InSb	mixed conductivity, the polished surface	4.9	0.18	40

REFERENCES

1. N.B. Hannay: *Semiconductors*. Reinhold Publ. Corp. New York, London, Chapman Hall Ltd., 1959.
2. B. Adamowicz: *Thesis doctor work*. Gliwice 1988
3. A. Many, Y. Goldstein, N.B. Grover: *Semiconductor surfaces*. North-Holland Publ. Co., Amsterdam 1965
4. B.A. Niesterenko, O.W. Snitko: *Fiziczeskije swojstwa*. ANUSSR. Kijów 1983
5. Sowriemiennyje Problemy Fiziczeskoj Chimii Powierchnosti Połuprowodnikow. Sbornik Naucznych Trudow. Akademia Nauk SSSR Novosybirsk, Nauka 1988
6. G.W. Gobeli, F.G. Allen: *Semiconductors and semimetals*. 1966. v. 2. ch. 11: *Photoelectric threshold and work function*. Academic Press, N.Y., London 1966
7. J. Wojas: *Fotoelektryczne badania półprzewodników*. Wyd. Nauk. SEMPER, Warszawa 1994
8. H. Ehrenreich: Phys. Rev. 1965, v. 20, pp. 1951·
9. J. Dougal, E. Stoner: Phil. Trans. 1968, v. 237, pp. 67

10. J. W o j a s: *Threshold photoemission from semiconducting real surfaces.* Kwart. Elektr. i Telekom. 1994, v. 40, z. 3, pp. 331–342
11. M. C a r d o n a: J. Phys. Chem. Solids. 1963, v. 24, pp. 1543
12. M. C a r d o n a: J. Phys. Chem. Solids. 1965, v. 26, pp. 1351
13. J. K o ł o d z i e j c z a k, S. Ż u k o t y ń s k i, H. S t r a m s k a: Phys. Stat. Sol. 1966, v. 14, pp. 471
14. J. K o ł o d z i e j c z a k, S. Ż u k o t y ń s k i: Phys. Stat. Sol. 1966, v. 16, K 55
15. J. W o j a s: Acta Phys. Pol. 1975, v. 40, pp. 603
16. J. W o j a s: Phys. Stat. Sol. 1969, v. 33, pp. K121
17. B. S e r o c z y ń s k a - W o j a s: Acta Phys. Pol. 1977, v. 52, pp. 893
18. J. W o j a s: Arch. Elektrotech. 1983, v. 32, pp. 697
19. J. W o j a s: *Energetyczna korelacja modelu fotoemisji z modelem pasmowym wnętrza kryształów Si, Ge i związków* $A^{III}-B^{V}$. Biuletyn WAT Elektronika, 1995, nr 7, pp. 73–93
20. M. S u b o t o w i c z: Annales Universitatis M. Curie-Skłodowska, Lublin, Poland, 1960, v. 13, Sectio AA. pp. 105
21. J. W o j a s: *Fizyka powierzchni półprzewodników.* Akadem. Oficyna Wyd. PLJ, Warszawa 1995
22. D. Haneman: J. Phys. Chem. Solids. 1959, v. 11, pp. 205; 1961, v. 121, pp. 1093

J. WOJAS, B. SEROCZYŃSKA-WOJAS

DOŚWIADCZALNE OKREŚLENIE ZAGIĘCIA PASM PRZY POWIERZCHNI PÓŁPRZEWODNIKA

S t r e s z c z e n i e

Został zdefiniowany współczynnik D określający względne zagięcie pasm przy powierzchni półprzewodnika. Za pomocą tego współczynnika można w przybliżeniu ocenić stopień zakłócenia (zaburzenia) obszaru przypowierzchniowego ładunku przestrzennego.

Słowa kluczowe: powierzchnia półprzewodnika, deformacja pasm, współczynnik zakłócenia.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1996. 42, z. 3, ss. 295–323

Układowe realizacje prądowych sztucznych sieci neuronowych

KRZYSZTOF WAWRYN

Instytut Elektroniki, Politechnika Koszalińska

Otrzymano 1996.05.10

Autoryzowano 1996.08.05

W artykule przedstawiono układowe realizacje prądowych sztucznych sieci neuronowych. Zostały zapisane prądowe układy: analogowej funkcji aktywacji, synapsy o stałym i przestrajanym współczynniku wagowym oraz opóźniające, z których buduje się różne struktury sztucznych sieci neuronowych. Wszystkie układy są przedstawione w dwóch wersjach: układów CMOS pracujących w klasie A i w klasie AB, a następnie przebadane symulacyjnie za pomocą programu VLSISPICE COMPASS DA na minikomputerze SUNSPARC 20. Wyniki badań zostały zamieszczone wraz ze schematami i analizą porównawczą.

Słowa kluczowe: sieci neuronowe, układy CMOS, układy symulacyjne.

1. WPROWADZENIE

W ostatnich latach wraz z postępem technologii wytwarzania scalonych układów analogowych wzrosło zainteresowanie sztucznymi sieciami neuronowymi. Sieci neuronowe są systemami analogowymi, w których przetwarzanie informacji (sygnałów) jest wzorowane na funkcjach wykonywanych przez system nerwowy człowieka. Duża liczba licznie połączonych ze sobą neuronów pozwala na jednoczesne (równoległe) wykonywanie wielu operacji. Tę cechę równoległego przetwarzania pozwalającą człowiekowi rozwiązywać złożone problemy w stosunkowo krótkim czasie projektanci chcą wykorzystać stosując sztuczne sieci neuronowe. Żądany proces przetwarzania informacji (sygnałów) przez zadaną strukturę sieci uzyskuje sie poprzez odpowiedni dobór współczynników wagowych połączeń synaptycznych (synaps) służących do przesyłania sygnałów między neuronami i wartości progów działania (pobudzenia, zahamowania) poszczególnych neuronów w sieci. Najczęściej dobór współczynników przeprowadza się za pomocą jednej z procedur adaptacyjnego lub samoorganizującego wyznaczania tych wartości. Procedury te nazywane są uczeniem

się neuronów. Wśród dotychczasowych metod uczenia się neuronów w sieci wyróżnia się dwie strategie uczenia z nauczycielem i bez nauczyciela.

W strategii uczenia neuronów z nauczycielem znana jest postać sygnału wyjściowego neuronu, a dobór wag jest prowadzony, tak aby sygnał wyjściowy neuronu był najbliższy zadanemu sygnałowi. Dla całej sieci jest to proces optymalizacji polegający na wyznaczeniu współczynników wagowych wszystkich neuronów. Do najbardziej znanych algorytmów uczenia z nauczycielem zaliczane są algorytmy: z regułą perceptronu [1], z regułą DELTA [2] i propagacji wstecznej [3].

W strategii uczenia neuronów bez nauczyciela dobór wag odbywa się na podstawie obserwacji informacji zawartych w zbiorach wyjściowych oraz zbiorach uczących. W tej strategii sieć musi posiadać cechy samoorganizacji polegającej na wyszukaniu istniejących wzorców, regularnosci lub wydzielonych cech przetwarzanych sygnałów. Do najbardziej znanych algorytmów uczenia bez nauczyciela zaliczane są algorytmy: Hebba [4], konkurencyjne i samoorganizujące Kohonena [5]. Algorytmy Grossberga: INSTAR [6] oraz OUTSTAR [7] mogą realizować uczenie zarówno z nauczycielem jak i bez nauczyciela.

Współczesne sztuczne sieci neuronowe dalekie są od żywego pierwowzoru, a pomimo to pozwalają na rozwiązanie wielu trudnych problemów z zakresu rozpoznawania i przetwarzania obrazów [8 – 13], automatyki i robotyki [14 – 17], automatycznego sterowania obiektami ruchomymi [14, 18 – 19] czy filtracji sygnałów [8, 20 – 22].

Rozwiązywanie różnorodnych zagadnień za pomocą sztucznych sieci neuronowych może odbywać się na dwa sposoby programowy i układowy. W pierwszej metodzie wykorzystuje się tradycyjne komputery lub neurokomputery ogólnego przeznaczenia zbudowane z programowanych tablic neuroprocesorów. Realizacja sieci neuronowej polega na emulacji określonej sieci neuronowej i programowej symulacji algorytmu uczenia. Realizacje programowe sieci neuronowych są bogato opisane w literaturze anglojęzycznej [23 – 25], a także i polskiej [26 – 28].

W drugiej metodzie wykorzystuje się sieci neuronowe zbudowane w postaci scalonego układu analogowego umożliwiającego układowe przetwarzanie określonego algorytmu uczenia. Ten rodzaj układowej realizacji sieci neuronowych jest bardziej efektywny pod względem czasu przetwarzania niż realizacja programowa, lecz wymaga bardzo zaawansowanych technologii wytwarzania precyzyjnych układów analogowych VLSI. Rozwój programowych metod realizacji sieci neuronowych znacznie wyprzedził rozwój metod układowych. Rezultaty tego rozwoju stanowią teoretyczne podstawy do tworzenia układów scalonych realizujących zadany algorytm uczenia sieci o znanej strukturze, a także neurokomputerów realizujących równoległe przetwarzanie informacji w sieciach zbudowanych z tysięcy jednakowych neuroprocesorów.

W ostatnich latach wraz z intensywnym rozwojem technologii półprzewodnikowych wzrosło zainteresowanie układowymi realizacjami sieci neuronowych. Powstały pierwsze rozwiązania dla określonych algorytmów uczenia [29 – 34]. Szczególnie intensywnie rozwinęły się sieci komórkowe, wprowadzone przez L.O. Chua [35 – 36]. W większości rozwiązań układowych sieci neuronowe przetwarzają sygnały

napięciowe [37-43]. Ostatnie rozwiązania dotyczą także układów prądowych [44 – 51], ze względu na ich zalety w porównaniu do układów napięciowych w układach scalonych o napięciu zasilania obniżonym do 3,3 V. Do zalet należą: większa dynamika, większa szybkość przetwarzania i mniejsza moc pobierana ze źródła zasilania.

W pracy zostaną przedstawione układowe realizacje sztucznych sieci neuronowych pracujących w trybie prądowym.

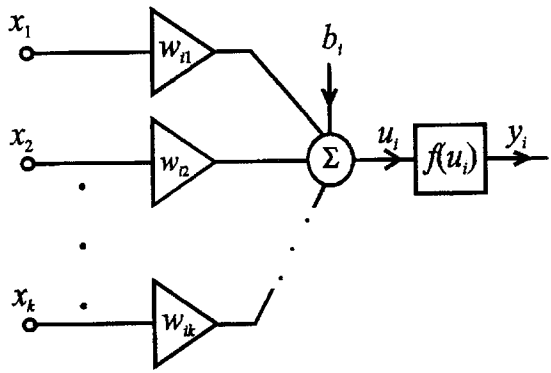
2. MODELE NEURONU I SIECI NEURONOWYCH

Sztuczne sieci neuronowe podobnie jak ich pierwowzór są zbudowane z dużej liczby neuronów. Neuron jest elementarną komórką przetwarzającą sygnały i pamiętającą przetwarzaną wielkość. W oparciu o zasadę działania rzeczywistej komórki neuronowej powstały różne modele neuronów. Wszystkie modele wykorzystują zasadę sumowania ważonych sygnałów pochodzących od neuronów połączonych synapsami i porównywania z wartością progową. W książce przyjęto model neuronu przedstawiony w postaci schematu blokowego na rysunku 1. Sygnał wyjściowy neuronu z rysunku 1 jest opisany następującymi zależnościami:

$$y_i = f(u_i) \tag{1}$$

$$u_i = \sum_{j=1}^{k} w_{ij} x_j + b_i, \tag{2}$$

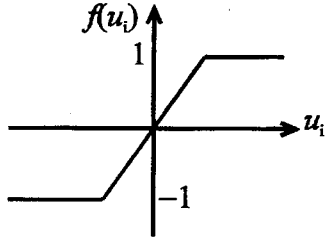
gdzie: wskaźnik i został użyty do oznaczenia neuronu w sieci, w_{ij} oznacza współczynnik wagowy synapsy łączącej wyjście j-tego neuronu z wejściem i-tego neuronu, b_i oznacza wartość progową neuronu i-tego nazywaną także polaryzacją, a $f(u_i)$ nieliniową funkcję aktywacji i-tego neuronu.



Rys. 1. Schemat blokowy komórki neuronowej

Analogowa funkcja aktywacji jest przedstawiona w postaci odcinkowo-liniowej na rysunku 2 i opisana zależnością

$$f(u_i)=\frac{1}{2}(\,|\,u_i+1\,|\,-\,|\,u_i-1\,|\,)=\begin{cases}1 & \text{dla} \quad u_i \geqslant 1\\ u_i & \text{dla} \quad -1 \leqslant u_i \leqslant 1\\ -1 & \text{dla} \quad u_i \leqslant -1\end{cases} \tag{3}$$



Rys. 2. Analogowa odcinkowo-liniowa funkcja aktywacji

Pojedyncze neurony zbudowane z synaps o różnych współczynnikach wagowych i nieliniowej funkcji aktywacji są połaczone ze sobą tworząc sieć neuronową. W zależności od rodzaju połączeń można wyróżnić kilka podstawowych typów sieci: jednokierunkową [52], rekurencyjną [53–55] i komórkową [35–36]. Przykłady sieci neuronowych są pokazane na rysunku 3.

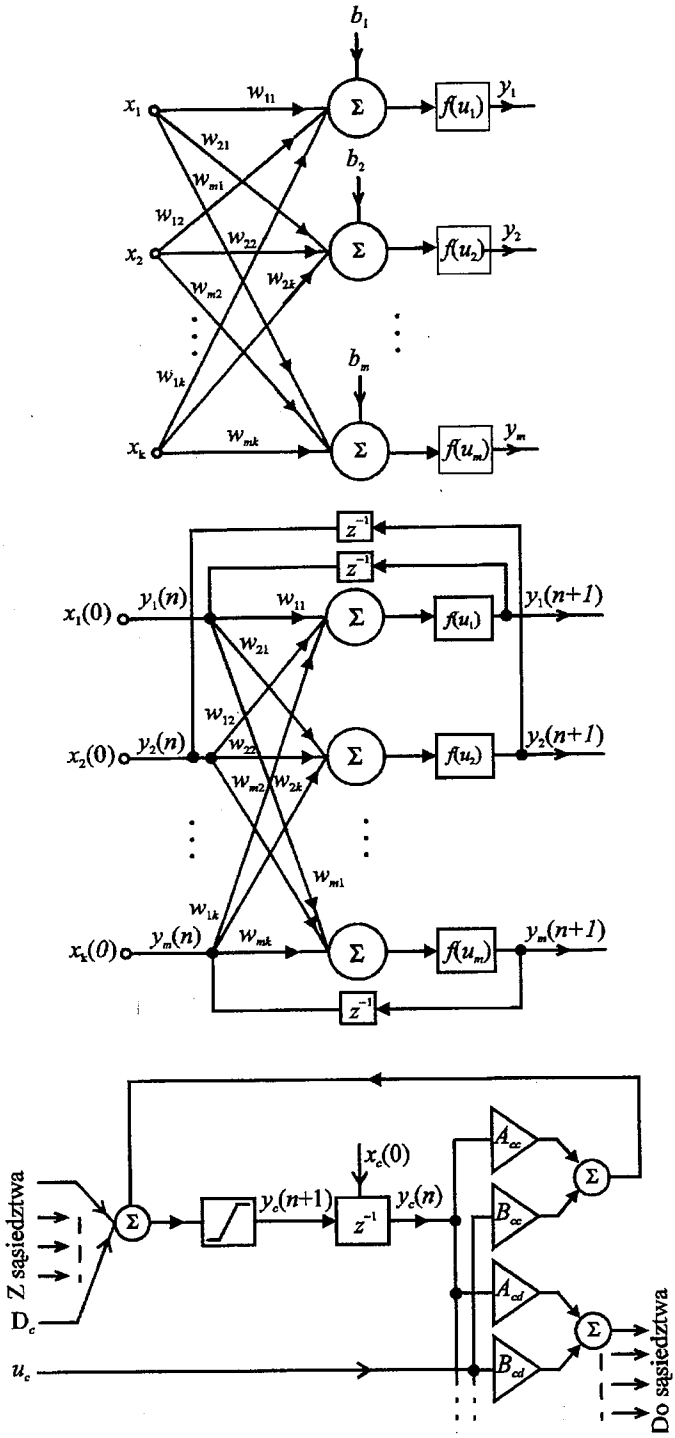
3. UKŁADOWE REALIZACJE SIECI NEURONOWYCH

Jak pokazano na rysunkach 2 i 3 pojedyncze neurony oraz sieci neuronowe są zbudowane z układów elementarnych takich jak: sumatory, układy realizujące nieliniową funkcję aktywacji, wzmacniacze i wzmacniacze o regulowanym wzmocnieniu realizujące połączenia synaptyczne, a sieci rekurencyjne i komórkowe dodatkowo z układów opóźniających o okres przełączania T_s.

Wszystkie elementarne układy przedstawione w pracy będą przetwarzały sygnały prądowe. Sumowanie sygnałów prądowych odbywa się zwykle w węźle wejściowym układu realizującego nieliniową funkcję aktywacji, stąd sumator prądów nie wymaga oddzielnego omówienia.

Wzmacniacze z regulowanym wzmocnieniem są stosowane w procesie uczenia sieci do realizacji doboru współczynników wagowych synaps. Współczynniki wagowe $w_{ij}=f(x_i^S)$ mogą być programowane, bowiem ich wartości są funkcją pewneo sygnału korygującego x_i^S nazywanego uczącym. Sygnałem uczącym w technice prądowej jest niezależnie od strategii uczenia sygnał prądowy oznaczany i_i^S. Sposób wytworzenia tego sygnału zalezy od przyjętej strategii uczenia. Sygnał ten zwykle jest kombinacją liniową np. sygnałów wyjściowych neuronów oraz sygnałów odniesienia i może być utworzony w układach zbudowanych ze wzmacniaczy i sumatorów.

Poniżej zostaną omówione podstawowe układy elementarne realizujące: analogową funkcję aktywacji, połączenia synaptyczne oraz układy opóźniające. Wszystkie układy elementarne będą przedstawiane w dwóch wersjach: układów pracujących w klasie A i w klasie AB. Zaletą układów pracujących w klasie A jest mała liczba

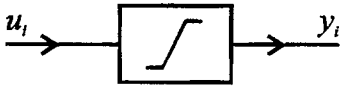


Rys. 3. Schematy blokowe neuronowej sieci, (a) jednowarstwowej jednokierunkowej, (b) jednowarstwowej rekurencyjnej, (c) komórkowej

elementów z których są zbudowane, a więc mała powierzchnia zajmowana w układzie scalonym, natomiast układów pracujących w klasie AB — mała moc pobierana ze źródła zasilania.

3.1. UKŁADY ANALOGOWEJ FUNKCJI AKTYWACJI

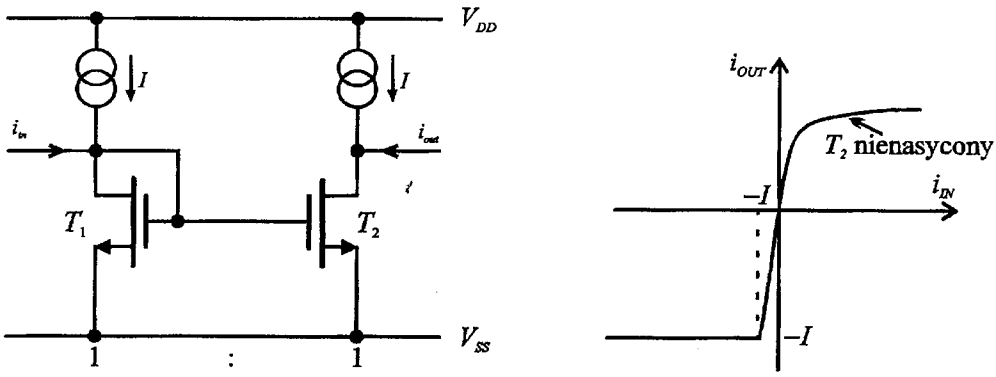
Analogowa funkcja aktywacji jest charakterystyczną przejściową układu. Symbol blokowy układu realizującego analogową liniowo-odcinkową funkcję aktywacji jest pokazany na rysunku 4.



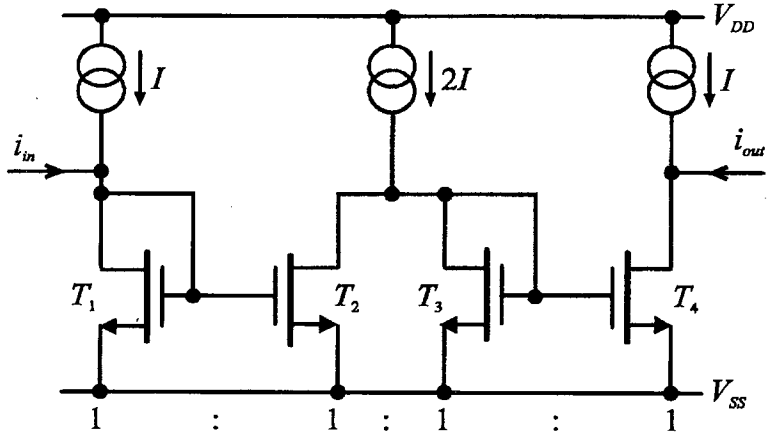
Rys. 4. Schemat blokowy układu analogowej odcinkowo-liniowej funckji aktywacji

3.1.1. Układy analogowej funkcji aktywacji pracujące w klasie A

Najprostsza struktura w technologii CMOS realizująca analogową funkcję aktywacji w technice prądowej jest pokazana na rysunku 5a. Układ ten jest zwierciadłem prądowym obciążonym jednakowymi źródłami prądowymi o wydajnosci I. Zasada pracy układu jest następująca. W zakresie prądu wejściowego (wypływającego) — $I < i_{in} < 0$ oba tranzystory pracują w nasyceniu i układ zachowuje się jak wzmacniacz o wzmocnieniu $A = 1$. Jeżeli $i_{in} \leqslant -I$ to układ zwierciadła prądowego jest odcięty już dla prądu $i_{in} = -I$ jak to pokazuje charakterystyka przejściowa z rysunku 5b. W zakresie dodatniego prądu wejściowego (wpływającego) $i_{in} > 0$ dla małych wartości prądu wejściowego oba tranzystory pracują w nasyceniu i układ zachowuje się jak zwierciadło prądowe. Zwiększenie prądu wejściowego powoduje przejście tranzystora T_2 z zakresu nasycenia do zakresu nienasycenia. Charakterystyka układu będzie nadal wznosić się, lecz z bardzo małym nachyleniem. Nie będzie ona tak ostro odcięta jak dla ujemnych prądów wejściowych.



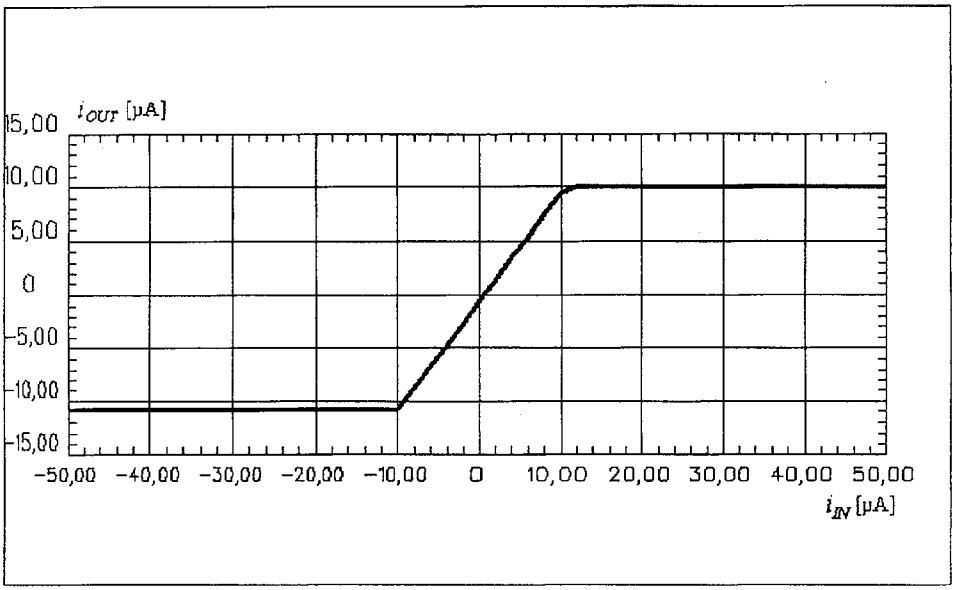
Rys. 5. Układ analogowej funkcji aktywacji wraz z jej charakterystyką



Rys. 6. Zmodyfikowany układ analogowej funkcji aktywacji

Podstawową wadą tego układu jest brak symetrii charakterystyki przejściowej tworzącej funkcję aktywacji. Wady tej nie posiada układ pokazany na rysunku 6 [45] będący kaskadowym połączeniem dwóch układów z rysunku 5. Pierwszy z układów obcina charakterystykę przejściową całego układu dla prądów wejściowych $i_{in} \leqslant -I$, a drugi dla prądów wejściowych $i_{in} \geqslant I$, bowiem prąd wejściowy drugiego układu identycznego jak pierwszy jest w przeciwfazie do prądu wejściowego pierwszego stopnia. Układ ten realizuje następującą funkcję aktywacji

$$i_{out} = f(i_{in}) = \frac{1}{2}(\,|\,i_{in}+I\,| - |\,i_{in}-I\,|\,). \tag{4}$$



Rys. 7. Charakterystyka przejściowa układu analogowej funkcji aktywacji pracującego w klasie A

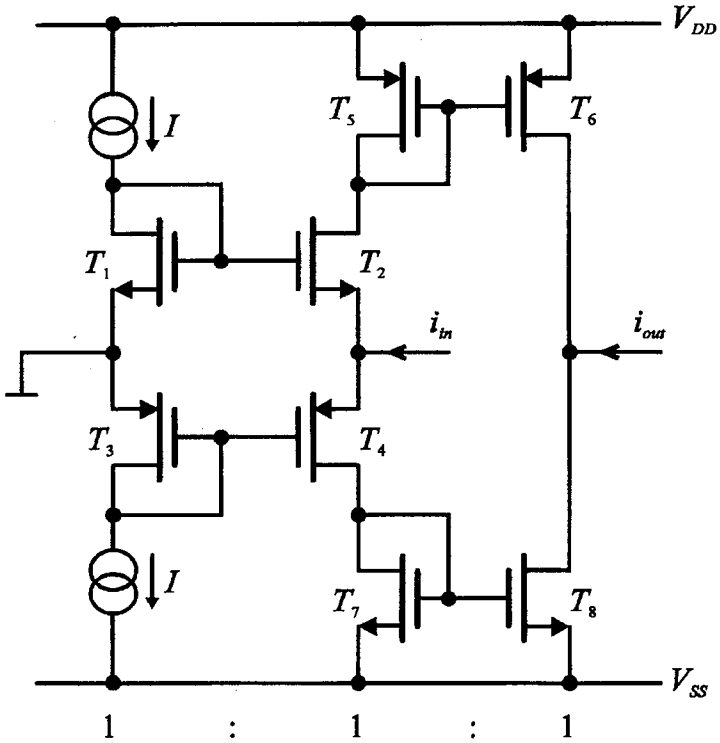
Zaletą układu jest jego prosta struktura i wynikająca stąd mała powierzchnia zajmowana w układzie scalonym. Wadą układu jest stały prąd pobierany ze źródła zasilania wynoszący $4I$ niezależnie od sygnału wejściowego. Możliwe jest obniżanie tego prądu jedynie do poziomu kilkakrotnie większego od prądów upływności tranzystorów MOS. Zmniejszenie prądów źródeł powoduje także zmniejszenie dynamiki prądu wejściowego, bowiem w klasie A prąd wejściowy nie może przekraczać prądu I. Budowa scalonych sieci neuronowych składających się z tysięcy albo nawet setek tysięcy neuronów wymaga zastosowania tej samej liczby układów realizujących funkcję aktywacji, a więc już tylko te układy wymagałyby dostarczenia prądu $4I$ dla każdego z neuronów oddzielnie nawet gdyby żaden z nich nie był aktualnie pobudzony sygnałem.

Układ przebadano symulacyjnie za pomocą programu VLSISPICE COMPASS DA na komputerze SUNSPARC20. Wszystkie tranzystory typu N mają te same rozmiary $W=3.6$ μm i $L=3.6$ μm, a tranzystory typu P $W=10$ μm i $L=3.6$ μm. Dla małych rozmiarów parametry transkonduktancyjne wszystkich tranzystorów są jednakowe, a współczynniki wzmocnienia prądowego w zakresie liniowym są równe i wynoszą $A=1$. Wyniki symulacji charakterystyki przejściowej układu są pokazane na rysunku 7.

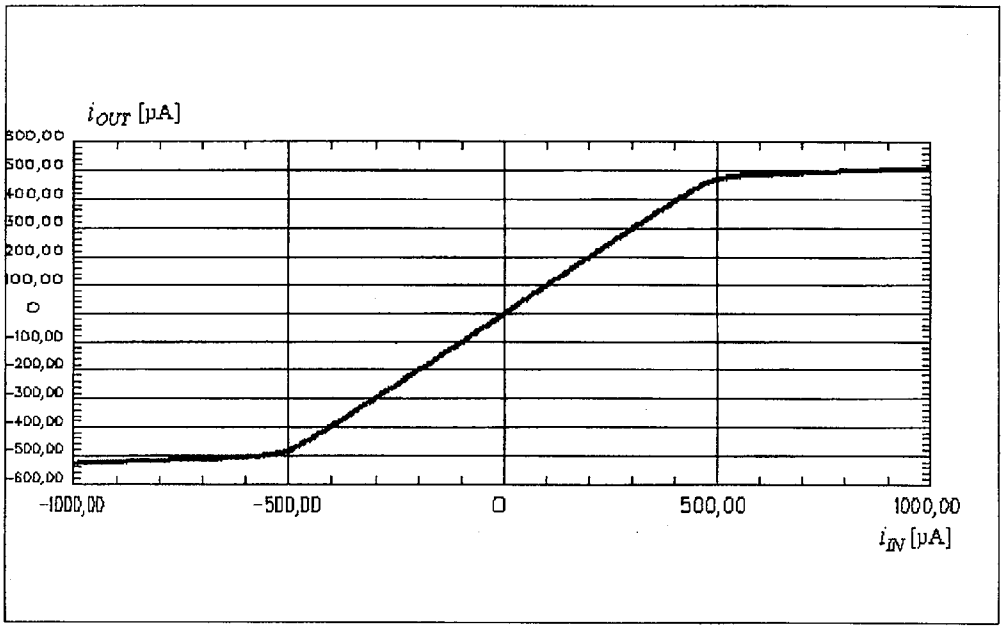
3.1.2. Układy analogowej funkcji aktywacji pracującej w klasie AB

Głównym ograniczeniem układu funkcji aktywacji pracującego w klasie A jest moc pobierana przez układ. W tej sytuacji poszukuje się rozwiązań redukujących pobór mocy ze źródła zasilania, szczególnie gdy neurony nie są pobudzone lub pobudzone sygnałami o małych wartościach prądu. Układem takim jest układ funkcji aktywacji pracujący w klasie AB pokazany na rysunku 8 [49]. Układ ten zawiera konwejor prądu wykonany w technologii CMOS [56]. Zaletą układu jest to że prądy wejściowy i wyjściowy mogą przekraczać prąd źródła odniesienia I. Stąd prąd źródła I może zostać ustalony na możliwie niskim poziomie, wielokrotnie niższym niż w układach pracujących w klasie A, bez konieczności ograniczania dynamiki układu. Mały prąd źródła odniesienia zapewnia mały pobór ze źródła zasilania, gdy komórka nie jest pobudzana sygnałem wejściowym. Wynosi on $2I+I$. Dodatkową zaletą układu jest to, że jedno źródło prądowe I może być użyte jako źródło odniesienia jednocześnie dla wielu komórek. Powoduje to dalsze zmniejszenie mocy pobieranej ze źródła zasilania w porównaniu z komórkami pracującymi w klasie A. Prąd pobierany ze źródła zasilania przypadający na jeden układ funkcji aktywacji nie pobudzony sygnałem wejściowym wynosi $2I+\frac{I}{n}$, gdzie n jest liczbą układów funkcji aktywacji jednocześnie zasilanych układem polaryzacji ze źródłami prądowymi o wydajności I.

Układ funkcji aktywacji pracujący w klasie AB przebadano symulacyjnie w tych samych warunkach co układ standardowy. Wyniki badań charakterystyki przejściowej są pokazane na rysunku 9. Przedstawiona charakterystyka różni się od idealnej nieliniowej funkcji aktywacji komórki neuronowej, bowiem w zakresie dużych



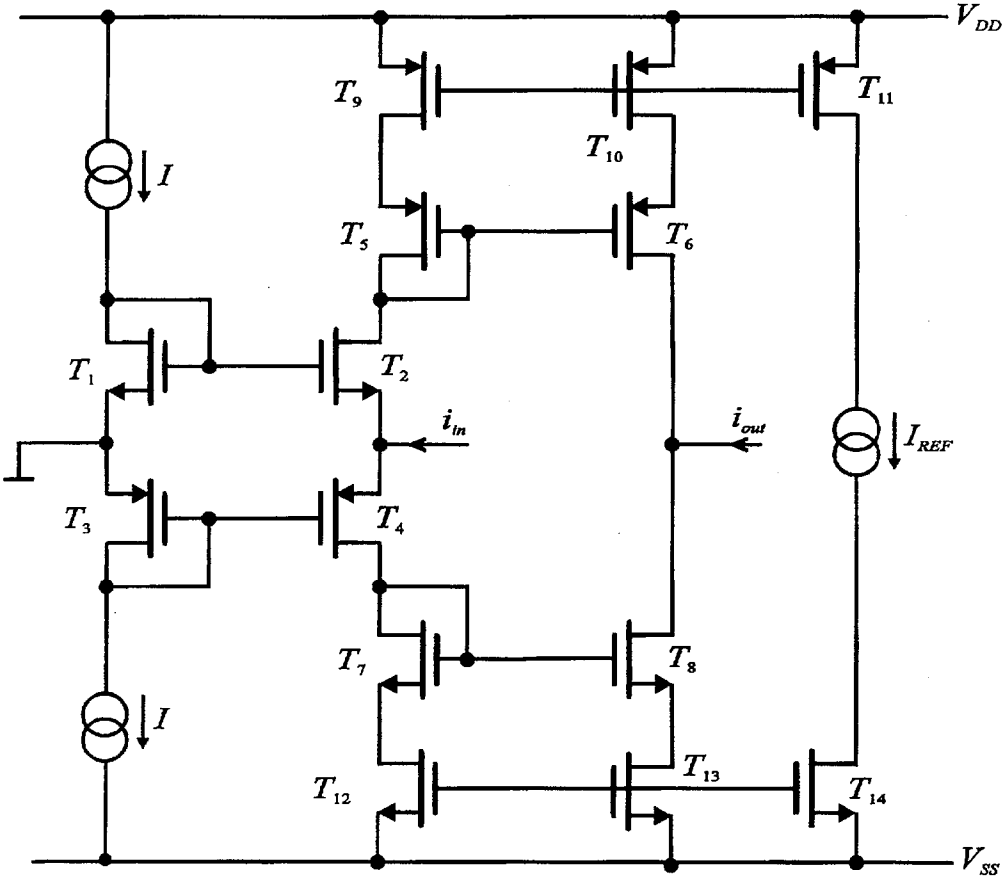
Rys. 8. Układ analogowej funkcji aktywacji pracujący w klasie AB



Rys. 9. Charakterystyka przejściowa układu funkcji aktywacji pracującego w klasie AB

bezwzględnych wartości prądu wejściowego nie jest równoległa do osi prądu wejściowego i_{IN}. Jest to spowodowane przejściem punktu pracy odpowiednio tranzystorów T_2 i T_4 z zakresu nasycenia do zakresu nienasycenia, w którym nie następuje ograniczenie prądu wejściowego lecz jedynie zmniejszenie szybkości narastania tego prądu. Stała wartość prądu wyjściowego może być uzyskana poprzez włączenie dodatkowego układu ograniczania prądu wyjściowego.

Układ funkcji aktywacji z ograniczaniem prądu wyjściowego jest pokazany na rysunku 10 [49]. Układ ograniczania prądu ujemnego i dodatniego jest zbudowany ze źródła prądowego I_{REF} oraz tranzystorów $T_9 \div T_{11}$ i $T_{12} \div T_{14}$. Układ odniesienia ze źródłem I_{REF} oraz tranzystorami T_{11} i T_{14} może być użyty jednocześnie do ograniczania charakterystyk funkcji aktywacji wielu układów. Prąd pobierany ze źródła zasilania przypadający na jeden układ funkcji aktywacji gdy $i_{in}=0$ wynosi $2I+\frac{I}{n}+\frac{I_{REF}}{m}$. Dwa pierwsze składniki sumy są takie same jak w układzie bez ogranicznika prądu, a trzeci określa wzrost prądu pobieranego ze źródła zasilania



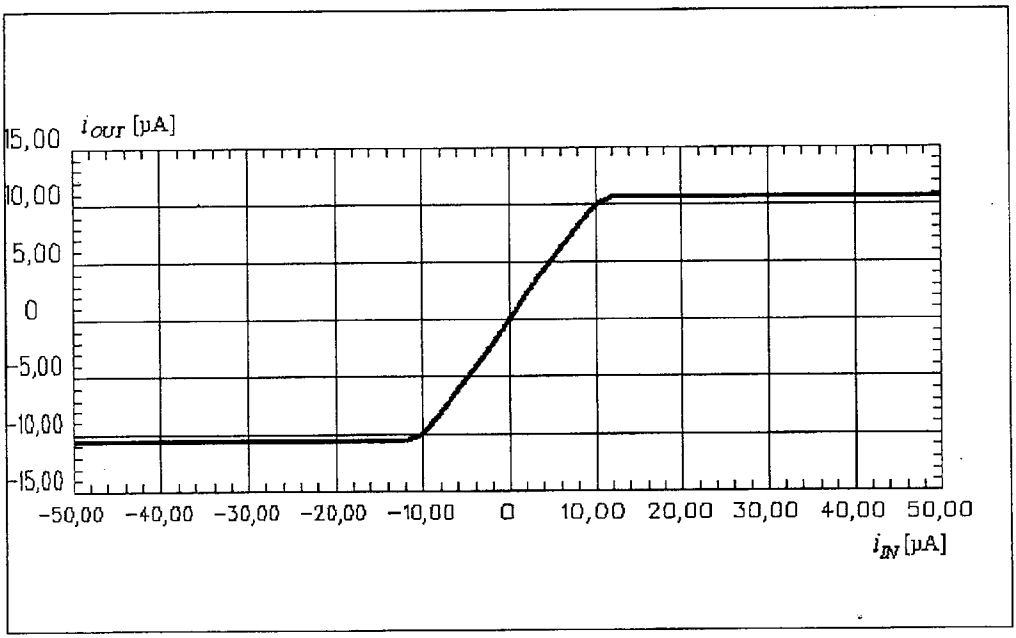
Rys. 10. Układ analogowej funkcji aktywacji pracujący w klasie AB z ograniczaniem prądu wyjściowego

przypadającego na jeden układ spowodowany włączeniem układu ograniczającego jednocześnie charakterystyki przejściowe m układów aktywacji.

Układ funkcji aktywacji pracujący w zakresie liniowym pobudzony prądem $0 < |i_{in}| < I_{REF}$, pobiera ze źródła zasilania prąd, przypadający na jeden układ, równy $2|i_{in}| + \frac{I}{n} + \frac{I_{REF}}{m}$.

Układ funkcji aktywacji pobudzony prądem $|i_{in}| \geqslant I_{REF}$, pobiera prąd, przypadający na jeden układ, równy $2I_{REF} + \frac{I}{n} + \frac{I_{REF}}{m}$. Prąd ten jest dla $n \gg 1$ i $m \gg 1$ prawie dwukrotnie mniejszy od prądu przypadającego na układ pracujący w klasie A. Wynika stąd wniosek, że prąd pobierany ze źródła zasilania przez układ funkcji aktywacji pracujący w klasie AB jest zawsze mniejszy, a w przypadku zerowych i małych wartości prądu wejściowego jest pomijalnie mały w porównaniu z prądem pobieranym ze źródła zasilania przez układ funkcji aktywacji o tej samej dynamice pracujący w klasie A.

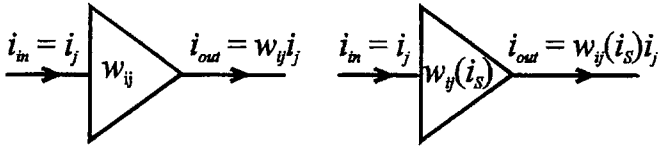
Układ z rysunku 10 przebadano symulacyjnie w tych samych warunkach co układ standardowy. Wyniki badań charakterystyki przejściowej układu są pokazane na rysunku 11. W zakresie dużych bezwzględnych prądów wejściowych jest ona równoległa do osi prądu wejściowego.



Rys. 11. Charakterystyka przejściowa układu analogowej funkcji aktywacji pracującego w klasie AB z ograniczaniem prądu wyjściowego

3.2. UKŁADY REALIZUJĄCE SYNAPSĘ

Neurony są połączone ze sobą synapsami o różnych współczynnikach wagowych w_{ij}. Synapsa jest realizowana za pomocą wzmacniacza prądowego, którego wzmocnienie reprezentuje współczynnik wagowy. Symbol blokowy synapsy o stałym współczynniku wagowym jest pokazany na rysunku 12a. Bardzo ważną cechą sieci neuronowych jest możliwość przestrajania (programowania) współczynników wagowych i to najlepiej sygnałem analogowym. W technice prądowej jest to sygnał prądowy, stąd $w_{ij}=f(i_S)$. Symbol blokowy synapsy o współczynniku wagowym przestrajanym sygnałem prądowym jest pokazany na rysunku 12b. Przestrajanie umożliwia realizację różnych algorytmów uczenia sieci neuronowej. Standardowy wzmacniacz prądowy CMOS posiada wzmocnienie proporcjonalne do stosunków $\frac{W}{L}$ odpowiednich tranzystorów. W ten sposób można zrealizować jedynie synapsę o stałym współczynniku wagowym, bowiem rozmiary tranzystorów nie mogą być zmieniane w trakcie pracy układu. W rozwiązaniach układowych umożliwiających przestrajanie wzmocnienia prądowego wzmacniacza używa się sygnału cyfrowego lub analogowego. Układ przestrajania wzmocnienia sygnałem cyfrowym jest prostszy, lecz zwykle dodatkowo wymagany jest przetwornik a/c, gdyż sygnał korygujący w większości algorytmów uczących jest analogowy. Układ przestrajania wzmocnienia sygnałem analogowym jest bardziej złożony, ale nie potrzebny jest przetwornik a/c.

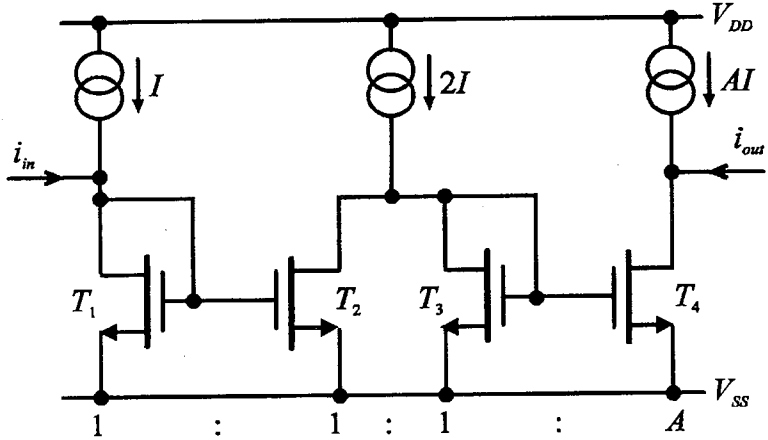


Rys. 12. Symbole blokowe synapsy, a) o stałym współczynniku wagowym, b) o przestrajanym współczynniku wagowym

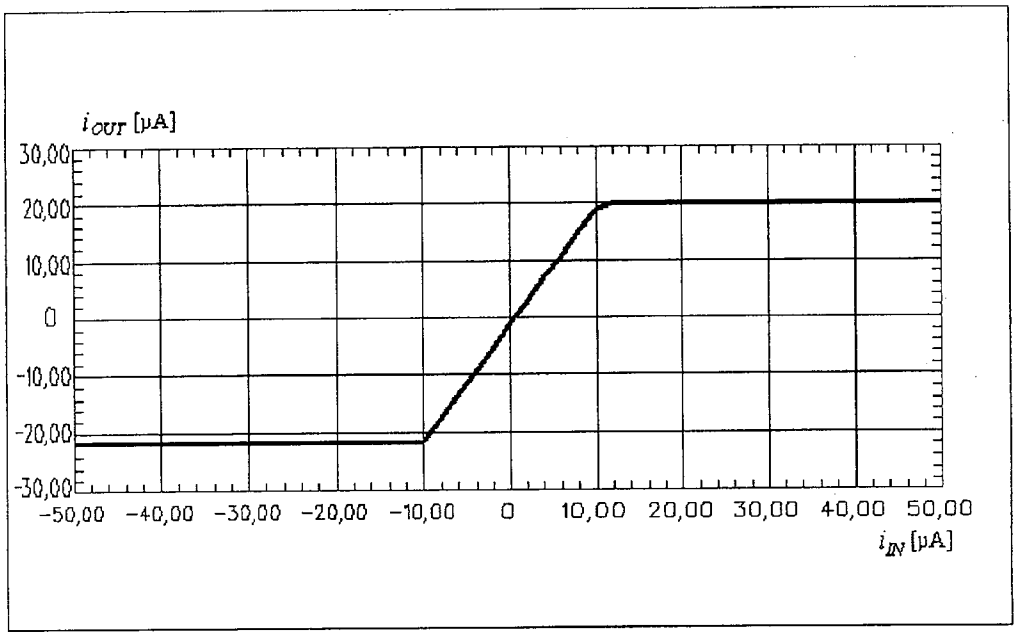
3.2.1. Układy synapsy o stałym współczynniku wagowym pracujące w klasie A

Układ synapsy o stałym współczynniku wagowym pracujący w klasie A jest pokazany na rysunku 13. Posiada on tę samą strukturę co układ nieliniowej funkcji aktywacji z rysunku 6, lecz wykorzystuje się w nim tylko obszar liniowej charakterystyki układu. Nachylenie charakterystyki przejściowej układu w zakresie liniowym jest dodatkowo kształtowane poprzez zastąpienie w drugim stopniu drugiego źródła prądowego o wydajnosci I źródłem o wydajności AI i A-krotne zwiększenie stosunku $\frac{W}{L}$ tranzystora T_4 w stosunku do tranzystora T_3. W zakresie prądu wejściowego $-I<i_{in}<I$ oba tranzystory pracują w nasyceniu i układ zachowuje się jak wzmacniacz o wzmocnieniu A.

Układ przebadano symulacyjnie za pomocą programu VLSISPICE COMPASS DA na komputerze SUNSPARC20. Parametry transkonduktancyjne tranzystorów $T_1 \div T_3$ są takie same jak w przypadku układu nieliniowej funkcji aktywacji, a tranzystora T_4 zostały dobrane tak, aby współczynnik wzmocnienia prądowego układu w zakresie liniowym wynosił $A=2$. Wyniki symulacji charakterystyki przejściowej układu są pokazane na rysunku 14.



Rys. 13. Układ synapsy o stałym współczynniku wagowym pracującej w klasie A

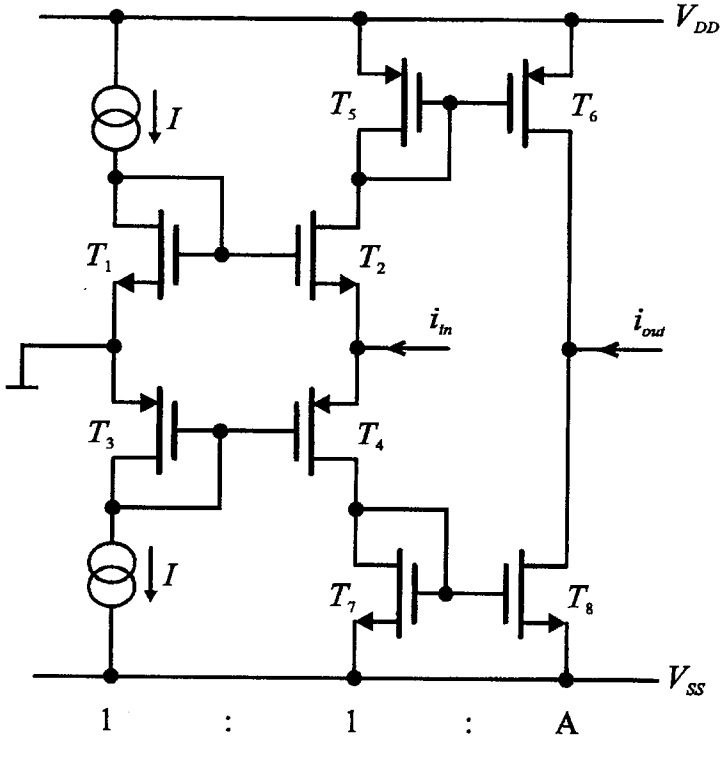


Rys. 14. Charakterystyka przejściowa synapsy o stałym współczynniku wagowym pracującej w klasie A

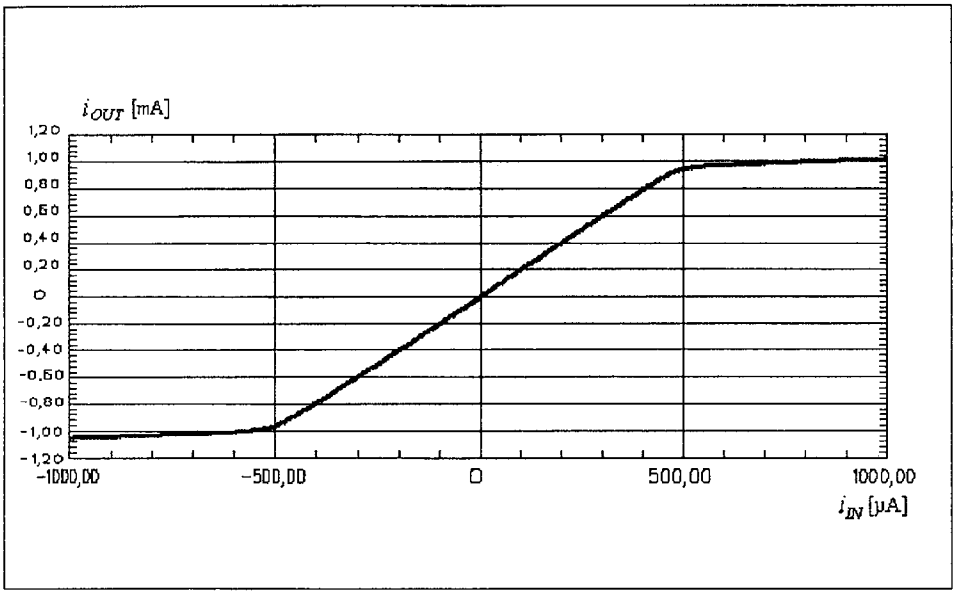
3.2.2. Układy synapsy o stałym współczynniku wagowym pracujące w klasie AB

Układ synapsy o stałym współczynniku wagowym pracujący w klasie AB jest pokazany na rysunku 15. Analogicznie jak w przypadku układu pracującego w klasie A, układ synapsy posiada tę samą strukturę co układ nieliniowej funkcji aktywacji z rysunku 8, lecz wykorzystuje się w nim tylko obszar liniowej charakterystyki układu. Nachylenie charakterystyki przejściowej układu w zakresie liniowym jest dodatkowo kształtowane poprzez A-krotne zwiększenie stosunku $\frac{W}{L}$ tranzystora T_6 w stosunku do tranzystora T_5 i tranzystora T_8 w stosunku do tranzystora T_7. W zakresie prądu wejściowego $-I < i_{in} < I$ oba tranzystory pracują w nasyceniu i układ zachowuje się jak wzmacniacz o wzmocnieniu A.

Układ przebadano symulacyjnie za pomocą programu VLSISPICE COMPASS DA na komputerze SUNSPARC20. Parametry transkonduktancyjne tranzystorów $T_1 \div T_5$ i T_7 są takie same jak w przypadku nieliniowej funkcji aktywacji, a tranzystorów T_6 i T_8 zostały dobrane tak, aby współczynnik wzmocnienia prądowego układu w zakresie liniowym wynosił $A=2$. Wyniki symulacji charakterystyki przejściowej układu są pokazane na rysunku 16.



Rys. 15. Układy synapsy o stałym współczynniku wagowym pracującej w klasie AB



Rys. 16. Charakterystyka przejściowa synapsy o stałym współczynniku wagowym pracującej w klasie AB

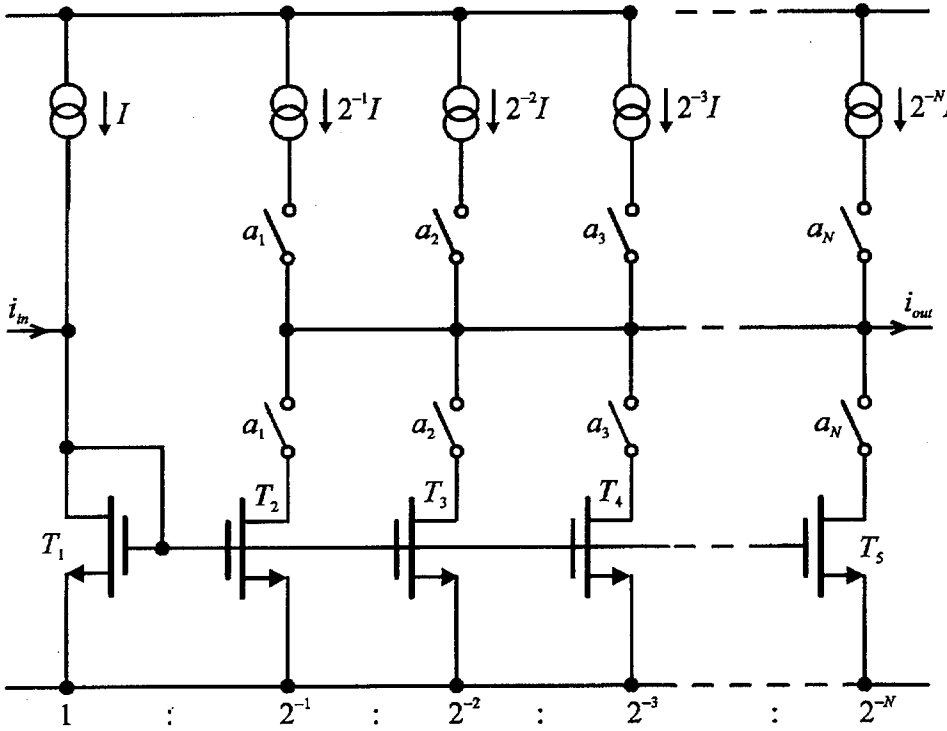
3.2.3. Układy przestrajanej synapsy pracującej w klasie A

Układ pracujący w klasie A o wzmocnieniu prądowym przestrajanym sygnałem cyfrowym $i_S(A)$, gdzie $A = a_1a_2a_3...a_N$ jest pokazany na rysunku 17. Układ ten sumuje na wyjściu prądy wzmacniaczy, których wzmocnienia są dobrane za pomocą rozmiarów tranzystorów i tworzą tablicę binarnych współczynników wagowych. Prąd wyjściowy układu jest opisany następującą zależnością

$$i_{out} = i_{in} \sum_{i=1}^{N} a_i 2^{-i} = f(a_1a_2a_3...a_N) i_{in} = f(A) i_{in} = f(i_S) i_{in}. \tag{5}$$

Wzmocnienie sygnału w układzie jest więc funkcją cyfrowego sygnału sterującego. Układ sterowania jest typowym przetwornikiem c/a.

Przestrajanie analogowe wzmocnienia prądowego układu pracującego w klasie A uzyskuje się za pomocą wzmacniaczy różnicowych, w których sygnał prądowy i_S zmienia prąd źródła pary różnicowej. Układ analogowego przestrajania wzmocnienia wykorzystujący układ Gilberta jest pokazany na rysunku 18 [50]. Tranzystory $T_1 \div T_{12}$ mają ten sam parametr transkonduktancyjny β, tranzystory T_{13} i T_{14} parametr $\beta' = A\beta$, tranzystory T_{n1} i T_{n2} parametr β_n, a tranzystory T_{p1} i T_{p2} parametr β_p. Wszystkie tranzystory mają te same napięcia progowe V_T. Jeżeli $|2i_{D5}| > \beta v_{id}$ i $|2i_{D8}| > \beta v_{id}$ gdzie v_{id} jest napięciem różnicowym to prądy drenów tranzystorów $T_1 \div T_8$ są opisane następującymi zależnościami



Rys. 17. Układ synapsy w klasie A przestrajanej sygnałem cyfrowym

$$i_{D1}=\frac{i_{D5}}{2}+\frac{i_{D5}v_{id}}{2}\left(\frac{\beta}{i_{D5}}-\frac{\beta^2 v_{id}^2}{4i_{D5}^2}\right)^{\frac{1}{2}},\qquad i_{D2}=\frac{i_{D5}}{2}-\frac{i_{D5}v_{id}}{2}\left(\frac{\beta}{i_{D5}}-\frac{\beta^2 v_{id}^2}{4i_{D5}^2}\right)^{\frac{1}{2}} \tag{6}$$

$$i_{D3}=\frac{i_{D8}}{2}-\frac{i_{D8}v_{id}}{2}\left(\frac{\beta}{i_{D8}}-\frac{\beta^2 v_{id}^2}{4i_{D8}^2}\right)^{\frac{1}{2}},\qquad i_{D4}=\frac{i_{D8}}{2}-\frac{i_{D8}v_{id}}{2}\left(\frac{\beta}{i_{D8}}-\frac{\beta^2 v_{id}^2}{4i_{D8}^2}\right)^{\frac{1}{2}} \tag{7}$$

$$i_{D5}=i_{D6}=I,\qquad i_{D8}=i_{D7}=I+i_S\,. \tag{8}$$

Prąd wyjściowy układu wynosi

$$i_{out}=i_{D14}-i_{D13}=Ai_{D11}-Ai_{D12}=A(i_{D9}-i_{D12})=A(i_{D1}+i_{D3}-i_{D2}-i_{D4}). \tag{9}$$

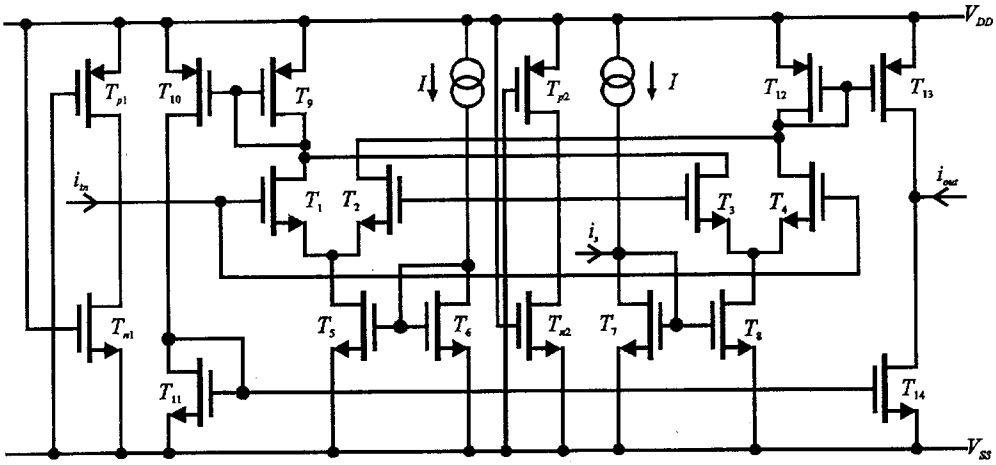
Uwzględniając zależności (6), (7) i (9) otrzymuje się

$$i_{out}=i_{D5}v_{id}\left(\frac{\beta}{i_{D5}}-\frac{\beta^2 v_{id}^2}{4i_{D5}^2}\right)^{\frac{1}{2}}-i_{D8}v_{id}\left(\frac{\beta}{i_{D8}}-\frac{\beta^2 v_{id}^2}{4i_{D8}^2}\right)^{\frac{1}{2}}. \tag{10}$$

Napięcie różnicowe można wyznaczyć z następujących równań

$$v_{id}=v_{DSn2}-v_{DSn1} \tag{11}$$

$$i_{Dp1}+i_{in}^{*}=i_{Dn1} \tag{12}$$



Rys. 18. Układ synapsy w klasie AB przestrajanej sygnałem analogowym

$$i_{Dp2} = i_{Dn2}\,. \tag{13}$$

Równania (12) i (13) można przekształcić do postaci

$$A_p(1+\lambda_p v_{SDp1}) + i_{in} = A_n(1+\lambda_n v_{DSn1}) \tag{14}$$

$$A_p(1+\lambda_p v_{SDp2}) = A_n(1+\lambda_n v_{DSn2}) \tag{15}$$

gdzie λ_n i λ_p są współczynnikami modulacji szerokości kanału tranzystorów typu N i P, a współczynniki A_n i A_p są wyrażone wzorami

$$A_n = \frac{\beta_n}{2}(\,|\,V_{DD} - V_{SS} - V_{Tn}\,|\,)^2 \tag{16}$$

$$A_p = \frac{\beta_p}{2}(\,|\,V_{SS} - V_{DD}\,| - |\,V_{Tp}\,|\,)^2 \tag{17}$$

oraz zachodzi równość

$$v_{SDp1} + v_{DSn1} = v_{SDp2} + v_{DSn2} = V_{DD} - V_{SS}\,. \tag{18}$$

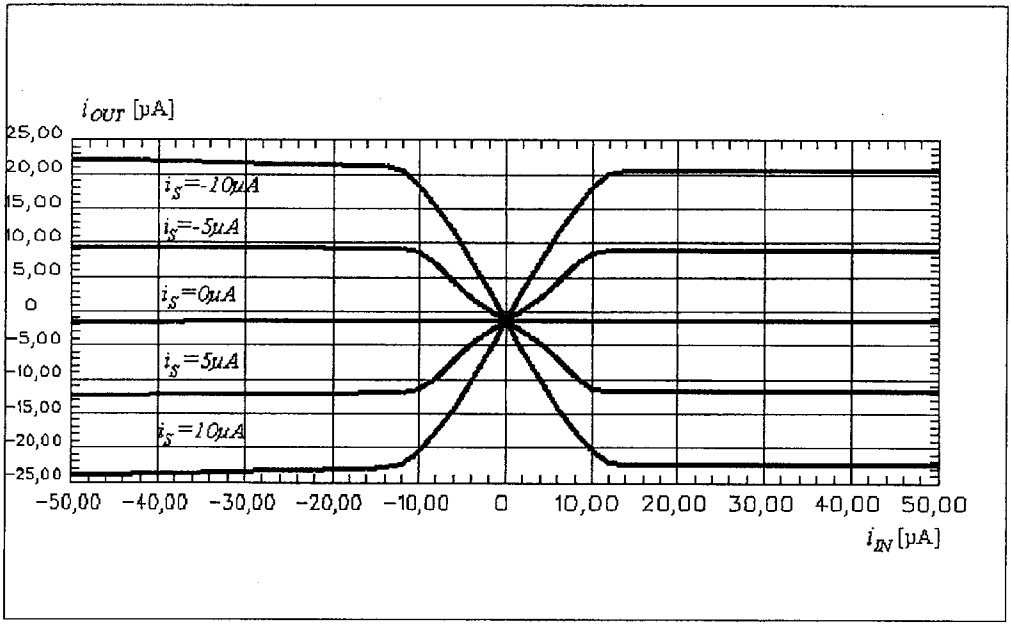
Uwzględniając zależności (14 – 18) w równaniu (11) dla $V_{DD} = |\,V_{SS}\,|$ otrzymuje się

$$v_{id} = -\frac{i_{in}}{A_n\lambda_n + A_p\lambda_p}\,. \tag{19}$$

Podstawiając równania (8) i (19) do równania (10) dla małych wartości prądu i_{in} oraz i_S otrzymuje się

$$i_{out} \cong -A\sqrt{\frac{\beta}{2I}}\,i_S v_{id} = \sqrt{\frac{\beta}{2I}}\,\frac{A i_S}{A_n\lambda_n + A_p\lambda_p}\,i_{in}\,. \tag{20}$$

Z zależności (20) wynika, że dla małych wartości prądu wejściowego w zakresie liniowej charakterystyki wzmocnienie układu jest proporcjonalne do prądu sterującego i_S. Układ zbadano symulacyjnie dla tych samych warunków co układy funkcji aktywacji. Rodzina charakterystyk przejściowych układu w funkcji sygnału i_S jest pokazana na rysunku 19.



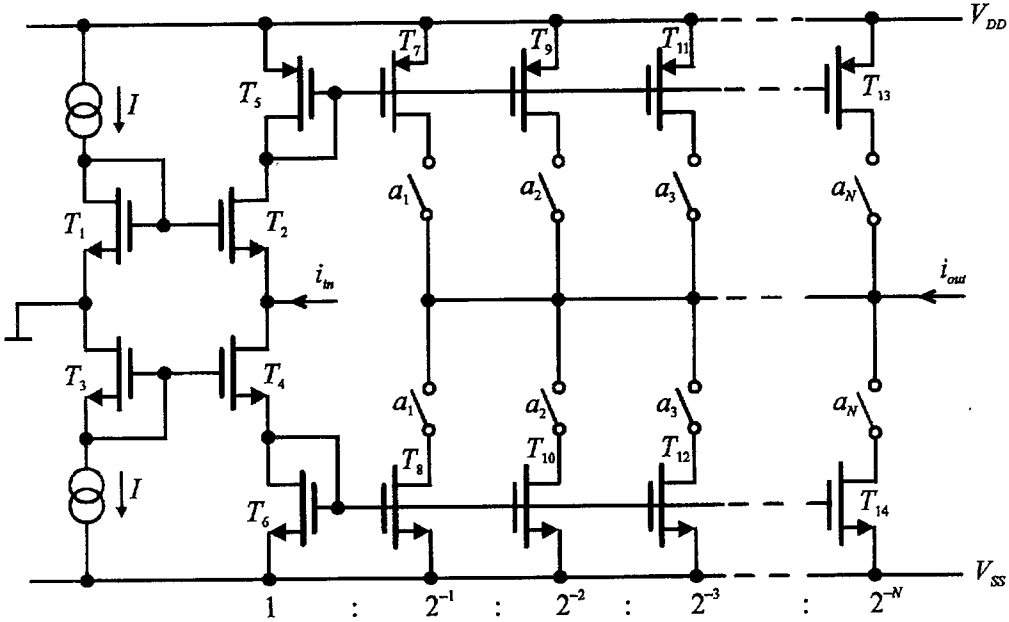
Rys. 19. Wyniki badań symulacyjnych synapsy w klasie A przestrajanej sygnałem analogowym

Dynamika prądu układów pracujących w klasie A zależy od prądu źródła I. Zwiększenie dynamiki, a więc i prądu I powoduje zwiększenie prądu pobieranego ze źródła zasilania niezależnie od sygnału wejściowego i_{in}. Dynamikę prądu niezależną od prądu źródła I otrzymuje się w układach pracujących w klasie AB.

3.2.4. Układy przestrajanej synapsy pracujące w klasie AB

Układ pracujący w klasie AB o przestrajanym wzmocnieniu prądowym sygnałem cyfrowym $i_S(A)$, gdzie $A=a_1a_2a_3...a_N$ jest pokazany na rysunku 20. Układ ten podobnie jak układ pracujący w klasie A sumuje prądy wyjściowe wzmiacniaczy, których wzmocnienia tworzą tablicę binarnych współczynników wagowych. Prąd wyjściowy układu pracującego w klasie AB tak samo jak układu pracującego w klasie A jest opisany zależnością (5).

Przestrajanie analogowe wzmocnienia prądowego układów pracujących w klasie AB uzyskuje się wykorzystując: właściwości funkcji kwadratowej opisującej charakterystykę wielkosygnałową tranzystora MOS lub właściwości wzmacniacza różnicowego, w którym za pomocą sygnału analogowego można regulować prąd źródła pary różnicowej.



Rys. 20. Układ synapsy w klasie AB przestrajanej sygnałem cyfrowym

Układ analogowego przestrajania wzmocnienia wykorzystujący właściwości funkcji kwadratowej jest pokazany na rysunku 21 [51]. Tranzystory $T_1 \div T_6$, $T_{d1} \div T_{d3}$, T_{d5}, T_{g1}, $T_{g3} \div T_{g5}$ mają ten sam parametr transkonduktancyjny β, a tranzystory T_{d4}, T_{d6}, T_{g2} i T_{g6} parametr $\beta' = A\beta$. Wszystkie parametry mają te same napięcia progowe V_T. Prądy drenów tranzystorów $T_3 \div T_6$ w nasyceniu oraz prąd wyjściowy są opisane następującymi zależnościami

$$i_{D3} = \frac{\left(I - \frac{i_{in} - i_S}{4}\right)^2}{I}, \qquad i_{D4} = \frac{\left(I + \frac{i_{in} - i_S}{4}\right)^2}{I}, \quad \text{dla} \quad |i_{in} - i_S| \leqslant 4I \tag{21}$$

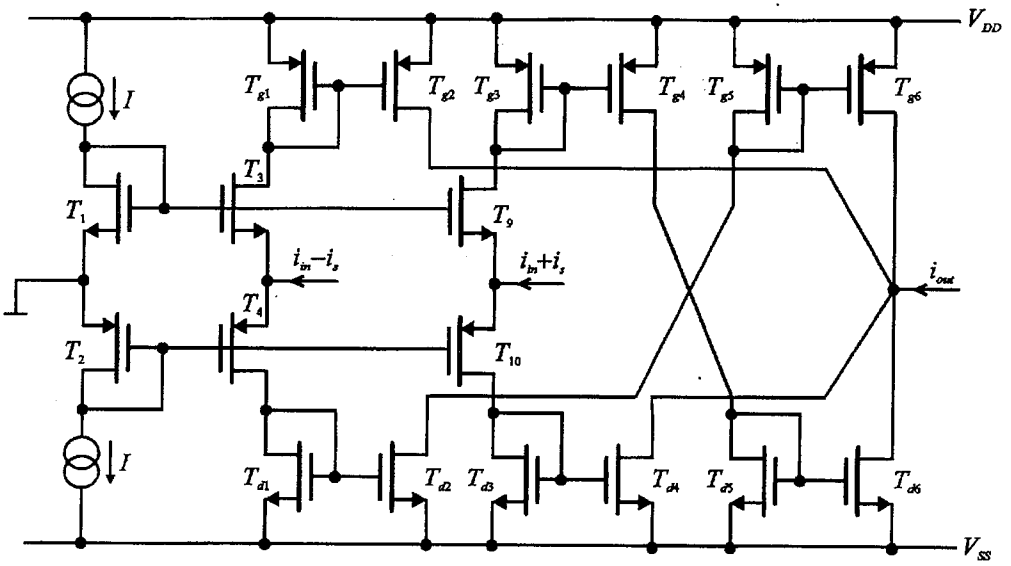
$$i_{D5} = \frac{\left(I - \frac{i_{in} + i_S}{4}\right)^2}{I}, \qquad i_{D6} = \frac{\left(I + \frac{i_{in} + i_S}{4}\right)^2}{I}, \quad \text{dla} \quad |i_{in} + i_S| \leqslant 4I \tag{22}$$

$$i_{out} = i_{Dd4} + i_{Dd6} - i_{Dg2} - i_{Dg6} = A(i_{D6} + i_{D5} - i_{D3} - i_{D4}). \tag{23}$$

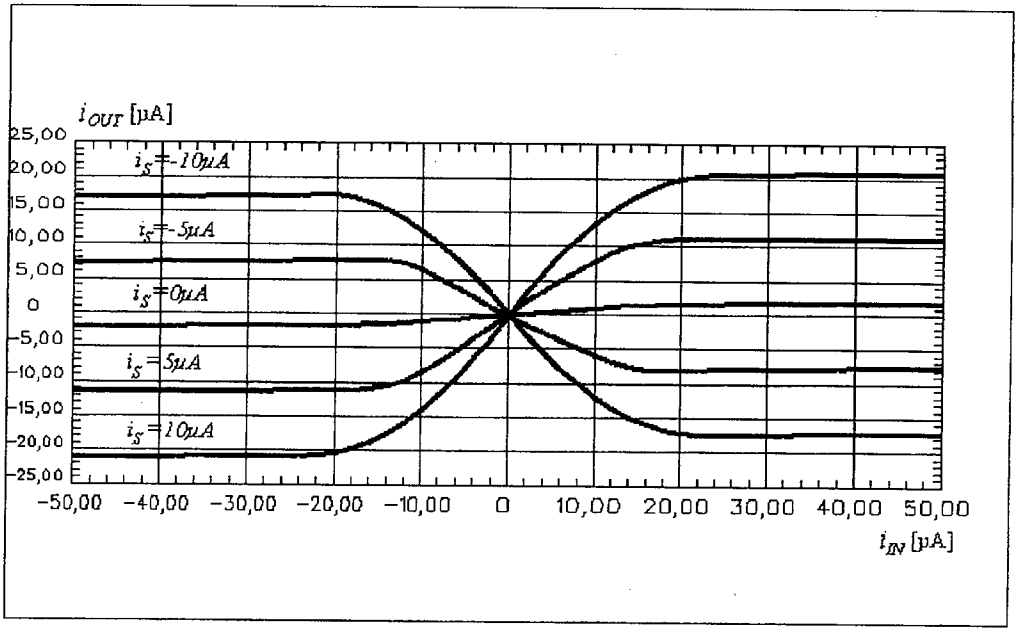
Podstawiając zależności (21) i (22) do równania (23) otrzymuje się

$$i_{out} = \frac{A i_S}{2I} i_{in}. \tag{24}$$

Z zależności (24) wynika, że dla małych wartości sygnału wejściowego w zakresie liniowej charakterystyki wzmocnienie układu jest proporcjonalne do prądu sterujące-



Rys. 21. Układ synapsy w klasie AB przestrajanej sygnałem analogowym z wykorzystaniem właściwości funkcji kwadratowej



Rys. 22. Wyniki badań symulacyjnych układu synapsy w klasie AB przestrajanej sygnałem analogowym z wykorzystaniem właściwości funkcji kwadratowej

go i_S. Układ zbadano symulacyjnie dla tych samych warunków co układy funkcji aktywacji. Rodzina charakterystyk przejściowych układu w funkcji sygnału i_S jest pokazana na rysunku 22.

Układ analogowego przestrajania wzmocnienia prądowego zbudowany z dwóch wzmacniaczy różnicowych pracujących w klasie AB, w których sygnał przestrajający i_S zmienia prądy źródeł obu par różnicowych wpływając w ten sposób na charakterystykę całego układu jest pokazany na rysunku 23 [51]. Tranzystory T_{g7}, T_{g8}, T_{n2} i T_{n4} oraz T_{d7}, T_{d8}, T_{p2} i T_{p4} sa przetwornikami prądu w napięcie. Zamieniają prądowy sygnał wejściowy na różnicowe napięcia wejściowe obu par różnicowych. Pozostałe w układzie tranzystory tworzą zwierciadła prądowe. Są one wykorzystane do zsumowania i wzmocnienia prądów wyjsciowych obu par w celu uformowania sygnału wyjściowego programowanej synapsy. Tranzystory T_{g2}, T_{g3}, $T_{g7} \div T_{g10}$, T_{d2}, T_{d3}, $T_{d7} \div T_{d10}$, $T_{n1} \div T_{n4}$ i $T_{p1} \div T_{p4}$ mają ten sam parametr transkonduktancyjny β, a tranzystory T_{g1}, $T_{g4} \div T_{g6}$ i $T_{d4} \div T_{d6}$ parametr $\beta' = A\beta$. Wszystkie tranzystory mają te same napięcia progowe V_T. Prądy tranzystorów T_{n4}, T_{p4}, T_{n3}, T_{p3} oraz wyjściowy są opisane następującymi zależnościami

$$i_{Dn4} = \frac{\left(I - \frac{i_{in}}{4}\right)^2}{I}, \qquad i_{Dp4} = \frac{\left(I + \frac{i_{in}}{4}\right)^2}{I}, \quad \text{dla} \quad |i_{in}| \leqslant 4I \tag{25}$$

$$i_{Dn3} = \frac{\left(I - \frac{i_S}{4}\right)^2}{I}, \qquad i_{Dp3} = \frac{\left(I + \frac{i_S}{4}\right)^2}{I}, \quad \text{dla} \quad |i_S| \leqslant 4I \tag{26}$$

$$\begin{gathered} i_{out} = i_{Dd6} - i_{Dg1} + i_{Dd4} - i_{Dg6} = \\ i_{Dg4} - i_{Dg1} + i_{Dd4} - i_{Dd1} = A(i_{Dg3} - i_{Dg2}) + A(i_{Dd3} - i_{Dd2}). \end{gathered} \tag{27}$$

Jeżeli $|2i_{Dn3}| > \beta v_{idg}$ i $|2i_{Dp3}| > \beta v_{idd}$ różnice prądów tranzystorów T_{g3} i T_{g2} oraz T_{d3} i T_{d2} można wyrazić następująco

$$i_{Dg3} - i_{Dg2} = i_{Dn3} v_{idg} \left(\frac{\beta}{i_{Dn3}} - \frac{\beta^2 v_{idg}^2}{4 i_{Dn3}^2} \right)^{\frac{1}{2}} \tag{28}$$

$$i_{Dd3} - i_{Dd2} = i_{Dp3} v_{idd} \left(\frac{\beta}{i_{Dp3}} - \frac{\beta^2 v_{idd}^2}{4 i_{Dp3}^2} \right)^{\frac{1}{2}} \tag{29}$$

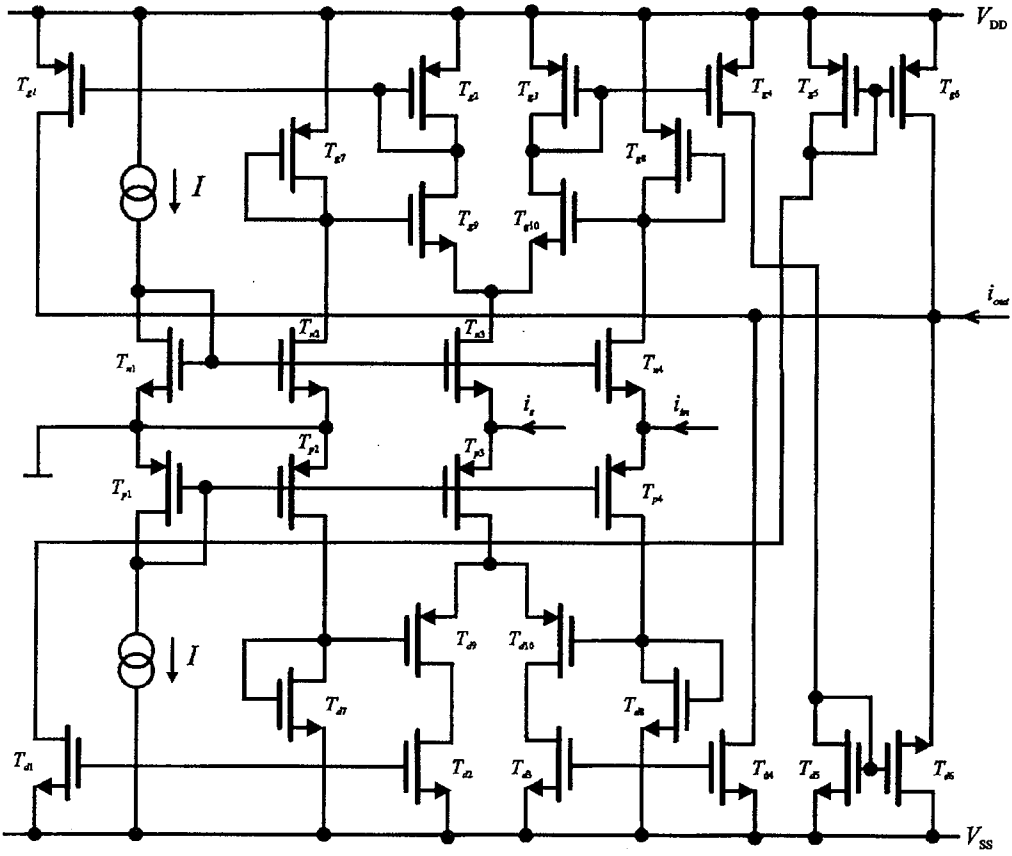
gdzie

$$v_{idg} = v_{GSg10} - v_{GSg9} = v_{SGg8} - v_{SGg7} \tag{30}$$

$$v_{idd} = v_{GSd10} - v_{GSd9} = v_{SGd8} - v_{SGd7}. \tag{31}$$

Napięcia różnicowe stopni: górnego v_{idg} i dolnego v_{idd} można wyznaczyć w następujący sposób

$$v_{idg} = \sqrt{\frac{2i_{Dg8}}{\beta_{g8}}} - \sqrt{\frac{2i_{Dg7}}{\beta_{g7}}} = \sqrt{\frac{2i_{Dn4}}{\beta}} - \sqrt{\frac{2i_{Dn2}}{\beta}} = \left(I - \frac{i_{in}}{4}\right)\sqrt{\frac{2}{\beta I}} - \sqrt{\frac{2I}{\beta}} = -\frac{i_{in}}{2\sqrt{2\beta I}} \tag{32}$$



Rys. 23. Układ synapsy w klasie AB przestrajanej sygnałem analogowym regulującym prąd pary różnicowej

$$v_{idd}=\sqrt{\frac{2i_{Dd8}}{\beta_{d8}}}-\sqrt{\frac{2i_{Dd7}}{\beta_{d7}}}=\sqrt{\frac{2i_{Dp4}}{\beta}}-\sqrt{\frac{2i_{Dp2}}{\beta}}=\left(I+\frac{i_{in}}{4}\right)\sqrt{\frac{2}{\beta I}}-\sqrt{\frac{2I}{\beta}}=-\frac{i_{in}}{2\sqrt{2\beta I}}. \tag{33}$$

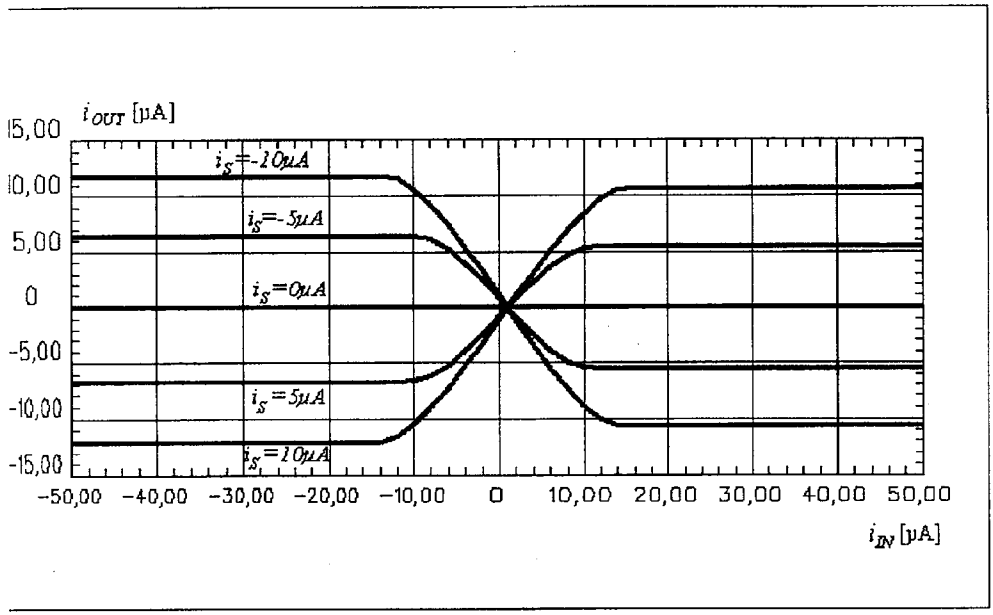
Podstawiając zależność (32÷33) do zależności (28÷29) otrzymuje się

$$i_{Dg3}-i_{Dg2}=-\frac{i_{Dn3}i_{in}}{2}\left(\frac{1}{2Ii_{Dn3}}-\frac{i_{in}^2}{64I^2i_{Dn3}^2}\right)^{\frac{1}{2}} \tag{34}$$

$$i_{Dd3}-i_{Dd2}=-\frac{i_{Dp3}i_{in}}{2}\left(\frac{1}{2Ii_{Dp3}}-\frac{i_{in}^2}{64I^2i_{Dp3}^2}\right)^{\frac{1}{2}}. \tag{35}$$

Uwzględniając zależności (25÷26) i (34÷35) dla małych zmian prądu i_S we wzorze (27) można przyjąć

$$i_{out}\sim Ai_{in}\sqrt{\frac{i_{Dp3}}{I}}-Ai_{in}\sqrt{\frac{i_{Dn3}}{I}}=Ai_{in}\sqrt{\frac{\left(I+\frac{i_S}{4}\right)^2}{I^2}}-Ai_{in}\sqrt{\frac{\left(I+\frac{i_S}{4}\right)^2}{I^2}}=\frac{Ai_s}{2I}i_{in}. \tag{36}$$

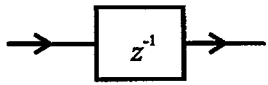


Rys. 24. Wyniki badań stymulacyjnych synapsy w klasie AB przestrajanej sygnałem analogowym regulującym prąd pary różnicowej

Z zależności (36) wynika, że dla małych wartości sygnału wejściowego w zakresie liniowej charakterystyki wzmocnienie układu jest proporcjonalne do prądu sterującego i_S. Układ zbadano symulacyjnie dla tych samych warunków co układy funkcji aktywacji. Rodzina charakterystyk przejściowych układu w funkcji sygnału i_S jest pokazana na rysunku 24. Wynika z niej możliwość programowania wzmocnienia za pomocą sygnału prądowego i_S. Zaletą układu jest bardzo mała moc pobierana ze źródła zasilania. Wadą układu jest duża powierzchnia zajmowana przez układ z powodu dużej liczby tranzystorów użytych do zrealizowania funkcji przestrajania.

3.3. UKŁADY OPÓŹNIAJĄCE O OKRES PRZEŁĄCZANIA T_S

W sieciach rekurencyjnych i komórkowych pracujących w czasie dyskretnym konieczne jest opóźnianie sygnału o okres przełączania T_S. Symbol blokowy układu opóźniającego sygnał o okres T_S jest pokazany na rysunku 25.



Rys. 25. Symbol bloków układu opóźniającego o okres T

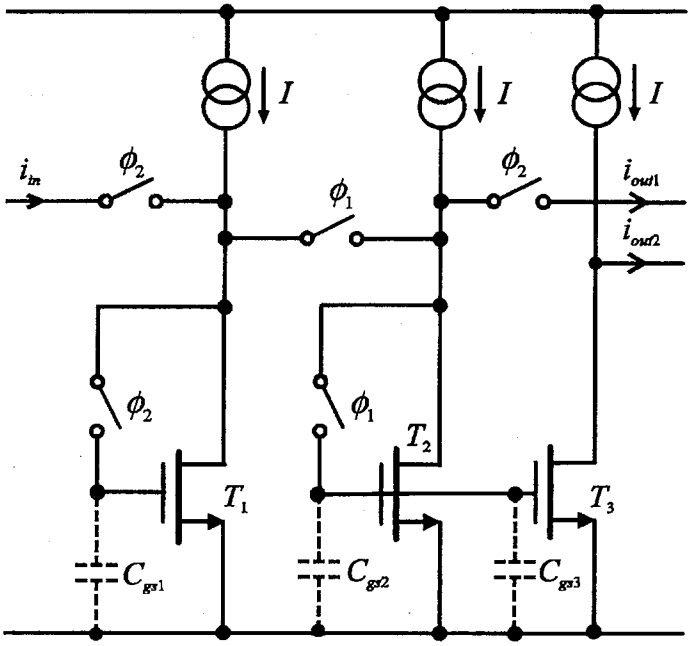
3.3.1. Układ opóźniający pracujący w klasie A

Układ opóźniający pracujący w klasie A jest przedstawiony na rysunku 26 [57]. Składa się on z dwóch kaskadowo połączonych układów opóźniających o połowę

okresu przełączania oraz stopnia wyjściowego umożliwiającego uzyskanie sygnału wyjściowego w całym okresie przełączania. Klucze analogowe są sterowane sygnałami z zegara dwufazowego o okresie przełączania T_S. Sygnały sterujące są oznaczone literą ϕ, a indeks oznacza, w której fazie klucz jest włączony. Zasada działania układu jest następująca.

W fazie 2 okresu $n-1$ pojemność C_{gs1} tranzystora T_1 ładuje się do napięcia v_{GS1} zapewniającego przepływ w tranzystorze T_1 prądu drenu

$$i_{D1}=I+i_{in}(n-1). \tag{37}$$



Rys. 26. Układ opóźniający pracujący w klasie A

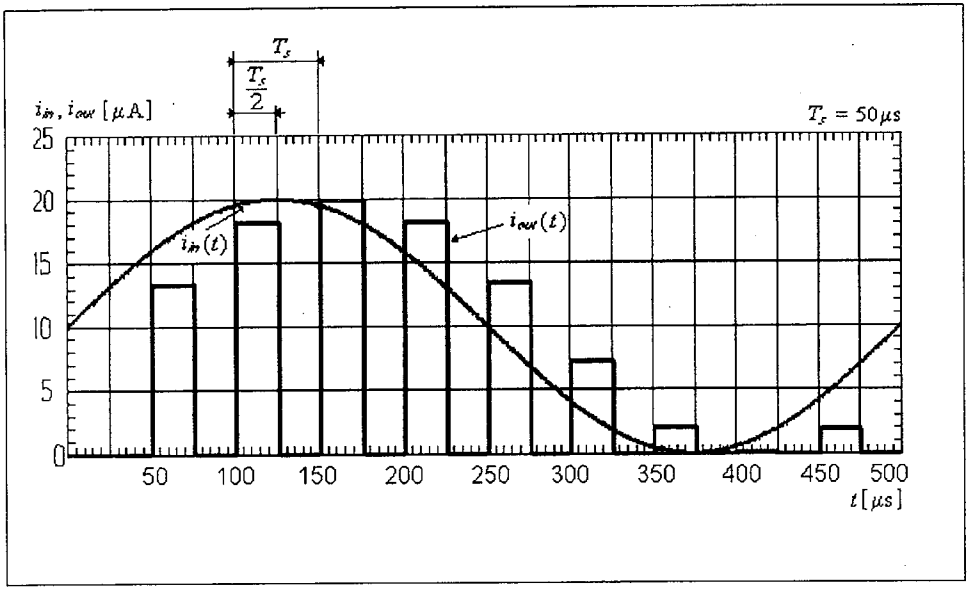
W fazie 1 okresu n napięcie v_{GS1} na pojemność C_{gs1} tranzystora T_1 nadal podtrzymuje przepływ prądu drenu i_{D1} w tranzystorze T_1, a pojemność C_{gs2} i C_{gs3} na bramkach tranzystorów T_2 i T_3 ładują się do napięcia $v_{GS2}=v_{GS3}$ zapewniającego przepływ przez tranzystory T_2 i T_3 prądów drenów

$$i_{D2}=i_{D3}=2I-i_{D1}=I-i_{in}(n-1). \tag{38}$$

Prąd na wyjściu 2 wynosi wówczas

$$i_{out2}(n)=I-i_{D3}=i_{in}(n-1). \tag{39}$$

W fazie 2 okresu n prądy i_{D2} i i_{D3} tranzystorów T_2 i T_3 oraz prąd na wyjściu 2 $i_{out2}(n)$ są nadal podtrzymywane przez napięcie $v_{GS2}=v_{GS3}$ na pojemnościach C_{gs2} i C_{gs3} tranzystorów T_2 i T_3. Prąd na wyjściu 1 wynosi



Rys. 27. Odpowiedź układu opóźniajacego drugiej generacji na pobudzenie sinusoidalne

$$i_{out1}(n) = I - i_{D2} = i_{in}(n-1). \tag{40}$$

W tym czasie ustala się nowy prąd tranzystora T_1 dla okresu $n+1$ równy $I + i_{in}(n)$, tak aby w okresie $n+1$ prąd wyjściowy wynosił $i_{in}(n)$. Prąd na wyjściu 1 jest dostępny tylko w fazie 1, a na wyjściu 2 w obu fazach przełączania. Opóźnienie prądu wyjściowego w stosunku do prądu wejściowego wynosi jeden okres przełączania.

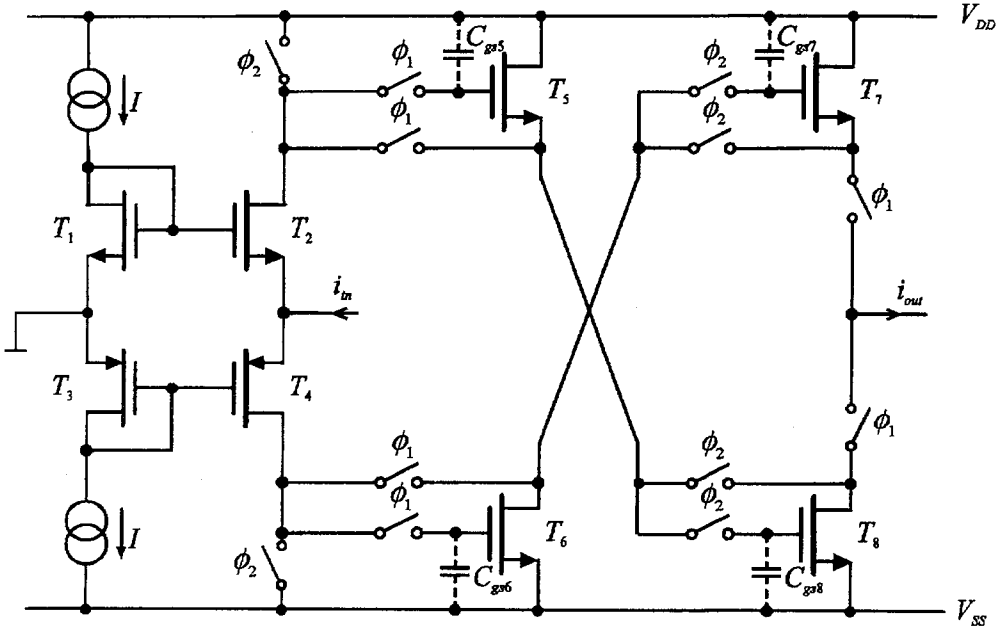
Układ opóźniający pracujący w klasie A zbadano symulacyjnie przy użyciu programu VLSISPICE COMPASS DA. Układ pobudzono sygnałem sinusoidalnym. Pobudzenie i odpowiedź układu są pokazane na rysunku 27.

3.3.2. Układ opóźniający pracujący w klasie AB

Układ opóźniający pracujący w klasie AB [57] jest przedstawiony na rysunku 28. Składa się on z jednego układu pamiętającego i krzyżowo włączonego stopnia wyjściowego drugiego układu pamiętającego. Zasada działania układu w obu fazach każdego cyklu jest następująca:

W fazie 2 okresu $n-1$ tranzystory T_7 i T_8 są w połączeniu diodowym i ich pojemności C_{gs7} i C_{gs8} ładują się do napięć v_{GS7} i v_{GS8} zapewniających przepływ prądów drenów i_{D7} i i_{D8} tranzystorów T_7 i T_8 wymuszonych odpowiednio prądami tranzystorów T_6 i T_5. Uwzględniając zależność (4.16) prądy drenów i_{D7} oraz i_{D8} wynoszą

$$i_{D7} = \frac{\left[I + \frac{i_{in}(n-1)}{4}\right]^2}{I} \quad \text{oraz} \quad i_{D8} = \frac{\left[I - \frac{i_{in}(n-1)}{4}\right]^2}{I} \quad \text{dla} \quad |i_{in}| \leqslant 4I. \tag{41}$$



Rys. 28. Układ opóźniajacy pracujący w klasie AB

W fazie 1 okresu n napięcia v_{GS7} i v_{GS8} na pojemnościach C_{gs7} i C_{gs8} tranzystorów T_7 i T_8 podtrzymują przepływ prądów drenów w tranzystorach T_7 i T_8. Stąd prąd wyjściowy wynosi

$$i_{out}(n) = i_{D7} - i_{D8} = i_{in}(n-1). \tag{42}$$

Jednocześnie tranzystory T_5 i T_6 są w połączeniu diodowym i ich pojemności wejściowe zostają naładowane do napięć v_{GS5} i v_{GS6} zapewniających przepływ w tranzystorach T_5 i T_6 następujących prądów

$$i_{D5} = \frac{\left[I - \frac{i_{in}(n)}{4}\right]^2}{I} \quad \text{oraz} \quad i_{D6} = \frac{\left[I + \frac{i_{in}(n)}{4}\right]^2}{I} \quad \text{dla} \quad |i_{in}| \leqslant 4I. \tag{43}$$

W fazie 2 okresu n tranzystory T_7 i T_8 są ponownie w połączeniu diodowym i ich pojemności C_{gs7} i C_{gs8} ładują się do napięć zapewniających przepływ nowych wartosci prądów i_{D5} oraz i_{D6} wyrażonych wzorem (43). Umożliwi to przeniesienie w fazie 1 okresu $n+1$ na wyjście prądu równego $i_{in}(n)$. Dlatego opóźnienie prądu wyjściowego do prądu wejściowego jest równe okresowi przełączania T_s. Układ opóźniający pracujący w klasie AB zbadano symulacyjnie przy użyciu programu VLSISPICE COMPASS DA w tych samych warunkach co układ pracujący w klasie A. Otrzymano takie same wyniki jak przedstawione na rysunku 27.

PODSUMOWANIE

W pracy przedstawiono modele neuronów, synaps oraz podstawowych rodzajów sieci neuronowych: jednokierunkowej, rekurencyjnej oraz komórkowej. Szczegółowo omówiono, a także przedstawiono wyniki badań symulacyjnych układów pracujących w trybie prądowym umożliwiających realizację neuronów, sieci neuronowych i algorytmów uczenia. Sieci neuronowe należą do systemów analogowych i ich układowe realizacje w technice prądowej są obecnie w fazie wykonywania pierwszych prototypów w postaci ukladów scalonych. Pierwsze sieci neuronowe w technice prądowej wykonane w postaci układu scalonego są sieciami komórkowymi. Wyniki pomiarów są zaprezentowane w pracach [45, 47÷48].

Rezultaty badań prototypów przedstawionych w pracach [45, 48] potwierdzają możliwość zastosowania układów prądowych do realizacji dużych sieci wykonywanych w wersji scalonej. Należy wkrótce oczekiwać nowych rozwiązań sieci neuronowych o dużej liczbie neuronów i dużej liczbie połączeń między nimi. Dziedzina ta ma przed sobą przyszłość z powodu zalet układów prądowych pozwalających na pracę z małymi napięciami zasilania przy zachowaniu pasma i dynamiki. Do problemów trudnych do rozwiązania należą moc pobierana ze źródła zasilania oraz powierzchnia zajmowana przez sieć w przypadku sieci składających się z tysięcy lub setek tysięcy neuronów.

BIBLIOGRAFIA

1. F. Rosenblat: *The perceptron: A probabilistic model for information storage and organization in the brain.* Psych. Rev. 1958, vol. 65, pp. 386–408
2. B. Widrow, M.E. Hoff: *Adaptive switching circuits.* Ire Wescon Conventional Record, New York, 1960, pp. 96–104
3. D.E. Rumelhart, G.E. Hinton, R.J. Williams: *Learning internal representation by error propagation.* Rumelhart D.E. and McClellan J.L: *Parallel Distributed Processing,* Cambridge: MIT Press, 1986
4. D.O. Hebb: *The organization of behaviour.* New York: J. Wiley & Sons, 1949
5. T. Kohonen: *Self-organizatoin and associative memory.* Berlin, Springer-Verlag, 1984
6. S. Grossberg: *Competitive learning from interactive activation to adaptive resonance.* Cognitive Science, 1987, vol. 11, pp. 23–63
7. G. Carpenter, S. Grossberg: *Massively parallel architecture for a self organizing neural pattern recognition machine.* Computer Vision, Graphic and Image Processing, 1983, vol. 37, pp. 54–115
8. B. Widrow, R. Winter: *Neural nets for adaptive filtering and adaptive pattern recognition.* IEEE Computer, 1990, pp. 25–39
9. L.O. Chua, B. Shi: *Multiply layer cellular neural network. A tutorial.* Algorithms and Parallel VLSI Architectures, North Holland, 1991
10. L.O. Chua et al.: *Signal processing using cellular neural networks.* J. of VLSI Signal Processing, 1991, vol. 3, pp. 25–51
11. L.O. Chua, B. Shi: *Exploiting cellular automata in the design in the cellular neural networks for binary image processing.* UCB/ERL M89/130, University of California at Berkeley, 1989
12. Y. Yao, W.J. Freeman, B. Burke, Q. Yang: *Pattern recognition by a distributed neural network: An industrial application.* Neural Networks, 1991, vol. 4, pp. 103–121
13. K. Ślot: *Cellular neural network design for solving specific image-processing problems.* Int. J. Circuit Theory and Applications, 1992, vol. 20, pp. 629–637

14. H. Hashimoto, T. Kubota, M. Kudon, F. Harashima: *Visual control of a robotic manipulator using neural networks*. Proc. 29th Conf. on Decision and Control, 1990, pp. 3295–3302
15. G. Josin, D Charney, D. Whiet: *Robot control using neural networks*. Proc. IEEE Int. Conf. on Neural Networks, 1988, pp. II-615–631
16. M. Kawato, Y. Uno, M. Isobe, R. Suzuki: *Hierarchical neural network model for voluntary movement with applications to robotics*. IEEE Control Systems Magazine, 1988, pp. 8–16
17. K.S. Narenda, K. Pathasarathy: *Identification and control of dynamical systems using neural networks*. IEEE Trans. Neural Networks, 1990, pp. 4–21
18. T. Roska et al.: *Detecting moving and standing objects using cellular neural networks*. Int. J. Circuit Theory and Applications, 1992, vol. 20, pp. 613–628
19. G. Eros et al.: *Optical tracking system for automatic guided vehicles using cellular neural networks*. Proc. IEEE Workshop on Cellular Neural Networks and Their Applications, 1992, pp. 216–221
20. R.P. Lippman: *Review of neural networks for speech recognition*. Neural Computation, 1989, pp. 1–38
21. T. Matsumoto et al.: *CNN cloning template: connected component detector*. IEEE Trans. Circuits and Systems, 1990, vol. CAS-37, pp. 633–635
22. T. Matsumoto et al.: *CNN cloning template: hole filter*. IEEE Trans. Circuits and Systems, 1990, vol. CAS-37, pp. 635–638
23. E. Sanches-Sinencio, C. Lau (Eds.): *Artificial Nerual Networks*, New York: IEEE Press, 1992
24. J.M. Żurada: *Introduction to Artificial Neural Networks*, St. Paul: West Publishing, 1992
25. J.A. Nossek (ed.): *Proc. of the IEEE Int. Workshop on Cellular Neural Networks and Their Applications*, 1992
26. R. Tadeusiewicz: *Sieci neuronowe*. Akademicka Oficyna Wydawnicza RM, Warszawa 1993
27. S. Osowski: *Sieci neuronowe*. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 1994
28. T. Kacprzak, K. Ślot: *Sieci neuronowe komórkowe*, Wydawnictwo Naukowe PWN, Warszawa–Łódź 1995
29. N.I. Khachab, M. Ismail: *A new continuous-time MOS implementation of feedback neural networks*. Proc. of 32nd Midwest Symp. on Circuits and Systems, 1989, pp. 221–224
30. F.M.A. Salam, M.R. Choi: *An all-MOS analog feedroward neural circuit with learning*, 1990, pp. 2508–2511
31. H.C. Card, C.R. Schneider, W.R. Moore: *Hebbian plasticity in MOS synapses*. IEE PRoc. Part G, 1991, vol. 137, pp. 175–184
32. M. Verleysen, P.G.A. Jespers: *An analog implementation of Hopfield's neural networks*. Electronics Letters, 1989, pp. 46–55
33. H. Wasaki, Y. Hario and S. Nakamura: *A localized learning rule for analog VLSI implementation of neural networks*. Proc. 33rd Midwest Symp. Circuits and Systems, 1990, pp. 1098–1101
34. T. Shima, T. Kimura, Y. Kamatani, T. Itakura, Y. Fujita, T. Iida: *Neuro chips with on-chip back-propagation and/or Hebbian learning*. IEEE J. of Solid-State Circuits, 1992, vol. SC-27, 1868–1876
35. L.O. Chua, L. Yang: *Cellular neural networks: theory*. IEEE Trans. Circuits and Systems, 1988, vol. CAS-35, pp. 1257–1272
36. L.O. Chua, L. Yang: *Cellular neural networks: applications*. IEEE Trans. Circuits and Systems, 1988, vol. CAS-35, pp. 1273–1290
37. M.A.C. Maher, S.P. Deweerth, M.A. Mahowald, C.A. Mead: *Implementing neural architectures using VLSI circuits*. IEEE Trans. Circuits and Systems, 1989, Vol. CAS-36, pp. 643–652
38. A.F. Murray: *Silicon implementation of neural netowrks*. IEE Proc. Part F, 1991, vol. 138, pp. 3–12
[39] J.J. Paulos, P.W. Hollis: *Neural networks using analog multipliers*. Proc. Int. Symp. Circuits and Systems, 1988, pp. 494–502
40. R.D. Reed, R.L. Geiger: *A multiply-input OTA circuit for neural networks*. IEEE Trans. Circuits and Systems, 1989, vo.. 36, pp. 767–770
41. O. Rossetto, C. Jutten, J. Herault, I. Kreutzer: *Analog VLSI synaptic matrices as building blocks for neural networks*. IEEE Micro., 1989, pp. 56–63
42. Y. Wang, F.M.A. Salam: *Design of neural network systems from custom analog VLSI chips*. Proc. Int. Symp. Circuits and Systems, 1990, pp. 1098–1102

43. J.M. Cruz, L.O. Chua: *Design of high speed, high density CNNs in CMOS technology*. Int. J. Circuit Theory and Applications, 1992, vol. 20, pp. 555 – 572
44. T.H. Borgstom, M. Ismail, S.B. Bibyk: *Programmable current mode neural network for implementation in analogue MOS VLSI*. IEE Proc. Part G, 1990, vo.. 137, pp. 175 – 184
45. J.E. Varrientos, E. Sanches-Sinencio, J. Ramirez-Angulo: *A current-mode cellular neural network implementation*. IEEE Trans. Circuits and Systems-II Analog and Digital Signal Processing, 1993, vol. CAS-40, pp. 147 – 155
46. A. Rodriguez-Vazquez, S. Espejo, R. Dominguez-Castro, J.L. Huertas, E. Sanchez-Sinencio: *Current-mode techniques for the implementation of continuous- and discrete-time cellular neural networks*. IEEE Trans. on Circuits and Systems-II Analog and Digital Signal Processing, 1993, vol. CAS-40, pp. 132 – 146
47. M. Anguita, F.J. Pelayo, A. Prieto, J. Ortega: *Analog CMOS implementation of a discrete time CNN with programmable cloning templates*, IEEE Trans. on Circuits and Systems-II Analog and Digital Signal Processing, 1993, vol. CAS-490, pp. 215 – 218
48. H. Harrer, J.A. Nossek, T. Roska, L. Chua: *A current-mode DTCNN universal chip*. Proc. Int. Symp. Circuits and Systems, 1994, vol. 5, pp. 135 – 138
49. K. Wawryn, B. Strzeszewski: *Neuron cells for low power VLSI neural networks*. 18th Nat. Conf. on Network Theory and Electronic Circuits, 1995, pp. 629 – 634
50. K. Wawryn, B. Strzeszewski: *Programmable low power VLSI current mode neuron cell*. Proc. of Int. Conf. Electronics, Circuits and Systems, 1996
51. K. Wawryn, B. Strzeszewski: *Low power VLSI neural cells for artifical neural networks*.Proc. Int. Symp. Circuits and Systems, 1996, vol. 3, pp. 372 – 375
52. B. Widrow, M.A. Lehr: *30 years of adaptive neural networks: Perceptron, madaline and backpropagation*. Proc. of the IEEE special issue on Neural Networks, 1990, vol. 78, pp. 1415 – 1442
53. J.J. Hopfield: *Neurons with graded response have collective computational properties like those of two state neurons*, Proc. National Academy of Sciences, 1984, pp. 3088 – 3092
54. J.J. Hopfield, D.W. Tank: *Neural computation of decisions in optimization problems*. Biolog. Cybern., 1985, pp. 141 – 154
55. J.J. Hopfield, D.W. Tank: *Computing with neural network circuits: A model*. Science, 1986, vol. 233, pp. 625 – 633
56. Z. Wang: *Wideband class AB (push-pull) current amplifier in CMOS technology*. Electronics Letters, 1990, vol. 25, pp. 543 – 545
57. N.C. Battersby, C. Toumazou: *Class AB switched-current techniques*. C. Toumazou, J.B. Hughes and N.c. Battersby: *Switched-currents an analogue technique for digital technology*, London, Peter Peregrinus Ltd., 1993

K. WAWRYN

CURRENT MODE ARTIFICIAL NEURAL NETWORKS

Summary

In this paper A and AB class current mode structures that set foundations for low power VLSI neural networks are proposed. The analog A and AB class building blocks for sigmoidal characteristic neuron cells, fixed and programmable weight synaptic connections and delay cells are presented and discussed. All structures has been designed for 0.8 μm CMOS technology, then simulated on fourth level VLSISPICE COMPASS DA program on SUNSPARC20 computer. A qualitative comparison is made betweeen A and AB class current mode structures.

Key words: neural networks, CMOS circuit , simulated circuits.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1996. 42, z. 3, ss. 325–334

Model idealny dla szacowania czasu opóźnienia w technice pomiarowej przepływów

ANDRZEJ BORYS

Wydział Telekomunikacji i Elektrotechniki, Akademia Techniczno-Rolnicza, Bydgoszcz

Received 1995.11.10

Authorized 1996.02.16

W przedstawionej pracy podano warunki dostateczne, które musi spełniać urządzenie do pomiaru czasu opóźnienia oraz przepływu, w którym dokonuje się pomiarów, aby były jednocześnie spełnione dwa równania postulowane przez Mescha i współpracowników [4], [5]. Rozpatrywany tutaj przypadek dotyczy urządzenia pomiarowego z sondą, w której umieszczono dwa czujniki w pewnej odległości jeden od drugiego. Natomiast rozpatrywany przepływ jest przepływem dwufazowym. Pierwsze z równań Mescha dla powyższego środowiska pomiarowego mówi o równości funkcji autokorelacji sygnałów rejestrowanych przez pierwszy i drugi czujnik. Natomiast drugie równanie określa zależność pomiędzy funkcjami korelacji i autokorelacji dla mierzonych sygnałów. Związek ten jest całką splotową, w której funkcję odpowiedzi impulsowej pełni funkcja gęstości prawdopodobieństwa dla czasów przelotu (opóźnienia) cząstek na drodze pomiędzy czujnikami. W przedstawionej pracy sformułowano idealizowany model, w którym powyższe równania są spełnione. Należy przypuszczać, że w każdym rzeczywistym modelu nie są one spełnione. Nie podano jednakże w tej pracy warunków koniecznych dla istnienia powyższych równości. Zastosowane tutaj metody analizy należą do repertuaru teorii przetwarzania sygnałów stochastycznych.

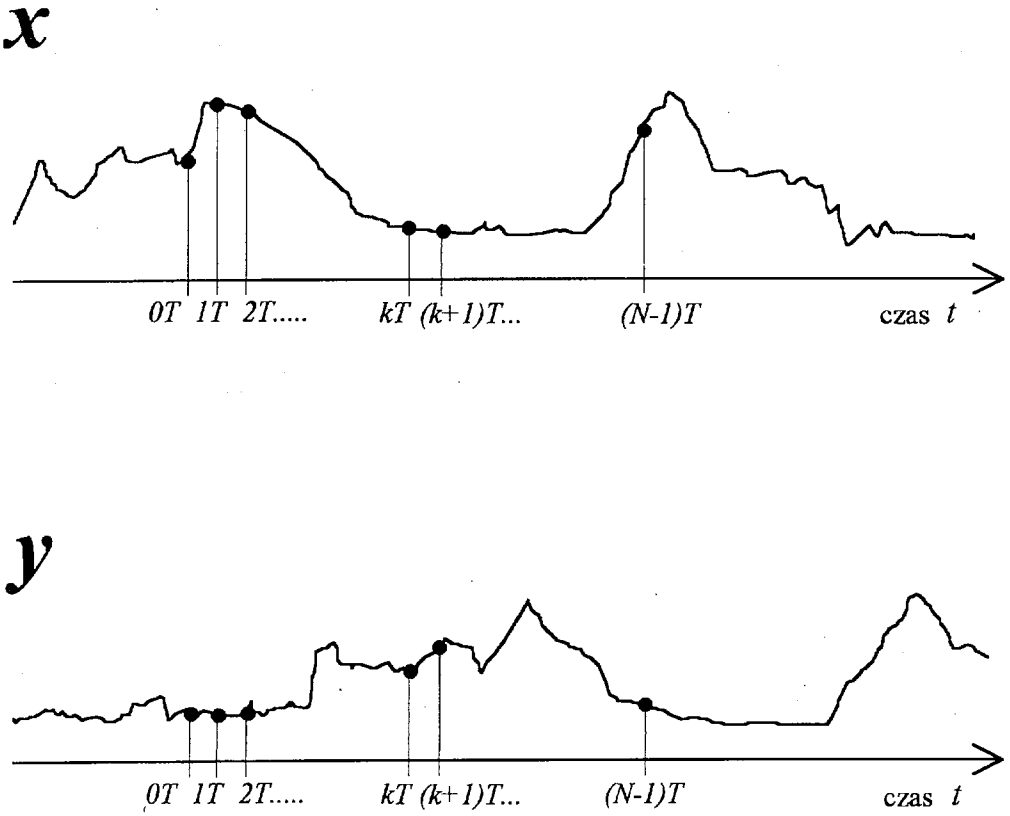
Słowa kluczowe: czas opóźnienia, sygnały stochastyczne, metoda przetwarzania sygnałów.

1. WPROWADZENIE

Rozpatrzmy dwa podobne do siebie sygnały stochastyczne, jak pokazano na rys. 1. Sygnały takie otrzymuje się między innymi w technice pomiarowej przepływów, gdy mierzy się czas przejścia cząstek między dwoma czujnikami pozostającymi w pewnej odległości od siebie [7].

Sygnał pochodzący z pierwszego czujnika oznaczono na rys. 1 literą x, natomiast sygnał pochodzący z drugiego czujnika literą y. Na rys. 1 widoczne jest przesunięcie czasowe pomiędzy charakterystycznymi partiami sygnałów x i y (pomiędzy „górkami" w sygnale x i sygnale y). Przesunięcie to jest miarą czasu przejścia cząstek

pomiędzy czujnikami pozostającymi w pewnej odległości od siebie [7]. W ogólności, powyższe przesunięcie czasowe zmienia się z czasem i jest wielkością przypadkową. Wielkość tę należy oszacować na podstawie sygnałów pomiarowych (takich sygnałów, jakie przedstawiono szkicowo na rys. 1). Jeżeli, tak jak to podano powyżej, przesunięcie czasowe jest pewnym procesem stochastycznym, to zadanie estymacyjne oznacza znalezienie odpowiednich histogramów. Histogramy te znajduje się sięgając do metod przetwarzania sygnałów stochastycznych.



Rys. 1. Sygnał pomiarowy (x) odbierany przez pierwszy czujnik i sygnał odbierany przez drugi czujnik (y) w urządzeniu pomiarowym służącym do estymacji czasu opóźnienia

Rodzaj sygnałów jakie przedstawiono na rys. 1 i zadanie estymacyjne jakie opisano powyżej występują w wielu problemach telekomunikacyjnych [1], [2]. Zwracamy w tym miejscu też uwagę, że przesunięcie czasowe, o którym była mowa powyżej, nazywa się w literaturze telekomunikacyjnej czasem opóźnienia [1], [2]. Na temat estymacji czasu opóźnienia i związanych z tym zagadnieniem algorytmów obliczeniowych napisano mnóstwo artykułów w czasopismach poświęconych problemom telekomunikacyjnym i problemom cyfrowego przetwarzania sygnałów. Każdy z opracowanych algorytmów estymacji czasu opóźnienia opiera się na jakimś modelu. W tej pracy jest omawiany pewien idealizowany model, który znajduje zastosowanie

w technice pomiarowej przepływów. Podane są warunki, przy których stosowanie powyższego modelu nie prowadzi do poważnych błędów w szacowanych wartościach czasu opóźnienia.

Przetwarzanie sygnałów ciągłych z reguły nie odbywa się bezpośrednio. Sygnały te są próbkowane w celu wykorzystania możliwości jakie daje technika cyfrowa. Powyższy proces pokazano szkicowo na rys. 1, gdzie okres próbkowania oznaczono literą T. Mesch i współpracownicy [4], [5] sformułowali swoje równania dla sygnałów ciągłych. Do równań tych odnoszą się wyprowadzenia przedstawione w tym artykule. Wyprowadzenia te są również sformułowane dla czasu ciągłego. Efekty związane z próbkowaniem nie są analizowane w tym artykule.

Równania Mescha mają następującą postać:

$$R_{xx}(\varphi)=R_{yy}(\varphi) \tag{1}$$

oraz

$$R_{xy}(\varphi)=\int_{-\infty}^{\infty} R_{xx}(\tau)p_D(\varphi-\tau)d\tau\,, \tag{2}$$

gdzie $R_{xx}(\varphi)$ i $R_{yy}(\varphi)$ są funkcjami autokorelacji dla sygnału, odpowiednio, x i y. Funkcja $R_{xy}(\varphi)$ jest funkcją korelacji wzajemnej (nazywanej w dalszej części artykyłu funkcją korelacji) pomiędzy sygnałami x i y, natomiast funkcja $p_D(.)$ jest funkcją gęstości prawdopodobieństwa zmiennej przypadkowej D (czas opóźnienia). Zmienne φ i τ we wzorach (1) i (2) są zmiennymi pomocniczymi.

Pierwsze z równań Mescha mówi, że własności statystyczne sygnałów odbieranych przez pierwszy i drugi czujnik, wyrażone przez funkcje autokorelacji (to jest wtedy, gdy każdy z sygnałów jest rozpatrywany oddzielnie), są takie same. Natomiast, gdy sygnały te są ze sobą „porównywane" (w sensie statystycznym), jak dzieje się to w funkcji korelacji, to wtedy jest spełniona zależność (2). Nadmieniamy też w tym miejscu, że równania (1) i (2) muszą być spełnione jednocześnie. Równania te tworzą model, na którym opiera się wiele algorytmów szacowania funkcji gęstości prawdopodobieństwa zmiennej: czas opóźnienia, między innymi algorytm z pracy [7].

Przedstawiona praca rozpoczyna się od wprowadzenia zawierającego krótkie omówienie rozpatrywanego problemu. W rozdziale drugim przedstawiony jest idealizowany model środowiska pomiarowego mającego zastosowanie w technice pomiarowej przepływów [4], [5], [7]. Podana jest też analiza tego modelu z wykorzystaniem teorii przetwarzania sygnałów. Korzystając z wyników tej analizy pokazuje się następnie, że równania Mescha są spełnione dla modelu idealizowanego. Jest to główny rezultat tego artykułu. Rozdział trzeci poświęcony jest problemowi wyjaśnienia pewnego nieporozumienia związanego z przedstawianiem równań (1) i (2) za pomocą modelu zawierającego liniowy układ (system) stacjonarny.

2. MODEL IDEALIZOWANY

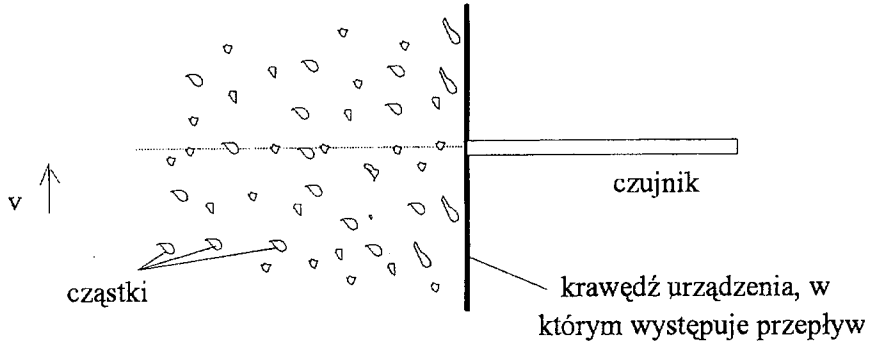
Nasze rozważania w tym rozdziale rozpoczniemy od sformułowania modelu idealizowanego, o którym była mowa we wprowadzeniu. A zatem modelem idealizowanym nazwiemy taki model, w którym

a) sygnały z czujników nie są obarczone efektem uśredniania przestrzennego (patrz rys. 2),

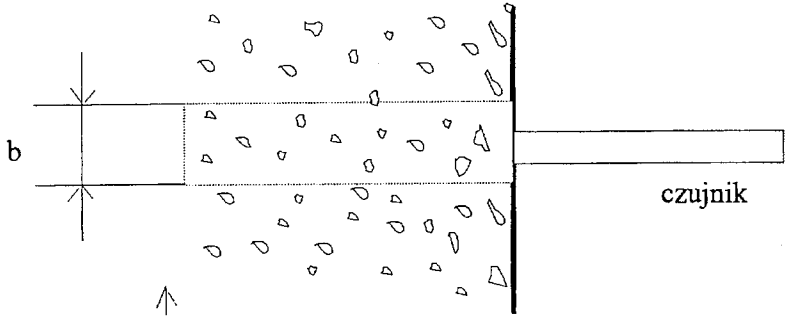
b) udziały w sygnale mierzonym pochodzące od zbiorów cząstek poruszających się z tą samą prędkoscią można modelować za pomocą szumu białego.

Czujnik idealny, jak pokazano na rys. 2a, jest to taki czujnik, który „omiata" poruszające się cząstki w pewnym obszarze płaszczyzny prostopadłej do wektora prędkosći tych cząstek. Fakt ten zaznaczono na rys. 2a linią przerywaną. W przeciwieństwie do tego czujnika, czujnik nieidealny dokonuje pomiaru w pewnej objętości [7] — na rys. 2b fakt ten jest zaznaczony prostokątem pociągniętym linią przerywaną (szerokość tego prostokąta wynosi *b*). O takim czujniku mówi się, że dokonuje uśredniania przestrzennego [7].

a) czujnik idealny (brak uśredniania przestrzennego)



b) czujnik nieidealny (czujnik uśrednia pomiar wykonany w pewnej objętości)



Rys. 2. Ilustracja do pojęcia czujnika idealnego (a) — bez uśredniania przestrzennego, i czujnika nieidealnego (b) — uśredniającego pomiar wykonany w pewnej objętości

Dla celów modelowania [7] można podzielić zbiór poruszających się z różnymi prędkościami cząstek z rys. 2 na podzbiory charakteryzujące się tym, że w każdym z tych podzbiorów prędkość cząstek jest taka sama. Dalej zakładamy, że każdy z powyższych podzbiorów ma swój udział w sumarycznym sygnale mierzonym przez czujnik. W założeniu b), podanym na początku rozdziału, zakładamy podobnie jak w pracy [7], że każdy z udziałów można opisać za pomocą szumu białego.

Wykorzystując założenie a) podane na początku rozdziału, można sygnały pomiarowe z czujników opisać za pomocą następujących równań:

$$x(t)=\sum_{i=1}^{N} x_i(t) \tag{3}$$

oraz

$$y(t)=\sum_{i=1}^{N} x_i(t-D_i), \tag{4}$$

gdzie $x_i(t)$, $i=1,...,N$, reprezentują udziały w sygnale sumarycznym pochodzące od poszczególnych podzbiorów cząstek poruszających się z tą samą prędkością. Zakładamy, że takich podzbiorów jest N, t oznacza czas, natomiast D_i reprezentuje czas przelotu (opóźnienie czasowe) cząstek na drodze pomiędzy czujnikami. Jeżeli odległość pomiędzy czujnikami wynosi d i podzbiór cząstek oznaczony indeksem i porusza się z prędkością v_i, to D_i dla tego podzbioru wynosi $D_i=\frac{d}{v_i}$.

Zauważymy w tym miejscu, że jeżeli czujniki nie są idealne, to znaczy dokonują uśredniania przestrzennego, to mierzone przez nie sygnały nie mogą być przedstawione w postaci podanej w równaniach (3) i (4). Wtedy w tych równaniach pojawia się dodatkowy operator odpowiedzialny za uśrednianie przestrzenne. Operator ten sprowadzono w pracy [7] do równoważnego operatora liniowego, reprezentującego filtr dolnoprzepustowy.

Teraz skorzystamy z założenia b), przyjmując dodatkowo, że procesy stochastyczne, których realizacjami są sygnały $x_i(t)$, $i=1,...,N$, są stacjonarne i niezależne. Wtedy możemy napisać

$$E\{[x_i(t_a)-E\{x_i(t_a)\}][x_j(t_b)-E\{x_j(t_b)\}]\}=\delta_{ij}\sigma_{x_i}^2\delta(t_b-t_a) \tag{5a}$$

oraz

$$\sigma_{x_i}^2=N_0 p_v(v)\Delta v, \tag{5b}$$

gdzie E oznacza wartość oczekiwaną, natomiast δ_{ij} jest deltą Kroneckera, a $\delta(t_b-t_a)$ impulsem Diraca. $\sigma_{x_i}^2$ jest wariancją sygnału $x_i(t)$. Ponadto N_0 jest pewną stałą, natomiast funkcja $p_v(v)=p_D(d/v)\cdot d/(v)^2$ jest funkcją gęstości prawdopodobieństwa zmiennej przypadkowej – prędkości. Δv oznacza przyrost prędkości (w przybliżeniu jest to różniczka zmiennej v). Zakładamy też, że zachodzi równość $\sum p_v(v)\Delta v=1$, wynikająca z własności funkcji gęstości prawdopodobieństwa.

Równanie (5b) określa jak szum biały, pochodzący od każdego z podzbiorów poruszających się cząstek, partycypuje w sygnale sumarycznym. Założono, że ten udział jest proporcjonalny do rozkładu prędkości. Zauważmy, że jest to całkiem naturalne założenie które mówi, że im więcej cząstek poruszających się z daną prędkością, to tym jest „większy udział (statystyczny) tej prędkości w sygnale rejestrowanym przez czujnik". Podstawiając zależności (3), (4) i (5) w równaniu

definicyjnym funkcji autokorelacji dla sygnałów x i y (jest to każdorazowo pierwsza część równań (6a) i (6b)) i wykonując następnie odpowiednie rachunki, dochodzi się do następujących równości:

$$R_{xx}(\varphi) \overset{\mathrm{df}}{=} E\{x(t)x(t+\varphi)\} = N_0\delta(\varphi) + A_0 \tag{6a}$$

$$R_{yy}(\varphi) \overset{\mathrm{df}}{=} E\{y(t)y(t+\varphi)\} = N_0\delta(\varphi) + A_0, \tag{6b}$$

gdzie stała A_0 wyraża się wzorem

$$A_0 = \sum_{i=1}^{N}\sum_{j=1}^{N} E\{x_i\}E\{x_j\} = \left\{\sum_{i=1}^{N} E\{x_i\}\right\}^2. \tag{6c}$$

Porównując ze sobą równania (6a) i (6b) widzimy, że rzeczywiście zachodzi równość (1), gdy są spełnione założenia a) i b) podane na początku tego rozdziału.

Sprawdzimy teraz, czy jednocześnie z równaniami (6) spełniona jest zależność (2). W tym celu, korzystając ponownie z równań (3), (4) i (5), wyprowadzimy wzór na funkcję korelacji pomiędzy sygnałami x i y. Rachunki są tutaj podobne jak przy wyprowadzeniu wzorów (6a) i (6b) i prowadzą do następującej zależności:

$$R_{xy}(\varphi) \overset{\mathrm{df}}{=} E\{x(t)y(t+\varphi)\} = N_0 p_D(\varphi) + A_0. \tag{7}$$

Następnym krokiem jest stwierdzenie, czy korzystając z zależności (6a) po prawej stronie równości (2) dochodzi się do tego samego wzoru, co wzór (7). A więc w tym wypadku będziemy mieli

$$\begin{aligned} &\int_{-\infty}^{\infty} [N_0\delta(\tau) + A_0] p_D(\varphi-\tau)d\tau = N_0 p_D(\varphi) + \\ &+ A_0 \int_{-\infty}^{\infty} p_D(\varphi-\tau)d\tau = N_0 p_D(\varphi) + A_0. \end{aligned} \tag{8}$$

Z równań (7) i (8) widać, że jeżeli są spełnione założenia a) i b) podane na początku tego rozdziału, to jednocześnie z równaniem (1) spełnione jest równanie (2).

Na zakończenie tego rozdziału podamy reprezentację równań (6a), (6b) i (7) w dziedzinie częstotliwości. Ma ona następującą postać:

$$G_{xx}(f) = G_{yy}(f) = N_0 + \delta(f) \tag{9a}$$

oraz

$$G_{xy}(f) = N_0 G_{p_D}(f) + A_0\delta(f), \tag{9b}$$

gdzie $G_{xx}(f)$ i $G_{yy}(f)$ są transformatami Fouriera funkcji autokorelacji, odpowiednio sygnału x i y. $G_{xy}(f)$ we wzorze (9b) jest transformatą Fouriera funkcji korelacji

sygnałów x i y, natomiast $G_{p_D}(f)$ funkcji $p_D(D)$; f w powyższych wzorach oznacza częstotliwość.

3. RÓWNOWAŻNY UKŁAD LINIOWY NIEZALEŻNY OD CZASU

Dwa podstawowe równania w teorii transmisji sygnałów stochastycznych przez liniowe systemy niezależne od czasu (stacjonarne) (ang. linear time-invariant (LTI) systems) to [6]:

$$G_{yy}(f)=G_{xx}(f)\cdot |H(f)|^2 \tag{10a}$$

oraz

$$G_{xy}(f)=G_{xx}(f)\cdot H(f), \tag{10b}$$

gdzie x oznacza sygnał wejściowy, natomiast y jest sygnałem wyjściowym. $G_{xx}(f)$ i $G_{yy}(f)$ są transformatami Fouriera funkcji autokorelacji, odpowiednio sygnału x i y. $G_{xy}(f)$ jest transformatą Fouriera funkcji korelacji sygnałów x i y, natomiast $H(f)$ jest transmitancją stacjonarnego układu (systemu) liniowego, przez który przesłano sygnał x.

Zauważmy też, że zastosowanie transformacji Fouriera do równań (1) i (2) przynosi

$$G_{yy}(f)=G_{xx}(f) \tag{11a}$$

oraz

$$G_{xy}(f)=G_{xx}(f)\cdot G_{p_D}(f). \tag{11b}$$

Spróbujmy teraz opisać „przejście" sygnału od pierwszego czujnika do drugiego w terminologii teletransmisyjnej. A więc załóżmy, że pomiędzy czujnikami istnieje hipotetyczny kanał (w sensie telekomunikacyjnym). Sygnał x (rejestrowany przez pierwszy czujnik) wchodzi do „kanału" i jest rejestrowany jako sygnał y przez drugi czujnik. Pytamy się jaka jest transmitacja powyższego kanału. Jeżeli porównamy ze sobą równania (10) i (11), to dojdziemy do wniosku, że transmitacja kanału musi spełniać następujący układ równań:

$$\begin{cases} |H(f)| = 1 \\ H(f)=G_{p_D} \end{cases} \quad \text{dla każdej częstotliwości } f. \tag{12}$$

Pokażemy teraz że nie istnieje taka funkcja $H(f)$, z wyjątkiem funkcji $\exp(-j2\pi f D_0)$, gdzie D_0 oznacza pewną stałą, która spełnia układ równań (12). W celu wykazania powyższego powołamy się na lemat 4 podany w pracy [3] na stronie 501. Przedtem jednak zdefiniujemy pojęcie funkcji charakterystycznej.

Funkcją charakterystyczną, $\psi(\xi)$, dla rozkładu prawdopodobieństwa F nazwiemy taką funkcję, która spełnia poniższe równanie [3]:

$$\psi(\xi)=\int_{-\infty}^{\infty} e^{j\xi z}F\{dz\}, \tag{13a}$$

gdzie z oznacza zmienną przypadkową. Dalej, jeżeli funkcja gęstości prawdopodobieństwa, $p(z)$, istnieje, to równanie (13a) może być przepisane w następującej postaci [3]:

$$\psi(\xi)=\int_{-\infty}^{\infty} e^{j\xi z} pz\{z\}dz. \tag{13b}$$

Lemat [3, str. 501]. Dla funkcji charakterystycznej istnieją tylko trzy poniższe możliwości:

a) $|\psi(\zeta)|<1$ dla wszystkich $\zeta\neq 0$.

b) $|\psi(\zeta)|=1$ i $|\psi(\zeta)|<1$ dla $0<\zeta<\lambda$. W tym przypadku $|\psi(\zeta)|$ jest funkcją okresową o okresie równym λ.

c) $|\psi(\zeta)|=1$ dla wszystkich ζ. W tym przypadku $\psi(\zeta)=\exp(jD_0\zeta)$, gdzie D_0 jest pewną stałą.

Zauważmy teraz, że wprowadzając $\zeta=-2\pi f$ w równaniu (13b), otrzymuje się zwykłą transformatę Fouriera funkcji $p(z)$. A zatem korzystając z powyższego faktu można warunki (12) przedstawić w równoważnej postaci

$$\begin{cases} |\psi_{p_D}(\zeta)|=1 \\ \psi^*(\zeta)=G_{p_D}(\zeta/(2\pi)) \end{cases} \quad \text{dla wszystkich } \zeta. \tag{14}$$

$\psi_{p_D}(\zeta)$ we wzorze (14) oznacza funkcję charakterystyczną dla funkcji gęstości prawdopodobieństwa p_D, natomiast symbol gwiazdki oznacza liczbę zespoloną sprzężoną.

Korzystając z zależności (14) i lematu podanego powyżej, można napisać

$$\begin{aligned} G_{p_D}(\zeta/(2\pi))&=H(-f)=(\exp(jD_0\zeta)^*= \\ &=(\exp(-j2\pi D_0 f)^*=\exp(j2\pi D_0 f) \end{aligned} \tag{15a}$$

Ostatecznie z (15a) wynika, że

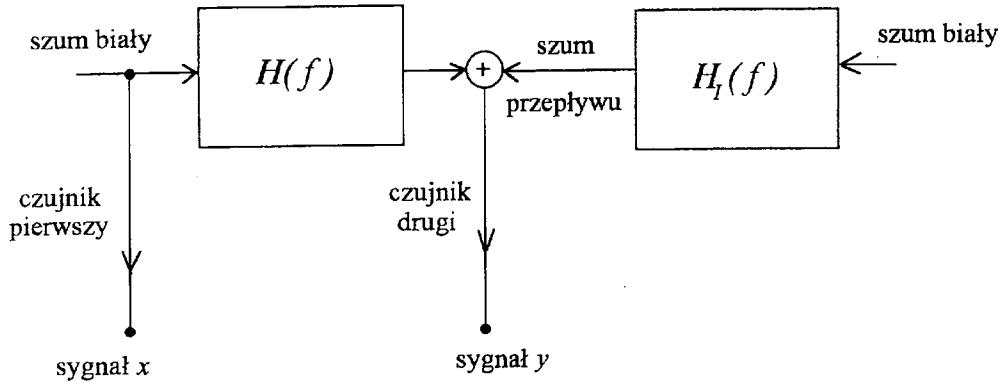
$$H(f)=\exp(-j2\pi D_0 f)=G_{p_D}(f). \tag{15b}$$

W ogólności funkcja gęstości prawdopodobieństwa p_D nie ma następującej postaci:

$$p_D=\delta(D-D_0). \tag{16}$$

Biorąc to pod uwagę oraz wynik dany równaniem (15b), stwierdzamy, że w ogólności nie jest możliwe modelowanie „kanału pomiędzy czujnikami" za pomocą stacjonarnego układu (systemu) liniowego. Możliwe jest to tylko wtedy, gdy wszystkie cząstki poruszają się między czujnikami z tą samą prędkością równą $v_0=\dfrac{d}{D_0}$.

Oczywiście koncepcję modelowania „kanału pomiędzy czujnikami" za pomocą stacjonarnego układu (systemu) liniowego można uratować, jeżeli wprowadzi się do modelu opisanego równaniami (3) i (4) dodatkowy szum, tzw. szum przepływu (*ang. flow noise*) [7]. Rys. 3 ilustruje tę koncepcję



Rys. 3. Ilustracja do opisu „kanału pomiędzy czujnikami" za pomocą stacjonarnego układu liniowego i dodatkowego szumu przepływu [7]

W pracy [7] pokazano, że transmitancja $H_I(f)$ z rys. 3 związana z szumem przepływu ma następującą postać:

$$H_I(f)=\sqrt{1-|H(f)|^2}. \tag{17}$$

Zauważmy też, że gdy funkcja $H(f)$ ma postać daną wzorem (15b), to zgodnie z wzorem (17) i rys. 3 moc szumu przepływu równa się zeru.

Koncepcja szumu przepływu jest dosyć sztuczna. Zauważmy, że służy ona li tylko zadaniu przedstawienia „kanału pomiędzy czujnikami" za pomocą stacjonarnego układu liniowego. Tylko przy wprowadzeniu powyższego szumu model ten jest poprawny. Zauważmy też, że moc szumu przepływu może być uważana za miarę odbiegania funkcji gęstości prawdopodobieństwa p_D od postaci danej wzorem (16).

W celu znalezienia funkcji gęstości prawdopodobieństwa p_D wystarczy oprzeć się na równaniu (2). W zadaniu znalezienia tej funkcji w niczym nie pomaga wprowadzenie pojęcia szumu przepływu.

PODSUMOWANIE

Warunki dostateczne, które musi spełniać urządzenie do pomiaru czasu opóźnienia, oraz przepływ, w którym dokonuje się pomiarów aby były jednocześnie spełnione dwa równania postulowane przez Mescha i współpracowników [4] [5], nie były dotychczas znane w literaturze. Zostały one podane w tej pracy.

W pracy sformułowano idealizowany model, w którym równania Mescha są spełnione. Należy przypuszczać, że w każdym rzeczywistym modelu nie są one spełnione. Nie podano jednakże w tej pracy warunków koniecznych dla istnienia powyższych równości. Jest to tematem do dalszych badań.

BIBLIOGRAFIA

1. C. Carter: *Time delay estimation for passive sonar signal processing.* IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-29, no. 3, 1981, pp. 463–470
2. C. Carter: *Coherence and time delay estimation.* Proceedings of the IEEE, vol. 75, no, 2, 1987, pp. 236–255
3. W. Feller: *An Introduction to Probability Theory and Its Applications,* John Wiley & Sons, New York, Volume II, 1966
4. H. Kipphan: *Verfahren zur störungsfreien Identifikation stochastischer Transporteknngrößen von Mehrphasen-Strömungen.* Ph.D. dissertation (w jęz. niemieckim), TU Karlsruhe, 1975
5. F. Mesch, R. Fritsche, H. Kipphan: *Transit time correlation — a survey on its applications to measuring transport phenomena.* Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control, vol. 96, 1974, pp. 414–420
6. A. Papoulis: *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, New York, 1965
7. R. Peters, F. Blischke, H. Meyr: *Transit-time estimation in turbulent flows.* IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-36, no. 1, 1988, pp. 20–28

A. Borys

An idealized model for time–delay estimation in turbulent flows

S u m m a r y

An idealized model describing the behaviour of two-sensor measuring setups for turbulent flows is presented. It is shown that the probability density function of transit – times can be found from a simple relation between the cross- and autocorrelation functions – eq. (2). In this context, the concepts of a „flow noise" and of an equivalent linear time–invariant system are shown to be irrelevant for the estimation of transit–times.

Key words: modeling time–delay, estimation of time–delay.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1996. 42, z. 3, ss. 335–342

Application of planar waveguide technique to investigation of liquid crystals

ANDRZEJ WALCZAK, ANDRZEJ KIEŻUN, EDWARD NOWINOWSKI-KRUSZELNICKI, LESZEK R. JAROSZEWICZ

Instytut Fizyki Technicznej, Wojskowa Akademia Techniczna

Received 1996.02.05

Authorized 1996.05.29

Modified method of prism coupler has been applied to obtain profiles of refractive index of liquid-crystalline film. Due to a special construction of a measuring device, studied film has been used in form of optical waveguide. The propagation constants of excited wavequide modes have been measured. The obtained family of the propagation constants is a data for refractive index calculation. For two arrangements of the liquid crystalline film, homogeneous and twisted by 90° one, the examples of experimental results are given.

Key words: liquid crystal waveguide, refractive index profile.

This research was supported by Polish State Committee for Scientific Research, Grant No 2P03B03308.

1. INTRODUCTION

In recent years a problem of guiding of optical waves in anisotropic medium, i.e. liquid crystal (LC) wavequide has been extensively studied[1]. This situation is caused by a possibility to construct a new class of devices using electro- or acoustoptic interactions. For this reason, measuring of essential parameters of the LC film is particularly important.

Two methods of measure of refractive index of the LC film have been developed in both mentioned methods the prism coupler is applied to excitation of leaky modes in LC film, either one-sided[2] or two-sided[3]. In the former case the studied film exhibits

---

[1] H.M. Boots: Light scattering in optically anisotropic medium, J. Opt. Soc. Am., 1994, vol. 11, no. 9, pp. 2599.

[2] R.Th. Kersten: The prism coupler as a precision instrument, Optica Acta, 1975, vol. 22, no. 6, pp. 515–521.

[3] Fuzi Yang, J.R. Sambles: Optical characterisation of liquid crystals by mans of half-leaky guided modees, J. Opt. Soc. Am., B, 1993, vol. 10, no. 5, pp. 858–866.

intermediate refractive index in relation to the edge layers, in the latter one lower than both edge layers.

The essence of the presented work is an application of the methods used in integrated optics for studying of planar wavequides[4] in order to measure refractive index profile of the LC film. The studied example, as in the above mentioned methods, is placed between two prisms, but due to a special prism's construction[5] it is possible to excite waveguided modes in the LC film. In order to obtain this effect a special set of layers have been designed. The studied film exhibits the highest refractive index in comparison to the edge layers. Such a configuration allows to measure propagation constants. Afterwards, the profile of refractive index may be calculated by means of procedure used in integrated optics[6].

The examples of experimental results for two different kinds of the LC layer are also illustrated.

2. MEASUREMENT SET-UP

The special construction of the measuring device, presented in Fig. 1, has been adopted to built planar waveguide form the studied LC film. The LC film is placed between two coupled prisms. Prism's bottom is covered with set of layers refractive index of which is lower than LC layer's one. Thickness of that buffers layer is properly chosen to ensure the tunnelling light coupling into the waveguide[5]. Parts of the buffer layer consist of the indium — tin oxide (ITO) to apply the driving voltage if needed. The LC layer is aligning with polyimide.

Light beam from He—Ne laser (4) after passing the polarizer falls at a certain angle θ_m on the coupling prism (2). This beam after reflection from the prism bottom falls on the screen (10). Simultaneous exponential decrease of electric field in the area of the buffer and aligning layers is present (A — A cross-section of the prism bottom). Since the refractive index in each point of the LC film is higher than that of the buffer, it is possible to choose such incidence angle, that guided modes are excited. The excited guided mode is radiated outside the waveguide by the same way through the buffer layers giving so called modal lines on the screen (8), (9).

For a perfectly isotropic and homogeneous waveguide only one mode would be excited with the same velocity as a proper projection of incident wave velocity into direction of propagation. Due to light scattering from LC medium the larger number of modes is excited causing more complicated image of modal lines.

---

[4] M.K. Barnoski: Introduction to the integrated optics, Plenum Press, New York and London, 1974, pp. 238—244.

[5] A. Kieżun, L.R. Jaroszewicz, E. Nowinowski-Kruszelnicki, A. Walczak: Głowica do pomiaru rozkładu współczynnika załamania w warstwach ciekłokrystalicznych, zgł. Patent RP P-310.765 z dnia 29.09.95.

[6] A. Kieżun, T. Patej, H. Działak: Metoda wyznaczania parametrów optycznych dyfuzyjnych wielomodowych falowodów planarnych, Biuletyn WAT, 1981, vol. 30, no. 3 (343), pp. 91—95.

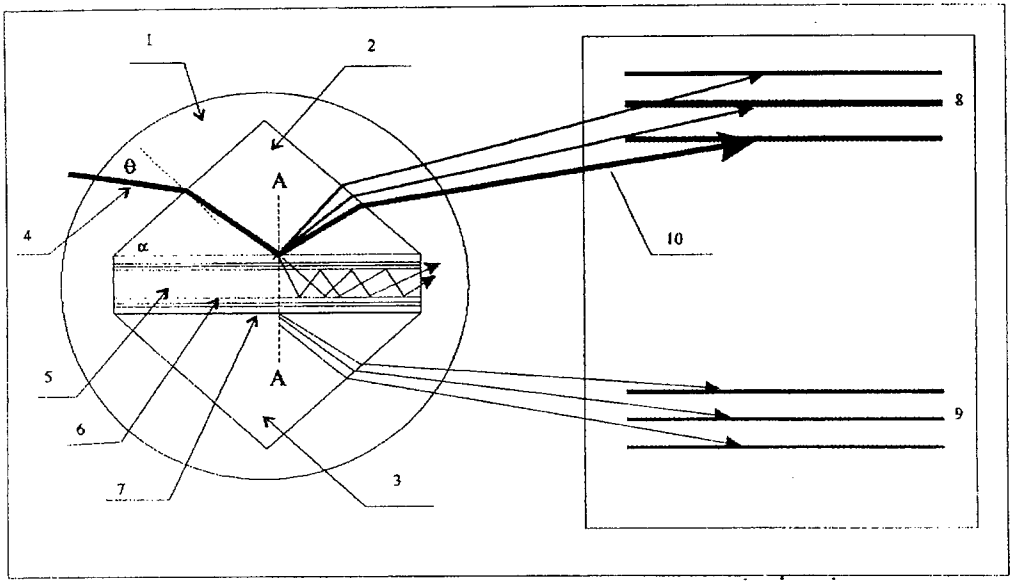


Fig. 1. Scheme of the measuring head[5]: I — revolving and sliding stage; 2, 3 — coupling prisms of high refractive index; 4 — incident light beam of known polarisation; 5 — investigated LC film being a planar waveguide; 6 — aligning layer, 7 — buffer layer, 8, 9 — modal lines, 10 — reflected beam

3. METHOD OF MEASUREMENT

From the condition of accordance between propagating wave phase velocity and incident wave velocity projection onto propagation direction one can obtain expression for normalised propagation constant of the mth mode.

$$N_m = \frac{\beta_m}{k_0} = n_p \sin\left(\alpha + \arcsin\frac{\sin\Theta_m}{n_p}\right), \tag{1}$$

where: β_m stands for propagation constants of the mth mode, $k_n = 2\pi/\lambda$ for wavenumber, and α for prism's angle. The Θ_m angle signifies incidence angle in which the mth mode is excited.

The measurement device construction allows one to measure modal lines in the case of nonsymmetric LC layer, too. The following data are obtained during measurement: N_m where m is changing from 0 to M, and N_k where k is changing from 0 to K. Characters m and k represent here numbers of the excited modes for coupling by one of the prisms in the measuring device. The coupling between TM and TE modes are present because of arbitrary orientation of the local optical axis in the LC layer. As a result, two families of the modal lines are observed independently from incidence light polarisation.

3.1. CALCULATION OF THE REFRACTIVE INDEX PROFILE

To calculate the refractive index profile of a separate mode the characteristic equation and measured propagation constants are used. This calculation consists in proper choice of the function adapted to the description of the profile to fulfil system of characteristic equations for given propagation. The characteristic equation may be written in the following form, in which the effect of prism and aligning layer on phase change on the waveguide-buffer border is neglected[4].

$$2k_0 \int_0^{x_m} (n^2(x) - N_m^2)^{1/2} \mathrm{d}x + 2\phi_{ca} + 2\phi_{cb} = 2\pi m, \tag{2}$$

where: m stands for a number of mode, x_m for quasi-classic turning point of the mth mode, and ϕ_{ca} and ϕ_{cb} for phase changes for total internal reflection in mode turning points. These phases are described by Fresnel's rules:

$$\phi_{ca,cb} = -\operatorname{arctg}\left(\frac{n^2(x_m)}{n_{a,b}^2}\right)^{\rho} \sqrt{\frac{N_m^2 - n_{a,b}^2}{n^2(x_m) - N_m^2}} \tag{3}$$

where: ρ is equal to 0 or 1 for TE and TM polarisation respectively, $n(x_m)$ stands for refractive index of the waveguide exactly on the turning point, n_a and n_b stands for a refractive indices of the neighbouring medium on the both sides of the waveguide. The system of equations (2) has been solved numerically. The algorithm of calculation is given in reference[6]. Results of that calculation have been applied to estimate the angular distribution of the director field in the investigated layer.

4. RESULTS

The measurements have been made for light wavelength λ equal to $0.6328*10^{-6}$m. The prism's refractive index for this wavelength is equal to 2.545. The buffer layers $1000*10^{-10}$ m thick for that wavelengths exhibits refractive index equal 1.46. The substance under investigation has refractive index $n_e = 1.6502$ and $n_0 = 1{,}5129$ (catalogue number W-602 from Prof. Dąbrowski laboratory in Institute of Chemistry MUT). To obtain homogeneous alignment of the nematic LC film rubbed polyimide layers has been applied. The LC layer thickness has been established on $6*10^{-6}$m by the glass spacers. The modal lines observed on the screen are illustrated in the Fig. 2.

The refractive index profile has been calculated by the procedure described above. Two different structures of the LC layer have been studied. The first one homogeneous (optical axis parallel to the waveguide walls and perpendicular to the direction of propagation), and the second twisted by 90°. The optical axes in twisted structure have been arranged in such a way, that on the one side of the waveguide optical axis is perpendicular to a direction of propagation while on the second is parallel. In the Fig. 3 the results for both structures have been presented.

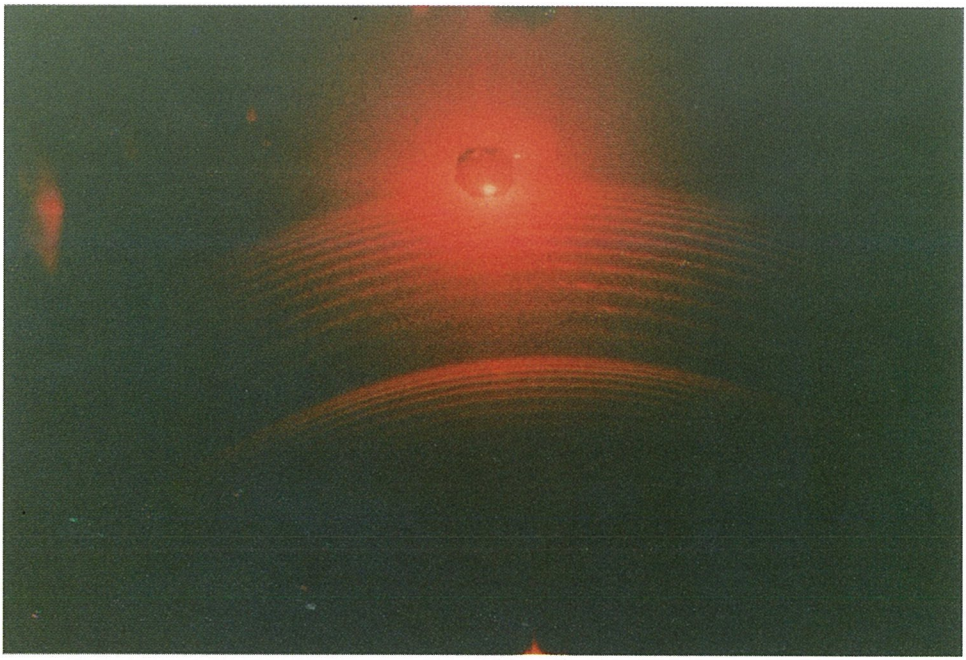


Fig. 2. The example of the modal lines oberved during measurement

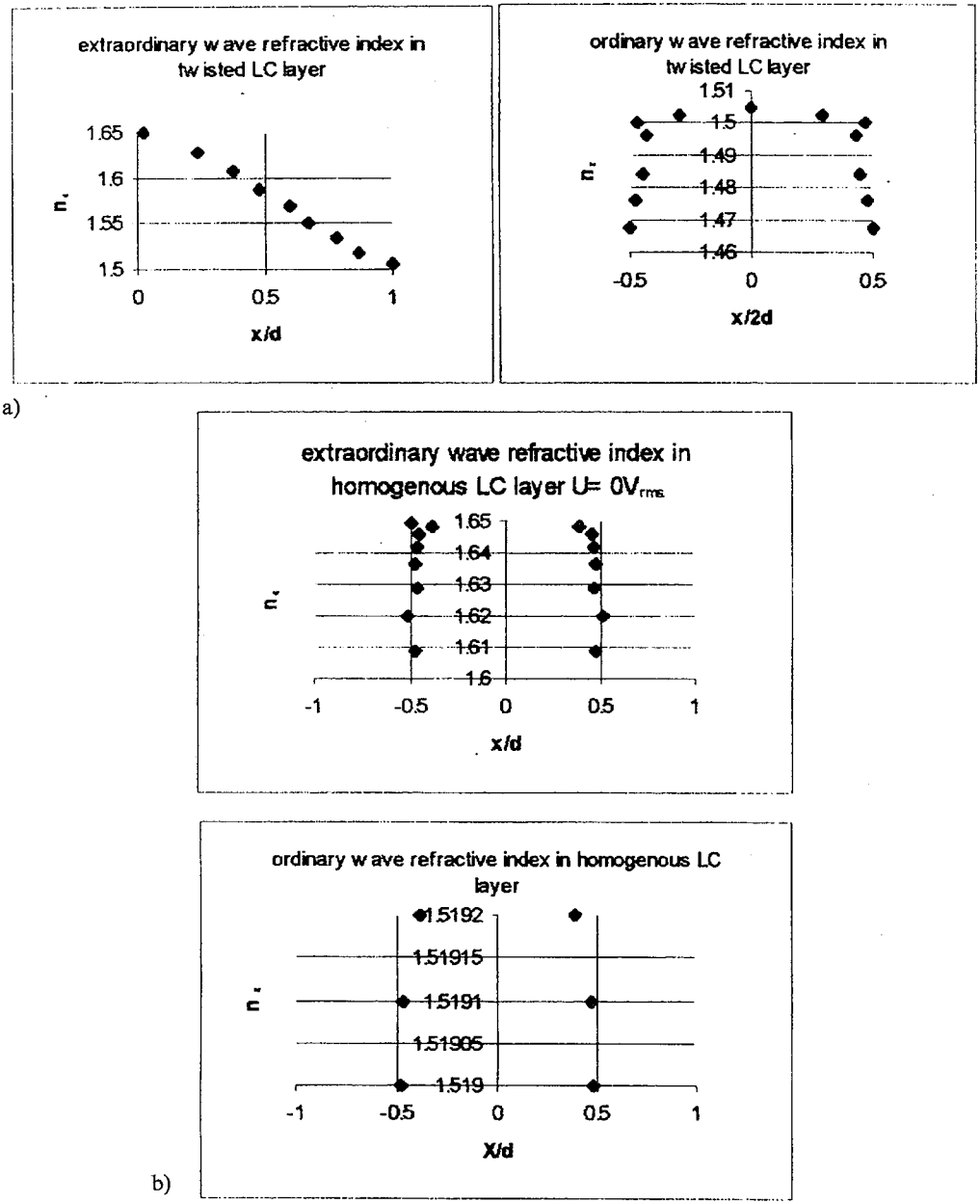
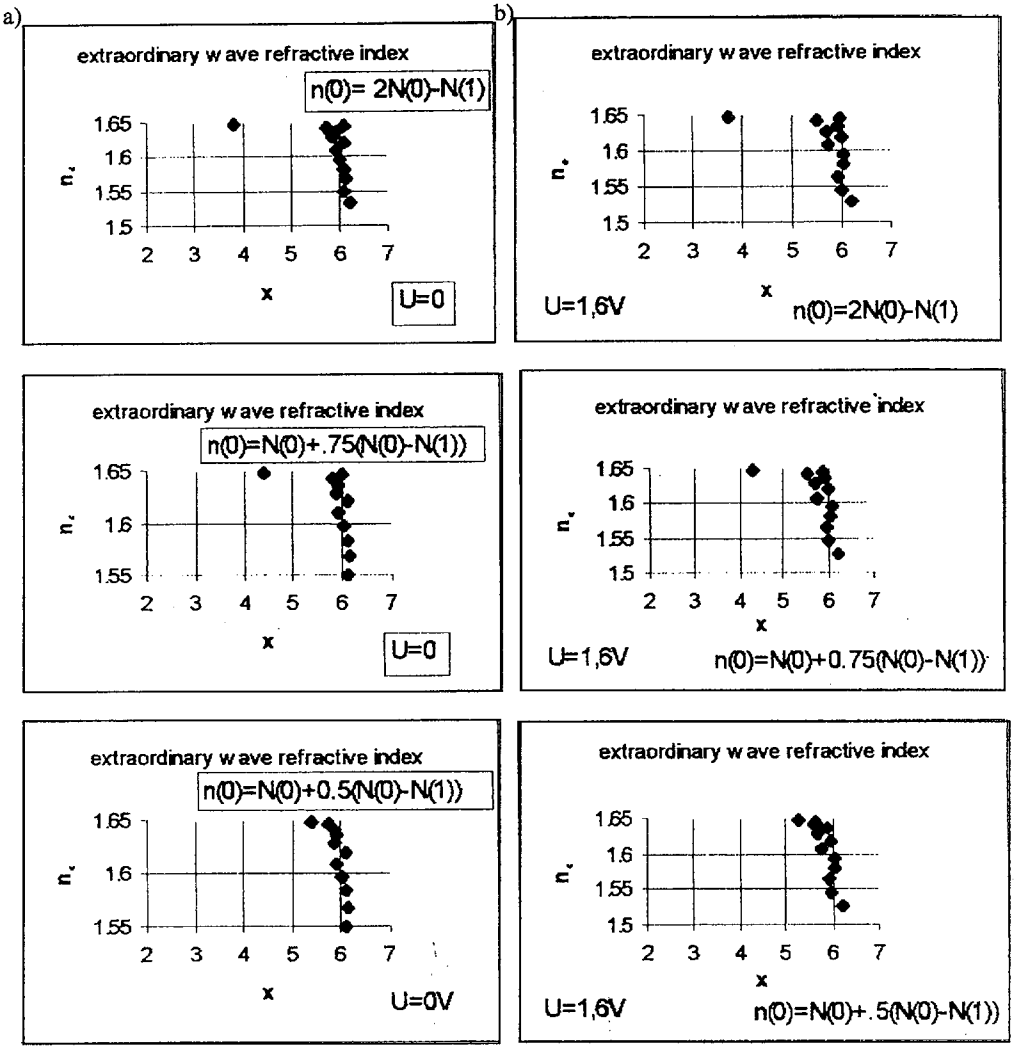


Fig. 3. The refractive indices for: a) twisted by 90° structure, b) homogeneous structure

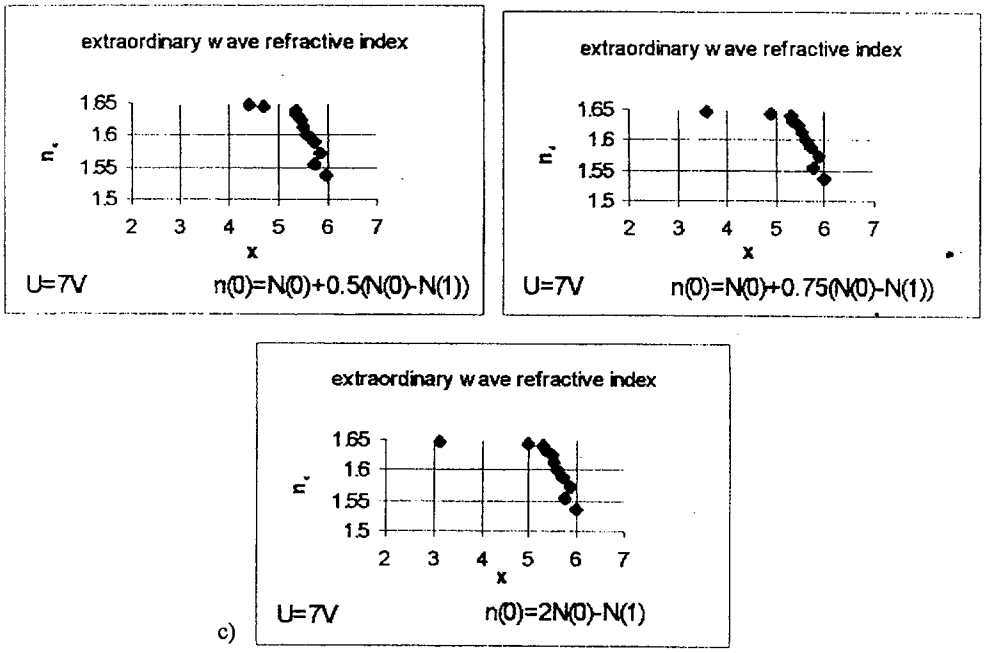
The refractive index profile detected by extraordinary wave is well visible in homoegenous layer but modal lines for ordinary wave in that layer are weak. The twisted layer refractive index profile illustrated in Fig. 3a is almost linear one for the extraordinary wave and almost constant for ordinary wave. It seems to be in accordance with our expectations. The driving voltage directed perpendicularly to



that layer has been applied to observe the profile dependency versus voltage amplitude.

Simultaneously the connection between $n(x)$ trial function and the final shape of the refractive index distribution have been analysed. It is illustrated in Fig. 4. In the driving voltage absence the refractive index profile in homogeneous layer is almost constant one. Only in the nearest vicinity of the waveguide wall there is a small profile disturbance. The $n(x)$ shape turned out to be really independent on the kind of trial function used during the profile calculation.

It is known that when amplitude of the driving voltage in homogeneous layer increases then the structure of the layers starts to be more and more homeotropic one. In the Fig. 4 observations above Fredericks transition (for that layer $V_{th}=0{,}8$ V) have been placed. The $n(x)$ profile in the vicinity of the waveguide edge when voltage amplitude increases up to 7 V and more. Visible shape of the $n(x)$ curve in the Fig. 4c



c)

Fig. 4. The $n(x)$ shape dependence on the trial function at U = 0 V (4a), U = 1,6 V (4b) and U = 7 V (4c). Formulae for trial function starting point are placed in

can be treated as resulting in interactions between liquid crystal and polyimide layer in terms of anchoring energy. It is known, that deformation induced in the LC layer by external voltage starts to change from the middle to the layer boundary and is growing towards the layer boundary. There exists near the layer boundary the part of the layer that is weakly reoriented especially in strong anchoring case. Its thickness decreases as the voltage amplitude increases and this is a reason for observed $n(x)$ shape.

5. REMARKS AND CONCLUSIONS

The error of the measurement of propagation constants is a result of the n_p and α angle estimation accuracy. For real measurement conditions an ultimate error of the measurement of mode propagation constant is of the order $5*10^{-4}$. The applied measuring set-up allows to measure propagation constant with the standard deviation δN of that order.

The second error source is an approximation of the characteristic equation (2). The turning point accuracy in presented calculation has been achieved not less than 10%. The effect of prism and intermodal coupling on the transversal resonance is not taken into account here. The obtained precision may be sufficient in many applications. Result for twisted layer and direct observations of the $n(x)$ profile in the waveguide edge vicinity seems to maintain the usefulness of the presented method to liquid crystalline waveguides investigation.

In the Fig. 4c it is easily seen that the pretilt angle treated as field unaffected in strong anchoring limit can not be assumed in this way especially in voltage range far above the Fredericks threshold. That result suggests the finite thickness of the boundary layer characterising by strong gradient of the tilt angle. This layer is field unaffected not only below the Fredericks transition but also just above it. Nevertheless in voltage range of proper value it is deformed.

A. WALCZAK, A. KIEŻUN, E. NOWINOWSKI-KRUSZELNICKI,
L.R. JAROSZEWICZ

ZASTOSOWANIE ŚWIATŁOWODU PLANARNEGO DO POMIARÓW WŁAŚCIWOŚCI WARSTW CIEKŁOKRYSTALICZNYCH

Streszczenie

Profil współczynnika załamania w cienkiej warstwie ciekłego kryształu został zmierzony przy zastosowaniu metody sprzęgacza pryzmatycznego. W celu wykonania pomiaru skonstruowano głowicę pomiarową, która umożliwiła wykorzystanie w tym pomiarze technik znanych w optyce zintegrowanej. Bezpośrednio mierzone są stałe propagacji modów falowodowych. W oparciu o ich wartości obliczany jest profil współczynnika załamania. Przedstawiono przykłady pomiarów dla warstwy homogenicznej oraz warstwy skręconego nematyka.

Słowa kluczowe: falowód ciekłokrystaliczny, profil współczynnika załamania.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1996. 42, z. 3, ss. 343–357

Nowe aspekty — antynomie — detekcji krawędzi obrazów i mikrometrii laserowej

BOHDAN WOŁCZAK

Instytut Elektroniki i Informatyki, Politechnika Szczecińska

Otrzymano 1996.06.24

Autoryzowano do druku 1996.08.08

W pracy przedstawiono nowe ograniczenia zastosowania kilku rozpowszechnionych kryteriów oceny szerokości pasm i detekcji krawędzi. Podano 60 współczynników szerokości pasm obliczonych dla dziesięciu typowych funkcji krawędzi $S(f)$ według sześciu zastosowanych definicji szerokości pasm.

Słowa kluczowe: detekcja krawędzi, mikrometria, kryteria oceny (definicje) szerokości pasm.

1. WSTĘP

Jednym z podstawowych pojęć elektrotechniki teoretycznej, optyki, łączności, automatyki, spektroskopii i mikroelektroniki jest określenie „Szerokości pasma zajmowanego przez sygnał". Szerokości pasm, nazywane często skrótowo „Pasmami" można różnorodnie definiować. Istniejące definicje pasm częstotliwości sygnałów nie są bezpośrednio wzajemnie powiązane, gdyż mają umowny charakter wynikający z rożnych międzynarodowo przyjętych konwencji. Z większymi trudnościami można się spotkać wtedy gdy na podstawie istniejących definicji należy udzielić jednoznacznej odpowiedzi na pozornie proste pytanie: „Czy przy określonej zmianie kształtu badanej krzywej $S(f)$ jej umowna szerokość (pasmo) ulegnie zawężeniu czy też rozszerzeniu?".

Na tak postawione pytanie uzyskuje się niejednoznaczne odpowiedzi. Dlatego złożona problematyka pomiarów lub obliczeń szerokości różnych możliwych krzywych $S(f)$ obrazujących rozkłady szarości krawędzi elementów struktur powinna być związana z badaniem zakresu przydatności wszystkich sposobów określania pasm.

Jakiekolwiek badania nad metodyką pomiarów zmian szerokości pasma mogą być efektywne tylko w tym przypadku, jeżeli definicje szerokości pasm, będące podstawowym punktem wyjścia do takich badań, będą niesprzeczne dla wszystkich fizykalnie realizowanych sygnałów. Rozważania dotyczące skutków zmian szerokości dowolnych elementów obrazujących rozkłady pól interferencyjnych

albo odwzorowujących zmiany drobnych elementów (na przykład ścieżek przewodzących prąd) w mikro-elektronicznych obwodach scalonych pokazują, że wyniki badań pozornych sprzeczności (antynomii) w ocenie zmian szerokości pasm [1] mogą być bezpośrednio przenoszone na ten nowy grunt w technologii obróbki obwodów elektronicznych i mikrometrii laserowej [2–7].

W artykule podano 60 nowych wartości współczynników szerokości pasm dla 10 typowych kształtów rozkładów szarości krawędzi (konturów) struktur (obrazów) przy 6 różnorodnych kryteriach pracy detektorów krawędzi, wraz z omówieniem nowych aspektów (antynomii) pomiarów położenia krawędzi.

2. BADANE FUNKCJE $S(f)$ ROZKŁADU SZAROŚCI KRAWĘDZI

W celu wykrycia i porównania ewentualnych antynomii różnorodnych kryteriów oceny zmian szerokości pasm opisanych na przykład w pracy [1] przyjęto następujące kształty funkcji $S(f)$ w podjętych doświadczeniach numerycznych:

$$S_1(f)=5.368\frac{\pi}{[1+10f^2]^3}; \tag{1}$$

$$S_2(f)=33.918\pi\exp(-16\pi f); \tag{2}$$

$$S_3(f)=\frac{1.25}{\pi}\exp\left(-\frac{f^2}{8}\right); \tag{3}$$

$$S_4(f)=\frac{1.1}{\pi}\left[\frac{\sin(0.5\pi f)}{f}\right]^3 \tag{4}$$

$$S_5(f)=\frac{0.03}{\pi}\left[\frac{\sin(\pi f)}{f}\right]^5; \tag{5}$$

$$S_6(f)=\frac{0.88}{\sqrt{(5+f)^2+(1+f)^2}}; \tag{6}$$

$$S_7(f)=\frac{\sqrt{(f^2+\pi)}}{2.8194055\sqrt{(f^5+2\pi)}}; \tag{7}$$

$$S_8(f)=\frac{0.56257}{\sqrt{(1-1.5515f^2)^2+1.3022f^2}}; \tag{8}$$

$$S_9(f)=\frac{1.4194722}{1+2\pi\sqrt{f^3}}; \tag{9}$$

$$S_{10}(f)=\frac{0.08127939(\pi+f)}{\sqrt{0.5+0.5f^2}\,\sqrt{0.5f^2+(0.5-f^2)^2}}; \tag{10}$$

które posiadają własność $\int_0^\infty S(f)\mathrm{d}f=1$, to znaczy że są znormowane.

Tablica 1

Wyniki obliczeń szerokości pasm (df) według różnych definicji — VI kryteriów

(df) $S(f)$	I kryterium „momentu"	II kryterium „Jedn. pola"	III kryterium „spadku o 3 dB"	IV kryterium „90% w paśmie"	V kryterium „99% w paśmie"	VI kryterium „informacyjne" W_0
dla $S_1(f)$	0,0912829	0,1862891	0,1612206	0,2850480	0,5698740	0,2722960
dla $S_2(f)$	0,0208485	0,0093846	0,0137897	0,0109943	0,0125190	0,0124068
dla $S_3(f)$	1,0000611	2,5061966	2,3548200	3,2884080	5,1394950	3,4498347
dla $S_4(f)$	0,2619660	0,7368813	2,0000000	0,8924397	1,1711064	0,9793580
dla $S_5(f)$	0,1149601	0,2994788	1,0000000	0,3820863	0,5714442	0,4070123
dla $S_6(f)$	14,830762	2,7796113	1,9921648	8,7588697	81,958303	4,3673730
dla $S_7(f)$	43,237718	3,7885461	2,1979785	41,239624	709,51547	6,0745698
dla $S_8(f)$	9,5204787	1,4483642	1,1330783	3,3402103	17,817839	2,0367373
dla $S_9(f)$	34,505200	0,7044872	0,7361562	17,173174	421,18951	1,8584000
dla $S_{10}(f)$	9,3242359	1,1549771	0,9525368	2,0049440	12,759913	1,5404640

3. WYNIKI OBLICZEŃ SZEROKOŚCI PASM

W celu obliczenia szerokości pasm badanych funkcji $S(f)$ według kryteriów „momentu znormowanego widma", „jednostkowego pola", „spadku o 3 dB", „energetycznego 0.9% i 0.99%" oraz „informacyjnego" [1] napisano program w języku C (TC++1.01) wykorzystujący metodę całkowania Gaussa. Otrzymane wyniki, liczone na komputerze typu IBM 386 SX przedstawiono w Tablicy 1.

4. OMÓWIENIE WYNIKÓW BADAŃ

Wyniki obliczeń podane w Tablicy 1 można ułożyć w określonej kolejności.

Tablica 2 pokazuje numery badanych sygnałów ułożone według wzrastających szerokości pasm dla poszczególnych definicji:

Tablica 2

Nowe antynomie detekcji krawędzi

$(df)_I$	2	1	5	4	3	10	8	6	9	7
$(df)_{II}$	2	1	5	9	4	10	8	3	6	7
$(df)_{III}$	2	1	9	10	5	8	6	4	7	3
$(df)_{IV90\%}$	2	1	5	4	10	3	8	6	9	7
$(df)_{V99\%}$	2	1	5	4	3	10	8	6	9	7
$(df)_{VI}$	2	1	5	4	10	9	8	3	6	7

Z tablicy tej widać, że kolejność wyżej podanych numerów sygnałów jest zgodna jedynie dla dwóch definicji: „momentu znormowanego widma" $(df)_I$ oraz „99% mocy w definiowanym pasmie" $(df)_{99\%}$. Pozostałe definicje szerokości pasm dały wyniki o sprzecznej wzajemnie kolejności uszeregowania numeracji sygnałów (od najwęższego do najszerszego pasma). Szczególnie interesującym przypadkiem antynomii jest (przy zastosowaniu energetycznych kryteriów mocy sygnału w definiowanym pasmie typu „90%" i „99%") przypadek zmiany kształtów $S(f)$ z kształtu Nr 3 na kształt Nr 10 i odwrotnie. Pozornym sprzecznościom, dla wszystkich definicji, nie podlegają zmiany kształtów rozkładów $S(f)$ Nr 1 oraz Nr 2. Inne antynomie zaznaczono w Tablicy 2 podwójnym obramowaniem.

Zgodność kryterium „99% mocy sygnału w definiowanym pasmie" z kryterium „momentu znormowanego widma" nie jest w ogólnym przypadku prawdziwa. W [1] pokazano, że zgodność wystąpiła tylko dla definicji: „90% mocy w definiowanym pasmie" oraz „informacyjnego kryterium W_0", z pewnym (antynomicznym) wyjątkiem. Liczne bardziej złożone antynomie wynikają z innych, zaznaczonych w Tablicy 2 par lub triad numerów rozkładów $S(f)$.

Wyniki badań pozornych sprzeczności (antynomii) w ocenie zmian szerokości pasm częstotliwości [1] mogą być bezpośrednio przenoszone na nowy grunt na przykład w technologii obróbki obwodów elektronicznych, co wiąże się z zagadnieniami mikrometrii laserowej.

5. WNIOSKI I ZAKOŃCZENIE

1) Pokazano, że zagadnienie zmiany szerokości pasm (przy jednoznacznej zmianie kształtu sygnału) nie może być rozwiązane przy pomocy jakiegokolwiek kryterium szerokości pasma — w sposób niesprzeczny z pozostałymi kryteriami określania pasm.
2) Opisane w literaturze fachowej sposoby zawężania pasm częstotliwości (lub wydzielania konturów z półtonowych obrazów), mogą według niektórych definicji okazać się w rzeczywistości, dla niektórych zmian kształtów sygnałów, systemami rozszerzającymi pasmo.
3) Niejednoznaczność pojęcia „szerokości pasma częstotliwości zajmowanego przez sygnał" powoduje, że zmiany szerokości pasm powinny być teoretycznie i eksperymentalnie badane w oparciu o wszystkie stosowane kryteria.
4) Celowym jest rozszerzenie badań przydatności dotychczas stosowanych kryteriów oceny zmian szerokości różnych kształtów krzywych rozkładu na te, które mogą być lub już są stosowane do detekcji krawędzi w mikroelektronice i optoelektronice.

Przy pomiarach deformacji obiektów, na skutek oddziaływania czynników zewnętrznych, krawędź jest podstawowym nośnikiem informacji o badanej strukturze. Należy ona do podstawowych cech obiektów podlegających operacjom analizy lub rozpoznawania obrazów. Obecnie stosuje się wiele metod detekcji krawędzi, jednakże w przypadku szybkich (lub znacznych) zmian kształtu detekcja ta powinna być wystarczająco szybka, precyzyjna i przede wszystkim — niesprzeczna.

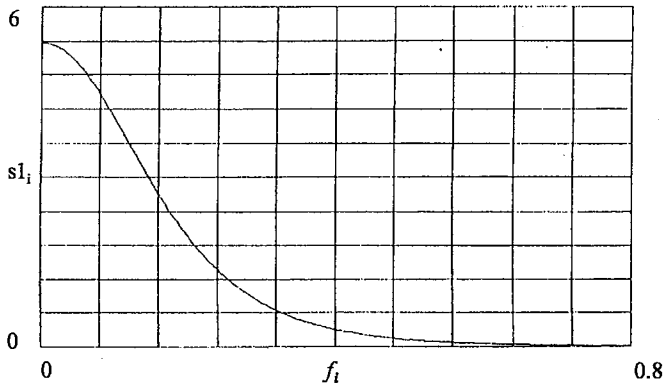
BIBLIOGRAFIA

1. B.K. Wołczak: *Antinomies of different criteria for evaluation bandwidths in images and laser measurements*. Proc. SPIE, vol. 2203, Laser Technology IV: Research Trends, Instrumentation and Application in Metrology and Material Processing, Szczecin – Świnoujście (Poland), September 1993, Ed. Bellingham, Washington USA 1995, pp. 582 – 587
2. C.P. Kirk: *Development and analysis of an automated dimension measuring microscope*. J. Microscopy 1987, vol. 147, pp 115 – 128
3. D. Nyyssonnen: *Linewidth measurement with an optical microscope: the effect of operating conditions on the image profile*. Applied Optics 1977, vol. 16, pp. 2223 – 2230
4. F. Schedewie, M. Kallmeyer, B. Solf, D. Wagner: *Ein neues elektrooptisches Kantefindermikroskop für die Halbleitertechnologie*. International Elektronische Rundschau 1971, Nr 4, S. 93 – 95
5. D. Nyyssonen: *Theory of optical edge detection and imaging of thick layers*. J. Opt. Soc.Am. 1982, vol. 72, pp. 1425 – 1436
6. D. Nyyssonen: *Practical method for edge detection and focusing for linewidth measurements on wafers*. Optical Engineering, 1987, vol. 26, pp. 81 – 85
7. F. Schedewie: *Kanteneinfangmikroskopie mit digitaler Objektabtastung*. Optik, 1973, vol. 37, S. 404 – 414
8. E.R. Kretzmer: *Statistics of television signals*. Bell System Technical Journal 1952, Nr 4
9. J. Seidler: *Statystyczna teoria odbioru sygnałów*, PWN, Warszawa – Wrocław 1963
10. B.K. Wołczak: *Porównanie rozmaitych kryteriów oceny zmian szerokości pasma losowych sygnałów*. Archiwum Elektrotechniki 1968, vol. 17, Nr 3, ss. 499 – 514

DODATEK 1

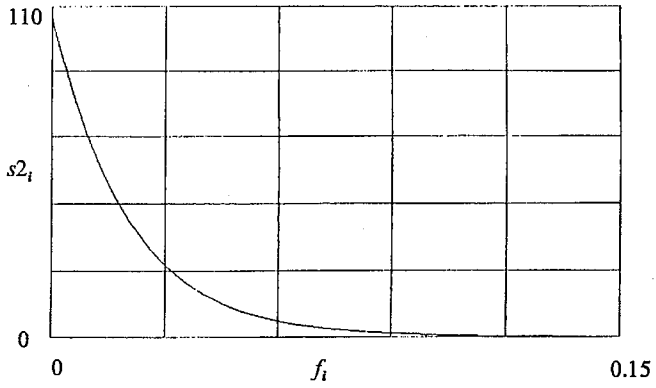
WYKRESY BADANYCH KRZYWYCH $S(f)$

$i:=0..300$ $\mathrm{dx}:=0.005$ $f_i:=\mathrm{dx}\cdot i$; Gdzie:
i — ilość obliczanych punktów
dx — przedział
f_i — zmienna, dla której obliczono wyrażenie: $s1_i:=\dfrac{5.368\cdot\pi}{[1+10\cdot f_i^2]^3}$



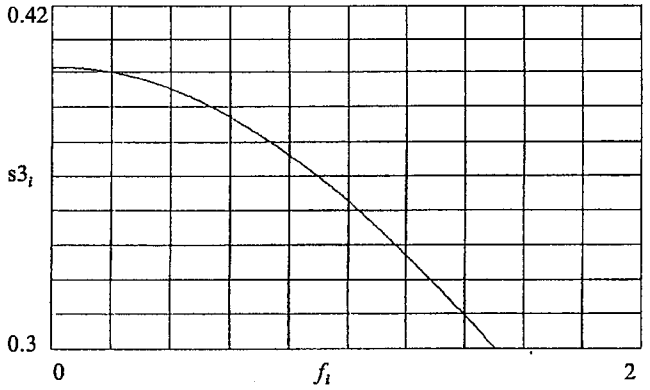
Rys. 1. Wykres widma $S1(f)$

$i:=0..300$ $\mathrm{dx}:=0.005$ $f_i:=\mathrm{dx}\cdot i$; Gdzie:
i — ilość obliczanych punktów
dx — przedział
f_i — zmienna, dla której obliczono wyrażenie: $s2_i:=33.91806\cdot\pi\cdot\exp[-16\cdot\pi\cdot f_i]$



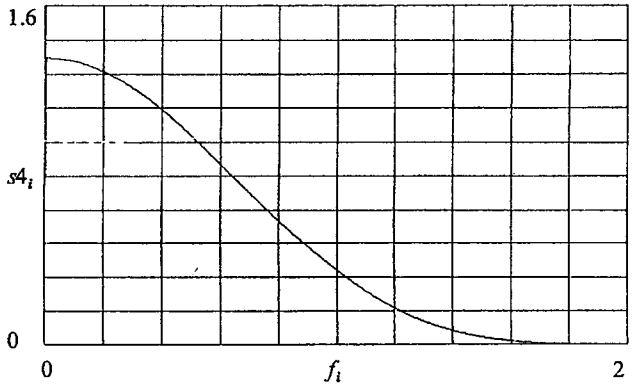
Rys. 2. Wykres widma $S2(f)$

$i := 0..300$ $dx := 0.005$ $f_i := dx \cdot i$; Gdzie:
i — ilość obliczanych punktów
dx — przedział
f_i — zmienna, dla której obliczono wyrażenie: $s3_i := \frac{1.25}{\pi} \cdot \exp\left[\frac{-f_i^2}{8}\right]$



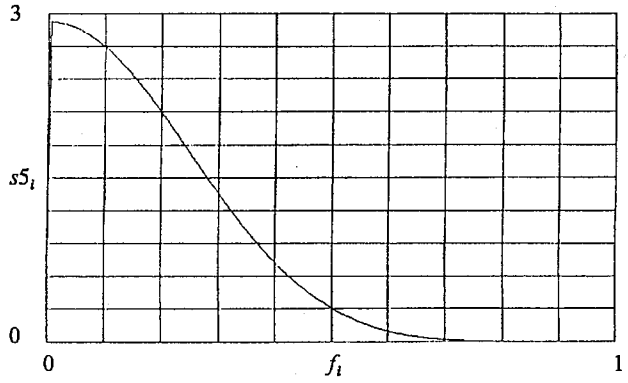
Rys. 3. Wykres widma $S3(f)$

$i := 0..400$ $dx := 0.005$ $f_i := dx \cdot i$; Gdzie:
i — ilość obliczanych punktów
dx — przedział
f_i — zmienna, dla której obliczono wyrażenie: $s4_i := \frac{1.1}{\pi} \cdot \left[\frac{\sin[0.5 \cdot \pi \cdot f_i]}{f_i}\right]^3$



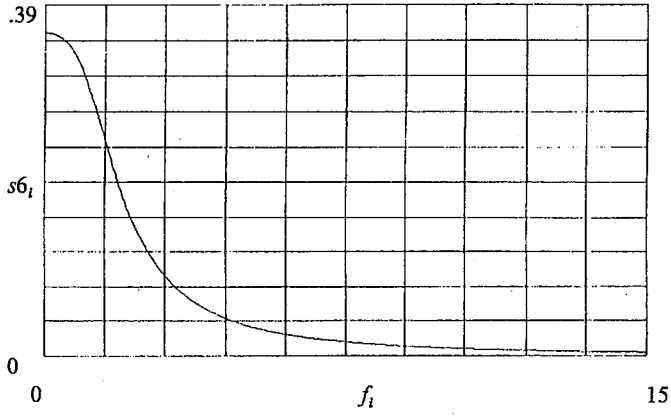
Rys. 4. Wykres widma $S4(f)$

$i := 0..300$ $dx := 0.005$ $f_i := dx \cdot i$; Gdzie:
i — ilość obliczanych punktów
dx — przedział
f_i — zmienna, dla której obliczono wyrażenie: $s5_i := \frac{0.03}{\pi} \cdot \left[\frac{\sin[\pi \cdot f_i]}{f_i}\right]^5$



Rys. 5. Wykres widma $S5(f)$

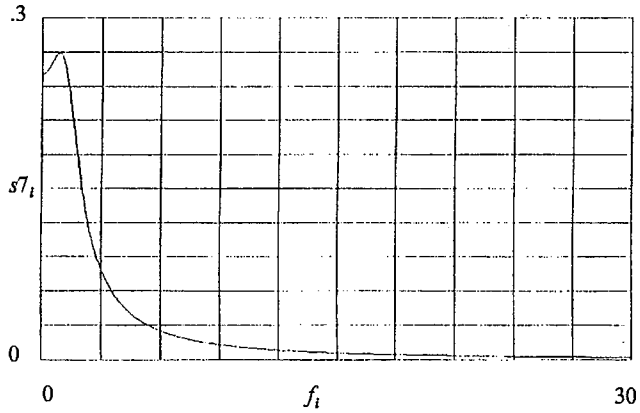
$i:=0..300$ $\mathrm{dx}:=0.05$ $f_i:=\mathrm{dx}\cdot i$; Gdzie:
i — ilość obliczanych punktów
dx — przedział
f_i — zmienna, dla której obliczono wyrażenie: $s6_i:=\dfrac{0.88}{\sqrt{[5+f_i]^2+[1+f_i]^2}}$



Rys. 6. Wykres widma $S6(f)$

$i:=0..300$ $\mathrm{dx}:=0.1$ $f_i:=\mathrm{dx}\cdot i$
Gdzie:
i — ilość obliczanych punktów
dx — przedział
f_i — zmienna, dla której obliczono wyrażenie:

$$s7_i:=\frac{1}{2.8194055}\cdot\frac{\sqrt{f_i^2+\pi}}{\sqrt{f_i^2+2\cdot\pi}}$$



Rys. 7. Wykres widma $S7(f)$

$i:=0..300$ $dx:=0.04$ $f_i:=dx\cdot i$

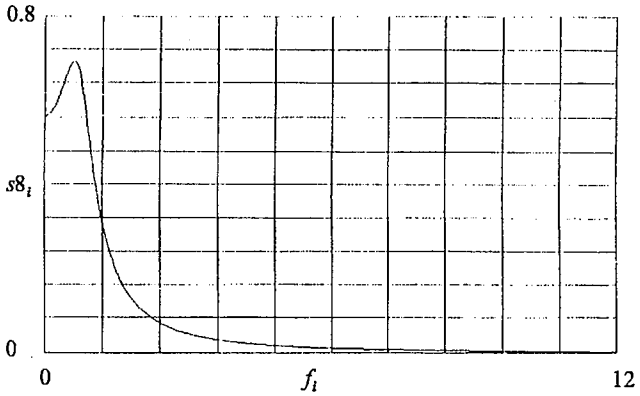
Gdzie: i — ilość obliczanych punktów

dx — przedział

f_i — zmienna, dla której obliczono wyrażenie:

$$s8_i:=\frac{0.56258235}{\sqrt{[1-1.5515\cdot f_i^2]^2+1.3022\cdot f_i^2}}$$

$i:=0..300$ $dx:=0.05$ $f_i:=dx\cdot i$



Rys. 8. Wykres widma $S8(f)$

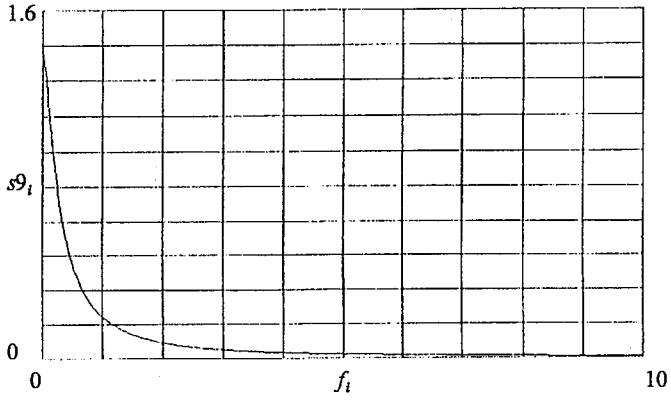
Gdzie:

i — ilość obliczanych punktów

dx — przedział

f_i — zmienna, dla której obliczono wyrażenie:

$$s9_i:=\frac{1.4294722}{1+2\cdot\pi\cdot\sqrt{f_i^3}}$$



Rys. 9. Wykres widma $S9(f)$

$i:=0..100 \quad dx:=0.05 \quad f_i:=dx\cdot i$

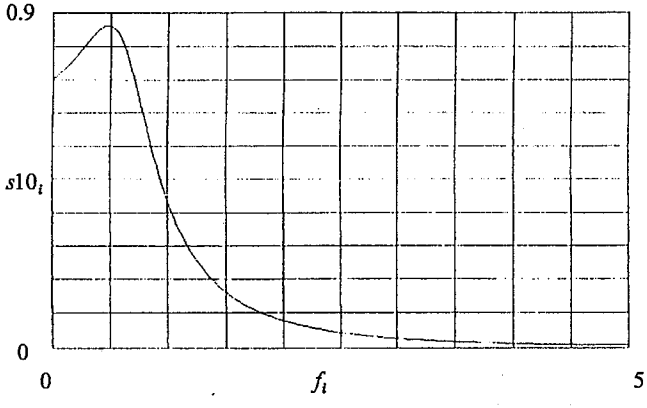
Gdzie:

i — ilość obliczanych punktów

dx — przedział

f_i — zmienna, dla której obliczono wyrażenie:

$$s10_i:=\frac{\pi+f_i}{\sqrt{0.5+0.5\cdot f_i^2}}\cdot\frac{0.08127939}{\sqrt{0.5\cdot f_i^2+[0.5-f_i^2]^2}}$$



Rys. 10. Wykres widma $S10(f)$

DODATEK 2

RÓŻNORODNOŚĆ DEFINICJI I KRYTERIÓW OCENY SZEROKOŚCI PASM

Problemy zawężenia pasm nie mogą być w pełni rozwiązane bez dokładniejszej teoretycznej analizy — które pasmo i w jakich warunkach może ulec zwężeniu. Szerokości pasma częstotliwości oblicza się z widma znormowanej gęstości mocy $S(f)$ posiadającego własność

$$\int_0^\infty S(f)\mathrm{d}f = 1 \tag{1}$$

i jest ono powiązane, jak podano w [10], z nieznormowanym widmem gęstości mocy $G(f)$ zależnościami:

$$S(f) = \frac{G(f)}{\int_0^\infty G(f)\mathrm{d}f}, \tag{2}$$

$$G(f) = 4\left[\int_0^\infty y(t)\cos 2\pi f\,\mathrm{d}t\right]^2, \tag{3}$$

gdzie $y(t)$ jest analizowanym sygnałem, t czasem, f częstotliwością. Jeżeli dla pewnych sygnałow, na przykład losowych, ich analityczna postać nie jest znana, znana jest natomiast ich znormowana funkcja autokorelacji $a(\tau)$ uzyskana teoretycznie lub eksperymentalnie zmierzona:

$$a(\tau) \overset{\mathrm{df}}{=} \frac{A(\tau) - (\bar{y})^2}{A(0) - (\bar{y})^2}, \tag{4}$$

gdzie wartość średnią sygnału $y(t)$ określa się

$$\bar{y} \overset{\mathrm{df}}{=} \lim_{T\to\infty} \frac{1}{T}\int_0^T y(t)\mathrm{d}t, \tag{5}$$

zaś nieznormowaną funkcję autokorelacji

$$A(\tau) \overset{\mathrm{df}}{=} \lim_{T\to\infty} \frac{1}{T}\int_0^T y(t)y(t+\tau)\mathrm{d}t, \tag{6}$$

to widmo znormowanej gęstości mocy $S(f)$ można obliczyć z przekształcenia Wienera-Kołmogorowa-Chinczyna [10]

$$S(f) = 4\int_0^\infty a(\tau)\cos 2\pi f\tau\,\mathrm{d}\tau \geqslant 0. \tag{7}$$

Definicje szerokości pasm podane w monografii Seidlera [9] można zapisać w uproszczonej postaci.

2.1. Kryteria momentu znormowanego widma

Szerokość pasma $(\mathrm{df})_{\mathrm{I}}$ jest równa połowie kwadratowego pierwiastka momentu znormowanego widma gęstości mocy $S(f)$".

$$(\mathrm{df})_{\mathrm{I}}=\frac{1}{2}\sqrt{\int_0^{\infty} f^2 S(f)\mathrm{df}}. \tag{8}$$

2.2. Kryterium jednostkowego pola w pasmie

„Iloczyn definiowanej szerokości pasma $(\mathrm{df})_{\mathrm{II}}$ i maksymalnej wartości znormowanego widma gęstości mocy $S(f)$ jest dla synałów dolnopasmowych równy jedności".

$$(\mathrm{df})_{\mathrm{II}}=\frac{1}{\max_f S(f)}. \tag{9}$$

2.3. Kryterium spadku widma do określonej wartości

„Jeżeli znormowane widmo gęstości mocy $S(f)$ monotonicznie maleje dla $0<f\to\infty$, to szerokość pasma $(\mathrm{df})_{\mathrm{III}}$ może być określona częstotliwoscią, przy której widmo to jest równe określonej części (zwykle połowie) maksymalnej wartości $S(f)$".
W skrócie jest ono zwykle nazywane kryterium spadku widma o około 3 dB.

$$0.5\max_f S(f)=S[(\mathrm{df})_{\mathrm{III}}]. \tag{10}$$

Dla widm osiągających okresowo wartości zerowe, szerokości pasm $(\mathrm{df})_{\mathrm{III}}$ są przedziałem częstotliwości zawartym między częstotliwością, której odpowiada *maksimum maksimorum* $S(f)$ a najbliższą częstotliwością przy której $S(f)=0$.

2.4. Energetyczne kryteria określonej części całkowitej mocy sygnału poza definiowanym pasmem

„Szerokość pasma $(\mathrm{df})_{\mathrm{IV,V}}$ określa się umowną częścią całkowitej mocy nadawania sygnału. Zwykle przyjmuje się 10% lub 1% całkowitej mocy sygnału poza definiowanym pasmem".

$$\int_0^{(\mathrm{df})_{\mathrm{IV}}} S(f)\mathrm{df}=0.9, \tag{11}$$

$$\int_0^{(\mathrm{df})_{\mathrm{V}}} S(f)\mathrm{df}=0.99. \tag{12}$$

2.5. Informacyjne kryterium likwidacji nadmiaru

„Szerokość pasma W_0 określa się taką górną częstotliwością przenoszenia kanału, przy której nieznormowany dolnopasmowy nadmiar informacji $R(W)$ zawartej w sygnale przechodzi przez wartość zerową".

$$R(W_0)=0. \tag{13}$$

Nieznormowany, dolnopasmowy nadmiar informacji $R(W)$ wyrażony w jednostkach dwójkowych [bit/element] można obliczyć z widma $S(f)$ według wzoru zapisanego przez Kretzmera [8] przy pomocy logarytmów dwójkowych

$$R(W)=\frac{1}{2W_0}\int_0^W \log_2 S(f)\mathrm{d}f+\frac{1}{2}\log_2(W)-\log_2\int_0^W S(f)\mathrm{d}f, \tag{14}$$

a przy oznaczeniu

$$N(W)\overset{\mathrm{df}}{=}R(W)\ln 2; \tag{15}$$

można operować nadmiarem „naturalnym":

$$N(W)=\frac{1}{2W}\int_0^W \ln S(f)\mathrm{d}f+\frac{1}{2}\ln W-\ln\int_0^W S(f)\mathrm{d}f. \tag{16}$$

Stosując znany wzór na pochodną górnej granicy całkowania

$$\frac{\mathrm{d}}{\mathrm{d}z}\int_a^z f(x)\mathrm{d}x=f(z), \tag{17}$$

w trakcie likwidowania nieoznaczoności regułą d'Hospitala, można bez trudu udowodnić:

$$N(0)=\infty. \tag{18}$$

Z obliczeń wykonanych w pracy [10] wynika, że dla niektórych widm spełnia się

$$N(\infty)=-\infty \tag{19}$$

i istnieje jedna wartość W_0, dla której

$$N(W_0)=0. \tag{20}$$

Jakkolwiek dowód istnienia jednoznaczności rozwiązań W_0 równań (13) lub (20) nie został przeprowadzony dla szerszej klasy fizykalnie realizowalnych widm $S(f)$, to jednak informacyjne kryterium szerokości pasma W_0 posia/da kilka zalet, z których najważniejszą jest ta, iż uwzględnia ono najistotniejszą cechę sygnału jaką jest ilość przenoszonej przez sygnał informacji. Zalety tej nie posiada żadne z dotychczas istniejących kryteriów oceny szerokości pasma.

B. WOŁCZAK

NEW ASPECTS — ANTINOMIES — OF EDGE DETECTION IN IMAGES AND LASER MICROMETRY

S u m m a r y

The aim of this work is to present a new limits of application of some often applied criterions of evaluation bandwidths and edge detections. There were given a 60 coefficients of bandwidths calculated for ten typical $S(f)$ edge functions according to the six definitions of bandwidths applied.

Key words: edge detection, micrometry, criterions of bandwidths.

SPIS TREŚCI

S. Koziel: Niejednorodne i niestacjonarne operatory mutacji w optymalizacji numerycznej przy użyciu algorytmów genetycznych 273
B. Seroczyńska-Wojas, J. Wojas: Doświadczalne określenie zagięcia pasm przy powierzchni półprzewodnika 287
K. Wawryn: Układowe realizacje prądowych sztucznych sieci neuronowych 295
A. Borys: Model idealny dla szacowania czasu opóźnienia w technice pomiarowej przepływów 325
A. Walczak, A. Kieżun, E. Nowinowski-Kruszelnickiu, L.R. Jaroszewicz: Zastosowanie światłowodu planarnego do pomiarów właściwości warstw ciekłokrystalicznych 335
B. Wołczak: Nowe aspekty detekcji krawędzi obrazów i mikrometrii laserowej 343

CONTENTS

S. Koziel: Non-uniform and non-stationary mutation in numerical optimization using genetic algorithm 273
B. Seroczyńska-Wojas, J. Wojas: Experimental determination of the band bending the surface of semiconductors 287
K. Wawryn: Current mode artificial neural networks 295
A. Borys: An idealized model for time-delay estimation in turbulent flows 325
A. Walczak, A. Kieżun, E. Nowinowski-Kruszelnickiu, L.R. Jaroszewicz: Application of planar waveguide technique to investigation of liquid crystals 335
B. Wołczak: The new aspects antinomies of edge detection in images and laser micrometry 343

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 50 stron po około 1800 znaków na stronie.

Wymagania podstawowe: Artykuły należy nadsyłać w maszynopisie pisanym jednostronnie lub na wyraźnym czarno-białym wydruku komputerowym, w 2 egzemplarzach, w języku polskim lub angielskim — wybranym przez autora. Tekst artykułu musi być poprzedzony tytułem pracy, imieniem i nazwiskiem Autora wraz z podaniem miejsca jego pracy. Wszystkie strony muszą mieć numerację ciągłą.

Sposob pisania tekstu: Tekst powinien być pisany bez używania wyróżnień, a w szczególności nie dopuszcza się spacjowania, podkreślania 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 powinen być napisany z podwójnym odstępem między wierszami, tytuły i podtytuły małymi literami. Marginesy z każdej strony powinny mieć około 35 mm. Przy podziale pracy na rozdziały i podrozdziały cyfrowe ich oznaczenia nie powinny być większe niż III stopnia (np. 4.1.1).

Sposób pisania tablic: Tablice powinny być pisane na oddzielnych stronach. Tytuły rubryk pionowych i poziomych powinny być napisane małymi literami z podwójnym odstępem między wierszami. Przypisy (notki) dotyczące tablic należy pisać bezpośrednio pod tablicą. Tablice należy numerować kolejno liczbami arabskimi, u góry każdej tablicy podać tytuł. Tablice umieścić na końcu maszynopisu. Przyjmowane są tablice algorytmów i programy na wydrukach komputerowych. W tym przypadku zachowany jest ich oryginalny układ.

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 (strzalki, linie, kropki, daszki) powinny być umieszczone dokładnie nad tymi elementami, do których się odnoszą. Numery wzorów cyframi arabskimi powinny być kolejne i umieszczone w nawiasach okrągłych z prawej strony. Nazwy jednostek, symbole literowe i graficzne powinny być zgodne z wytycznymi IEC (International Electronical Commission) oraz ISO (International 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 treści artykułu. Przykłady opisu publikacji:

— periodycznej: F.Valdoni: A new millimetre wave satelite. E.T.T. 1990, vol. 2, nr 5, p. 553
— nieperiodycznej: K. Andersen: A resourse allocation framework. XVI International Symposium. Stockholm (Sweden), May 1991, paper A2.4
— książki: Y.P. Tvidis: Operation and modelling of the MOS transistors. Mc Graw-Hill, New York 1987, p. 141–148.

Material ilustracyjny: Rysunki powinny być wykonane wyraźnie, na papierze gładkim, lub milimetrowym w formacie nie mniejszym niż 9x12 cm. Mogą być także w postaci wydruku komputerowego. Fotografie lub diapozytywy przyjmowane są raczej kolorowe lub czarno-białe w formacie nie przekraczającym 10 × 15 cm. Na marginesie każdego rysunku i na odwrocie fotografii powinno być napisane ołówkiem imię i nazwisko Autora oraz skrót tytułu artykułu, do którego są przeznaczone.

Spis podpisów pod rysunki i fotografie powinien być umieszczony na oddzielnej stronie.

Streszczenie. Do każdego artykułu musi być dołączone streszczenie z tytułem artykułu. Streszczenia oraz tytuły (Summary) muszą być w języku polskim i angielskim. Streszczenie powinno wyjaśniać główny cel pracy, wskazywać korzyści i ograniczenia, możliwe zastosowania i zalecenia dla dalszego rozwoju danej gałęzi techniki. Objętość streszczenia nie powinna przekraczać 100 słów, a jego treść nie może być identyczna ze „Wstępem", lub „Zakończeniem". Pod streszczeniami powinny być podane słowa kluczowe.

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 i zwrócić osobiście, lub listownie pod adresem 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.

Subscription for external subscribers:

The promotional subscription price in 1996 is $ 90 including postage for institutions. Subscriptions should be sent to the publisher, Polish Scientific Publishers PWN Ltd, Journal Division, Miodowa 10, 00-250 Warsaw, POLAND, fax (48) (22) 26 09 50, (48) (22) 26 71 63, with a copy by fast to Electronics and Telecommunications Quarterly 00-665 Warsaw, ul. Nowowiejska 15/19, p. 470. Subscription is occupated after showing a checque or transfer documents. Our bank account is as follows: Bank account: PBK VIII O/Warszawa nr 370028-1052. At subscripber's request this journal will be air mailed at additional postage to 50% dross price to European countries and 65% overseas.

Subscription orders available through the local press distribution or through the Foreign Trade Enterprise ARS Polona, 00-068 Warszawa, Krakowskie Przedmieście 7, Poland.

Cena $\frac{\text{7 zł}}{\text{70 000 zł}}$

Kwartalnik Elektroniki i Telekomunikacji

WARUNKI PRENUMERATY

Wpłaty na prenumeratę przyjmowane są na okresy roczne

Na teren kraju prenumeratę przyjmują jednostki kolportażowe RUCH S.A. i wszystkie urzędy pocztowe na terenie całego kraju, właściwe dla miejsca zamieszkania lub siedziby prenumeratora oraz doręczyciele w miejscowościach, gdzie dostęp do urzędu jest utrudniony, a także Zakład Kolportażu Wydawnictwa SIGMA-NOT

Termin przyjmowania przez RUCH S.A. prenumeraty krajowej i na zagranicę oraz przez Pocztę Polską (tylko prenumerata krajowa) są kwartalne: do 20.XI na I kw., do 20 II na II kw., do 20 V na III kw., i do 20 VIII na IV kw.
Zakład Kolportażu Wydawnictwa SIGMA-NOT przyjmuje prenumeratę do 5 XII roku poprzedzającego

Dostawa zamówionej prasy następuje w sposób uzgodniony z zamawiającym, pod wskazanym adresem, w ramach opłaconej prenumeraty.

Adresy przedsiębiorstw kolportażowych i ich konta bankowe:

Zakład Kolportażu Wydawnictwa SIGMA-NOT, 00-716 Warszawa, ul. Bartycka 20, skr. poczt. 1004
Konto PBK III Oddział Warszawa nr 370015-1573-2720-3-67
Prenumerata ze zleceniem wysyłki za granicę jest dwukrotnie wyższa od ceny krajowej. Zlecający powinien podać dokładny adres odbiorcy. Dodatkowe informacje można otrzymać pod nr telefonu 49-30-86 lub 40-00-21 wew. 249, 295, 299.

Ruch S.A. Oddział Krajowej Dystrybucji Prasy, 00-958 Warszawa, ul. Towarowa 28
Konto PBK S.A. XIII Oddział Warszawa nr 370044-16551
Dostawa odbywa się pocztą zwykłą w ramach opłaconej prenumeraty, z wyjątkiem zlecenia dostawy pocztą lotniczą, której koszt w pełni pokrywa zleceniodawca. Prenumerata ze zleceniem dostawy za granicę jest o 100% wyższa od krajowej. Od osób lub instytucji, zamieszkałych lub mieszczących się w miejscowościach, w których nie ma jednostek kolportażowych RUCH prenumeratę Kwartalnika można zamówić w Oddziale Warsząwskim RUCH na konto: PBK S.A. XIII Oddział Warszawa, nr 370044-16551.

Bieżące numery można nabyć w **Księgarni Wydawnictwa Naukowego PWN,** ul. Miodowa 10, 00-251 Warszawa. Również można je nabyć, a także zamówić (przesyłka za zaliczeniem pocztowym) we Wzorcowni Ośrodka Rozpowszechniania Wydawnictw Naukowych PAN, Pałac Kultury i Nauki, 00-901 Warszawa w cenie podanej na okładce kwartalnika.

W roku 1997	
prenumerata roczna w kraju	34,00 zł (340.000 zł)
prenumerata roczna za granicę	90 $ USA